Chapter 5
The Trust–Loyalty Nexus in AI-Mediated Hyper-Personalized Services:
Transparency, Authenticity, and Vulnerability

Azize Şahin
http://orcid.org/0000-0002-3115-6812
İstanbul University, Turkey

Aysun Sahin
http://orcid.org/0000-0003-3285-1232
Duzce University, Turkey

ABSTRACT

AI-driven hyper-personalization is transforming service encounters by tailoring information, offers, and interactions to individual consumers in real time. Across retail, banking, healthcare, and hospitality, providers increasingly rely on predictive models, recommender systems, automated decision tools, and conversational interfaces to adapt service content at the level of the single customer. Yet the rapid expansion of algorithmic personalization has not been matched by a comparable growth in consumer trust, and this widening gap raises a central question for service marketing: when does AI-enabled personalization deepen durable loyalty, and when does it produce shallow, fragile, or actively eroding relationships?

DOI: 10.4018/979-8-3373-9159-5.ch005

Copyright © 2027, IGI Global Scientific Publishing. Copying or distributing in print or electronic forms without written permission of IGI Global Scientific Publishing is prohibited. Use of this chapter to train generative artificial intelligence (AI) technologies is expressly prohibited. The publisher reserves all rights to license its use for generative AI training and machine learning model development.

1. INTRODUCTION

Artificial intelligence has moved from experimental service innovation to the core infrastructure of customer experience management. IBM's Global AI Adoption Index reported that 42% of enterprise-scale companies had already deployed AI, while another 40% were still exploring or experimenting with AI without full deployment (IBM, 2024). More recent evidence from McKinsey shows that organizations are also moving toward agentic AI systems: 23% of surveyed organizations were scaling AI agents in at least one business function, and 39% were experimenting with them (McKinsey & Company, 2025). These figures indicate that AI-mediated service encounters are becoming routine rather than exceptional. For service researchers, this shift is theoretically consequential because AI does not merely accelerate existing service processes; it restructures how information is collected, options are curated, recommendations are justified, and consumers decide whether a provider deserves continued trust.

However, the expansion of AI-mediated personalization has not been matched by equivalent growth in consumer trust. Salesforce's State of the AI Connected Customer report, based on more than 16,000 consumers and business buyers, shows that only 42% of customers trust businesses to use AI ethically, with the figure declining from 58% in 2023 (Salesforce, 2024). This decline is especially important for hyper-personalized services because consumers are asked to exchange data, attention, and decision autonomy for relevance, convenience, and adaptive service experiences. The central managerial problem is therefore not whether AI can personalize service encounters, but whether such personalization is perceived as competent, transparent, authentic, and respectful of consumer vulnerability. This trust gap provides the conceptual starting point for the present study: AI-driven hyper-personalization can strengthen loyalty only when consumers interpret algorithmic adaptation as legitimate, explainable, and aligned with their interests, rather than as opaque surveillance or relational manipulation.

Artificial intelligence–driven hyper-personalization is reshaping the architecture of contemporary service encounters. Service organizations are increasingly using predictive models, recommender systems, automated decision tools, conversational interfaces, and data-rich analytics to tailor service content, offers, and interactions at the level of the individual consumer. Across retail, banking, healthcare, and hospitality, such systems infer preferences, combine behavioral and financial data, and tailor information, offers, and service pathways to the individual—applications examined in detail in Section 6. These developments indicate a shift from standardized service propositions to dynamically adaptive, algorithmically mediated service ecosystems (Huang & Rust, 2018; Wedel & Kannan, 2016; Wirtz et al., 2018).

Within these service environments, consumer trust becomes a central mechanism linking AI-enabled personalization to satisfaction, continued use, and loyalty. Trust has long been recognized as a core antecedent of relational exchange and digital service adoption (Gefen et al., 2003; McKnight et al., 2002). Mayer et al. (1995) define trust as the willingness to accept vulnerability based on positive expectations about another party. This definition is particularly relevant for AI-mediated services because consumers often cannot fully monitor how data are collected, how recommendations are generated, or how algorithmic decisions affect service outcomes. Hyper-personalization increases the need for trust.

This chapter is motivated by three limitations of the existing literature. The first limitation concerns the dimensionality of trust. Classical trust theory distinguishes competence-related beliefs from benevolence- and integrity-related beliefs (Mayer et al., 1995; Schoorman et al., 2007). In service settings, these dimensions map onto cognitive trust and affective or relational trust. AI-driven personalization may strengthen cognitive trust when it delivers accurate and timely recommendations; however, it may weaken affective trust when personalization feels intrusive, manipulative, or misaligned with consumers' interests. Therefore, treating trust as a single construct can conceal theoretically important asymmetries in consumer responses.

The second limitation concerns the transparency of algorithms. Research on explainable AI often conceptualizes transparency as an informational intervention that helps users understand system logic and calibrate trust to performance (Rai, 2020). However, in consumer services, transparency is a market signal. Consumers rarely possess the technical expertise required to evaluate an underlying model. Instead, they interpret explanations, data use disclosures, and opt-out controls as signs of organizational openness, fairness, and respect for autonomy. Therefore, signaling theory is useful because it explains how observable cues reduce uncertainty under information asymmetry (Connelly et al., 2011; Spence, 1973).

Third, the role of consumer vulnerability remains underdeveloped. Vulnerability is not a fixed demographic label. It is a contextually activated condition that arises when consumers face power imbalances, resource constraints, limited knowledge, situational stress, or dependency in marketplace interactions (Baker et al., 2005; OECD, 2023). Hyper-personalized AI systems can intensify consumer vulnerability by making inferences that consumers do not expect, cannot contest, or understand. In high-stakes sectors such as healthcare and financial services, vulnerability makes trust more fragile and raises ethical thresholds for personalization.

This chapter develops an integrative conceptual framework for understanding the trust–loyalty nexus in AI-mediated hyper-personalized service. It combines cognitive–affective trust theory, signaling theory, and consumer vulnerability literature to explain when AI-enabled personalization strengthens durable loyalty and when it creates shallow or eroding relationships with consumers. It contributes to

service marketing by treating trust as a dynamic relational resource that is actively shaped by the design and governance of AI-mediated personalization. It also introduces a trust trajectory typology that distinguishes trust-building, trust-neutral, and trust-eroding paths over time. The core argument can be stated in three linked claims. First, AI-driven hyper-personalization initially generates cognitive trust by demonstrating relevance, accuracy, and responsiveness. Second, whether this cognitive trust converts into affective trust depends on algorithmic transparency and perceived authenticity, which together signal benevolence and integrity. Third, consumer vulnerability sets the threshold condition for this conversion and amplifies the risk of trust erosion when personalization is opaque, intrusive, or misaligned with consumer interests. This conditional conversion logic—from cognitive to affective trust, gated by transparency, authenticity, and vulnerability—is the chapter's central theoretical contribution and the mechanism that the trust trajectory typology operationalizes.

This chapter is conceptual rather than empirical; its purpose is to synthesize fragmented literatures on trust, transparency, authenticity, and vulnerability into a single integrative framework. The propositions developed below are therefore offered as scaffolding for future empirical work rather than as a formally tested model. This chapter proceeds as follows. Section 2 reviews consumer trust in AI-mediated services. Section 3 develops the theoretical foundations, including cognitive–affective trust theory, signaling theory, consumer vulnerability, and perceived authenticity. Section 4 presents the integrative conceptual framework, first specifying its core processes (4.1) and then the boundary conditions that qualify the trust–loyalty nexus (4.2). Section 5 introduces the trust trajectory typology. Section 6 provides sectoral illustrations in retail, banking, healthcare, and hospitality. Section 7 outlines the research agenda. Section 8 presents managerial and ethical implications. Section 9 concludes.

2. CONSUMER TRUST IN AI-MEDIATED SERVICES

Consumer trust in AI-mediated services has become a central construct for explaining adoption, continued use, and loyalty across contemporary service industries. Trust is particularly salient because AI systems are typically opaque to lay users. Consumers cannot directly observe how algorithms process data, prioritize alternatives, generate recommendations, or align service outputs with organizational incentives. They must therefore rely on indirect cues when deciding whether to accept automated recommendations, disclose personal data, use algorithmic advice, or allow an AI-enabled interface to structure their decision processes. In this context, trust operates as a complexity-reducing mechanism that enables consumers to act under

conditions of limited technical understanding, asymmetric information, and uncertain accountability (Gefen et al., 2003; Mayer et al., 1995; McKnight et al., 2002).

