Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 6 Iss. 1 January 2026, pp: 241-251
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Behavioral Manipulation in Big Data Implementation: Systematic Literature Review

Nancy Vanessa[1*], Hendi Sama[2], Mangapul Siahaan[3]

[1,2,3]Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Internasional Batam, Indonesia

E-Mail: [1]2231021.nancy@uib.edu, [2]hendi@uib.ac.id, [3]mangapul.siahaan@uib.ac.id

Received Nov 05th 2025; Revised Dec 25th 2025; Accepted Jan 28th 2026; Available Online Jan 31th 2006
Corresponding Author: Nancy Vanessa

Abstract

This study examines behavioral manipulation in big data implementation through a systematic literature review of thirty peer-reviewed articles published between 2020 and 2025. The review aims to provide a clear understanding of the mechanisms, impacts, and mitigation strategies related to the use of big data to influence human behavior. The PRISMA 2020 framework was applied, starting with 250 identified records, and after screening based on inclusion and exclusion criteria, 30 studies were selected for full analysis. The results indicate that behavioral manipulation most frequently occurs through algorithmic recommendation systems, price personalization, deceptive interface designs (dark patterns), and data-driven persuasion techniques. These mechanisms were consistently associated with reduced user autonomy, biased decision-making, psychological pressure, and widening social inequalities. Several studies further reveal that algorithmic transparency alone is insufficient to prevent manipulation when users lack meaningful understanding or control over automated systems. The review also identifies emerging mitigation strategies, including dynamic consent mechanisms, independent algorithmic audits, ethical-by-design interfaces, and adaptive regulatory frameworks. However, the findings suggest that such interventions remain fragmented and unevenly implemented across sectors. Approximately 83.3% of the reviewed studies conclude that addressing behavioral manipulation in big data requires an integrated response combining technical safeguards, ethical system design, regulatory oversight, and strengthened digital literacy.

Keywords: Behavioral Manipulation, Big Data Implementation, Decision Making, Systematic Literature Review

1. INTRODUCTION

The advancement of digital technology has positioned big data as a vital element across multiple sectors, including business, healthcare, technology, and public policy. Big data enables the efficient processing of vast amounts of information, offering significant benefits such as personalized user experiences, operational optimization, and predictive analytics. However, its use also presents ethical and social challenges, particularly when employed for behavioral manipulation. In practical terms, behavioral manipulation manifests in several concrete forms, including personalized recommendation systems that strategically prioritize certain content to steer attention, dynamic pricing schemes that adjust prices based on inferred willingness to pay, interface designs that employ dark patterns such as forced continuity or confirmshaming, and targeted political or commercial messaging optimized to exploit emotional vulnerabilities. These practices go beyond neutral personalization because they deliberately shape choice architectures in ways that benefit system operators while limiting users' reflective decision-making. Big data–driven behavioral manipulation involves sophisticated algorithms designed to predict and influence individual decisions without their awareness, encompassing consumer activities, political orientation, and other forms of online interaction [1]. Scholars have increasingly emphasized that manipulation can occur not only through data-driven personalization but also through interface-level interventions, such as dark patterns, which exploit cognitive biases and systematically steer user decisions [2], [3]. For example, in a field experiment in educational technology, personalized recommendations were found to increase content consumption by approximately 60%, and overall app usage by 14%, compared to non-personalized systems [4]. These findings suggest that behavioral manipulation is embedded within both algorithmic and design-based infrastructures, making it a multifaceted phenomenon.

The urgency of this issue has grown alongside the integration of artificial intelligence into digital systems, which further strengthens the capacity of big data to analyze and influence human behavior. The combination of machine learning models, predictive analytics, and real-time personalization allows organizations to generate highly adaptive manipulation strategies that users may not recognize as such. This creates significant risks, including privacy violations, erosion of individual autonomy, and reinforcement of existing social biases. In Bangladesh, for instance, 18.3% of e-commerce websites analyzed were found to contain one or more dark patterns, indicating that manipulative interface design is not just theoretical but already widespread in developing digital markets [5]. Likewise, a study involving recommendation AI for dietary habit improvement in Japan revealed that trust and usage of AI were significantly affected by how data management, user communication, and transparency were designed [6]. Similar concerns have been observed in online learning systems in Indonesia, where local studies note risks such as data breaches and unauthorized access undermining user trust [7].

Although numerous studies have examined the potential of big data for innovation, systematic academic discussions concerning the consequences of behavioral manipulation and strategies for ethical mitigation remain limited. Prior research has shown that manipulation not only influences short-term decision-making but also generates long-term psychological, social, and economic impacts. Additionally, the expansion of online-based management systems in education and organizational settings underscores the increasing reliance on digital infrastructures, which, while improving efficiency, also introduces risks of misuse and manipulation if not managed responsibly [8]. Theoretical work on dark patterns has emphasized that millions of users are daily exposed to such interface designs, yet user awareness remains low [3]. This study therefore seeks to provide a deeper understanding of the mechanisms, consequences, and mitigation strategies of behavioral manipulation through big data. The research was designed to address several key questions, including the ethical and practical implications for individuals and society. Furthermore, it aims to propose strategic recommendations to minimize negative impacts while ensuring responsible use of big data. In this way, the study is expected to serve as a robust academic foundation for the development of more sustainable and ethically grounded technology policies in the future.

