https://journals.insightpub.org/index.php/imj

Insight Management Journal

Volume 6, Issue 2, November 2026, pages 17-31

The Dark Side of AI-Driven Hyper-Personalization: Customer Irritation, Affective Accumulation, and Value Co-Destruction

Eva Nirmala Sari Chaniago*, Muhammad Dharma Tuah Putra Nasution, Fadli

Universitas Sumatera Utara, Medan, Indonesia

Article info

Article history:
Submitted [02-06-2026]
Revised [02-07-2026]
Accepted [03-10-2026]

Keywords:
Customer Irritation
Artificial Intelligence
Hyper-Personalization
Affective Customer Experience
Value Co-Destruction

A b s t r a c t

Firms increasingly invest in AI-driven hyper-personalization to enhance customer experience (CX), yet emerging evidence indicates that these technologies may also generate unintended negative emotional outcomes. This study aims to develop a conceptual framework explaining how failures in AI-driven hyper-personalization lead to customer irritation and subsequent value co-destruction in AI-mediated service interactions. Using a conceptual research approach based on an integrative review and synthesis of recent literature, the study draws upon Expectation–Disconfirmation Theory, Psychological Reactance Theory, and Affective Events Theory to construct a theoretically grounded framework. The proposed framework identifies AI response irrelevance and perceived intrusiveness as key antecedents of customer irritation, which subsequently mediates the effects of hyper-personalization failure on diminished trust, switching intention, and customer disengagement. The framework further advances five research propositions, identifies six boundary-condition moderators, and explains three structural paradoxes that clarify how hyper-personalization may systematically generate negative customer experiences despite its intended benefits. These findings extend value co-destruction theory by positioning customer irritation as a distinct cumulative affective mechanism underlying CX degradation and provide a stronger theoretical foundation for future empirical research on AI-mediated services. Future studies are recommended to empirically validate the proposed framework using longitudinal, experimental, or cross-cultural research designs across diverse AI-enabled service contexts.

Introduction

Firms invest billions of dollars in AI-driven hyper-personalization on the premise that knowing customers better leads to serving them better. Yet a growing body of evidence reveals that the same systems designed to enhance relevance and deepen relationships are, under predictable conditions, systematically generating frustration, discomfort, and disengagement (Pavone et al., 2023). The theoretical gap that allows this pattern to remain poorly understood is the focus of this study (Castillo et al., 2021).

Artificial intelligence now underpins the core infrastructure of contemporary customer experience (CX), including chatbots, recommender systems, virtual assistants, and algorithmically personalized journeys across channels (Hardcastle et al., 2025). Hyper-personalization is widely framed as a source of superior efficiency, relevance, and engagement that enhances convenience, satisfaction, and relationship quality (Ming-Hui Huang & Rust, 2024). When it works, it deepens affective attachment and strengthens purchase intentions beyond what functional efficiency alone produces (Bilal et al., 2024). When it fails, however, it does not merely disappoint: it intrudes, misrecognizes, and accumulates. Customers increasingly report encounters with irrelevant responses, poor contextual understanding, and personalization that oversteps perceived boundaries of privacy and autonomy. These failures are particularly consequential on digital platforms where AI mediates high-frequency, real-time interactions, and where sub-threshold frictions accumulate rapidly and often pass beneath the threshold of explicit complaint (Ranieri et al., 2024). Meta-analytic evidence across physical robots, chatbots, and other AI confirms that service AI interactions systematically generate distinct negative affective reactions that reduce engagement when design features fall below expectations (Blut et al., 2021).

*Corresponding author
Email address: Eva Nirmala Sari Chaniago (evac307@gmail.com)
ISSN: 2774-1737 DOI: 10.47065/imj.v6i2.541

Despite the growing practical salience of these negative experiences, academic discourse remains skewed toward the positive dimensions of AI, emphasizing adoption, trust, and personalization benefits (Hollebeek et al., 2021). A scientometric analysis of AI in marketing confirms that the field has been dominated by themes of efficiency, engagement, and value creation, with negative experiential dimensions receiving comparatively little theoretical attention (Mustak et al., 2021). Even when risks such as privacy or fairness are acknowledged, they are typically treated as secondary outcomes rather than central mechanisms (Liu-Thompkins et al., 2022). Positive orientations similarly dominate research that integrates AI into CX frameworks (Ameen et al., 2021), and even studies that foreground service failure tend to prioritize cognitive over affective explanations (C. Chen, 2024). Existing work on AI service failure highlights emotions such as anger and disappointment, yet it rarely theorizes customer irritation as a distinct, cumulative, and proximal response to hyper-personalization breakdowns. The CX literature therefore remains underdeveloped in explaining how everyday AI misfires translate into disengagement and value loss.

This gap is not merely empirical but theoretical. Prior frameworks for negative affect in AI-CX, including the experiential perspective of Puntoni et al. (2021) and the service co-destruction model of Castillo et al. (2021), conceptualize negative emotion primarily as a reaction to salient, discrete failures. They do not account for the gradual, sub-threshold accumulation of minor affective events that characterize routine hyper-personalization encounters. Pavone et al. (2023) advance understanding of acute emotional responses such as rage and indignation but stop short of theorizing the lower-intensity, slower-building emotional states that may prove equally consequential over time. The present study identifies this as the central theoretical gap in the literature: the absence of a construct and mechanism that explains how unremarkable, individually tolerable AI missteps aggregate into a durable emotional state capable of driving trust erosion and behavioral disengagement.