In AI-mediated services, trust has multiple objects. Consumers may trust the AI system, service provider, human professionals supervising the system, platform infrastructure, or institutional safeguards that govern data use. These trust layers are interdependent rather than isolated. Trust in the provider can transfer to the AI application when the organization is perceived as competent, benevolent, and ethically responsible. Conversely, a poorly explained, inaccurate, or unfair AI experience can feed back into distrust toward the organization itself. This layered trust architecture is central to hyper-personalized services because AI systems are not experienced as detached technical artifacts. They are interpreted as extensions of the provider's service logic, data governance practices, and relational intentions (Flavián et al., 2022; Gefen et al., 2003; McKnight et al., 2002).

Existing research on AI in services shows that perceived reliability, technology readiness, awareness, perceived usefulness, and service context influence consumers' willingness to adopt and use AI-enabled services (Flavián et al., 2022; Huang & Rust, 2018). However, adoption alone does not adequately explain customer loyalty. Consumers may initially use AI-mediated services because they are convenient, novel, embedded in existing platforms, or difficult to avoid. Long-term loyalty, however, requires a more durable perception that the provider uses AI in ways that are accurate, fair, beneficial, and respectful of consumer interests. Therefore, the trust–loyalty nexus depends not only on functional performance but also on consumers' relational interpretation of how and why AI is used.

The personalization–privacy paradox illustrates this ambivalence. Personalization can increase relevance, reduce search costs, and improve decision efficiency; however, it can also heighten concerns about information collection, surveillance, profiling, and unfair use of personal data (Aguirre et al., 2015; Awad & Krishnan, 2006; Bleier & Eisenbeiss, 2015; Cloarec et al., 2022). This paradox becomes sharper under hyper-personalization because AI systems infer not only stated preferences but also latent interests, behavioral patterns, habits, vulnerabilities, and propensities. What appears to the firm as improved customer understanding may be experienced by the consumer as an intrusion into private life. Thus, the same personalization capability can strengthen cognitive trust by increasing perceived accuracy while weakening affective trust if consumers question the provider's motives or respect for boundaries.

Recommendation agents provide an early illustration of this tension. Personalization and familiarity can increase trust in recommendation agents; however, such trust depends on how consumers interpret the agent's role, competence, and intentions (Komiak & Benbasat, 2006). Contemporary AI systems extend this problem because they are more adaptive, data-intensive, and deeply embedded in the

service journey. Consumers may revise their trust judgments as they encounter new algorithmic actions, explanation cues, personalization errors, perceived boundary violations, or recovery efforts. Therefore, trust in AI-mediated services should be treated as a dynamic and revisable judgment rather than a stable attitude formed at the moment of adoption.

Healthcare further illustrates the fragility of trust in AI-mediated services. Patients may resist medical AI, even when algorithmic systems are statistically accurate, because healthcare decisions are deeply personal, emotionally charged, and tied to expectations of human care (Longoni et al., 2019). Trust in healthcare AI also depends on implementation processes, perceived human oversight, explainability, and the alignment between AI recommendations and professional judgment (Steerling et al., 2023). These insights are relevant beyond healthcare because they show that technical competence alone rarely guarantees relational acceptance. In high-stakes services, consumers expect organizations to remain accountable for AI-mediated outcomes, especially when the consequences affect health, finances, identity, or personal autonomy.

Service robot research offers a similar insight from frontline service contexts. Automation changes service experience by altering expectations about social presence, empathy, accountability, and recovery (van Doorn et al., 2017; Wirtz et al., 2018). Although service robots differ from back-end personalization algorithms, both forms of AI mediate consumer experience and reconfigure the locus of agency in service delivery. The common implication is that trust must be theorized not only as belief in technological capability but also as belief in the provider's relational and ethical stance.

3. THEORETICAL FOUNDATIONS

3.1. Cognitive and Affective Trust Theory

Cognitive–affective trust theory distinguishes between competence-based beliefs and relational beliefs. Mayer et al. (1995) conceptualize trust as rooted in perceptions of ability, benevolence, and integrity. McKnight et al. (2002) further distinguish trusting beliefs, institution-based trust, and trusting intentions in e-commerce. Cognitive trust refers to beliefs about whether the trustee can perform reliably and competently. Affective or relational trust refers to beliefs about whether the trustee is benevolent, honest, and concerned with the trustor's interests.

This distinction is essential for AI-mediated services. Cognitive trust concerns whether the AI system works, whether recommendations are accurate, whether service predictions are useful, whether automated decisions are consistent, and whether the

interface reduces effort. Affective trust concerns whether the organization behind the AI acts with care and integrity, whether data use respects boundaries, whether personalization is fair, whether vulnerable consumers are protected, and whether human support remains available when algorithmic decisions matter.

Hyper-personalization can create trust asymmetry. A highly accurate recommender may increase cognitive trust while reducing affective trust if consumers feel that the system knows too much or uses sensitive inferences without permission. A healthcare AI tool may appear technically sophisticated while failing to convey compassion or human accountability. A banking algorithm may correctly identify financial needs while simultaneously raising concerns about exploitation if it targets high-cost products to financially fragile consumers. The key theoretical point is that competence and benevolence do not necessarily move together.

This asymmetry has direct loyalty implications. Cognitive trust may support initial adoption and routine use, especially in low-risk service contexts. Affective trust is more important for durable loyalty because it anchors commitment, forgiveness, and resistance to switching. If AI-mediated personalization produces cognitive trust without affective trust, loyalty may remain shallow and convenience-based. However, if it builds both, loyalty becomes more resilient. Recent empirical work supports this two-dimensional view and extends it beyond Western samples. In a study of 632 Chinese consumers, Zhao et al. (2025) found that cognitive and affective trust jointly form overall trust in AI agents, which in turn drives acceptance and ethical expectations. Because trust in AI is culturally contingent—shaped by power distance, uncertainty avoidance, and prevailing attitudes toward automation—evidence from non-Western contexts is important for assessing how widely the cognitive–affective structure travels, a point developed further in the research agenda.

3.2. Signaling Theory and Algorithmic Transparency

Signaling theory explains how actors reduce uncertainty when one party holds information that the other party cannot directly observe. Spence (1973) originally developed the logic of signaling under information asymmetry, and Connelly et al. (2011) extended signaling theory to management contexts. In AI-mediated services, providers and their technical teams know much more than consumers about model design, training data, inference logic, and commercial objectives. Consumers must therefore interpret signals that indicate whether the provider is trustworthy.

Algorithmic transparency operates as such a signal. Transparency may include data use notices, explanation interfaces, reason codes, model limitations, privacy dashboards, human review options, and third-party audit indicators. These signals do not require consumers to be technical experts. Their relational function is to communicate openness, accountability, and respect for consumer agency. Rai (2020)

argues that explainable AI moves systems from black-box to glass-box conditions; however, the consumer-facing meaning of explanation is not only technical. It is also symbolic and relational. This distinction marks a difference between how transparency is treated in computer science and how it operates in consumer research. In the explainable AI literature, transparency is largely a technical property of the model, achieved through interpretability methods and post-hoc explanations that expose how an output was produced (Mersha et al., 2024). In consumer services, by contrast, transparency is a relational signal whose meaning depends on whether disclosures are perceived as credible, usable, and sincere rather than merely available. Park and Yoon (2024) show empirically that algorithm transparency signaling increases user trust and relational satisfaction, indicating that the value of explanation lies as much in its signaling function as in its technical content. The framework advanced here bridges these views: it treats the technical interpretability emphasized in computer science as the raw material that, once interpreted by consumers, becomes a relational signal of accountability.

Transparency signals are credible when they are specific, usable, and consistent with observable practice. Generic privacy statements or complex legal disclosures may fail because they signal compliance rather than care. In contrast, clear explanations of why a recommendation appears, which data categories were used, how consumers can modify preferences, and how they can contest an AI-mediated decision can support both cognitive and affective trust. In this sense, transparency is a governance interface between organizational intention and consumer interpretation.

