Journal of Intercultural Management
Vol. **18** | No. **2** | **2026**
pp. **66–90**
DOI **10.2478/joim-2026-0007**

Nasser Bouchareb
Ferhat Abbas University Setif 1, Algeria
Nacir007@yahoo.fr

Ismail Morad
University of Ain Temouchent, Algeria
morad.ismail@univ-temouchent.edu.dz

Yacine Herizi
University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj, Algeria
Yacine.herizi@univ-bba.dz

Data Treatment, Consumer Profiling, and Privacy: Toward a Unified Conceptual Framework

Abstract

Objective: To develop a unified conceptual framework explaining how AI-driven data treatment (collecting, integrating, transforming, and inferring from consumer traces) produces consumer profiles, and how those profiles translate into distinct privacy impacts, while showing where governance controls can intervene.

Methodology: A conceptual, theory-building approach that synthesizes insights across marketing, information systems, and privacy/AI governance literature to model privacy risk as a mechanism chain: data treatment dimensions, profiling intensity, privacy impacts, with governance operating as a moderating layer across the chain.

Findings: The paper identifies a small set of structural dimensions that largely determine privacy risk: the extent to which data are linkable across contexts, the depth of inference, the persistence of representations over time, and the degree of coupling to consequential actions. These dimensions shape profiling intensity (granularity, profile stability, automation, actionability) and generate multiple harm pathways – informational, autonomy-related, distributive/fairness-related, and security-related – whose severity increases when systems are opaque and hard to contest.

Value Added: The paper offers a parsimonious vocabulary that connects upstream data practices to downstream harms via profiling intensity, helping researchers and practitioners move beyond narrow "consent/anonymization" framings and better diagnose *why* privacy harm can occur even when data are legally obtained or "de-identified".

Recommendations: Firms should limit high-risk profiling by reducing inference depth, cross-context linkage, and persistence, and by making consequential profiles transparent and contestable; regulators should focus on enforceable rules for transparency/contestability and restrict high-impact profiling; researchers should operationalize key constructs (e.g., inference depth, persistence, actionability) and test the framework empirically across sectors.

Key words: data treatment, consumer profiling, AI governance, privacy harms, contextual integrity, algorithmic decision-making, digital marketing

JEL codes: M31, L86, D83

Introduction

Artificial intelligence (AI) has become a central infrastructure for contemporary consumer understanding. Across digital platforms and data brokerage ecosystems, firms increasingly rely on "data treatment", the systematic collection, integration, transformation and analysis of heterogeneous data traces, to infer preferences, predict behavior, and optimize decision-making at scale. These practices promise efficiency and relevance through personalization, recommender systems, dynamic pricing, fraud detection, and targeted advertising. Yet the same pipeline that converts raw traces into actionable insight also expands the privacy surface of consumer life: it increases the volume of data processed, deepens the inferential reach of analytics, and amplifies the capacity to classify individuals into consequential categories that shape opportunities, information exposure, and market treatment.

Existing privacy debates often focus on discrete issues, notice-and-consent, anonymization, data minimization, or cybersecurity, without fully accounting for the way profiling emerges from the end-to-end treatment of data. In practice, the privacy risk is not limited to the disclosure of explicit identifiers; it also arises from the production of *inferred attributes*, the recombination of seemingly innocuous datasets, and the persistence of profiles that outlive the contexts in which data were generated. Moreover, AI-enabled profiling shifts privacy concerns from individual disclosure to broader socio-technical effects (Mavriki & Karyda, 2019), including autonomy erosion through behavioral targeting, differential treatment via scoring and segmentation, and "group privacy" harms where outcomes are determined by membership in inferred categories rather than by revealed personal facts (Mann & Matzner, 2019). This creates

a conceptual mismatch: traditional privacy framings built around data ownership or secrecy are often ill-suited to explain harms that occur even when data are legally obtained, technically "de-identified", or never directly exposed to third parties.

Against this background, this paper develops a unified conceptual framework that links data treatment practices to consumer profiling dynamics and to privacy impacts. The objective is not to provide a technical taxonomy of machine-learning methods, but to clarify the conceptual mechanisms by which data treatment becomes profiling, and profiling becomes privacy risk. We propose that privacy outcomes are shaped by a small number of analytically separable dimensions, such as the scope and sensitivity of data inputs, the depth of inference enabled by transformations, the granularity and stability of profiles, and the degree to which profiles are operationalized in automated or semi-automated decisions. By articulating these dimensions and their relationships, the framework integrates insights that are often treated in isolation across information systems, marketing, law and policy, and AI ethics.

Although the mechanisms linking AI-driven data treatment, consumer profiling, and privacy harms are broadly applicable across digital markets, their practical implications are shaped by cultural and institutional contexts. Societies differ in their expectations regarding personal privacy, the acceptability of algorithmic profiling, and the level of trust placed in organizations and public institutions. Likewise, regulatory traditions vary considerably in the degree to which they emphasize individual rights, organizational flexibility, or public oversight. Consequently, the governance mechanisms proposed in this framework should not be understood as universally identical in their implementation or effectiveness. Instead, their legitimacy and practical impact are expected to depend on the broader cultural and regulatory environment in which AI-driven consumer profiling occurs.

The paper makes three contributions. First, it offers a precise conceptualization of "data treatment" as a lifecycle of transformations whose privacy implications can exceed those associated with data collection alone. Second, it theorizes consumer profiling as the construction and operationalization of behavioral and categorical representations ("data doubles") and embeds them into market

interactions. Third, it synthesizes major privacy harm pathways, informational, autonomy-related, distributive/fairness-related, and security-related, showing how each can be generated or intensified by particular combinations of treatment and profiling intensity. The framework also yields propositions and a research agenda for future empirical work on when and why AI-driven consumer understanding crosses the threshold from beneficial personalization to privacy-invasive surveillance.