2. LITERATURE REVIEW

Previous studies have demonstrated that big data techniques employ a combination of behavioral analysis, demographic information, and machine learning algorithms to influence users' decision-making processes. While such practices offer opportunities for innovation, they also raise pressing ethical concerns, including privacy violations and inequities in data usage. This phenomenon is frequently observed on social media platforms, where algorithms are designed to target users with specific content to maximize engagement, often altering individual preferences in the process. The Facebook, Cambridge Analytica scandal stands as an iconic example of how big data can be exploited for large-scale behavioral manipulation, triggering global debates on transparency and the ethics of data use. Such practices have generated critical questions regarding public trust in personal data management and the risks of privacy erosion [9].

This phenomenon is often found on social media platforms, where algorithms are used to target users with specific content to maximize engagement, even to the extent of altering user preferences. The Facebook–Cambridge Analytica scandal has become an iconic example of how big data can be used for behavioral manipulation on a passive scale, creating global concerns regarding transparency and the ethics of big data usage [10]. This practice has raised critical questions about public trust in the management of personal data and the risk of privacy erosion.

In the context of the digital economy, behavioral manipulation is also frequently used in predictive marketing. Research has shown that companies can increase consumer conversion rates by presenting advertisements tailored to individual preferences based on behavioral analysis. Although this strategy enhances marketing efficiency, criticism may arise due to the lack of disclosure to users regarding how their data will be processed and used. In the public sector, the use of big data to guide user behavior through data-driven policies such as the social credit system in China has generated global discussion about the boundary between behavioral governance and the violation of individual rights [11].

In addition, the development of machine learning and AI technologies has expanded the scope of behavioral manipulation through big data. These algorithms not only predict actions but can also actively influence decisions through recommendations designed to affect emotions, as seen on e-commerce and streaming platforms. Recent studies have shown that such algorithms often employ inherent biases in training data, which may worsen social inequalities, also referred to as discrimination [12].

Although many benefits can be derived from the implementation of big data, the ethical challenges it raises require serious attention. International organizations such as UNESCO and OECD have advocated the development of regulations that balance innovation with the protection of individual rights. This has become increasingly important as the adoption of big data continues to rise across global sectors. Therefore, it is essential to gain a deeper understanding of how big data can be used for behavioral manipulation and its implications for modern society.

This study aims to explore the phenomenon of behavioral manipulation through the implementation of big data, which is becoming increasingly widespread across various sectors. Big data refers to large-scale, high-velocity, and high-variety datasets that are continuously generated through digital interactions, such as online transactions, social media activities, sensor data, and platform usage logs. Beyond volume, big data is characterized by its analytical potential, enabling advanced machine learning and predictive modeling techniques to extract behavioral patterns and infer individual preferences. This capability has transformed decision-making processes across sectors, while simultaneously increasing the power asymmetry between data holders and data subjects. The main focus of the study is to understand the methods used to influence individual behavior, such as predictive algorithms, machine learning, and content personalization designed based on user behavior patterns. In this context, the study will analyze the impact of behavioral manipulation on individual privacy, transparency, and social justice. Prior scholarship has investigated behavioral manipulation enabled by big data in areas such as recommender systems, targeted advertising, political microtargeting, and deceptive interface design. Empirical and conceptual studies consistently show that data-driven personalization can steer attention, shape preferences, and amplify engagement, while raising concerns regarding autonomy, privacy, and algorithmic bias. However, this body of work remains dispersed across disciplines and often examines isolated mechanisms or ethical issues rather than offering an integrative synthesis across technical, social, and regulatory dimensions. This fragmentation motivates the need for a systematic literature review that consolidates findings and clarifies overarching patterns and gaps. This is particularly important given the criticisms of algorithmic bias and the potential misuse of data that may exacerbate inequality.

3. RESEARCH METHOD

This study employed the Systematic Literature Review (SLR) method to identify, evaluate, and synthesize all relevant research evidence regarding behavioral manipulation in the implementation of big data. Systematic Literature Review (SLR) has become an established methodological approach for synthesizing evidence in multidisciplinary fields involving complex socio-technical systems. Unlike traditional narrative reviews, SLR relies on transparent search strategies, explicit inclusion and exclusion criteria, and structured analytical procedures, thereby enhancing reproducibility and reducing selection bias. In research on big data ethics and behavioral manipulation, where studies span computational techniques, psychological effects, and regulatory debates, SLR provides a rigorous means to integrate heterogeneous findings, identify thematic patterns, and articulate theoretical contributions. The SLR approach was chosen because it enables researchers to systematically and objectively collect information from a wide range of scientific literature, thereby producing a comprehensive understanding of the existing trends, techniques, and challenges in this field.

The SLR process was designed based on the framework adapted from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), ensuring that each stage was conducted transparently and could be reproduced [13]. The stages of the SLR process were designed as follows.

1. Literature Search Process
   The literature search was carried out using several academic databases, namely Google Scholar, PubMed, Scopus, and Frontiers. The keywords applied included combinations of terms such as “behavioral manipulation,” “big data implementation,” and “decision making.” The search was conducted within the time frame of 2020–2025 to ensure the relevance of the studies to recent technological developments. Each search result was examined through its abstract and keywords to confirm its relevance to the research objectives. Studies that were not available in full text or lacked a direct relationship with the main topic were excluded.

2. Selection Criteria
   The selection of literature was carried out in two stages, namely the initial screening and the advanced screening. In the initial screening stage, documents that did not meet the following criteria were eliminated:
   a. Written in English to ensure accessibility and analysis.
   b. Published in indexed journals or conference proceedings to guarantee quality.
   c. Thematic relevance to behavioral manipulation and big data, in accordance with the predetermined keywords.