Addressing this gap requires integrating theories that foreground expectations, autonomy, and affective processes. Expectation–Disconfirmation Theory explains how discrepancies between promised personalization and experienced irrelevance generate negative evaluations. Psychological Reactance Theory accounts for how perceived intrusiveness threatens autonomy and motivates resistance behaviors. Affective Events Theory provides a dynamic account of how repeated negative AI interactions accumulate into enduring emotional states and behavioral outcomes. Taken together, these three theoretical perspectives converge on a mechanism, namely the cumulative build-up of low-intensity irritation, that none of them addresses individually and that prior AI-CX research has not yet formalized.

Building on these perspectives, this study develops a conceptual framework in which AI response irrelevance and perceived intrusiveness act as antecedents of customer irritation, which in turn mediates the relationship between hyper-personalization failure and three adverse outcomes: diminished trust, increased switching intention, and disengagement behavior. The framework advances five formal propositions and identifies six boundary-condition moderators (AI literacy, privacy concern, prior AI experience, algorithmic interpretability, perceived human-likeness, and consumer–machine relationship orientation) that govern when and for whom the irritation mechanism is most consequential.

This study makes four contributions. First, it introduces AI-induced customer irritation as a theoretically distinct affective construct, defined by its low intensity, cumulative character, and origin in repeated micro-level personalization failures, distinguishing it from the episodic anger and acute frustration documented in prior AI service failure research. Second, it reframes AI-enabled CX as a dual-valence phenomenon in which the emotional risks of hyper-personalization arise through a qualitatively different mechanism than its benefits: affective accumulation rather than single-event disconfirmation. Third, it identifies three structural paradoxes (personalization versus autonomy, intelligence versus control, and relevance versus intrusiveness) as the generative conditions under which hyper-personalization systematically produces irritation, providing a theoretical vocabulary for future research to examine where, when, and for whom each paradox is most consequential. Fourth, by positioning irritation as the mediating affective pathway between AI design failures and value co-destruction, this study advances the co-destruction literature Castillo et al. (2021) by specifying the mechanism, rather than merely cataloguing the outcomes, through which AI-mediated interactions destroy rather than create value.

Theoretical Foundation

Customer Experience Theory in the AI Context

The dominant theoretical tradition in AI-CX research rests on a value-creation premise: AI improves service quality, personalization enhances relevance, and automation scales efficiency. What this tradition has not adequately theorized is the affective cost of these same processes, specifically the

discomfort, resentment, and cumulative fatigue that AI-mediated interactions can generate even when no single interaction constitutes an identifiable failure. Recent work on AI-enabled CX reveals systematic affective costs, including discomfort, anger, and alienation, that current theory insufficiently captures (Castillo et al., 2021). Research on automated service interactions further demonstrates that when engagement quality deteriorates, customers disengage rapidly and the relational benefits of AI are quickly reversed (Hollebeek et al., 2021), and chatbot service quality studies in retail contexts confirm that when AI fails to meet customer expectations, satisfaction erodes markedly even in high-involvement hedonic categories (Chung et al., 2020).

CX is now widely conceptualized as a multidimensional constellation of cognitive, affective, sensory, behavioral, and social responses to brand-related touchpoints (Becker & Jaakkola, 2020). In AI-mediated environments, algorithmic service quality and personalization shape both cognitive appraisals and affective states, with downstream consequences for satisfaction and loyalty (Ranieri et al., 2024). However, existing literature tends to privilege positive outcomes while under-theorizing negative affective responses emerging from AI interactions (Puntoni et al., 2021).

This imbalance is particularly problematic because existing negative-affect accounts in AI-CX focus disproportionately on responses to salient, identifiable failures: a chatbot that misroutes a complaint, a recommendation engine that produces an obviously inappropriate suggestion, or a virtual assistant that fails to resolve a query. These accounts treat negative affect as event-contingent and episodic. What they do not capture is the class of affective responses that emerge not from a single identifiable failure but from the cumulative weight of sub-threshold frictions across many ordinary interactions.

Conceptual work on the dark side of AI in services has begun to catalog systemic threats, including bias, lack of empathy, privacy loss, and anxiety, as recurring features of AI-mediated encounters rather than isolated incidents (Belanche et al., 2024). Research on AI service failures confirms that automated systems can trigger frustration, confusion, and stress (Castillo et al., 2021), but the mechanisms by which low-grade, repeated misalignments produce durable affective states remain underspecified. These findings indicate the need to reposition CX theory toward capturing negative, interaction-level affect, with irritation proposed here as the construct that fills this theoretical gap.