The remainder of the paper first clarifies the meaning and scope of data treatment in AI-mediated consumer analytics, emphasizing how transformation and inference reshape what data "are" in privacy terms. It then examines consumer profiling as a socio-technical practice, how profiles are constructed, stabilized, and made actionable in market settings. Building on these foundations, the paper maps privacy harms and the pathways through which they emerge, before integrating the elements into a unified framework that supports theory development, governance thinking, and a targeted agenda for empirical validation.

Conceptualizing "Data Treatment" in AI Companies

In AI-mediated consumer analytics, *data treatment* is often reduced, especially in legal or operational discourse, to a broad notion of "processing". For theory-building, that definition is too thin. A more analytically useful concept treats data treatment as an end-to-end pipeline of transformation and inference through which organizations convert dispersed traces into decision-relevant knowledge about consumers. The privacy stakes, therefore, are not located only in whether data are collected or transferred, but in how data are rendered comparable, linkable, and predictive, and then embedded into market action. Several works implicitly adopt this pipeline perspective, where firms: Collect and integrate dispersed traces; Transform data into analysis-ready representations; Inference and modeling for consumer understanding; Deployment into decision and action loops.

Table 1. Stages linking dispersed traces to marketing decisions

Pipeline Stage	Consumer-focused Example	Source
Collection & Integration	Online/offline, social, and transaction data unified for big data consumer analytics	(Erevelles et al., 2016; Fernández-Rovira et al., 2021; Kitchens et al., 2018)
Transformation	Pre-processing and feature engineering for customer models	(Giovanelli et al., 2022; Khattach et al., 2025; Sivakolundhu, 2024)
Inference	Segmentation, recommendation, CLV, churn prediction	(Erevelles et al., 2016; Gupta et al., 2020; Kitchens et al., 2018; Kodi, 2024)
Decision/Action	Personalized offers, dynamic pricing, real-time interventions	(Hajli et al., 2020; Khattach et al., 2025; Shah, 2025; Naphade, 2025)

A minimal conceptual model can be expressed through four tightly coupled operations. First, firms capture traces by instrumenting interaction: clicks, dwell time, purchases, device signals, location pings, and customer-service logs (Bujlow et al., 2017; Sakalauskas & Kriksciuniene, 2024; Schweidel et al., 2022). The resulting "raw data" are not neutral; they reflect design choices about what is measurable, which behaviors count as signals, and what remains unrecorded, an insight consistent with scholarship on datafication and platform governance (Gillespie, 2010, 2014; Van Dijck, 2014). Second, traces are standardized and integrated: they are cleaned, normalized, and merged across services, devices, and time (Wang et al., 2014; S. Xu et al., 2015). Identity resolution, whether deterministic (logins, persistent identifiers) or probabilistic (matching and fingerprinting), is pivotal because it converts fragmented episodes into a coherent behavioral narrative. At this point, privacy concerns begin to exceed classic disclosure models: the system can become "person-aware" through linkability even when explicit identifiers are minimized.

Third, integrated traces are converted into representations suitable for computation. This is the move from data as records to data as features: variables, embeddings, segments, and scores that compress complex life into analytic proxies (e.g., "price sensitivity", "likely churn", "affinity for category X") (Bastarache et al., 2022). This representational step is where many privacy issues become structurally "built in", because the consumer is no longer treated as a participant in a transaction but as an object of continuous measurement and prediction.

Fourth, systems produce inferences and activate them: models estimate propensities and classifications, and organizations operationalize them through ranking, targeting, personalization, dynamic pricing, risk scoring, and attention allocation (Ban & Keskin, 2021; Jha et al., 2025). Here, privacy becomes inseparable from power. As theorists of surveillance and market governance have argued, the critical question is not merely what a firm knows, but what it can do with that knowledge to shape exposure, choice architectures, and access to goods or opportunities (Zuboff, 2019).

This pipeline perspective clarifies why privacy cannot be fully assessed at the point of collection alone. Two mechanisms are especially central. The first is context collapse, captured well by Helen Nissenbaum's notion of contextual integrity: information generated in one social setting (browsing, entertainment, health queries, mobility) is made usable in another (advertising, credit-like scoring, eligibility decisions) once it is merged and abstracted (Nissenbaum, 2004). The second is inferential escalation: privacy-relevant attributes may be produced rather than disclosed. In other words, "sensitivity" is not only a property of inputs; it can be an emergent property of the treatment process. This logic also explains why de-identification is an unstable privacy guarantee in rich consumer ecosystems: linkage across datasets and high-dimensional representations can enable re-identification or attribute reconstruction, a recurring lesson in re-identification research associated with (Narayanan & Shmatikov, 2008).

To develop a unified framework linking data treatment to profiling and privacy, it is therefore helpful to characterize treatment not by technical minutiae but by a small set of structural dimensions that shape downstream risk (Quach et al., 2022; Salem et al., 2023). Four dimensions are particularly diagnostic across sectors. Linkability refers to the capacity to merge traces across contexts and identities; it is the main driver of "whole-of-life" visibility (Zheng et al., 2018). Inferential depth captures the distance between what is observed and what is predicted (or imputed) about preferences, vulnerabilities, or life circumstances (Bindschaedler et al., 2018). Persistence describes the temporal reach of stored representations and the extent to which past traces constrain present and future interactions. Operational coupling denotes how tightly inferences are connected to consequential actions, whether predictions merely inform content ordering

or directly gate opportunities, adjust prices, or allocate scrutiny (El Mokhtari et al., 2021; Mustafa et al., 2019). Together, these dimensions explain why two systems may appear similar at the level of "processing" yet differ radically in privacy implications: privacy risk increases as treatment becomes more linkable, more inferential, more persistent, and more tightly coupled to market decisions.