   In the advanced screening stage, the literature was selected based on the following criteria:
   a. A clear focus on big data applications that influence user behavior, such as digital marketing, social media, or data-driven decision-making.
   b. Empirical studies or conceptual reviews that address ethical aspects, algorithmic bias, or transparency in the use of big data.
   c. Publications within the time range of 2020–2025.

3. Quality Assessment
   Each piece of literature that passed the selection stage was evaluated for quality using the Critical Appraisal Skills Programme (CASP) framework. The assessment criteria included:
   a. Clarity of research objectives.
   b. Appropriateness of methods in relation to the research objectives.
   c. Relevance of findings to behavioral manipulation in big data.
   d. Strength of arguments, methodological validity, and contribution to the research field.

   Each article was assigned a score based on these criteria, and only literature with high-quality scores was included in the analysis.

4. Data Analysis Method
   The data obtained from the selected literature were analyzed thematically using a narrative synthesis approach. The analysis process involved three main steps:
   a. Data extraction, where key information from each article, such as methods, findings, and implications, was recorded in a standardized data sheet.
   b. Categorization, in which the extracted data were grouped based on major themes, such as manipulation techniques, ethical impacts, and policy responses.
   c. Synthesis, where each theme was analyzed to identify patterns, differences, and research gaps. This step contributed to building a coherent understanding of the phenomenon of behavioral manipulation in big data.

Table 1. Inclusion and Exclusion for the Implementation of Behavioral Manipulation in Big Data

	Inclusion Criteria	Exclusion Criteria
From File	Articles published in English-language journals within the period 2020–2025.	Books, proceedings, theses, dissertations, literature review articles, non-English works, or publications outside 2020–2025.
From Title	Articles with titles addressing behavioral manipulation in the implementation of big data.	Articles with titles addressing behavioral manipulation but not in the context of big data implementation.
From Abstract	Articles with abstracts discussing behavioral manipulation in the implementation of big data.	Articles with abstracts that do not discuss behavioral manipulation in the implementation of big data.
From Full Text	Articles whose full text discusses behavioral manipulation in the implementation of big data.	Articles whose full text does not discuss behavioral manipulation in the implementation of big data.

Articles selected for review were required to meet the inclusion criteria outlined in Table 1. Ultimately, the researcher identified a set of final articles to use in the research process. In contrast, articles eliminated from the list did not meet the inclusion criteria due to inaccuracy or insufficient alignment with the topic of behavioral manipulation in big data implementation. The following diagram illustrates the flow of the inclusion and exclusion process within the PRISMA framework (n: number of articles) (Figure 1).

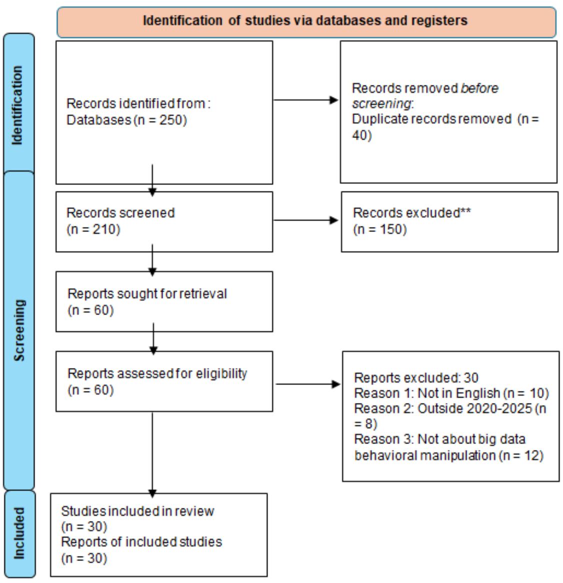

Figure 1. PRISMA Framework for the Implementation of Behavioral Manipulation in Big Data

The article selection process in this study followed the PRISMA 2020 flow. A total of 250 articles were initially identified from academic databases. After removing 40 duplicates, 210 articles remained for the screening stage. At this stage, 150 articles were eliminated after reviewing titles and abstracts because they were considered irrelevant, leaving 60 articles for the eligibility assessment. Of these, 30 articles were excluded, consisting of 10 that were not written in English, 8 published outside the 2020–2025 period, and 12 that did not specifically address behavioral manipulation in the implementation of big data. Consequently, 30 articles were included in the final stage for further analysis in this systematic literature review.

4. RESULTS AND DISCUSSION

A systematic review of 30 relevant articles on behavioral manipulation in the context of big data revealed that this phenomenon encompasses highly complex dimensions, including technical mechanisms, psychological and social impacts, ethical dilemmas and transparency, as well as regulatory frameworks and resistance strategies. From the literature analyzed, four major themes were identified: the mechanisms of manipulation, the impacts on user behavior and autonomy, issues of transparency and bias, and the role of regulation and ethical interventions.

4.1. Mechanisms of Manipulation in Big Data and Algorithms

The literature indicates that mechanisms of manipulation in big data operate through a combination of recommendation algorithms, price personalization, behavioral modification, and digital interface design incorporating dark patterns. Acemoglu and colleagues (2023, 2025) emphasized that companies with access to behavioral data can use surface attributes or glossy attributes to alter consumer perceptions even without changing the substantive qualities of a product [14], [15]. A classical study on price personalization by Li et al. (2022) further demonstrated that companies exploit granular data to differentiate prices among consumers based on their willingness to pay, creating a systematic form of economic manipulation [16]. Bandy (2021) reinforced this argument through an audit of algorithms showing how content visibility is distorted, leading users to be subtly steered toward certain decisions [17].