3.3. Consumer Vulnerability

Consumer vulnerability literature adds a crucial boundary condition to the trust–loyalty nexus. Baker et al. (2005) define vulnerability as a state of powerlessness arising from imbalances in marketplace interactions. Vulnerability can be informational, financial, emotional, physical, situational, or structural. The OECD (2023) emphasizes that digital environments can amplify vulnerability because data-driven systems may be difficult for consumers to observe, understand, or contest.

AI-mediated hyper-personalization can intensify vulnerability in three ways. First, it can expand informational asymmetry by allowing firms to infer sensitive traits or conditions that consumers did not intentionally disclose. Second, it can reduce perceived control when consumers cannot understand or contest personalization logic. Third, it can create dependency when consumers rely on a service provider in high-stakes contexts, such as health, finance, mobility, or care. Under these conditions, the same personalization practice may be experienced as helpful by one consumer and exploitative by another.

Therefore, vulnerability moderates the influence of transparency and personalization on trust. Low-vulnerability consumers may tolerate limited transparency if personalization is useful and low-risk. High-vulnerability consumers require stronger safeguards, clearer explanations, more conservative data use, and human support. In the absence of such safeguards, trust can erode, even when the algorithmic system performs efficiently.

Table 1. Theoretical Lenses for the Trust-Loyalty Nexus in AI-Mediated Services

Theoretical Lens	Core Concept	Role in the Chapter	Main Risk if Used Alone
Cognitive and Affective Trust Theory	Trust includes competence, benevolence, and integrity beliefs.	Explains how AI personalization can build technical confidence while weakening relational trust.	May understate information asymmetry and market signaling.
Signaling Theory	Observable signals reduce uncertainty in the presence of information asymmetry.	Explains how transparency, explanations, and controls communicate credibility and accountability.	May overemphasize cues without examining vulnerable consumers.
Consumer Vulnerability Literature	Vulnerability arises from contextual power, information, and resource imbalances.	Explains why the same personalization practice can empower some consumers and threaten others.	May understate the technical and relational complexity of AI systems.
Perceived Authenticity	Authenticity reflects perceived genuineness, consistency, and alignment with values.	Explains why competence alone does not guarantee relational trust; personalization must feel sincere.	May overemphasize subjective perception without structural safeguards.

3.4. Perceived Authenticity as a Relational Trust Cue

Perceived authenticity provides an important relational bridge between hyper-personalization and consumer trust. In AI-mediated services, personalization may signal competence; however, competence alone does not guarantee that consumers will perceive the provider as sincere, benevolent, or worthy of loyalty. A recommendation can be accurate yet still feel exploitative. A chatbot can be responsive yet still feel scripted. A personalized offer can be relevant yet still raise doubts about whether the provider is genuinely serving the consumer's interests. For this reason, perceived authenticity should be treated as a relational trust cue that shapes whether AI-driven hyper-personalization strengthens or weakens affective trust. To avoid conceptual overlap, the chapter draws a clear boundary: perceived authenticity is not synonymous with affective trust, nor is it a sub-dimension of it. Rather, it is an interpretive cue through which consumers infer the benevolence, sincerity, and relational alignment that constitute affective trust. In the framework advanced here,

authenticity functions as a mediating interpretation that links transparency signals to affective trust, whereas affective trust is the resulting relational belief itself.

Brand authenticity refers to the extent to which consumers perceive a brand to be genuine, credible, consistent, and true to its values (Morhart et al., 2015). In branding research, authenticity is linked to perceptions of sincerity, continuity, integrity, and symbolic value (Beverland, 2005; Napoli et al., 2014). These dimensions are especially relevant in AI-mediated service settings because hyper-personalization creates a form of simulated intimacy. AI systems can address consumers by name, remember their prior interactions, infer their preferences, and anticipate their needs. Such behaviors may be interpreted as attentive service when they are consistent with the provider's identity, service promises, and ethical commitments. However, these same behaviors may be interpreted as artificial closeness or manipulative familiarity when consumers perceive a gap between personalized interaction and genuine organizational care.

This distinction is central to the cognitive–affective structure of trust. Hyper-personalization strengthens cognitive trust by demonstrating technical accuracy, predictive capability, and service efficiency. However, affective trust depends on whether consumers believe that the provider is acting with benevolence and integrity (Mayer et al., 1995; McKnight et al., 2002). Perceived authenticity operates as a cue through which consumers infer relational qualities. When personalization is consistent with the brand's stated values, prior conduct, and service role, it reinforces affective trust. For example, a healthcare provider that uses AI to tailor follow-up reminders, explain treatment options, and support patient understanding may be perceived as authentically patient-centered if the personalization is transparent and clinically appropriate. By contrast, the same level of personalization may undermine authenticity if it appears to prioritize cost reduction, cross-selling, or behavioral control.

Perceived authenticity also explains why transparency alone is insufficient. Transparency functions as a governance interface that discloses how personalization works; however, authenticity concerns why personalization is being used and whose interests it serves. A service provider can technically explain an algorithm while still appearing opportunistic if the personalization feels misaligned with consumer welfare. This is particularly important in the personalization–privacy paradox, where consumers value relevance but become uncomfortable when firms infer sensitive information or use personal data in unexpected ways (Aguirre et al., 2015; Awad & Krishnan, 2006; Bleier & Eisenbeiss, 2015). In such contexts, authenticity depends not only on disclosure but also on restraint. Consumers are more likely to perceive personalization as authentic when the provider uses data in ways that are proportionate, contextually appropriate, and visibly connected to service improvement rather than purely for commercial extraction.

Perceived authenticity further clarifies the difference between trust-building and trust-neutral trajectories. In a trust-building trajectory, personalization is not merely useful; it is experienced as a credible expression of the provider's commitment to service. Consumers perceive that AI is being used to support care, convenience, fairness, or empowerment in ways that fit the provider's identity. In a trust-neutral trajectory, personalization may remain functionally effective, but relationally thin. The system works, but it does not generate a stronger sense of sincerity, care, or identification. In a trust-eroding trajectory, personalization may be interpreted as inauthentic because it simulates concern while serving hidden commercial or surveillance goals. This is where hyper-personalization becomes relationally risky: the more intimate the interaction feels, the more damaging perceived inauthenticity becomes.

Consumer vulnerability intensifies the role of authenticity. Vulnerable consumers may be especially sensitive to whether AI-mediated personalization feels supportive or exploitative. In high-stakes contexts such as healthcare, banking, insurance, and elder services, consumers often depend on providers and may lack the ability to evaluate algorithmic decisions fully. For these consumers, authenticity is not a decorative brand attribute; it is a protective trust cue. Personalization that acknowledges vulnerability, preserves dignity, and provides meaningful human oversight can strengthen affective trust. Personalization that targets vulnerability for persuasion, upselling, or behavioral influence can erode trust, even when the immediate recommendation is technically accurate (Baker et al., 2005; Trinh, 2025).

Accordingly, perceived authenticity should be embedded in the framework of this chapter as a relational interpretation mechanism. AI-driven hyper-personalization affects consumer trust not only through relevance, accuracy, and transparency, but also through the perceived authenticity of the provider's motives and conduct. The central question is not whether the AI system personalizes effectively. The stronger theoretical question is whether personalization is experienced as a genuine extension of the provider's service promise or as an algorithmically produced simulation of care. This distinction is critical for understanding why some AI-mediated services deepen loyalty, whereas others generate discomfort, distrust, and resistance.

Because these constructs are closely related, it is useful to fix their boundaries before turning to the framework. In this chapter, cognitive trust refers to beliefs about competence, reliability, and accuracy—whether the AI system works. Affective trust refers to beliefs about benevolence, integrity, and care—whether the provider acts in the consumer's interest; it is the relational core of what some authors term relational trust, and the two terms are used interchangeably here. Perceived authenticity is not a form of trust but an interpretive cue through which consumers infer benevolence and sincerity, thereby shaping affective trust. Transparency is a signal that simultaneously supports cognitive trust, by making personalization predictable,

and affective trust, by communicating accountability. Consumer vulnerability is a contextually activated condition that moderates how strongly these cues translate into trust. Holding these distinctions fixed allows the framework to specify not merely that these constructs matter, but how each one enters the conversion from cognitive to affective trust.