Conceptualizing data treatment in this way positions it as the upstream mechanism through which consumer profiling becomes possible. Profiling is not an optional add-on that occurs after neutral processing; it is the predictable outcome of a pipeline designed to transform traces into actionable predictions. This is precisely why a privacy analysis anchored in data treatment, rather than isolated moments of disclosure, offers a stronger theoretical basis for understanding contemporary AI companies' relationships to consumers.

Consumer Profiling as a Socio-Technical Process

Consumer profiling is commonly described as a technical outcome, "the model's prediction" or "the segment assignment". For a theoretical account, this is insufficient. Profiling is better conceptualized as a socio-technical process that produces *actionable representations* of consumers and embeds them into organizational routines, market infrastructures, and interface environments (Heiberg et al., 2022; Syrjälä et al., 2025; Vespignani, 2012). Profiles are not merely descriptive summaries; they are operational objects that coordinate how firms allocate attention, set terms of exchange, and structure what consumers see, pay, or can access.

A profile emerges when data treatment outputs are translated into stable categories or scores that can travel across decision points. This translation typically involves three interlocking moves. First, the consumer is rendered into a representation (features, embeddings, "propensity", "lifetime value", "risk", "affinity"). Second, that representation is evaluated through ranking or scoring, making the consumer comparable to others in a system optimized for selection. Third, the evaluation is

activated through interventions, recommendations, targeting, friction, prioritization, price steering, eligibility gating, or differential service. The practical meaning of a profile lies in this activation: a profile matters because it changes what the system does to or for the consumer (Barbosa et al., 2021; Trusov et al., 2016).

Profiling is socio-technical because its categories and thresholds are not determined by data alone (Heiberg et al., 2022; Syrjälä et al., 2025; Vespignani, 2012). They reflect organizational objectives (growth, retention, risk reduction), institutional constraints (compliance, reputation), and the economic logic of platform ecosystems (advertising auctions, recommender optimization, brokerage). What appears as a neutral prediction is therefore shaped by design decisions: what counts as a "conversion", which outcomes are rewarded, which proxies are acceptable, how long histories are retained, and where uncertainty is tolerated. In this sense, profiling is a governance practice conducted through computation: it standardizes consumers into system-readable types, then allocates differentiated treatment accordingly (Eke et al., 2019; Kim et al., 2025; Y. Xu et al., 2025).

Two theoretical lenses clarify what is distinctive about contemporary consumer profiling. The first is the idea of the surveillant assemblage, which highlights how profiles are assembled from multiple data flows that were originally produced in different contexts; the consumer becomes a "data double" that can be acted upon across settings (Haggerty & Ericson, 2000). The second is the concept of algorithmic identity, emphasizing that identity categories are increasingly assigned through probabilistic classification rather than self-description (Cheney-Lippold, 2011). Together, these perspectives foreground a key analytical point: profiling is not only about *knowing* consumers but about *making* consumers legible in ways that support systematic sorting and intervention.

This view also explains why profiling creates privacy concerns even without data disclosure. The central mechanism is not necessarily exposure, but consequential classification under opacity and asymmetry (Büchi et al., 2020; King & Forder, 2016). Consumers rarely observe the full profile, the features that shape it, or the counterfactual alternatives ("what would I have seen/paid/received if I were classified differently?"). Profiles can also be sticky: once created, they persist through retention practices and are reinforced by feedback loops (Reed & Forehand, 2016; Trusov et al., 2016). Recommendations influence behavior; behavior updates the profile;

the updated profile further shapes exposure. Profiling thus becomes *recursive*: it does not simply reflect preferences; it can actively participate in producing them.

For the unified framework developed in this paper, profiling can be analytically characterized by a small set of properties that determine its privacy relevance and its potential for harm. Granularity captures how fine the distinctions are (broad segment vs individualized score). Actionability captures whether the profile merely informs content ordering or triggers materially consequential decisions. Portability captures whether profiles travel across services, partners, or contexts through shared identifiers and ecosystem integrations. Stability captures persistence over time and resistance to revision. Contestability captures whether the consumer can understand, challenge, or correct the classification in meaningful ways. Profiling becomes most privacy-invasive when it is granular, highly actionable, portable across contexts, stable over long horizons, and weakly contestable (Mann & Matzner, 2019), conditions that effectively convert consumer data treatment into an infrastructure for continuous sorting.

Consumer profiling should be treated as the bridge between data treatment and privacy impact: it is the point where transformed traces become institutionalized representations and where those representations are operationalized into differentiated treatment (Valletti & Wu, 2020). This conceptualization aligns with the paper's broader argument that privacy is not only threatened by "having data", but by the systematic capacity to classify, predict, and intervene in consumer life through scalable socio-technical systems.

Privacy Harms and Risk Pathways

Privacy harms in AI-enabled consumer analytics cannot be reduced to "data leakage" or unauthorized disclosure. In the context of data treatment and profiling, harm more often arises from systematic visibility, inference, and differential treatment, including situations where the data never leave the firm or are nominally "de-identified" (Custers, 2018; Solove, 2006). Conceptually, the relevant question is not only what information is collected, but how the treatment pipeline and

profiling apparatus convert traces into actionable classifications that shape what consumers can see, pay, or access (Acquisti et al., 2015). This shifts privacy from a narrow informational concern to a governance problem that involves autonomy, fairness, and power asymmetries (Solove, 2006).