Technical manipulation also emerges in the context of security. Nguyen et al. (2024) discussed poisoning attacks against recommendation systems, whereby external actors inject false data to bias algorithms in favor of specific groups [18]. Jagielski et al. (2018) demonstrated that data poisoning can alter machine learning models in ways that are difficult to detect [19]. Meanwhile, Shmueli and Tafti (2023) explained how big data–driven predictions are often reinforced through behavioral modification techniques, thereby blurring the line between neutral prediction and manipulative intervention [20]. The literature on dark patterns adds another critical dimension. Punetha (2024), in a systematic review, found that interface designs such as confirmshaming, obstruction, and forced action are manipulative tactics deliberately employed to direct user behavior [21]. The official report *Patterns in the Dark* (DSB, 2024) empirically showed how governments must address dark patterns as part of manipulative digital infrastructure [22]. Collectively, the evidence demonstrates that manipulation in big data extends beyond prediction and recommendation into system design, pricing models, and security frameworks that can be exploited.

4.2. Impacts on Autonomy, Decision-Making, and User Experience

Data-driven manipulation has tangible consequences for individual autonomy and decision-making. Carroll et al. (2023) and Sabour et al. (2025) showed in experimental studies that humans are highly prone to following algorithmic suggestions, even when such suggestions are demonstrably suboptimal [23], [24]. This was further supported by Fan and Liu (2022), who found that algorithmic autonomy influences consumer decision-making in a non-linear pattern, where moderate algorithmic control exerted the strongest impact on purchasing decisions [25].

The psychological and social impacts are also significant. Hu (2025) investigated social media and found that algorithmic recommendations enhance engagement while simultaneously producing frustration and mental fatigue [26]. Arora et al. (2024) warned about the long-term effects of social media algorithms on adolescents, including heightened stress, anxiety, and susceptibility to manipulative content[27]. Experimental work by Bogert et al. (2021) showed that as tasks become more difficult, individuals increasingly rely on algorithms over social influence, indicating a tendency to delegate decision-making to systems [28].

From a design perspective, Chang, Seaborn, and Adams (2024) explained the psychological mechanisms behind dark patterns, which operate by exploiting cognitive biases and human perceptual weaknesses [29]. Fagan (2024), through a review of persuasion psychology, added that tactics such as framing and seduction create an illusion of free choice when options are, in fact, constrained [30]. Ethnographic research by Overbye-Thompson (2025) showed that some users attempted to resist manipulation through workaround strategies, though such resistance was not always effective [31].

Other studies link manipulation to broader social vulnerabilities. Padarha (2023) described a dystopian data environment in which algorithmic manipulation perpetuates ongoing ethical violations and deepens structural inequities [32]. Stella, Ferrara, and De Domenico (2018) demonstrated how social bots amplify the spread of negative content, producing harmful psychological and social consequences [33]. Taken together, these findings underscore that algorithmic manipulation not only affects immediate behavior but also erodes psychological autonomy and long-term social trust.

4.3. Transparency, Bias, and Ethical Dilemmas

One of the central themes in literature is the paradox of algorithmic transparency. Klenk (2024) argued that transparency can itself be manipulative when it is merely formal and fails to grant users meaningful control [34]. Wang (2022) extended this argument by showing that transparency often functions as a normative instrument of power, reinforcing algorithmic dominance [35]. Ulrik Franke (2022), in his reflection, also questioned the extent to which society should care about transparency, as disclosure does not necessarily translate into meaningful understanding [36].

The issue of bias receives significant attention. Starke et al. (2022), in a systematic review of public perceptions of algorithms, found that people frequently perceive algorithms as unfair and nontransparent, especially when automated outcomes lack sufficient explanation [37]. Saura (2022) demonstrated that big data in the governance context poses serious challenges related to privacy and bias, undermining public trust [38]. Hacker (2023) further emphasized that algorithmic manipulation often manifests in commercial contexts as a form of unfair business practice [39].

The ethical discussion is reinforced by Hosseini et al. (2022), who found that the use of big data in social research introduces critical problems related to data reuse, methodological bias, and the absence of ethical regulation [40]. Padarha (2023) underscored that algorithmic manipulation and large-scale data use have created what amounts to near-permanent ethical violations [32]. Cellard (2022) proposed the concept of surfacing algorithms as a method to enhance accountability by making algorithms visible through documentation and representations that can be publicly scrutinized [41]. Thus, while transparency is often presented as a solution, the literature demonstrates that without interpretive capacity and genuine control, transparency may serve as a tool of legitimizing manipulation.

4.4. Regulation, Ethical Design Interventions, and Social Resistance

The final theme highlights the necessity of regulation and ethical design interventions to limit algorithmic manipulation. Yi and Li (2024), in a systematic review of dark pattern regulation, showed that legal interventions are beginning to take shape but remain partial and do not adequately address the technical dimensions of algorithms [2]. Fagan (2024) stressed that the psychology of digital persuasion requires regulatory measures that protect users from subconscious tactics [30]. Reports by the OECD (2024) and the Open Government Partnership (2023) on algorithmic transparency in the public sector further underscore the importance of policy instruments for ensuring openness and accountability [42], [43].