Hyper-Personalization and AI-Driven Interaction Tensions

Hyper-personalization refers to AI-driven, real-time tailoring of content and interactions based on individual-level data. Foundational work on the future of AI in marketing anticipated that this shift would bring not only unprecedented efficiency and personalization capability but also substantial challenges around consumer privacy, autonomy, and the erosion of human judgment in service encounters (Davenport et al., 2020). In retail contexts specifically, AI-driven personalization is projected to reshape the entire customer journey from discovery through post-purchase, yet this transformation simultaneously introduces new points of friction when algorithmic outputs misalign with customer expectations or preferences (Guha et al., 2021). Theoretically, personalization enhances relevance, reduces cognitive effort, and increases perceived value (Hoyer et al., 2020). Research on AI-driven consumer engagement further demonstrates that when personalization succeeds, it can deepen affective attachment and strengthen purchase intentions beyond what functional efficiency alone produces (Bilal et al., 2024).

However, emerging literature highlights a critical paradox: the same mechanisms that enable precision and efficiency may also generate perceptions of surveillance, manipulation, and overreach when personalization becomes excessive or contextually inappropriate (Puntoni et al., 2021). Rather than merely reducing service quality, failures in hyper-personalization introduce a deeper psychological conflict. Customers experience a tension between promised relevance and perceived intrusion, particularly when AI outputs misalign with expectations or reveal excessive data inference (Nguyen et al., 2022). This suggests that hyper-personalization failure should be understood not only as a functional issue but as a violation of autonomy and identity boundaries (Ranieri et al., 2024).

What distinguishes this study's treatment of hyper-personalization failure from prior accounts is the recognition that such violations need not be dramatic to be consequential. A recommendation that is slightly off-target, a message that arrives at an inopportune moment, or a prompt that implies a degree of inference the customer finds unsettling: none of these constitutes a failure in the conventional sense, yet each contributes a small affective cost. It is the accumulation of these costs, rather than any individual incident, that this framework theorizes as the primary driver of irritation-induced value co-destruction.

Critically, AI-driven personalization operates not at a single touchpoint but across the entire customer journey, meaning that these micro-costs accumulate through pre-purchase, purchase, and post-purchase stages simultaneously (Gao & Liu, 2022). This multi-stage exposure amplifies the cumulative

irritation mechanism proposed here, as customers face repeated personalization attempts at every phase of their interaction with the firm.

Expectation–Disconfirmation Theory and Cognitive Triggers of Irritation

Expectation–Disconfirmation Theory (EDT) explains how discrepancies between expected and perceived performance shape customer evaluations (Y. Chen & Prentice, 2025). In AI-mediated CX, expectations are often elevated precisely because automated interactions are designed and marketed as superior to human alternatives (Hollebeek et al., 2021). Meta-analytic evidence further confirms that human acceptance of AI, spanning tools, agents, and intermediary forms, is consistently shaped by elevated performance expectations that, when unmet, generate negative evaluative responses disproportionate to equivalent failures by human agents (Bingqing Li et al., 2026). Notably, AI errors tend to produce stronger reactions than human errors because they violate expectations of technological precision (Hardcastle et al., 2025), and algorithmic failures activate distinct blame attributions precisely because algorithms are held to a standard of objectivity and infallibility that humans are not (Srinivasan & Sarial-Abi, 2021).

In hyper-personalization contexts, this dynamic is amplified in a way that prior EDT applications have not fully theorized. Promises of individual-level relevance create a high benchmark, making even small mismatches highly salient. More importantly, these mismatches are not experienced in isolation: each sub-threshold disconfirmation event lowers the customer's tolerance for subsequent ones. As a result, customers interpret irrelevant outputs as evidence that the system has fundamentally misrecognized them, triggering negative cognitive appraisals that evolve into irritation through repeated exposure rather than through the intensity of any single instance (Gao & Liu, 2022). This is the mechanism through which EDT, originally developed to explain episodic satisfaction judgments, is extended here to account for the temporal dynamics of affective accumulation.

Psychological Reactance Theory and Perceived Intrusion

Psychological Reactance Theory conceptualizes negative affect and resistance as responses to perceived threats to freedom and autonomy. In AI-mediated CX, personalization relies on extensive data capture, classification, and inference about customers' preferences and behaviors (Liu-Thompkins et al., 2022). When these processes become overly visible or misaligned with customer expectations of privacy and control, through over-targeting, unsolicited recommendations, persistent prompts, or excessive familiarity, customers may infer that their behavioral freedoms are being constrained or manipulated (Puntoni et al., 2021). Research on AI-powered advertising and dataveillance emphasizes that hyper-personalized journeys can blur the line between relevance and coercion, undermining transparency and autonomy (Hardcastle et al., 2025). More broadly, scholars have argued that the opacity of algorithmic inference places customers in an asymmetric informational relationship that further intensifies reactance (Puntoni et al., 2021). Research spanning predictive and generative AI contexts further demonstrates that consumer reactions to AI are inherently dual-valenced: the same features that generate fascination and improved decision outcomes simultaneously trigger anxiety, a diminished sense of agency, and identity-related concerns, precisely the conditions under which reactance-driven irritation is most likely to accumulate (Hermann & Puntoni, 2024).

Reactance manifests as defensive responses, including adversarial behaviors toward AI systems, data obfuscation, channel avoidance, and resistance to engagement (Puntoni et al., 2021). CX and chatbot studies document how intrusive data requests, opaque inferences, and forced self-service interactions produce alienation, reduced trust, and abandonment of the customer journey (Ranieri et al., 2024). Reactance-driven disengagement is further amplified when customers perceive that AI systems preempt their choices rather than support them (Hoyer et al., 2020).