A useful theoretical organization distinguishes four families of privacy harms that are tightly coupled to profiling intensity. Informational harms include loss of control over personal information flows, unexpected secondary uses, and the creation of comprehensive consumer "portraits" through linkage and enrichment (Lupton, 2020; Solove, 2006). These harms include re-identification and attribute reconstruction, but also the more ordinary (and pervasive) experience of being tracked across contexts (Acquisti et al., 2015). Autonomy harms arise when profiling is used to steer attention, preferences, and choices: targeting and ranking systems can exploit situational vulnerabilities, produce "choice architectures" optimized for firm objectives, and narrow the informational environment in ways that are difficult to perceive from within (Susser et al., 2019). Distributive and fairness harms occur when profiles become the basis for unequal treatment, exclusion from offers, price steering, differential service quality, or systematically higher friction for certain inferred groups (Barocas & Selbst, 2016; Wachter, 2022). Importantly, these harms do not require explicit use of sensitive categories; they can emerge through proxies and correlated behavioral signals (Barocas & Selbst, 2016). Finally, security harms increase as treatment and profiling intensify: aggregation creates high-value targets, and the compromise of a profile repository can reveal not only what consumers did, but what the system inferred about them (Solove, 2006).

These harms are generated through recurring risk pathways that connect data treatment to profiling outcomes. One pathway is context transposition: traces generated in one context (browsing, mobility, social interaction) are repurposed in another (marketing eligibility, risk scoring, pricing) once they are integrated and made machine-legible, undermining context-appropriate information flows (Solove, 2006). Another pathway is inferential escalation, where seemingly non-sensitive inputs are transformed into sensitive conclusions (e.g., vulnerability, life events, financial stress) and treated as operational truths (Acquisti et al., 2015). A third pathway is function creep in which data and models built

for one purpose become infrastructural assets for additional purposes because marginal reuse is cheap and organizational incentives favor expansion (Solove, 2006). These pathways matter because they produce privacy impacts even under "lawful collection" narratives, precisely because the harm mechanism is transformation and activation, not simply collection (Acquisti et al., 2015).

A distinct pathway concerns opacity and contestability. Profiling systems typically operate under strong informational asymmetry: consumers rarely see the features driving their classification, the uncertainty of the model, or the counterfactual treatment they would have received under a different profile (Wachter, 2022; Wachter et al., 2017). This opacity makes harms difficult to detect and correct, and it can normalize a governance regime where important decisions are made through unobservable ranking and scoring (de Laat, 2018; Vaassen, 2022; Wachter et al., 2017). The result is a privacy problem with institutional characteristics: the individual cannot reasonably manage it through isolated consent decisions, because the relevant processes are continuous, adaptive, and distributed across ecosystems (Acquisti et al., 2015).

Profiling systems introduce feedback loops that can deepen harms over time. Interventions derived from profiles (recommendations, targeting, selective exposure, friction) shape behavior; the resulting behavior becomes new training data; the updated model further stabilizes the profile and intensifies intervention (Susser, 2019; Susser et al., 2019). This recursive dynamic is central to autonomy and fairness harms: systems may gradually narrow a consumer's informational environment, amplify certain behaviors, and lock individuals into trajectories that appear "personalized" but function as a form of soft constraint (Susser et al., 2019; Barocas & Selbst, 2016). For a unified conceptual framework, this means privacy impact should be analyzed not as a one-time event but as an emergent outcome of repeated cycles of: treatment; profiling; intervention; learning.

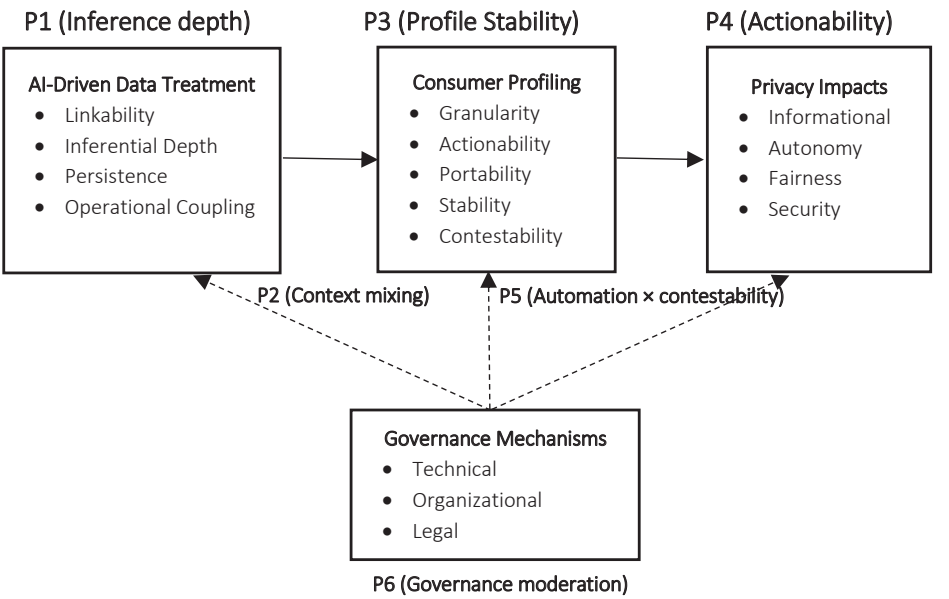
Taken together, these harm families and pathways motivate a shift in what counts as privacy-relevant evaluation. The most consequential risks arise when treatment produces high linkability, high inferential depth, strong persistence, and tight operational coupling between profiles and decisions (Solove, 2006; Wachter et al., 2017). Under these conditions, privacy is no longer primarily about secrecy; it becomes about the legitimacy and accountability of a system

that continuously classifies consumers and makes those classifications consequential (Barocas & Selbst, 2016; Susser et al., 2019).

Toward a Unified Conceptual Framework

A unified account of privacy in AI-enabled consumer analytics becomes possible when the analysis is organized as a causal chain rather than as a checklist of compliance items. The chain links *data treatment* (what the system can observe and infer) to *profiling intensity* (how strongly those inferences govern consumer treatment), to *privacy impact* (what kinds of harms occur and how severe they are), with *governance controls* operating as moderators that can weaken, or unintentionally strengthen, each link.