Lee et al. (2024) demonstrated the success of dynamic consent in healthcare data contexts, where users could tailor data permissions to specific situations [44]. This study highlighted the promise of participatory design approaches in reducing manipulation. Grimmelikhuijsen (2023) found that algorithmic transparency in the public sector enhanced perceptions of governmental legitimacy, even though users did not always comprehend technical details [45]. At the same time, user resistance to manipulation also emerges in the literature. Yuan (2025) observed that users fall along a spectrum, ranging from full compliance with algorithms to active rejection through personal strategies [46]. Ullah (2025) studied online reviews and found that dark patterns affect genders differently, indicating that certain groups are more vulnerable to manipulation [47]. Stella et al. (2018) reaffirmed that manipulation is not merely an individual interaction problem but also a systemic issue within information ecosystems amplified by bots and automation [33].

Overall, the evidence makes clear that solutions to behavioral manipulation through big data cannot rely on a single instrument. Public regulation, algorithm audits, ethical interface design, and digital literacy initiatives must operate in tandem. The *Patterns in the Dark* report (2024) illustrates how public policy has begun moving in this direction but also emphasizes the need for greater synergy among policymakers, researchers, and civil society [22].

In addition to the thematic synthesis presented above, recent studies provide more nuanced insights into whether personalization and recommendation systems generate benefits or harm in practice. Aridor *et al.* [48] conducted a large-scale field experiment using the MovieLens dataset and found that algorithmic recommendations significantly influenced user behavior, with approximately 40% of the content consumed being directly attributable to the recommendation mechanism. This indicates that personalization systems can meaningfully shape decision pathways by exposing users to previously unexplored options, thereby improving informational efficiency. However, empirical research also highlights substantial drawbacks. Mansoury *et al.* [49] demonstrated that feedback loops in recommender systems amplify popularity bias over time popular items receive disproportionate visibility while niche content becomes increasingly marginalized, reducing exposure diversity and user autonomy. Similarly, Kowald *et al.* [50] found that in the entertainment domain, popularity amplification led to a twofold increase in exposure inequality, with users receiving progressively narrower recommendation sets. These findings support the view that algorithmic personalization can unintentionally reinforce social and informational stratification when left unchecked.

Complementing this critical perspective, Ribeiro *et al.* [51] introduced the concept of the "amplification paradox," arguing that algorithmic influence does not always escalate linearly; rather, behavioral moderation from users and content saturation effects may limit the extent of manipulation. Their simulation studies revealed that algorithmic amplification plateaus after repeated exposures, implying that user agency and contextual engagement can buffer against total behavioral convergence. Taken together, these contrasting results underscore the dual nature of algorithmic personalization: while it enhances efficiency and user satisfaction under transparent and well-calibrated conditions, it can also perpetuate bias, limit diversity, and erode fairness in opaque systems. Beyond summarizing prior studies, this review contributes a structured synthesis of how behavioral manipulation through big data is operationalized across digital systems. The analysis identifies three dominant clusters of mechanisms, namely algorithmic personalization through recommender systems, interface based nudging such as dark patterns, and data driven persuasion through microtargeting and price discrimination. Across these clusters, the reviewed literature consistently links manipulation to four recurring outcome dimensions, namely erosion of user autonomy, amplification of social bias, reduced informational diversity, and psychological pressure manifested as stress or compulsive engagement. In addition, this review highlights several systematic gaps in existing research. Empirical studies remain heavily concentrated in developed economies and platform based contexts such as e commerce and social media, while longitudinal evidence on long term behavioral effects is scarce. Research on mitigation strategies is also uneven, with regulatory and technical proposals often discussed normatively rather than evaluated through controlled experiments or real world deployments. By mapping these patterns and limitations, this study provides an integrative framework that connects technical design choices, behavioral outcomes, and governance responses, thereby offering a clearer agenda for future interdisciplinary research on responsible big data implementation. Consequently, ongoing efforts to design accountable algorithms must integrate fairness metrics, transparency audits, and participatory oversight to mitigate manipulative risks while preserving user benefit.

Table 2 provides a structured overview of the core studies included in this review, summarizing their domains of application, methodological approaches, forms of behavioral manipulation examined, reported impacts, and proposed mitigation strategies. This comparative mapping supports the thematic synthesis presented in Sections A–D and highlights cross-cutting patterns across technical, psychological, ethical, and regulatory dimensions.

Table 2. Summary of Studies on Behavioral Manipulation and Big Data

No	Author	Domain	Method	Manipulation Mechanism	Impact	Mitigation
14	Acemoglu et al.	Markets	Modeling	Attribute framing	Distorted choice	Policy design
15	Acemoglu et al.	Markets	Empirical theory	Microtargeting	Preference shaping	Regulation
16	Li et al.	Manufacturing	Review	Data driven pricing	Economic manipulation	Governance
17	Bandy	Platforms	Audit review	Ranking distortion	Behavior steering	Algorithm audits
18	Nguyen et al.	Recommenders	Survey	Poisoning attacks	Bias injection	Defense systems
19	Jagielski et al.	ML systems	Experiment	Data poisoning	Model corruption	Security
20	Shmueli & Tafti	Prediction	Conceptual	Behavior modification	Autonomy erosion	Ethical limits
21	Punetha	Interfaces	SLR	Dark patterns	Coercion	Regulation
22	DSB Report	Platforms	Policy study	Dark infrastructure	Manipulation	Public law
23	Sabour et al.	Decision making	Experiment	AI persuasion	Compliance	User safeguards
24	Carroll et al.	AI systems	Conceptual/exp	Manipulative framing	Reduced agency	Transparency
25	Fan & Liu	Commerce	Experiment	Algorithmic autonomy	Purchase distortion	Design
26	Hu	Social media	Empirical	Recommendation nudges	Fatigue	Ethical design
27	Arora et al.	Youth social media	Review	Engagement loops	Stress	Policy
28	Bogert et al.	Decision tasks	Experiment	Automation	Delegation	Training