4. INTEGRATIVE CONCEPTUAL FRAMEWORK

4.1. Core Processes in the Integrative Framework

The proposed framework conceptualizes AI-driven hyper-personalization as a sociotechnical configuration that continuously tailors service encounters through algorithms, data infrastructures, interfaces, and organizational routines. This is not merely a technical capability. It is a relational system that shapes how consumers interpret providers' competence, benevolence, fairness, and authenticity. The framework specifies three core processes: cognitive and affective trust formation, transparency signaling, and vulnerability moderation.

The first process is cognitive and affective trust formation. AI-driven hyper-personalization influences cognitive trust through perceived relevance, accuracy, timeliness, and reliability. A system that anticipates needs, reduces search costs, and adapts smoothly to the context demonstrates competence. Simultaneously, hyper-personalization influences affective trust through perceived motives, authenticity, respect for privacy, and fairness. Consumers not only ask whether the system works but also whether it works for them or on them.

The second process is transparency signaling. Algorithmic transparency, explainability, consent interfaces, and contestability mechanisms shape how consumers interpret a provider's hidden practices. Transparency can strengthen cognitive trust by making personalization more predictable and procedurally understandable. It can also strengthen affective trust by signaling that the provider respects consumer autonomy and is willing to be accountable. However, transparency can backfire if it is vague, excessive, inconsistent, or presented as legal protection rather than consumer support.

The third process is vulnerability moderation. Consumer vulnerability shapes the threshold at which personalization becomes intrusive, the degree of explanation needed to sustain trust, and the extent to which consumers can act on the available controls. Vulnerable consumers may require stronger safeguards, even when personalization is accurate. In high-stakes services, vulnerability-sensitive design is essential. This is part of the legitimacy of AI-mediated service innovation.

The framework links these processes to loyalty outcomes. When hyper-personalization enhances cognitive and affective trust, and when transparency signals and vulnerability safeguards are credible, loyalty is likely to deepen. When personalization enhances only cognitive trust, loyalty may remain functional and fragile. When personalization is opaque, intrusive, or unfair, especially for vulnerable consumers, trust erosion can lead to avoidance, complaints, switching, negative word of mouth, and regulatory concerns.

Table 2. Core Processes in the Integrative Framework

Core Process	Trust Dimension Affected	Mechanism	Loyalty Implication
Cognitive and affective trust formation	Personalization is interpreted through both competence beliefs and relational beliefs about the provider.	Relevance, accuracy, and timeliness build cognitive trust, while perceived motives, authenticity, and privacy respect shape affective trust.	Cognitive gains alone yield fragile, convenience-based loyalty; combined cognitive and affective gains deepen it.
Transparency signaling	Explanations, consent interfaces, and contestability disclose otherwise hidden personalization practices.	Credible, usable signals make personalization predictable and communicate accountability and respect for autonomy.	Specific and consistent signals reinforce loyalty; vague or compliance-only signals can backfire and raise suspicion.
Vulnerability moderation	Consumer vulnerability conditions the threshold at which personalization becomes intrusive and the safeguards required.	Vulnerability sets how much explanation, conservative data use, and human oversight are needed to sustain trust.	Unaddressed vulnerability accelerates trust erosion even when the system performs accurately.

4.2. Boundary Conditions of the Trust-Loyalty Nexus

The trust–loyalty nexus in AI-mediated hyper-personalized services is not uniform across consumers, technologies, and service settings. Therefore, the framework developed in this study should be interpreted conditionally rather than universally. AI-driven personalization may strengthen trust and loyalty when it improves relevance, reduces effort, and communicates respect for consumer agency. However, the same personalization logic may weaken trust when the service context is high-risk, the data used are sensitive, the consumer is vulnerable, transparency is weak, or human oversight is absent. Boundary conditions clarify when hyper-personalization becomes a trust-building resource and when it becomes a source of perceived manipulation, surveillance, or relational breach.

Service risk is a primary boundary condition. In low-risk services, such as routine retail recommendations or entertainment suggestions, consumers may tolerate limited transparency if personalization is convenient and accurate. In high-risk services,

such as healthcare, financial advice, insurance, credit scoring, or safety-related hospitality services, the tolerance for opacity is much lower. Errors in these contexts may produce material, emotional, or social harm. Consequently, consumers are more likely to demand explanations, accountability, and human reviews. Prior research on medical AI has shown that consumers may resist algorithmic recommendations even when AI systems perform accurately because health-related decisions are closely tied to human care, empathy, and perceived uniqueness (Longoni et al., 2019). This suggests that the same AI capability may be evaluated differently depending on the perceived consequences of accepting its output.

Data sensitivity is another critical condition. Hyper-personalization based on purchase histories, browsing behavior, or location data may already raise privacy concerns; however, personalization based on inferred health status, financial distress, emotional vulnerability, family circumstances, or identity-related attributes is more likely to trigger discomfort. The personalization–privacy paradox demonstrates that consumers value relevance but also resist profiling when they perceive limited control over personal information (Aguirre et al., 2015; Awad & Krishnan, 2006; Cloarec et al., 2024). Experimental evidence reinforces this boundary: Kim and Han (2025) show that highly personalized messages can backfire when situational privacy concern is high, so that the same personalization depth that helps under low concern reduces favorable responses under high concern. Under high data sensitivity, transparency must move beyond generic privacy notices. Consumers need to understand what data are used, why the data are relevant, how inferences are generated, and whether they can contest or disable personalization. Without such controls, personalization may erode affective trust, even when it improves functional outcomes.

Consumer vulnerability further influences trust formation. Vulnerability may arise from age, limited digital literacy, financial fragility, illness, dependency, emotional stress, or restricted access to alternatives (Baker et al., 2005). Vulnerable consumers may experience AI-mediated personalization as more threatening because they have fewer resources to interpret, resist, or correct the algorithmic influence. For example, a financially secure consumer may perceive personalized credit offers as convenient, whereas a financially distressed consumer may perceive the same targeting as exploitative. Similarly, a digitally skilled patient may appreciate AI-supported health monitoring, whereas a patient with low health literacy may feel overwhelmed or disempowered by it. Therefore, vulnerability intensifies the need for accessible explanations, conservative personalization boundaries, and visible human support.

Quality of transparency is also central. Transparency is not merely the amount of information disclosed; it is the extent to which explanations are understandable, timely, relevant, and actionable. Long legalistic disclosures may satisfy formal requirements while doing little to support trust. From a signaling perspective, transparency works when it credibly communicates openness, fairness, and respect for

consumer agency (Connelly et al., 2011; Spence, 1973). High-quality transparency explains the basis of personalization in language that consumers can understand, provides meaningful choices, and allows for correction or contestation. Low-quality transparency, by contrast, can become performative. It may even increase distrust if consumers interpret it as a defensive or symbolic gesture rather than a genuine accountability mechanism.

Human oversight moderates the relationship between AI-mediated personalization and trust. In routine, low-risk services, consumers may fully accept fully automated personalization if the outcomes are useful and reversible. In high-stakes or emotionally sensitive services, visible human oversight becomes a trust-preserving safeguard. Human professionals can contextualize algorithmic outputs, interpret exceptions, and provide empathy when consumers experience uncertainty or harm. Service robot and AI service research has shown that automation alters expectations regarding accountability, empathy, and recovery (van Doorn et al., 2017; Wirtz et al., 2018). Thus, human oversight should not be treated as a technical add-on. It is a relational signal that the provider remains accountable for AI-mediated decisions.

The regulatory context also shapes trust and loyalty. Stronger regulatory environments may increase consumer confidence by establishing baseline protections for data rights, automated decision-making, and accountability. In the European Union, for example, the General Data Protection Regulation establishes rights to information and to meaningful explanation of automated decisions, while the AI Act introduces risk-based obligations and transparency requirements for higher-risk AI systems; together they illustrate how data-protection and AI-specific rules increasingly govern personalization. However, compliance alone is insufficient to generate trust. Consumers may not know the details of regulatory protections, and formal compliance may not address perceived fairness, authenticity, or emotional discomfort. In AI-mediated services, regulation provides a floor, but trust requires relationally meaningful practices. Firms that treat regulation as the minimum standard and add clearer explanations, stronger consent mechanisms, and accessible human reviews are more likely to sustain trust-building trajectories.