Figure 1. Unified conceptual framework linking AI-driven data treatment, consumer profiling, governance mechanisms, and privacy impacts



Source: Developed by the Authors.

Figure 1 summarizes the proposed conceptual framework. It illustrates how the structural properties of AI-driven data treatment shape consumer profiling characteristics, which in turn influence the severity and nature of privacy harms, while governance mechanisms moderate these relationships across the entire process.

At the upstream end of the framework, the structural properties of AI-driven data treatment – particularly linkability, inferential depth, persistence, and operational coupling – shape the informational capacity available for consumer profiling (Chierici et al., 2019; Churchill, 2016; Lyon, 2014). These upstream characteristics influence the extent to which consumer profiles become granular, actionable, portable, stable, and contestable, thereby determining the likelihood and severity of downstream privacy harms.

These treatment properties feed into profiling intensity, which describes how forcefully treated data are converted into operational classifications. Profiling intensity is captured by granularity (broad segment vs individual scoring), profile stability (temporal continuity of consumer classifications), automation (degree of human discretion/oversight), and actionability (how tightly profiles drive differential outcomes) (Battaglini & Rasmussen, 2019; Karagiannis et al., 2007). Actionability is the key "severity amplifier": a profile that merely changes content ordering is qualitatively different from a profile that triggers eligibility gating, price steering, or differential service. This is where privacy becomes inseparable from governance because the harms depend on how classification is embedded in decisions and whether those decisions are explainable and contestable.

Downstream, privacy impact can be theorized as "*harm type × severity*". Harm types include informational harms (unexpected secondary use, loss of control, comprehensive portrait-building), autonomy harms (manipulative steering, constrained informational environments), distributive/fairness harms (exclusion, disadvantage, proxy discrimination), and security harms (aggregation and breach amplification) (Mann & Matzner, 2019). Severity increases when impacts are (i) difficult to detect from the consumer's vantage point, (ii) hard to reverse once profiles stabilize, and (iii) consequential for access, price, visibility, or dignity. The framework also anticipates that harms can be systematically unequal even

without explicit sensitive attributes, because profiling often operationalizes proxies that reproduce disparate outcomes (Barocas & Selbst, 2016).

Governance controls operate throughout the chain rather than at a single point. Technical controls (data minimization, privacy-enhancing technologies, access controls, audit logs), organizational controls (purpose limitation in product governance, retention rules, model-risk management, independent review), and legal controls (rights, transparency duties, constraints on profiling and secondary use) can reduce risk, but only when matched to the actual mechanism producing harm. For instance, consent alone is structurally weak against harms produced by inference depth and persistence, whereas contestability and decision transparency are more directly targeted at profiling-driven impacts (Bouchareb, 2026; Wachter et al., 2017). Likewise, limits on cross-context linkage address the context-mixing mechanism emphasized by contextual integrity (Nissenbaum, 2004).

The effectiveness of governance controls should also be interpreted within their cultural and institutional context. While the structural mechanisms of the framework remain constant, consumer expectations regarding privacy, transparency, automated decision-making, and organizational accountability differ across societies. Similarly, regulatory environments vary in the rights they grant individuals, the obligations imposed on organizations, and the enforcement of profiling restrictions. These differences influence how governance mechanisms moderate the relationship between data treatment, profiling intensity, and privacy harms. Effective governance therefore depends not only on technical safeguards but also on transparent organizational practices and accountability mechanisms that foster trust in AI-enabled decision environments (Bouchareb & Moussaoui, 2026). Consequently, the proposed framework is intended to provide a common conceptual structure while allowing its governance dimension to adapt to diverse cultural and regulatory settings.

Testable propositions

The framework yields propositions that can be evaluated across sectors and platform types:

- **P1 (Inference depth).** Greater **inference depth** increases privacy impact even when raw inputs are nominally non-sensitive, because treatment can produce sensitive attributes downstream (Acquisti et al., 2015).
- **P2 (Context mixing).** Greater **linkability** and cross-context integration increases perceived privacy violation and informational harms by disrupting context-appropriate information flows (Nissenbaum, 2004, 2009; Susser et al., 2019).
- **P3 (Stability).** Higher profile **stability** increases autonomy and fairness harms by making historical traces consequential and by stabilizing classifications that are difficult to escape.
- **P4 (Actionability).** The effect of profiling on harm severity is stronger when **actionability** is high (e.g., gating, price steering, differential service) than when profiling is used only for low-stakes personalization.
- **P5 (Automation × contestability).** Higher **automation** increases harm severity when transparency and contestability are weak; strong contestability attenuates this effect (Wachter et al., 2017).
- **P6 (Governance moderation).** Multi-layer governance attenuates the effect of profiling intensity on privacy impact; in the absence of such governance, profiling intensity becomes a primary driver of disparate outcomes (Barocas & Selbst, 2016).

Implications for future research

Three empirical directions follow directly from the framework. First, develop measurement constructs for inference depth, context mixing, profile stability, and actionability that are comparable across firms and sectors (moving beyond "data collected" as the primary metric). Second, test causal pathways using multi-method designs: survey/conjoint experiments for perceived violation under controlled variations of scope/linkability; audit studies and impact assessments to detect differential treatment and proxy-based disparity; and longitudinal designs to capture persistence and feedback loops. Third, evaluate governance as a moderator through comparative and quasi-experimental approaches (policy changes, platform rule shifts, retention-limit interventions) to estimate which

controls meaningfully reduce harm and under what conditions. A fourth direction is to examine the framework across different cultural and institutional environments. Comparative studies could investigate whether consumer perceptions of profiling intensity, privacy harms, transparency, and contestability differ across countries with distinct privacy norms and regulatory traditions. Such research would help determine which elements of the framework are universal and which are shaped by local cultural expectations and governance practices.