The present framework extends Psychological Reactance Theory's application to AI-CX in one important respect. Prior accounts treat reactance as triggered primarily by overt or identifiable intrusions: a notification that reveals data tracking, a recommendation that feels presumptuous, or an explicitly coercive prompt. The irritation construct proposed here captures a more diffuse form of reactance, one that builds incrementally through repeated encounters with personalization that is neither clearly intrusive nor clearly appropriate but that, over time, registers as a cumulative infringement on the customer's sense of autonomy. In this view, irritation is not simply disappointment with performance; it is a motivated affective reaction to perceived control threats inherent in hyper-personalization practices, particularly when customers feel pressured to comply with AI-driven choices or data demands.

Affective Events Theory and the Accumulation of Irritation

Affective Events Theory (AET) posits that discrete events occurring within service environments trigger immediate emotional reactions, which accumulate over time to shape broader attitudes and behavioral responses. Within AI-mediated CX, each interaction with chatbots, recommendation systems, or algorithmic interfaces constitutes an affective event, the valence of which depends on perceived accuracy, responsiveness, empathy, and procedural fairness (Ameen et al., 2021). Empirical research indicates that AI-enabled interactions frequently underperform on affective and relational dimensions. Compared to human service encounters, AI systems are often perceived as impersonal, rigid, and lacking emotional intelligence, resulting in psychological discomfort and reduced perceived warmth (Liu-Thompkins et al., 2022). This affective gap has practical consequences: customers who find AI interactions relationally thin are less likely to remain engaged and more susceptible to switching, even when functional service quality is adequate (Prentice & Nguyen, 2020). These affective deficiencies have been linked to diminished well-being, lower satisfaction, and weakened customer-brand relationships, particularly in contexts where emotional support or nuanced understanding is expected (Castillo et al., 2021).

In hyper-personalization journeys, customer experience is rarely shaped by a single critical failure. Instead, it emerges from a sequence of micro-level interactions, including slightly irrelevant recommendations, repetitive targeting messages, misinterpreted intents, and ineffective chatbot responses. While each individual failure may appear trivial, their cumulative presence generates a pattern of recurring friction within the customer journey. Prior research demonstrates that such repeated inconsistencies increase cognitive load, confusion, and perceived effort, ultimately fostering distrust and disengagement (Ranieri et al., 2024). Corroborating evidence from voice AI contexts further shows that service failures trigger negative customer emotions that, in turn, directly activate complaint behaviors, confirming that affect, not cognition alone, is the proximal driver of post-failure behavioral responses (B. Li et al., 2023). From an AET perspective, these repeated AI misalignments function as negative affective events that progressively accumulate rather than dissipate (Castillo et al., 2021). Unlike isolated service failures, which may be forgiven or rationalized, recurring micro-failures create a compounding emotional burden. Over time, this accumulation transforms situational annoyance into a more stable and persistent affective state.

AET has been applied extensively in organizational behavior and service research, typically to explain how discrete workplace or service events shape employee and customer affect. Its application here is extended in a theoretically specific direction: whereas most AET-based accounts examine events that are discrete and identifiable, this study applies the theory to events that are, individually, beneath the threshold of conscious complaint. This is the distinctive theoretical move: irritation as defined here is not the sum of several grievances large enough to be articulated but the residue of many events small enough to be dismissed in isolation yet consequential in aggregate. This positions the present framework as an extension of AET into the sub-threshold affective domain of AI-mediated service.

Accordingly, customer irritation can be conceptualized as a cumulative affective response emerging from repeated exposure to minor yet unresolved AI-related disruptions. Initially experienced as low-intensity annoyance, these events gradually coalesce into a sustained emotional condition characterized by tension, frustration, and heightened sensitivity toward subsequent interactions. This accumulated irritation not only shapes immediate experiential evaluations but also influences downstream behavioral outcomes, including reduced engagement, avoidance of AI-mediated channels, and an increased propensity to switch or disengage from the service ecosystem (Puntoni et al., 2021). By framing hyper-personalization failures through AET, this study highlights that the dark side of AI-driven CX is not solely rooted in major breakdowns but in the systematic accumulation of small, frequent, and often overlooked interactional flaws, a perspective that shifts the analytical focus from isolated service errors to the temporal dynamics of emotional build-up (Ameen et al., 2021).

Research Methods

Research Design

This study adopts a conceptual research design aimed at developing theory rather than empirically testing causal relationships. Conceptual research advances knowledge by integrating established theories and prior empirical insights into a coherent explanatory framework for emerging phenomena (Jaakkola, 2020; MacInnis, 2011). Accordingly, this study employs theory synthesis and conceptual integration to explain how failures in AI-driven hyper-personalization generate customer irritation as the central affective mechanism underlying value co-destruction in AI-mediated customer experiences.