Service sector differences further qualify the framework. Retail personalization often centers on convenience, product discovery, and promotional relevance. Banking personalization involves financial risk, fairness, and long-term welfare. Healthcare personalization involves bodily vulnerability, emotional sensitivity, and professional responsibility. Hospitality personalization occupies a relational and experiential space in which comfort, privacy, and personal recognition are central. These sectoral differences mean that trust cues are not interchangeable. A recommendation explanation that is sufficient in retail may be inadequate in banking, and an automated reminder that is helpful in hospitality may feel intrusive in healthcare.

Finally, consumer AI literacy and prior trust in the provider shape how consumers interpret personalization. Consumers with higher AI literacy may be better able to understand algorithmic limits, evaluate explanations, and distinguish useful personalization from manipulative targeting. Consumers with lower AI literacy may rely more heavily on brand reputation, interface cues and institutional signals. Trust in the provider can also be transferred to the AI system, especially when consumers lack direct knowledge of technology. However, this transfer can operate in reverse; a poorly explained AI failure may damage trust in the organization. Therefore, the trust–loyalty nexus is recursive in nature. Provider trust shapes AI trust, and AI-mediated experiences reshape this trust.

Overall, these boundary conditions show that AI-driven hyper-personalization should not be evaluated based solely on accuracy, relevance, or conversion. Its trust implications depend on the interaction among service risk, data sensitivity, vulnerability, transparency quality, human oversight, regulatory context, sector characteristics, AI literacy, and prior provider trust. This conditional view strengthens the framework by explaining why the same personalization practice may generate loyalty in one context, indifference in another, and distrust in a third.

5. TRUST TRAJECTORY TYPOLOGY

Drawing on the integrative framework, this section introduces a typology of trust trajectories that captures how trust in AI-mediated services evolves under different configurations of hyper-personalization, transparency signaling, and consumer vulnerability. Rather than treating trust as a static level, the typology emphasizes dynamic patterns over time, drawing inspiration from longitudinal research on trust development and repair in marketing and relationship science. The typology distinguishes among three ideal-type trajectories: trust-building, trust-neutral, and trust-eroding. These trajectories represent analytically distinct patterns that can coexist within the same service ecosystem across different segments and contexts. To keep the categories conceptually disciplined, each trajectory is defined by a specific configuration of the two trust dimensions rather than by surface features alone. A trust-building trajectory combines high cognitive trust with high affective trust. A trust-neutral trajectory combines moderate-to-high cognitive trust with low affective trust, which is why loyalty in this state remains functional and easily disrupted rather than relational. A trust-eroding trajectory is marked by declining affective trust followed by declining cognitive trust. Defining the categories through these paired levels distinguishes trust-neutral, in particular, from a mere label for

convenience dependence: it is the specific state in which competence beliefs are satisfied but relational beliefs are not.

The following propositions formalize the relationships proposed in the framework.

P1. When AI-driven hyper-personalization is perceived as relevant and respectful and transparency signals are credible, cognitive and affective trust reinforce each other, producing a trust-building trajectory.

P2. When personalization enhances functional outcomes but transparency signals are weak or compliance-oriented, cognitive trust develops without corresponding affective trust, producing a trust-neutral trajectory.

P3. When personalization is perceived as intrusive or opaque, especially under high vulnerability, both cognitive and affective trust deteriorate, producing a trust-eroding trajectory.

5.1. Trust-Building Trajectory

The trust-building trajectory describes a reinforcing cycle in which AI-driven personalization strengthens both cognitive and affective trust. Consumers experience personalization as relevant, respectful, and aligned with their goals. Recommendations reduce effort without crossing perceived privacy boundaries. Dynamic offers are perceived as fair, rather than opportunistic. Service interfaces make the experience easier without displacing meaningful consumer controls.

Transparency is central to this trajectory in the study. Clear explanations of why a recommendation appears, what data were used, and how preferences can be modified help consumers understand personalization as legitimate. These signals support cognitive trust by clarifying system logic and affective trust by communicating respect for users. Vulnerability is either low or actively mitigated through conservative personalization, accessible support, and human oversight.

Loyalty in this trajectory is robust. Consumers are more likely to continue using the service, recommend it, forgive minor errors, and view AI-enabled personalization as part of the provider's value proposition. The provider's AI capability becomes a relational asset rather than a purely functional feature.

5.2. Trust-Neutral Trajectory

Trust-neutral trajectory arises when personalization improves functional outcomes but fails to generate deeper relational trust. Consumers may believe that the system works; however, they do not necessarily perceive the provider as caring, transparent, or authentic. Personalization remains a convenience feature. It reduces search costs

and increases efficiency; however, it does not create emotional attachment or strong relational commitment.

This trajectory often appears in low-involvement service contexts, where consumers value speed and convenience over relationship depth. This may also occur when transparency is present only at the compliance level. Consumers are neither alarmed nor reassured by this. Therefore, loyalty is shallow, driven by habit, switching costs, or the lack of a better alternative.

Trust-neutral conditions can be risky for vulnerable consumers. Continued use may mask anxiety or dependence. When cognitive trust exists without affective trust, a single failure, unexpected data use, or unexplained decision can quickly move consumers into a trust-eroding trajectory.

5.3. Trust-Eroding Trajectory

Trust-eroding trajectory occurs when hyper-personalization is experienced as intrusive, opaque, unfair, or manipulative. Consumers may feel that the provider is exploiting superior information rather than creating mutual value. Highly targeted messages that reveal sensitive inferences can create discomfort. Dynamic pricing or personalized offers may be interpreted as discriminatory when consumers cannot understand the basis of the variation.

Affective trust usually deteriorates first because consumers question the provider's benevolence and integrity. Cognitive trust can decline when recommendations are inconsistent, biased or difficult to explain. In the absence of credible transparency signals, consumers may infer that the system is incompetent or exploitative. Vulnerability amplifies the severity of this trajectory because vulnerable consumers incur greater costs when service decisions are inaccurate, unfair, or difficult to challenge.

The loyalty outcomes in this trajectory are negative. Consumers may reduce their use, switch providers, complain publicly, resist data sharing, or seek regulatory protection. Trust repair requires more than technical corrections. It requires acknowledgement, explanation, compensation where appropriate, and visible changes in governance practices.

Table 3. Trust Trajectory Typology in AI-Mediated Hyper-Personalized Services

Trajectory	Personalization Experience	Transparency Condition	Vulnerability Condition	Expected Loyalty Outcome
Trust-Building	Relevant, respectful, accurate, and authenticity-preserving.	Specific, usable, and credible.	Low or actively mitigated.	Strong commitment, continued use, advocacy, and forgiveness.
Trust-Neutral	Efficient and useful but relationally thin.	Minimal or compliance-oriented.	Low to moderate, or under-recognized.	Habitual or convenience-based loyalty that is easy to disrupt.
Trust-Eroding	Intrusive, opaque, unfair, or manipulative.	Absent, vague, or inconsistent.	High or ignored.	Avoidance, churn, negative word of mouth, and regulatory concern.

6. SECTORAL ILLUSTRATIONS

The following sectoral illustrations demonstrate how the integrative framework and trust trajectory typology operate in practice. These four sectors—retail, banking, healthcare, and hospitality—were selected because they represent distinct combinations of service risk, data sensitivity, consumer vulnerability, and regulatory intensity. Together, they show how the same AI capabilities can produce different trust trajectories, depending on contextual conditions and design choices.

6.1. Retail

Retail firms are intensive adopters of AI-driven personalization. Recommendation engines, targeted promotions, personalized homepages, and dynamic service communications use purchase records, browsing histories, clickstream data, and contextual cues. These systems strengthen cognitive trust when consumers repeatedly receive useful recommendations and experience reduced search costs (Aguirre et al., 2015; Bleier & Eisenbeiss, 2015; Wedel & Kannan, 2016).