Conclusion

AI-enabled consumer analytics has shifted the privacy problem from isolated moments of disclosure to an integrated regime of data treatment, profiling, and intervention. Privacy is increasingly shaped not by whether data are collected in the first place, but by how traces are transformed, linked, abstracted, and converted into profiles that govern what consumers see, pay, and can access. This paper argues that the core analytical challenge is therefore structural: privacy outcomes emerge from a chain that begins with upstream treatment practices (scope, linkability, inference depth), intensifies through downstream profiling properties (granularity, profile stability, automation, actionability), and materializes as heterogeneous harms whose severity depends on opacity, contestability, and the consequential nature of differential treatment.

By synthesizing these elements into a unified conceptual framework, the paper contributes a parsimonious vocabulary for describing why privacy risks can remain high even under familiar compliance narratives. In particular, the framework highlights three dynamics that are often underestimated in fragmented debates. First, inference is a privacy event: systems can produce sensitive attributes from non-sensitive inputs and operationalize them without explicit disclosure. Second, context mixing is a risk mechanism: cross-domain linkage transforms socially situated information into portable assets that undermine reasonable expectations of contextual integrity. Third, profiling becomes consequential through actionability and stability: when profiles are durable and

tightly coupled to decisions, privacy harms extend beyond informational control to autonomy, fairness, and security.

The framework also clarifies practical implications for governance. Effective privacy protection cannot be reduced to transparency statements or individual choice mechanisms; it requires controls that are proportionate to the mechanisms producing harm. Firms should treat inference depth and profile actionability as first-order risk variables, not as technical details, and should build constraints on persistence, cross-context linkage, and secondary use into product and model governance. Regulators should prioritize enforceable requirements that strengthen transparency and contestability for consequential profiling, while limiting high-risk combinations of linkability, persistence, and actionability. For researchers, the framework identifies profiling intensity as a key mediating construct that connects upstream data practices to downstream impacts and thus helps structure empirical work across disciplines.

Limitations. As a theoretical contribution, the framework necessarily abstracts from sector-specific institutional constraints and heterogeneous data environments. Risk mechanisms and appropriate controls may vary across domains such as advertising, retail, finance, health-adjacent consumer apps, and public services, where stakes, incentives, and regulatory baselines differ. In addition, privacy expectations are not culturally invariant: cross-cultural norms about sensitivity, acceptability of profiling, and trust in institutions can shape both perceived violation and the legitimacy of governance interventions. These considerations caution against treating the framework as a one-size-fits-all diagnostic and motivate comparative studies that adapt constructs to sectoral contexts and diverse normative settings.

The paper's propositions translate the framework into an empirical program: they invite future research to measure inference depth and context mixing, to test how profile stability and actionability amplify autonomy and fairness harms, and to evaluate which governance interventions measurably reduce impact across sectors and platform ecosystems. As AI systems increasingly mediate consumer life, the fundamental privacy question becomes one of legitimacy: under what conditions is it acceptable for organizations to continuously infer,

classify, and intervene, and what institutional safeguards are necessary to ensure that consumer understanding does not quietly become consumer control.

Declaration of competing interest

The author(s) declare no competing interests.

Declaration Of Generative AI and AI-Assisted Technologies In The Writing Process

During the preparation of this work, the authors used an LLMs tool in order to improve the language and readability of the article. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Data Availability Statement

Not Applicable.

Ethical Approval

Not Applicable.

Financial Disclosure statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465.

Ban, G.-Y., & Keskin, N. B. (2021). Personalized Dynamic Pricing with Machine Learning: High-Dimensional Features and Heterogeneous Elasticity. *Management Science*, *67*(9), 5549–5568. https://doi.org/10.1287/mnsc.2020.3680.

Barbosa, N. M., Wang, G., Ur, B., & Wang, Y. (2021). Who Am I?: A Design Probe Exploring Real-Time Transparency about Online and Offline User Profiling Underlying Targeted Ads. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(3), 1–32. https://doi.org/10.1145/3478122.

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, *104*(3), 671–732. https://doi.org/10.15779/Z38BG31.

Bastarache, L., Brown, J. S., Cimino, J. J., Dorr, D. A., Embi, P. J., Payne, P. R. O., Wilcox, A. B., & Weiner, M. G. (2022). Developing real-world evidence from real-world data: Transforming raw data into analytical datasets. *Learning Health Systems*, *6*(1), e10293. https://doi.org/10.1002/lrh2.10293.

Battaglini, M., & Rasmussen, S. (2019). Transparency, automated decision-making processes and personal profiling. *Journal of Data Protection & Privacy*, *2*(4), 331–349. https://doi.org/10.69554/STNB6990.

Bindschaedler, V., Grubbs, P., Cash, D., Ristenpart, T., & Shmatikov, V. (2018). The tao of inference in privacy-protected databases. *Proceedings of the VLDB Endowment*, *11*(11), 1715–1728. https://doi.org/10.14778/3236187.3236217.

Bouchareb, N. (2026). When algorithms sell: rethinking consumer behavior in the AI-enhanced marketplace. *AI & SOCIETY*, *41*, 469–470. https://doi.org/10.1007/s00146-025--02453-0.

Bouchareb, N., & Moussaoui, I. (2026). AI-driven scientific advertising: ethics, visibility, and research integrity. *Telematics and Informatics Reports*, *22*, 100314. https://doi.org/10.1016/j.teler.2026.100314.

Büchi, M., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., Velidi, S., & Viljoen, S. (2020). The chilling effects of algorithmic profiling: Mapping the issues. *Computer Law & Security Review*, *36*, 105367. https://doi.org/10.1016/j.clsr.2019.105367.