Literature Identification and Theory Synthesis

The conceptual framework was developed through the identification and synthesis of peer-reviewed literature retrieved from major academic databases, including Scopus, Web of Science, ScienceDirect, SpringerLink, Emerald Insight, and Google Scholar. The reviewed literature covered artificial intelligence, hyper-personalization, customer experience, customer irritation, trust, switching intention, disengagement behavior, and value co-destruction, while incorporating seminal works on Expectation–Disconfirmation Theory, Psychological Reactance Theory, and Affective Events Theory. Following a theory synthesis approach (Jaakkola, 2020), relevant theories and empirical findings were critically examined and integrated to identify conceptual gaps and establish the theoretical mechanisms linking AI-driven hyper-personalization failure with customer irritation and its behavioral consequences.

Conceptual Framework Development

The conceptual framework was developed by integrating Expectation–Disconfirmation Theory, Psychological Reactance Theory, and Affective Events Theory to explain customer responses to AI-driven hyper-personalization failures. The framework positions AI response irrelevance and perceived intrusiveness as antecedents of customer irritation, which subsequently influences trust, switching intention, and disengagement behavior. Based on these theoretically derived relationships, five research propositions, six boundary-condition moderators, and three structural paradoxes were formulated following established theory-building principles (Wacker, 1998), providing a foundation for future empirical research.

Research Process

The research process consisted of five sequential stages: (1) literature identification from reputable academic sources; (2) theory synthesis through critical review and integration of relevant studies; (3) theoretical integration to establish the relationships among the proposed constructs; (4) conceptual framework development and proposition formulation; and (5) derivation of theoretical and managerial implications. The overall research process is illustrated in Figure 1.

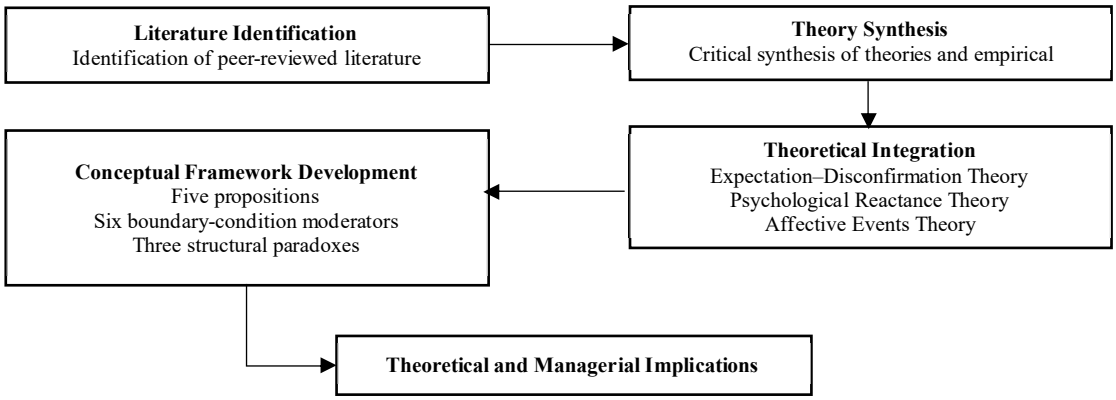


Figure 1. Research Process for Developing the Proposed Conceptual Framework

Results and Discussion

Conceptual Framework

This study develops a conceptual framework explaining the dark-side mechanism of AI in customer experience by modeling how failures in hyper-personalization trigger customer irritation, which subsequently leads to adverse behavioral outcomes. Three key constructs anchor the framework. AI response irrelevance refers to the perceived mismatch between customer needs and AI-generated outputs, reflecting failures in accuracy, contextual fit, and task support (Hoyer et al., 2020). Perceived intrusiveness captures the extent to which AI-driven personalization is experienced as excessive, privacy-invasive, or autonomy-threatening (Puntoni et al., 2021). Customer irritation is conceptualized as a cumulative negative affective state characterized by annoyance, frustration, and agitation arising from repeated or salient AI failures, and is distinguished from frustration and anger by its low intensity, gradual onset, and persistence across interaction episodes rather than its concentration in a single event (Ameen et al., 2021). The framework further incorporates three key outcomes, namely reduced trust, switching

intention, and disengagement/avoidance behavior, representing distinct behavioral manifestations of value co-destruction in AI-mediated CX (Ranieri et al., 2024).

Building on prior research on AI-enabled service failure and value co-destruction, the framework integrates cognitive, motivational, and affective processes linking AI characteristics to customer responses (Castillo et al., 2021). A diagnostic perspective on AI value destruction further supports this framing by demonstrating that specific AI characteristics, including rigidity, opacity, and misalignment with customer goals, can systematically undermine value across inputs, processes, and outcomes rather than merely at moments of overt failure (Canhoto & Clear, 2020). The present framework extends this logic by identifying irritation as the affective mechanism through which such systemic misalignments translate into behavioral harm.

Propositions Development

Grounded in Expectation–Disconfirmation Theory, AI response irrelevance is expected to generate negative evaluations when system outputs fail to meet personalization expectations (Y. Chen & Prentice, 2025). Critically, this effect is theorized here as cumulative rather than episodic: it is the repeated pattern of disconfirmation, not any single instance, that reaches the threshold of irritation. Promises of individual-level relevance elevate the performance benchmark, making even minor mismatches cognitively salient and emotionally consequential. Each sub-threshold failure is interpreted not as an isolated error but as further evidence of systemic misrecognition, progressively intensifying the affective burden on the customer.