No	Author	Domain	Method	Manipulation Mechanism	Impact	Mitigation
				reliance		
29	Chang et al.	UX	Scoping review	Cognitive exploitation	Manipulation	Design ethics
30	Fagan	Persuasion	Review	Framing/illusion	Loss autonomy	Regulation
31	Overbye-Thompson	Users	Ethnography	Workarounds	Partial resistance	Literacy
32	Padarha	Society	Critical	Data exploitation	Structural harm	Governance
33	Stella et al.	Social bots	Network analysis	Amplification	Polarization	Platform control
34	Franke	Theory	Philosophy	Formal transparency	Legitimization	Meaningful control
35	Wang	Governance	Theory	Normative transparency	Power asymmetry	User rights
36	Franke	Ethics	Theory	Disclosure limits	False empowerment	Interpretability
37	Starke et al.	Public views	SLR	Opaque automation	Distrust	Explainability
38	Saura	Gov AI	Empirical	Data misuse	Bias	Privacy law
39	Hacker	Commerce	Legal analysis	Unfair practices	Consumer harm	Legal reform
40	Hosseini et al.	Research	Ethics review	Data reuse	Normative gaps	Ethics boards
41	Cellard	Methods	Qualitative	Algorithm surfacing	Accountability	Documentation
42	OGP	Public sector	Policy	Opaque systems	Distrust	Oversight
43	OECD	Public AI	Policy	Automation opacity	Risk	Regulation
44	Lee et al.	Health	Field study	Consent architectures	Reduced manipulation	Participation
45	Grimmelikhuijsen	Gov AI	Experiment	Explainable systems	Trust gains	Transparency
46	Yuan et al.	Browsing	Survey	Awareness effects	Resistance	Literacy
47	TEMU study	E-commerce	Behavioral	Dark patterns	Gender disparity	Consumer law
48	Aridor et al.	Movies	Field experiment	Recommenders	Behavior shift	Calibration
49	Mansoury et al.	Platforms	Simulation	Feedback loops	Bias amplification	Diversification
50	Kowald	Entertainment	Empirical	Popularity bias	Inequality	System redesign
51	Ribeiro et al.	Recommenders	Simulation	Amplification paradox	Plateau effects	User agency

Figure 2 illustrates how technical design choices enable manipulation mechanisms that generate downstream psychological and social impacts, which in turn motivate regulatory, ethical design, and resistance responses.

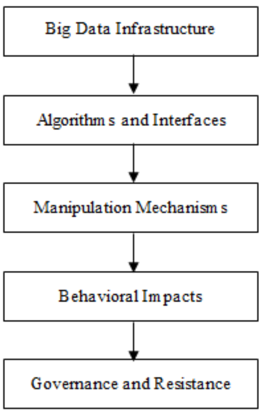

Figure 2. Integrated Ethical Framework of Behavioral Manipulation in Big Data

5. CONCLUSION AND RECOMMENDATIONS

A systematic review of thirty articles demonstrated that behavioral manipulation through big data is a multidimensional phenomenon involving technical mechanisms, psychological and social impacts, ethical dilemmas, and regulatory responses. On the one hand, recommendation algorithms, price personalization, dark patterns, and data poisoning attacks enable digital actors to subtly yet effectively shape user preferences and behaviors. Studies by Acemoglu, Li, Bandy, and Nguyen revealed how granular data and interface design can be directed to influence the decisions of both consumers and citizens. The impacts of such practices are evident in reduced individual autonomy, increased digital stress and fatigue, and the emergence of social biases that exacerbate structural inequalities, as described by Carroll, Hu, Arora, and Stella. Although transparency is often proposed as a solution, research by Klenk, Wang, Franke, and Starke highlighted that openness without the capacity for interpretation merely legitimizes manipulation. Mitigation efforts have emerged through regulation, participatory design, and social resistance for instance, through dynamic consent (Lee), OECD and OGP reports on transparency, and the findings of Yuan and Ullah on user resistance strategies, yet significant gaps remain in their broader implementation.

Based on these findings, a comprehensive strategy is required that integrates technical, regulatory, design, and social dimensions to limit behavioral manipulation in big data implementation. From the research perspective, future studies should focus more on developing countries and employ longitudinal approaches to capture long-term impacts more accurately. From the policy perspective, regulations must become more adaptive, not only demanding transparency but also granting users practical rights to control algorithms, reject recommendations, or opt out of personalization systems, accompanied by independent audit mechanisms. From the design perspective, user empowerment principles through participatory approaches such as dynamic consent should be expanded across sectors, while digital literacy must be strengthened so that individuals can recognize and resist manipulative patterns such as dark designs. With such a combination of strategies, algorithmic manipulation can be mitigated, and the use of big data can be more closely aligned with the principles of individual autonomy, social justice, and ethical sustainability.

REFERENCES

[1] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, “When Big Data Enables Behavioral Manipulation *,” 2024.

[2] M. Li *et al.*, “A Comprehensive Study on Dark Patterns,” *Proc. Make sure to enter correct Conf. title from your rights confirmation emai (Conference Acron. ’XX)*, vol. 1, no. 1, 2024, [Online]. Available: http://arxiv.org/abs/2412.09147

[3] T. Kollmer and A. Eckhardt, “Dark Patterns: Conceptualization and Future Research Directions,” *Bus. Inf. Syst. Eng.*, vol. 65, no. 2, pp. 201–208, 2023, doi: 10.1007/s12599-022-00783-7.