Bujlow, T., Carela-Español, V., Solé-Pareta, J., & Barlet-Ros, P. (2017). A Survey on Web Tracking: Mechanisms, Implications, and Defenses. *Proceedings of the IEEE*, *105*(8), 1476–1510. https://doi.org/10.1109/JPROC.2016.2637878.

Cheney-Lippold, J. (2011). A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society*, *28*(6), 164–181. https://doi.org/10.1177/0263276411424420.

Chierici, R., Mazzucchelli, A., Garcia-Perez, A., & Vrontis, D. (2019). Transforming big data into knowledge: the role of knowledge management practice. *Management Decision*, *57*(8), 1902–1922. https://doi.org/10.1108/MD-07-2018-0834.

Churchill, E. F. (2016). Designing data practices. *Interactions*, *23*(5), 20–21. https://doi.org/10.1145/2983401.

Custers, B. (2018). Profiling As Inferred Data. Amplifier Effects and Positive Feedback Loops. In: E. Bayamlıoğlu, I. Baraluic, L. Janssens & M. Hildebrandt (Eds.), BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen (pp. 112–115), Amsterdam University Press. https://doi.org/10.2139/ssrn.3466857.

de Laat, P. B. (2018). Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability? *Philosophy & Technology*, *31*(4), 525–541. https://doi.org/10.1007/s13347-017-0293-z.

Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions. *IEEE Access*, *7*, 144907–144924. https://doi.org/10.1109/ACCESS.2019.2944243.

El Mokhtari, J., El Kalam, A. A., Benhaddou, S., & Leroy, J.-P. (2021). Coupling of Inference and Access Controls to Ensure Privacy Protection. *International Journal of Safety and Security Engineering*, *11*(5), 529–535. https://doi.org/10.18280/ijsse.110504.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001.

Fernández-Rovira, C., Álvarez Valdés, J., Molleví, G., & Nicolas-Sans, R. (2021). The digital transformation of business. Towards the datafication of the relationship with customers. *Technological Forecasting and Social Change*, *162*, 120339. https://doi.org/10.1016/j.techfore.2020.120339.

Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, *12*(3), 347–364. https://doi.org/10.1177/1461444809342738.

Gillespie, T. (2014). The Relevance of Algorithms. In: *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0009.

Giovanelli, J., Bilalli, B., & Abelló, A. (2022). Data pre-processing pipeline generation for AutoETL. *Information Systems*, *108*, 101957. https://doi.org/10.1016/j.is.2021.101957.

Gupta, S., Leszkiewicz, A., Kumar, V., Bijmolt, T., & Potapov, D. (2020). Digital Analytics: Modeling for Insights and New Methods. *Journal of Interactive Marketing*, *51*(1), 26–43. https://doi.org/10.1016/j.intmar.2020.04.003.

Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology*, *51*(4), 605–622. https://doi.org/10.1080/00071310020015280.

Hajli, N., Tajvidi, M., Gbadamosi, A., & Nadeem, W. (2020). Understanding market agility for new product success with big data analytics. *Industrial Marketing Management*, *86*, 135–143. https://doi.org/10.1016/j.indmarman.2019.09.010.

Heiberg, J., Truffer, B., & Binz, C. (2022). Assessing transitions through socio-technical configuration analysis – a methodological framework and a case study in the water sector. *Research Policy*, *51*(1), 104363. https://doi.org/10.1016/j.respol.2021.104363.

Jha, A., Bhatia, A., & Tiwari, K. (2025). Bayesian Deep Learning Meets Self-Attention: A Risk-Aware Approach to Advertisement Optimization. *IEEE Access*, *13*, 86413–86429. https://doi.org/10.1109/ACCESS.2025.3570537.

Karagiannis, T., Papagiannaki, K., Taft, N., & Faloutsos, M. (2007). Profiling the End Host. In: *Passive and Active Network Measurement* (pp. 186–196). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71617-4_19.

Khattach, O., Moussaoui, O., & Hassine, M. (2025). End-to-End Architecture for Real-Time IoT Analytics and Predictive Maintenance Using Stream Processing and ML Pipelines. *Sensors*, *25*(9), 2945. https://doi.org/10.3390/s25092945.

Kim, D., Lee, D. H., & Hwang, M. K. (2025). A Comprehensive Profiling System Integrating Myers-Briggs Type Indicator (MBTI) and Dominance, Influence, Steadiness, and Conscientiousness (DISC) for Personalized Health Training: Correlational Analysis and Usability Evaluation. *JMIR Human Factors*, *12*, e73397–e73397. https://doi.org/10.2196/73397.

King, N. J., & Forder, J. (2016). Data analytics and consumer profiling: Finding appropriate privacy principles for discovered data. *Computer Law & Security Review*, *32*(5), 696–714. https://doi.org/10.1016/j.clsr.2016.05.002.

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. *Journal of Management Information Systems*, *35*(2), 540–574. https://doi.org/10.1080/07421222.2018.1451957.

Kodi, D. (2024). Designing Real-time Data Pipelines for Predictive Analytics in Large-scale Systems. *FMDB Transactions on Sustainable Computing Systems*, *2*(4), 178–188. https://doi.org/10.69888/FTSCS.2024.000294.

Lupton, D. (2020). Thinking with care about personal data profiling: A more-than-human approach. *International Journal of Communication*, *14*, 3165–3183. https://ijoc.org/index.php/ijoc/article/view/13540.

Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, *1*(2). https://doi.org/10.1177/2053951714541861.

Mann, M., & Matzner, T. (2019). Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data & Society*, *6*(2), 205395171989580. https://doi.org/10.1177/2053951719895805.

Mavriki, P., & Karyda, M. (2019). Automated data-driven profiling: threats for group privacy. *Information & Computer Security*, *28*(2), 183–197. https://doi.org/10.1108/ICS-04-2019-0048.