P1: AI response irrelevance positively influences customer irritation.

Drawing on Psychological Reactance Theory, perceived intrusiveness is expected to evoke irritation by threatening customers' autonomy and sense of control (Puntoni et al., 2021). Even when individual intrusions fall below the level that would provoke explicit complaint, repeated exposure to such events is theorized to accumulate into a persistent reactance-derived irritation state. Over-targeting, unsolicited inference, and excessive familiarity cumulatively signal to the customer that their behavioral space is being managed rather than supported, a perception that intensifies with each successive encounter.

P2: Perceived intrusiveness positively influences customer irritation.

Customer irritation subsequently shapes downstream evaluations and behaviors through three distinct pathways, each representing a different dimension of relational harm. These pathways are theoretically separable because they operate at different stages of the customer-firm relationship and implicate different psychological mechanisms: trust reflects a cognitive-evaluative judgment, switching intention reflects a motivational-behavioral orientation, and disengagement reflects a habituation-avoidance process. Together, they capture the full range through which accumulated irritation translates into value co-destruction.

Persistent irritation signals misalignment between customer expectations and firm practices, thereby undermining trust and weakening customer-brand relationships over time (Nguyen et al., 2022). Empirical evidence from AI chatbot settings further demonstrates that when service quality deteriorates below expectations, trust declines significantly and customer retention becomes substantially harder to sustain (Q. Chen et al., 2023). Unlike the sharp trust decline triggered by a single high-severity failure, the trust erosion proposed here operates gradually, reflecting the chronic rather than acute nature of irritation-based affect.

P3: Customer irritation negatively influences trust.

Irritation also activates approach–avoidance motivational dynamics, leading customers to evaluate the costs of continued engagement against the perceived effort of switching. Research on AI-enabled service failure and customer behavioral responses indicates that accumulated negative affect raises the psychological threshold for continued engagement, making switching a progressively more attractive alternative as irritation intensifies (Castillo et al., 2021). The frustration-aggression dynamic further reinforces this trajectory: when customers repeatedly encounter aversive stimuli without recourse, redirected behavioral energy manifests as provider substitution (Ozuem et al., 2025).

P4: Customer irritation positively influences switching intention.

Distinct from switching, which implies active re-evaluation and provider change, disengagement and avoidance behavior reflect a more passive withdrawal from AI-mediated interactions within the

existing service relationship. Customers may remain nominally loyal while systematically avoiding the AI channels, recommendation interfaces, or personalization features that have generated irritation. This pattern has been documented in contexts where consumers suppress negative word-of-mouth and mask irritation behind relational attachment, yet concurrently reduce their depth of engagement with AI-mediated touchpoints (B. Huang & Philp, 2021). The behavioral consequence is not departure but diminished interaction quality, a silent erosion of value co-creation potential that firms are poorly positioned to detect through conventional satisfaction metrics.

P5: Customer irritation positively influences disengagement and avoidance behavior.

Table 1 presents the five formal propositions alongside their theoretical grounding and the mechanisms through which each is expected to operate**.**

Table 1. Propositions, theoretical bases, and mechanisms

Proposition	Statement	Theoretical basis	Mechanism
P1	AI response irrelevance positively influences customer irritation.	Expectation–Disconfirmation Theory	Repeated sub-threshold disconfirmation lowers tolerance and escalates cumulative negative affect across successive interactions.
P2	Perceived intrusiveness positively influences customer irritation.	Psychological Reactance Theory	Incremental autonomy threats, each individually below the level of explicit complaint, accumulate into a persistent reactance-derived irritation state.
P3	Customer irritation negatively influences trust.	Affective Events Theory; Value Co-Destruction Theory	Chronic irritation signals systemic misrecognition, gradually eroding cognitive and affective trust rather than triggering acute relational collapse.
P4	Customer irritation positively influences switching intention.	Frustration–Aggression Theory; Affective Events Theory	Accumulated negative affect raises the psychological cost of continued engagement, making provider substitution progressively more attractive.
P5	Customer irritation positively influences disengagement and avoidance behavior.	Affective Events Theory; Psychological Reactance Theory	Customers suppress exit but withdraw from AI-mediated touchpoints, producing a pattern of silent disengagement that conventional satisfaction metrics cannot detect.

Mediating Role of Customer Irritation

A central premise of the framework is that customer irritation operates as the core mediating mechanism linking AI failures to behavioral outcomes. Rather than exerting direct effects, AI response irrelevance and perceived intrusiveness influence customer behavior primarily through irritation. This mediated structure reflects a theoretically important claim: that the behavioral consequences of hyper-personalization failure are not immediate or proportional responses to individual incidents but are instead mediated by an accumulating affective state that amplifies the impact of each subsequent failure.

Specifically, AI failures first trigger cognitive disconfirmation and perceived autonomy threats, which then evolve into irritation (Y. Chen & Prentice, 2025). This affective state subsequently drives reduced trust, switching intention, and disengagement (Ranieri et al., 2024). By positioning irritation as the central causal mechanism, the framework advances prior research that has treated emotions as secondary outcomes rather than primary explanatory drivers (Pavone et al., 2023). The key theoretical advance is the specification of irritation as the proximal affective cause of behavioral outcomes, replacing the implicit assumption in much of the literature that negative CX outcomes are driven by the severity of individual failures rather than the accumulation of minor ones.