[4] G. N. Dec and K. Agrawal, “Personalized Recommendations in EdTech : Evidence from a,” 2022.

[5] Y. Sazid and K. Sakib, “Prevalence and User Perception of Dark Patterns: A Case Study on E-Commerce Websites of Bangladesh,” *Int. Conf. Eval. Nov. Approaches to Softw. Eng. ENASE - Proc.*, no. Enase, pp. 238–249, 2024, doi: 10.5220/0012734900003687.

[6] S. Genovesi, K. Kaesling, and S. Robbins, *The International Library of Ethics, Law and Technology 40 Recommender Systems: Legal and Ethical Issues*. 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-34804-4

[7] S. Tjahyadi and E. Titoni, “Risk Analysis In Indonesian Educational Online Learning Systems: A Systematic Literature Review,” *JITE (Journal Informatics Telecommun. Eng.*, vol. 8, no. 2, pp. 240–247, 2025, doi: 10.31289/jite.v8i2.13239.

[8] H. Sama, I. Deu, and S. Tjahyadi, “Development of an Online-Based Management System to Facilitate School Events,” *Conf. Manag. Bus. Innov. Educ. Soc. Sci.*, vol. 4, no. 1, pp. 26–39, 2024, [Online]. Available: https://journal.uib.ac.id/index.php/combines

[9] R. A. García-Hernández *et al.*, “A Systematic Literature Review of Modalities, Trends, and Limitations in Emotion Recognition, Affective Computing, and Sentiment Analysis,” *Appl. Sci.*, vol. 14, no. 16, 2024, doi: 10.3390/app14167165.

[10] Greenfield Patrick, “The Cambridge Analytica files: the story so far | News | The Guardian,” *TheGuardian.com*. 2018.

[11] H. R. Kwon and E. A. Silva, “Mapping the Landscape of Behavioral Theories: Systematic Literature Review,” *J. Plan. Lit.*, vol. 35, no. 2, pp. 161–179, 2020, doi: 10.1177/0885412219881135.

[12] I. D. Raji, T. Gebru, M. Mitchell, J. Buolamwini, J. Lee, and E. Denton, “Saving Face: Investigating the ethical concerns of facial recognition auditing,” *AIES 2020 - Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, pp. 145–151, 2020, doi: 10.1145/3375627.3375820.

[13] M. J. Page *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n71.

[14] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. E. Ozdaglar, “A Model of Behavioral Manipulation,” *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4638206.

[15] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "When Big Data Enables Behavioral Manipulation," *Am. Econ. Rev. Insights*, vol. 7, no. 1, pp. 19–38, 2025, doi: 10.1257/aeri.20230589.

[16] C. Li, Y. Chen, and Y. Shang, "A review of industrial big data for decision making in intelligent manufacturing," *Eng. Sci. Technol. an Int. J.*, vol. 29, no. xxxx, 2022, doi: 10.1016/j.jestch.2021.06.001.

[17] J. Bandy, "Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits," *Proc. ACM Human-Computer Interact.*, vol. 5, no. CSCW1, pp. 1–34, 2021, doi: 10.1145/3449148.

[18] T. T. Nguyen *et al.*, "Manipulating Recommender Systems: A Survey of Poisoning Attacks and Countermeasures," *ACM Comput. Surv.*, vol. 57, no. 1, 2024, doi: 10.1145/3677328.

[19] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, and B. Li, "Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning," *Proc. - IEEE Symp. Secur. Priv.*, vol. 2018-May, no. 1, pp. 19–35, 2018, doi: 10.1109/SP.2018.00057.

[20] G. Shmueli and A. Tafti, "Rejoinder: How to 'improve' prediction using behavior modification," *Int. J. Forecast.*, vol. 39, no. 2, pp. 566–569, 2023, doi: 10.1016/j.ijforecast.2022.12.007.

[21] A. Punetha, "Dark Patterns in User Interfaces : A Systematic and Meta-Analysis," 2024.

[22] J. Baumeister Ji-Young Park Andrew Cunningham Stewart Von Itzstein Ian Gwilt Aaron Davis James Walsh, "Deceptive Practices In Online Interactions Patterns In The Dark A Report to the Data Standards Chair Heading Two Line Heading Landscape Assessment: Dark Patterns A Report to the Data Standards Chair".

[23] S. Sabour *et al.*, "Human Decision-making is Susceptible to AI-driven Manipulation," 2025, [Online]. Available: http://arxiv.org/abs/2502.07663

[24] M. Carroll, A. Chan, H. Ashton, and D. Krueger, "Characterizing Manipulation from AI Systems," *ACM Int. Conf. Proceeding Ser.*, 2023, doi: 10.1145/3617694.3623226.

[25] Y. Fan and X. Liu, "Exploring the role of AI algorithmic agents: The impact of algorithmic decision autonomy on consumer purchase decisions," *Front. Psychol.*, vol. 13, no. October, pp. 1–17, 2022, doi: 10.3389/fpsyg.2022.1009173.

[26] Z. Hu, "Research on the Impact of Social Media Algorithmic on User Decision-making: Focus on Algorithmic Transparent and Ethical Design," *Appl. Comput. Eng.*, vol. 174, no. 1, pp. 18–22, 2025, doi: 10.54254/2755-2721/2025.po24665.

[27] S. Arora, S. Arora, and J. Hastings, "The Psychological Impacts of Algorithmic and AI-Driven Social Media on Teenagers: A Call to Action," *2024 IEEE Digit. Platforms Soc. Harms, DPSH 2024*, 2024, doi: 10.1109/DPSH60098.2024.10774922.