Mustafa, M. A., Cleemput, S., Aly, A., & Abidin, A. (2019). A Secure and Privacy-Preserving Protocol for Smart Metering Operational Data Collection. *IEEE Transactions on Smart Grid*, *10*(6), 6481–6490. https://doi.org/10.1109/TSG.2019.2906016.

Naphade, D. (2025). The Evolution and Modernization of Data Pipeline Architectures. *European Journal of Computer Science and Information Technology*, *13*(6), 42–53. https://doi.org/10.37745/ejcsit.2013/vol13n64253.

Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. *2008 IEEE Symposium on Security and Privacy (Sp 2008)*, 111–125. https://doi.org/10.1109/SP.2008.33.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*(1), 119–158.

Nissenbaum, H. (2009). *Privacy in Context*. Stanford University Press. https://doi.org/10.1515/9780804772891.

Quach, S., Thaichon, P., Martin, K. D., Weaven, S., & Palmatier, R. W. (2022). Digital technologies: tensions in privacy and data. *Journal of the Academy of Marketing Science*, *50*(6), 1299–1323. https://doi.org/10.1007/s11747-022-00845-y.

Reed, A., & Forehand, M. R. (2016). The ebb and flow of consumer identities: the role of memory, emotions and threats. *Current Opinion in Psychology*, *10*, 94-100. https://doi.org/10.1016/j.copsyc.2015.12.015.

Sakalauskas, V., & Kriksciuniene, D. (2024). Personalized Advertising in E-Commerce: Using Clickstream Data to Target High-Value Customers. *Algorithms*, *17*(1), 27. https://doi.org/10.3390/a17010027.

Salem, A., Cherubin, G., Evans, D., Köpf, B., Paverd, A., Suri, A., Tople, S., & Zanella-Béguelin, S. (2023). SoK: Let the Privacy Games Begin! A Unified Treatment of Data Inference Privacy in Machine Learning. *2023 IEEE Symposium on Security and Privacy (SP)*, 327–345. https://doi.org/10.1109/SP46215.2023.10179281.

Schweidel, D. A., Bart, Y., Inman, J. J., Stephen, A. T., Libai, B., Andrews, M., Rosario, A. B., Chae, I., Chen, Z., Kupor, D., Longoni, C., & Thomaz, F. (2022). How consumer digital signals are reshaping the customer journey. *Journal of the Academy of Marketing Science*, *50*(6), 1257–1276. https://doi.org/10.1007/s11747-022-00839-w.

Shah, K. (2025). Real-Time Analytics in E-commerce: Strategies for Implementing Near Real-Time ETL Pipelines. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *11*(2), 2654–2675. https://doi.org/10.32628/CSEIT25112746.

Sivakolundhu, R. (2024). From Raw Data to Actionable Insights: How Data Pipelines Enable Effective Data Observability for Informed Decision Making. *International Journal of Science and Research (IJSR)*, *13*(7), 273–276. https://doi.org/10.21275/SR24705081043.

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, *154*(3), 477–564. https://doi.org/10.2307/40041279.

Susser, D. (2019). Invisible Influence: Artificial Intelligence and the Ethics of Adaptive Choice Architectures. In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 403–408. https://doi.org/10.1145/3306618.3314286.

Susser, D., Roessler, B., & Nissenbaum, H. F. (2019). Online Manipulation: Hidden Influences in a Digital World. *4 Georgetown Law Technology Review*, *1*, 1–45. https://doi.org/10.2139/ssrn.3306006.

Syrjälä, H., Ruiz, C. D., Leipämaa-Leskinen, H., & Luomala, H. T. (2025). From consumers to consumption: The socio-technical assemblage of the persona in market segmentation. *Journal of Business Research*, *194*, 115387. https://doi.org/10.1016/j.jbusres.2025.115387.

Trusov, M., Ma, L., & Jamal, Z. (2016). Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting. *Marketing Science*, *35*(3), 405–426. https://doi.org/10.1287/mksc.2015.0956.

Vaassen, B. (2022). AI, Opacity, and Personal Autonomy. *Philosophy & Technology*, *35*(4), 88. https://doi.org/10.1007/s13347-022-00577-5.

Valletti, T., & Wu, J. (2020). Consumer Profiling with Data Requirements: Structure and Policy Implications. *Production and Operations Management*, *29*(2), 309–329. https://doi.org/10.1111/poms.13108.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776.

Vespignani, A. (2012). Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, *8*(1), 32–39. https://doi.org/10.1038/nphys2160.

Wachter, S. (2022). The Theory of Artificial Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law. *Tulane Law Review*, *97*, 1–50. https://dx.doi.org/10.2139/ssrn.4099100.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005.

Wang, L., Xu, L. Da, Bi, Z., & Xu, Y. (2014). Data Cleaning for RFID and WSN Integration. *IEEE Transactions on Industrial Informatics*, *10*(1), 408–418. https://doi.org/10.1109/TII.2013.2250510.

Xu, S., Lu, B., Baldea, M., Edgar, T. F., Wojsznis, W., Blevins, T., & Nixon, M. (2015). Data cleaning in the process industries. *Reviews in Chemical Engineering*, *31*(5), 453–490. https://doi.org/10.1515/revce-2015-0022.

Xu, Y., Wang, D., Liu, X., & Pan, Y. (2025). Research on user profiling based on Apriori algorithm for power marketing big data. In: W. Pedrycz & S.S. Agaian (Eds.), *Fifth International Conference on Telecommunications, Optics, and Computer Science (TOCS 2024)* (p. 51). SPIE, *13629*. https://doi.org/10.1117/12.3067611.

Zheng, X., Cai, Z., & Li, Y. (2018). Data Linkage in Smart Internet of Things Systems: A Consideration from a Privacy Perspective. *IEEE Communications Magazine*, *56*(9), 55–61. https://doi.org/10.1109/MCOM.2018.1701245.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* PublicAffairs.