Moderating Factors

The relationship between hyper-personalization failure and customer irritation is not uniform across individuals or contexts. Six moderating variables are proposed, each governing the rate at which irritation accumulates or the threshold at which it crosses into behaviorally consequential territory. AI

literacy is expected to weaken the relationship between AI failures and irritation. More knowledgeable users are better able to interpret system limitations and attribute failures to technical constraints rather than to indifference or manipulation (Blut et al., 2021). By providing a benign causal schema for AI errors, higher AI literacy effectively raises the irritation threshold. Privacy concern, by contrast, is expected to strengthen the effect of perceived intrusiveness: highly sensitive individuals are more reactive to data-driven personalization practices and cross the irritation threshold at lower levels of perceived intrusion (Liu-Thompkins et al., 2022). Together, these two moderators define how customers interpret AI behavior at the cognitive level.

Prior experience with AI may operate bidirectionally. Positive prior experiences can buffer irritation by establishing a base of goodwill that absorbs individual failures without triggering cumulative affect, whereas a history of negative interactions lowers the affective threshold and accelerates the accumulation process (Hardcastle et al., 2025). Algorithmic interpretability operates through a related mechanism: when customers can understand why an AI system produced a particular output, they are less likely to attribute irrelevance or intrusiveness to malicious intent, which dampens irritation intensity (C. Chen, 2024). Transparency in AI design is therefore not only an ethical imperative but a functional lever for interrupting the irritation accumulation cycle.

Perceived human-likeness of the AI agent shapes how consumers assign blame and formulate emotional responses when failures occur. Research on humanoid versus non-humanoid service robots shows that less human-like agents tend to redirect blame toward the firm rather than the technology itself (Choi et al., 2021), while evidence on agent type and conversational style demonstrates that design choices around tone, language, and interaction modality meaningfully moderate consumer reactions to service failures (Kuanr et al., 2025). Highly humanized AI interfaces may thus partially deflect the affective burden of micro-failures onto the AI entity, buffering firm-level trust erosion. Finally, consumer–machine relationship orientation, specifically the degree to which customers approach AI agents as quasi-social rather than purely instrumental partners, shapes how ambiguous AI behaviors are interpreted. Customers with more relational orientations are more likely to experience boundary violations as personal affronts, amplifying irritation, while instrumentally oriented customers more readily absorb the same behaviors as technical limitations (Pentina et al., 2023).

Positioning the Framework Against Prior Accounts

The proposed framework is distinguishable from prior accounts of negative affect in AI-CX on three specific dimensions. First, whereas existing frameworks, including Puntoni et al. (2021) experiential perspective and Castillo et al. (2021) co-destruction model, conceptualize negative emotion as a response to identifiable, discrete failures, the present framework theorizes negative affect as emergent from interaction sequences rather than individual events. The unit of analysis is the trajectory, not the episode. Second, whereas prior emotion-focused accounts have examined high-arousal states such as anger and frustration that are immediately salient and self-reported (Pavone et al., 2023), the present framework addresses a low-arousal, sub-threshold state that accumulates beneath the level of articulable complaint, a distinction underscored by meta-analytic evidence showing that different negative emotions follow entirely different antecedent and consequential pathways (Valentini et al., 2020). This distinction has direct methodological implications: irritation requires longitudinal and behavioral measurement strategies that post-encounter surveys are structurally unable to capture. Third, whereas value co-destruction accounts have catalogued outcomes without specifying the affective pathway through which they are produced (Canhoto & Clear, 2020), the present framework supplies this missing mechanism, positioning irritation as the link between AI design features and behavioral harm. Taken together, these three differentiators establish the irritation framework not as a refinement of existing theory but as a structurally distinct account of a previously underspecified phenomenon.

Discussion

Theoretical Contributions

The central theoretical claim of this study is that the predominant account of negative affect in AI-CX, namely episodic, high-intensity, and event-contingent, is not wrong so much as incomplete. It captures one end of the negative affect spectrum while leaving the other end theoretically unoccupied. The three contributions below address that gap by introducing a construct, extending a theory, and naming a set of structural conditions that, taken together, provide a new analytical language for the dark side of AI-mediated customer experience.

The first contribution is the construct of AI-induced customer irritation as a theoretically distinct affective state. The nearest related constructs in prior research are frustration, anger, and dissatisfaction.