The central retail risk is the personalization–privacy paradox. Consumers may appreciate relevance but become uneasy when personalization reveals unexpected inferences. A well-documented case involves a major U.S. retailer that used purchase pattern analysis to infer pregnancy status and send targeted coupons, resulting in consumer backlash when the inference was revealed to a family member before the consumer had disclosed the pregnancy (Duhigg, 2012). This case illustrates how technically accurate personalization can trigger trust erosion when it crosses perceived privacy boundaries without adequate transparency and consent. More recent evidence confirms that this tension persists in contemporary AI-enabled retail. Canhoto et al. (2024) find that AI-enabled personalization in physical retail

environments produces a recurring "snakes and ladders" dynamic, in which gains in perceived relevance can be rapidly reversed when consumers sense covert data use or loss of control, underscoring that the paradox intensifies rather than disappears as personalization becomes more sophisticated.

Trust-building retail personalization requires clear explanations, preference controls, conservative use of sensitive categories, and visible privacy safeguards. For example, Amazon's recommendation system provides explicit "why recommended" explanations and allows users to remove items from their browsing history that influence recommendations. These features support cognitive trust through accuracy and affective trust through perceived control. Retailers that treat personalization as a pure conversion tool risk eroding affective trust, even when their algorithms perform well.

6.2. Banking and Financial Services

Banks use AI for fraud detection, credit scoring, robo-advisory services, risk profiling, and personalized financial advice. These applications can build cognitive trust by increasing speed, accuracy, and proactive support. However, financial services are high-stakes and vulnerability-sensitive. Consumers may face significant consequences if automated credit, pricing, or advisory systems are opaque or unfair.

In this sector, transparency signaling is not merely a communication problem. It is central to perceived procedural justice. Consumers must understand why an offer, denial, risk score, or recommendation was generated and how they can seek a review. The European Union's General Data Protection Regulation (GDPR) and the EU AI Act require explainability for consequential automated decisions, reflecting this normative expectation. However, regulatory compliance alone does not guarantee trust. A robo-advisor that discloses its fee structure and investment logic in accessible language may build trust, whereas one that buries this information in legal disclaimers may trigger skepticism. Empirical work on fintech adoption is consistent with this view; Sabir et al. (2023) find that perceived usefulness, ease of use, and convenience drive consumer attitudes toward robo-advisors, with trust operating as an influential factor in the acceptance model, which suggests that functional performance and relational trust must be cultivated together in financial AI.

Vulnerability-sensitive banking design should avoid exploiting financial distress, provide understandable explanations, and maintain human review for consequential decision-making. For example, some banks have implemented "financial health" dashboards that proactively alert customers to potential cash flow problems and offer conservative solutions rather than high-interest credit products. Such designs can shift vulnerable consumers from trust-neutral or trust-eroding trajectories toward trust-building ones.

6.3. Healthcare

Healthcare AI personalizes risk assessment, diagnosis support, monitoring, reminders, and care pathways. These applications can enhance cognitive trust by improving the accuracy, timeliness, and coordination of care. However, healthcare is deeply relational. Patients often evaluate AI through expectations of empathy, professional judgment, and moral responsibility (Longoni et al., 2019; Steerling et al., 2023).

A trust-building healthcare trajectory requires human oversight, accessible explanations, and clear boundaries regarding AI's role. For example, IBM Watson for Oncology was initially deployed with high expectations but faced resistance from clinicians who found its recommendations inconsistent with local practice norms and patient preferences (Ross & Swetlitz, 2018). This case illustrates that technical competence alone does not ensure trust when a system fails to align with professional and relational expectations. Patient-level evidence points in the same direction. In a prospective study of patients facing localized prostate cancer decisions, Rodler et al. (2024) found that patients trusted physicians more than AI for individualized communication, yet strongly preferred AI-assisted physicians over physicians alone (66.74% versus 29.61%). This pattern indicates that trust in healthcare AI is conditional and relational: it is highest when AI augments, rather than displaces, the clinician.

AI should be framed as augmenting, rather than replacing, clinical judgment. Patients should be aware of when AI is involved, what data are used, and who remains accountable. Some hospitals now use “AI disclosure” protocols in radiology, in which patients are informed that an AI system assisted in analyzing their scan but that a radiologist reviewed and confirmed the results. Without these safeguards, personalization can feel like surveillance, cost-cutting, or depersonalization, especially for patients with vulnerable conditions.

6.4. Hospitality

Hospitality personalization includes pre-arrival preference capture, smart room configuration, personalized recommendations, automated check-in, and loyalty communications. In the hospitality industry, affective trust is particularly important because service experiences are often hedonic, relational, and identity-relevant. Guests may interpret personalization as care when it respects comfort, privacy, and cultural expectations.

Hospitality creates intimate data environments. Smart rooms, location-based recommendations, and ambient personalization can raise privacy concerns if guests do not know what is collected or how it is used. The Marriott International data

breach (2018), which exposed the personal information of approximately 500 million guests, illustrates how data security failures can catastrophically erode trust in hospitality personalization systems, even when technical service delivery remains intact. Research on smart, connected environments helps explain why such settings are so trust-sensitive. Chan-Olmsted et al. (2024) show that consumers continuously weigh the convenience and control benefits of smart device personalization against perceived loss of control and the risk of feeling surveilled, with trust serving as the pivotal factor in this privacy calculus. In hospitality's intimate spaces, where the "device" is effectively the guest room itself, this calculus becomes especially consequential for affective trust.

Trust-preserving hospitality design should provide clear in-room controls, opt-out options, conservative data use in personal spaces, and easy human assistance when automation fails. Some boutique hotels now provide "privacy mode" switches that disable all smart sensors and data collection, accompanied by clear signage explaining what features remain active. Such designs signal respect for guest boundaries and can sustain affective trust, even among privacy-sensitive travelers.

Table 4. Sectoral Implications of the Framework

Sector	Typical AI Application	Main Trust Opportunity	Main Trust Risk	Governance Requirement
Retail	Recommendations, promotions, and personalized interfaces.	Reduced effort and higher relevance.	Creepiness, covert inference, and unfair personalization.	Preference controls, data-use explanations, and sensitive-category limits.
Banking	Credit scoring, robo-advice, fraud detection.	Faster, more tailored financial support.	Opaque or exploitative decisions affecting vulnerable consumers.	Explainability, contestability, and human review.
Healthcare	Decision support, risk prediction, patient engagement.	Timely, personalized, and coordinated care.	Depersonalization, surveillance, or loss of clinical accountability.	Clinician oversight, clear disclosure, and accessible explanations.
Hospitality	Smart rooms, personalized recommendations, loyalty offers.	Personal care, comfort, and convenience.	Privacy concerns in intimate service spaces.	Opt-out controls, clear notices, and human fallback.

7. RESEARCH AGENDA

Future research should longitudinally examine trust trajectories. Most studies measure trust at one point in time, yet AI-mediated services evolve through multiple interactions. Panel studies, experience sampling, and longitudinal field data can

identify when trust-building, trust-neutral, and trust-eroding trajectories emerge and how they change after service failures, privacy events, or transparency interventions.

A second priority is measurement development. Existing trust scales provide a strong foundation (Mayer et al., 1995; McKnight et al., 2002); however, AI-mediated services require measures that separate cognitive trust in system competence from affective trust in provider benevolence and integrity. Measures of transparency signaling quality should capture clarity, specificity, usability, sincerity, and contestability. Measures of AI-induced vulnerability should capture perceived powerlessness, dependence, anxiety, and limited ability to contest outcomes.

Experimental research should examine how transparency designs interact with personalization depth and vulnerability. Studies can manipulate recommendation explanations, data-use notices, human oversight, and opt-out controls across low-risk and high-risk scenarios. The outcomes should include cognitive trust, affective trust, perceived authenticity, privacy concern, perceived control, loyalty intention, and avoidance behavior.

To make these directions actionable, Table 5 maps the three propositions onto candidate variables, research designs, and recent empirical exemplars. The table demonstrates that the framework is not only conceptual but also empirically testable, and it offers concrete starting points for operationalizing the trust trajectories.