[28] E. Bogert, A. Schecter, and R. T. Watson, "Humans rely more on algorithms than social influence as a task becomes more difficult," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-021-87480-9.

[29] W. J. Chang, K. Seaborn, and A. A. Adams, "Theorizing Deception: A Scoping Review of Theory in Research on Dark Patterns and Deceptive Design," *Conf. Hum. Factors Comput. Syst. - Proc.*, 2024, doi: 10.1145/3613905.3650997.

[30] P. Fagan, "Clicks and tricks: The dark art of online persuasion," *Curr. Opin. Psychol.*, vol. 58, no. July 2024, p. 101844, 2024, doi: 10.1016/j.copsyc.2024.101844.

[31] H. Overbye-Thompson and R. E. Rice, "Understanding how users may work around algorithmic bias," *AI Soc.*, no. Hamilton 2019, 2025, doi: 10.1007/s00146-025-02498-1.

[32] S. Padarha, "Data-Driven Dystopia : an uninterrupted breach of ethics [3] Reckless misuse of sensitive data by large firms," no. march 2021.

[33] M. Stella, E. Ferrara, and M. De Domenico, "Bots increase exposure to negative and inflammatory content in online social systems," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 49, pp. 12435–12440, 2018, doi: 10.1073/pnas.1803470115.

[34] U. Franke, "Algorithmic Transparency, Manipulation, and Two Concepts of Liberty," *Philos. Technol.*, vol. 37, no. 1, pp. 2–7, 2024, doi: 10.1007/s13347-024-00713-3.

[35] H. Wang, "Transparency as Manipulation? Uncovering the Disciplinary Power of Algorithmic Transparency," *Philos. Technol.*, vol. 35, no. 3, pp. 1–25, 2022, doi: 10.1007/s13347-022-00564-w.

[36] U. Franke, "How Much Should You Care About Algorithmic Transparency as Manipulation?," *Philos. Technol.*, vol. 35, no. 4, pp. 1–7, 2022, doi: 10.1007/s13347-022-00586-4.

[37] C. Starke, J. Baleis, B. Keller, and F. Marcinkowski, "Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature," *Big Data Soc.*, vol. 9, no. 2, 2022, doi: 10.1177/20539517221115189.

[38] J. R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marqués, "Assessing behavioral data science privacy issues in government artificial intelligence deployment," *Gov. Inf. Q.*, vol. 39, no. 4, 2022, doi: 10.1016/j.giq.2022.101679.

[39] P. Hacker, "Manipulation by algorithms. Exploring the triangle of unfair commercial practice, data protection, and privacy law," *Eur. Law J.*, vol. 29, no. 1–2, pp. 142–175, 2023, doi:

10.1111/eulj.12389.
[40] M. Hosseini, M. Wieczorek, and B. Gordijn, "Ethical Issues in Social Science Research Employing Big Data," *Sci. Eng. Ethics*, vol. 28, no. 3, pp. 1–21, 2022, doi: 10.1007/s11948-022-00380-7.
[41] L. Cellard, "Surfacing Algorithms: An Inventive Method for Accountability," *Qual. Inq.*, vol. 28, no. 7, pp. 798–813, 2022, doi: 10.1177/10778004221097055.
[42] M. Valderrama, M. P. Hermosilla, and R. Garrido, "State of the Evidence: Algorithmic Transparency," *Open Gov. Partnersh.*, no. May, 2023.
[43] J. David Gutiérrez and A. Gillwald, "Algorithmic Transparency in the Public Sector A state-of-the-art report of algorithmic transparency instruments Acknowledgements," no. November, 2024.
[44] A. R. Lee, D. Koo, I. K. Kim, E. Lee, S. Yoo, and H. Y. Lee, "Opportunities and challenges of a dynamic consent-based application: personalized options for personal health data sharing and utilization," *BMC Med. Ethics*, vol. 25, no. 1, 2024, doi: 10.1186/s12910-024-01091-3.
[45] S. Grimmelikhuijsen, "Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making," *Public Adm. Rev.*, vol. 83, no. 2, pp. 241–262, 2023, doi: 10.1111/puar.13483.
[46] Y. Yuan, Y. Shi, T. Su, and H. Zhang, "Resistance or compliance? The impact of algorithmic awareness on people's attitudes toward online information browsing," *Front. Psychol.*, vol. 16, 2025, doi: 10.3389/fpsyg.2025.1563592.
[47] "Dark Patterns , Online Reviews , and Gender : A Behavioural Analysis of Consumer Decision Making on TEMU in the E- Commerce Context," pp. 1–61.
[48] G. Aridor, D. Goncalves, D. Kluver, R. Kong, and J. Konstan, "The Economics of Recommender Systems: Evidence from a Field Experiment on MovieLens," *EC 2023 - Proc. 24th ACM Conf. Econ. Comput.*, no. 10129, p. 117, 2023, doi: 10.1145/3580507.3597677.
[49] M. Mansoury, H. Abdollahpouri, M. Pechenizkiy, B. Mobasher, and R. Burke, "Feedback Loop and Bias Amplification in Recommender Systems," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 2145–2148, 2020, doi: 10.1145/3340531.3412152.
[50] D. Kowald, "Investigating Popularity Bias Amplification in Recommender Systems Employed in the Entertainment Domain," *Ewaf*, p. 3, 2025.
[51] M. Horta Ribeiro, V. Veselovsky, and R. West, "The Amplification Paradox in Recommender Systems," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 17, pp. 1138–1142, 2023, doi: 10.1609/icwsm.v17i1.22223.