Table 5. Empirical Pathways for Testing the Propositions

Proposition	Key Variables	Suggested Design	Illustrative Reference
P1 (trust-building): relevant, respectful personalization plus credible transparency reinforces cognitive and affective trust.	IVs: personalization relevance, transparency credibility. DVs: cognitive trust, affective trust, loyalty intention.	2 (relevance: high/low) x 2 (transparency: high/low) between-subjects experiment, measuring both trust dimensions separately.	Park & Yoon (2024); Zhao et al. (2025).
P2 (trust-neutral): functional personalization with weak or compliance-oriented transparency yields cognitive but not affective trust.	IVs: transparency quality (substantive vs. compliance-only). DVs: cognitive trust, affective trust, switching ease.	Cross-sectional survey with mediation analysis (PLS-SEM) separating the two trust paths to loyalty.	Chan-Olmsted et al. (2024); Cloarec et al. (2024).
P3 (trust-eroding): intrusive or opaque personalization under high vulnerability deteriorates both trust dimensions.	IVs: intrusiveness, data sensitivity. Moderator: consumer vulnerability. DVs: affective trust, avoidance, complaint intention.	Moderated experiment (3 personalization x 2 vulnerability), plus longitudinal panel to trace erosion after a privacy event.	Kim & Han (2025); Canhoto et al. (2024).

Cross-cultural research is also needed. Privacy expectations, institutional trust, power distance, uncertainty avoidance, and attitudes toward automation differ across societies. The same transparency signal may be interpreted as reassuring in one

context and insufficient in another. Comparative studies can clarify how cultural norms shape the personalization threshold at which relevance turns into an intrusion.

Sector-specific research should examine whether the framework operates differently in retail, banking, healthcare, hospitality, education, transportation, and public service sectors. High-stakes sectors likely require stronger transparency and human oversight than low-stakes convenience. Future research should also investigate how regulations, professional norms, and service criticality moderate the trust–loyalty nexus.

Qualitative and mixed-methods research can deepen the understanding of consumer sense-making. Interviews, diaries, netnography, and digital ethnography can reveal how consumers narrate AI-mediated personalization, how they identify creepiness or authenticity, and how vulnerable consumers cope with opaque systems. These methods are particularly useful for uncovering silent disengagement, resistance, and trust repair expectations.

Finally, research should examine the organizational implementation. Trust-preserving AI personalization depends on governance routines, cross-functional collaboration, data stewardship, ethical reviews, frontline training, and accountability structures. Case studies of successful and failed deployments can clarify the organizational capabilities that support transparency and vulnerability-sensitive AI services.

A further priority concerns the shift from recommendation-based personalization to agentic AI. As noted at the outset, organizations are increasingly deploying autonomous agents that do not merely suggest options but plan, decide, and act on the consumer's behalf (McKinsey & Company, 2025). This shift reshapes the trust–loyalty nexus in ways the present framework only begins to address. When an agent transacts autonomously, the locus of consumer control moves from approving individual recommendations to delegating authority over a stream of future actions, which raises the stakes of both cognitive trust (will the agent act competently and within its mandate?) and affective trust (will it act in my interest when I am not watching?). Early evidence that consumers form combined cognitive and affective trust in AI agents (Zhao et al., 2025) suggests that the two-dimensional structure remains relevant, but the conditions for transparency, contestability, and human override are likely to be more demanding under delegation than under recommendation. Future research should therefore examine how the trust trajectories described here unfold when consumers grant agents standing authority, and how providers can preserve meaningful oversight without negating the convenience that motivates delegation in the first place.

PRACTICAL, MANAGERIAL, AND ETHICAL IMPLICATIONS

Service managers should treat trust as a design objective rather than a by-product of personalization accuracy. They should decide which trust trajectory they want to cultivate in different service contexts and consumer segments. Convenience-based personalization may be sufficient for low-risk services. In high-risk or high-vulnerability contexts, managers must deliberately build cognitive and affective trust through transparency, restraint, and human oversight.

Customer experience designers should calibrate the depth of personalization. More personalization is not automatically better. Designers should consider whether the recommendation or offer uses data that consumers expect the firm to use, whether the inference is sensitive, whether the consumer can understand the reason for personalization, and whether there is an easy way to modify or refuse it. Preference dashboards, micro-explanations, and privacy-preserving defaults can reduce perceived intrusion.

AI and data analytics teams should incorporate explainability, fairness, and vulnerability assessment into model development. Data minimization, purpose limitation, model documentation, and bias monitoring should be treated as part of service quality. High-stakes AI-mediated decisions should include human review and contestability. Model performance metrics should be balanced with trust metrics, such as perceived fairness, complaint patterns, opt-out rates, and trust recovery after failures. These practices should be organized around three governance commitments that are central to contemporary AI ethics: fairness, bias mitigation, and accountability. Fairness requires that personalization not produce systematically disadvantageous outcomes for protected or vulnerable groups, for example through discriminatory pricing or differential access to favorable offers. Bias mitigation requires ongoing auditing of training data and model outputs, because personalization models can encode and amplify historical inequities even when no discriminatory intent exists. Accountability requires that responsibility for AI-mediated outcomes remain clearly assigned to the organization rather than diffused into the algorithm, with documented escalation paths and the capacity to explain and reverse consequential decisions. Treating these three commitments as design requirements, rather than as after-the-fact compliance, is what allows transparency and authenticity to be experienced by consumers as genuine.

Healthcare and financial services managers should apply stricter safeguards because consumer vulnerability and decision consequences are higher. In healthcare, AI should augment, rather than displace, clinical relationships. In financial services, AI should not be used to target vulnerable consumers with exploitative offers. Clear explanations, conservative personalization, and human escalation paths are essential.

Retail and hospitality managers should focus on transparency in everyday service encounters. In retail, consumers should be able to see and modify the basis of recommendations. In hospitality, personal spaces require heightened privacy standards. Smart-room personalization should be controllable and clearly explained. Across both sectors, trust-preserving personalization requires restraint as much as technological sophistication.

Policymakers and regulators should recognize that AI-mediated hyper-personalization creates new forms of information asymmetry and vulnerability. Disclosure rules should encourage usable explanations rather than dense legal statements. Consumer protection should address not only data collection but also inference, targeting, dynamic pricing, and contestability in consequential service decisions. The High-Level Expert Group on AI (2019) provides useful principles; however, service-specific operationalization remains necessary for their implementation.

CONCLUSION

This chapter has examined the trust–loyalty nexus in AI-mediated hyper-personalized services through the lenses of cognitive–affective trust theory, signaling theory, and consumer vulnerability literature. The central argument is that hyper-personalization does not simply improve relevance or convenience. Rather, it restructures the relational conditions under which consumers decide whether to trust, continue using, or remain loyal to service providers.

The analysis demonstrates that AI-driven personalization can strengthen cognitive trust by demonstrating technical competence, accuracy, and responsiveness. Simultaneously, it can weaken affective trust if it is perceived as intrusive, opaque, opportunistic, or insensitive to vulnerability. Therefore, algorithmic transparency is not merely a technical issue. It functions as a credibility signal that shapes how consumers interpret providers' intentions in the context of information asymmetry.

Trust trajectory typology clarifies why personalization does not yield uniform loyalty outcomes. Trust-building trajectories emerge when personalization is relevant, transparent, respectful, and vulnerability-sensitive. Trust-neutral trajectories emerge when personalization is useful but relationally thin. Trust-eroding trajectories arise when personalization is intrusive, opaque, unfair, or exploitative, especially in contexts of high vulnerability.

The long-term viability of AI-mediated service ecosystems will depend on whether organizations use hyper-personalization to deepen mutual value or exploit informational advantage. Trust should therefore be treated as a central design lens for AI service innovation. Service organizations that combine technical competence with transparency, human oversight, privacy protection, and vulnerability-sensitive

governance are more likely to build durable loyalty. Organizations that maximize personalization without relational accountability may gain short-term efficiency while undermining the trust on which sustainable service relationships depend.

ARTIFICIAL INTELLIGENCE DISCLOSURE STATEMENT

AI-assisted tools were used only to support language refinement, structural review, formatting consistency, and readability improvement. The authors bear full responsibility for the conceptual development, scholarly argumentation, source selection, citation verification, and final content of the chapter.

REFERENCES

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, *91*(1), 34–49. DOI: 10.1016/j.jretai.2014.09.005

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *Management Information Systems Quarterly*, *30*(1), 13–28. DOI: 10.2307/25148715

Baker, S. M., Gentry, J. W., & Rittenburg, T. L. (2005). Building understanding of the domain of consumer vulnerability. *Journal of Macromarketing*, *25*(2), 128–139. DOI: 10.1177/0276146705280622

Beverland, M. B. (2005). Crafting brand authenticity: The case of luxury wines. *Journal of Management Studies*, *42*(5), 1003–1029. DOI: 10.1111/j.1467-6486.2005.00530.x

Bleier, A., & Eisenbeiss, M. (2015). Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science*, *34*(5), 669–688. DOI: 10.1287/mksc.2015.0930

Canhoto, A. I., Keegan, B. J., & Ryzhikh, M. (2024). Snakes and ladders: Unpacking the personalisation-privacy paradox in the context of AI-enabled personalisation in the physical retail environment. *Information Systems Frontiers*, *26*(3), 1005–1024. DOI: 10.1007/s10796-023-10369-7 PMID: 36684411

Chan-Olmsted, S., Chen, H., & Kim, H. J. (2024). In smartness we trust: Consumer experience, smart device personalization and privacy balance. *Journal of Consumer Marketing*, *41*(6), 597–609. DOI: 10.1108/JCM-12-2021-5072

Cloarec, J., Meyer-Waarden, L., & Munzel, A. (2022). The personalization–privacy paradox at the nexus of social exchange and construal level theories. *Psychology and Marketing*, *39*(3), 647–661. DOI: 10.1002/mar.21587

Cloarec, J., Meyer-Waarden, L., & Munzel, A. (2024). Transformative privacy calculus: Conceptualizing the personalization–privacy paradox on social media. *Psychology and Marketing*, *41*(7), 1574–1596. DOI: 10.1002/mar.21998

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67. DOI: 10.1177/0149206310388419

Duhigg, C. (2012, February 16). How companies learn your secrets. *The New York Times Magazine*. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Flavián, C., Pérez-Rueda, A., Belanche, D., & Casaló, L. V. (2022). Intention to use analytical artificial intelligence in services: The effect of technology readiness and awareness. *Journal of Service Management*, *33*(2), 293–320. DOI: 10.1108/JOSM-10-2020-0378

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *Management Information Systems Quarterly*, *27*(1), 51–90. DOI: 10.2307/30036519

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, *21*(2), 155–172. DOI: 10.1177/1094670517752459

IBM. (2024). *IBM Global AI Adoption Index: Enterprise report*. IBM. https://filecache.mediaroom.com/mr5mr_ibmspgi/179414/download/IBM%20Global%20AI%20Adoption%20Index%20Report%20Dec.%202023.pdf

Kim, H., & Han, S. (2025). Triggering the personalization backfire effect: The moderating role of situational privacy concern. *Behavioral Sciences (Basel, Switzerland)*, *15*(10), 1323. Advance online publication. DOI: 10.3390/bs15101323 PMID: 41153114

Komiak, S. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *Management Information Systems Quarterly*, *30*(4), 941–960. DOI: 10.2307/25148760

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *The Journal of Consumer Research*, *46*(4), 629–650. DOI: 10.1093/jcr/ucz013

Marriott International. (2018, November 30). *Marriott International announces Starwood guest reservation database security incident*. https://news.marriott.com/news/2018/11/30/marriott-international-announces-starwood-guest-reservation-database-security-incident

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. DOI: 10.2307/258792

McKinsey & Company. (2025). *The state of AI in 2025: Agents, innovation, and transformation*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359. DOI: 10.1287/isre.13.3.334.81

Mersha, M., Lam, K., Wood, J., AlShami, A. K., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing*, *599*, 128111. Advance online publication. DOI: 10.1016/j.neucom.2024.128111

Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. *Journal of Consumer Psychology*, *25*(2), 200–218. DOI: 10.1016/j.jcps.2014.11.006

Napoli, J., Dickinson, S. J., Beverland, M. B., & Farrelly, F. (2014). Measuring consumer-based brand authenticity. *Journal of Business Research*, *67*(6), 1090–1098. DOI: 10.1016/j.jbusres.2013.06.001

OECD. (2023). *Consumer vulnerability in the digital age*. OECD Publishing., DOI: 10.1787/4d013cc5-

Park, K., & Yoon, H. Y. (2024). Beyond the code: The impact of AI algorithm transparency signaling on user trust and relational satisfaction. *Public Relations Review*, *50*(5), 102507. Advance online publication. DOI: 10.1016/j.pubrev.2024.102507

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. DOI: 10.1007/s11747-019-00710-5

Rodler, S., Kopliku, R., Ulrich, D., Kaltenhauser, A., Casuscelli, J., Eismann, L., Waidelich, R., Buchner, A., Butz, A., Cacciamani, G. E., Stief, C. G., & Westhofen, T. (2024). Patients' trust in artificial intelligence–based decision-making for localized prostate cancer: Results from a prospective trial. *European Urology Focus*, *10*(4), 654–661. DOI: 10.1016/j.euf.2023.10.020 PMID: 37923632

Ross, C., & Swetlitz, I. (2018, July 25). IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show. *Stat News*. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-cancer-treatments/

Sabir, A. A., Ahmad, I., Ahmad, H., Rafiq, M., Khan, M. A., & Noreen, N. (2023). Consumer acceptance and adoption of AI robo-advisors in fintech industry. *Mathematics*, *11*(6), 1311. Advance online publication. DOI: 10.3390/math11061311

Salesforce. (2024). *State of the AI connected customer*. Salesforce Research. https://www.salesforce.com/en-us/wp-content/uploads/sites/4/documents/research/State-of-the-Connected-Customer.pdf

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, *32*(2), 344–354. DOI: 10.5465/amr.2007.24348410

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374. DOI: 10.2307/1882010

Steerling, E., Siira, E., Nilsen, P., Svedberg, P., & Nygren, J. (2023). Implementing AI in healthcare: The relevance of trust: A scoping review. *Frontiers in Health Services*, *3*, 1211150. Advance online publication. DOI: 10.3389/frhs.2023.1211150 PMID: 37693234

Trinh, M. (2025). AI-induced consumer vulnerability: Psychological mechanisms and marketplace implications. *Journal of Consumer Policy*, *48*(1), 45–68. DOI: 10.1007/s10603-024-09532-7

van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research*, *20*(1), 43–58. DOI: 10.1177/1094670516679272

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97–121. DOI: 10.1509/jm.15.0413

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931. DOI: 10.1108/JOSM-04-2018-0119

Zhao, X., You, W., Zheng, Z., Shi, S., Lu, Y., & Sun, L. (2025). How do consumers trust and accept AI agents? An extended theoretical framework and empirical evidence. *Behavioral Sciences (Basel, Switzerland)*, *15*(3), 337. Advance online publication. DOI: 10.3390/bs15030337 PMID: 40150232

KEY TERMS AND DEFINITIONS

Affective Trust: Trust grounded in beliefs about benevolence, integrity, care, and relational alignment.

AI-Mediated Service: A service encounter in which artificial intelligence systems shape interaction, recommendation, communication, or decision support between the provider and consumer.

Algorithmic Transparency: The extent to which consumers receive understandable information about how AI-enabled recommendations or decisions are generated.

Cognitive Trust: Trust grounded in beliefs about competence, reliability, accuracy, and technical capability.

Consumer Trust: A consumer's willingness to rely on a service provider or AI system under conditions of uncertainty and limited control.

Consumer Vulnerability: A contextually activated condition in which consumers face power, information, resource, or dependency imbalances in marketplace interactions.

Customer loyalty: A consumer's behavioral and attitudinal commitment to continue using, repurchasing from, or recommending a service provider.

Explainable AI: AI design and communication practices that make algorithmic outputs more understandable and contestable by users and stakeholders.

Hyper-Personalization: The real-time tailoring of service content, offers, interfaces, or interactions using behavioral, contextual, transactional, and predictive data.

Personalization-Privacy Paradox: The tension between consumers' desire for relevant personalized services and their concern about data collection, profiling, and privacy loss.

Transparency Signaling: The use of explanations, disclosures, controls, and audit cues to communicate credibility, accountability, and respect for consumer agency.

Trust trajectory: The direction and pattern through which trust develops, stabilizes, or deteriorates across repeated AI-mediated service encounters.

Trust-Preserving Design: A service design approach that uses transparency, data restraint, human oversight, and contestability to protect consumer trust.

Vulnerability-Sensitive Personalization: Personalization that adjusts data use, explanation, oversight, and control mechanisms according to the consumer's risk and vulnerability context.