Frustration has been documented in response to chatbot failures and scripted rigidity (Pavone et al., 2023), and recent work grounding chatbot service failure recovery explicitly in Frustration–Aggression Theory confirms that the dominant theoretical treatment of negative AI affect has focused on high-arousal, acute emotional responses triggered by identifiable failure events (Ozuem et al., 2025). Anger has been theorized as a high-arousal response to perceived injustice in AI service encounters (Castillo et al., 2021). Dissatisfaction with automated interactions is typically modeled as a cognitive evaluative judgment that emerges when engagement quality falls below expectations (Hollebeek et al., 2021). Irritation, as theorized here, differs from all three on two dimensions: arousal level and temporal structure. It is lower in arousal than frustration and anger, making it less immediately salient but more persistent. It is not a judgment, as dissatisfaction is, but an affective state, one that shapes subsequent judgments rather than constituting a judgment itself. This construct-level differentiation is not terminological but substantive: it implies a different measurement approach (sustained affect scales rather than post-encounter evaluations), different antecedent conditions (micro-failure frequency rather than failure severity), and different intervention logics (interaction design rather than service recovery). Empirical support for this differentiation comes from research showing that different AI types elicit qualitatively distinct emotional profiles, including happiness, fear, and anger, confirming that negative AI affect is not a single undifferentiated state but a family of related yet distinct constructs that require separate theoretical treatment (Pantano & Scarpi, 2022). Meta-analytic evidence from service failure research further substantiates this differentiation: a systematic synthesis of customer emotions across service failure and recovery contexts demonstrates that anger, frustration, and disappointment operate through distinct antecedent pathways and produce meaningfully different behavioral consequences, underscoring that treating negative affect as a unidimensional construct leads to theoretical imprecision and misspecified intervention strategies (Valentini et al., 2020).

The second contribution is the temporal extension of value co-destruction theory in AI-CX. Prior accounts of AI-mediated value co-destruction, most notably Castillo et al. (2021), identify the process through which customers actively or passively contribute to value loss during AI service interactions and treat the individual encounter as the unit of analysis. This account is theoretically adequate for high-salience failures, those that are discrete, attributable, and sufficiently severe to trigger conscious evaluation. What it cannot explain is the value erosion that occurs outside these conditions: the gradual, interaction-by-interaction accumulation of affective residue that leaves no single incident worthy of complaint yet leaves the relationship meaningfully degraded. The present framework shifts the unit of analysis from the interaction to the interaction sequence, arguing that value co-destruction in hyper-personalization contexts is better understood as a cumulative process driven by the affective residue of repeated sub-threshold failures. This temporal reframing has implications not only for theory but for how firms should monitor CX: longitudinal affect tracking and interaction-sequence analysis become more informative than post-encounter satisfaction surveys, and the appropriate intervention point is the quality of ordinary interactions rather than the management of exceptional ones.

The third contribution is the identification of three structural paradoxes, namely personalization versus autonomy, intelligence versus control, and relevance versus intrusiveness, as the generative conditions for irritation-based value co-destruction. Each paradox represents a situation in which the feature of AI-driven personalization that produces its greatest benefit simultaneously creates the conditions for significant harm. Prior research has noted individual instances of these tensions in the context of AI-driven consumer experiences (Puntoni et al., 2021), and separately in the affective and social dimensions of AI-mediated marketing interactions (Liu-Thompkins et al., 2022), but has not synthesized them into a unified paradox-based account of why hyper-personalization systematically generates the conditions for irritation. By naming and formalizing these paradoxes, the present framework provides a theoretical vocabulary for future research to examine where, when, and for whom each paradox is most consequential.

Non-Linear Logic and Core Paradoxes

A central theoretical claim of this study is that hyper-personalization operates through non-linear dynamics, characterized by an inverted U-shaped effect on customer value (Hardcastle et al., 2025). While moderate levels of personalization enhance relevance and efficiency, excessive personalization can lead to perceived overreach, loss of control, and escalating irritation (Puntoni et al., 2021). This non-linearity is reinforced through emotional accumulation mechanisms, whereby irritation lowers customers' tolerance thresholds and amplifies the perceived impact of subsequent failures over time (Ameen et al., 2021). The implication is that the relationship between personalization intensity and customer value is not stable but

dynamic: the same degree of personalization that produces satisfaction in an early interaction may produce irritation after repeated exposure. Within this logic, three core paradoxes emerge.

The Personalization–Autonomy Paradox. Although personalization is intended to enhance relevance, it may simultaneously threaten perceived autonomy (Hoyer et al., 2020). Even when customers formally consent to data use, irritation can arise when personalization is experienced as directive or controlling rather than autonomy-supporting (Puntoni et al., 2021). The paradox is that increasing the precision of personalization, the very feature that is supposed to demonstrate respect for individual preferences, can simultaneously intensify the experience of being managed rather than served.

The Intelligence–Control Paradox. Advances in AI capability are typically associated with improved performance (Ming-Hui Huang & Rust, 2024). However, beyond a certain threshold, heightened predictive accuracy may be interpreted as anticipatory overreach, leading to discomfort and a perceived loss of agency (Hardcastle et al., 2025). The paradox is that the more accurately an AI anticipates a customer's needs, the more the customer may feel that their behavior has been modeled and pre-empted in ways they did not sanction. This dynamic is further complicated by evidence that consumers respond differently to AI and human agents depending on the emotional valence of the interaction: AI is more acceptable as a bearer of unwelcome information precisely because its perceived lack of agency creates emotional distance, yet this same quality renders it less trusted when consumers seek genuine understanding and responsiveness (Garvey et al., 2023). The same non-judgmental quality that makes service robots preferable in embarrassment-inducing encounters simultaneously limits their capacity to build the relational warmth that sustains long-term engagement (Holthöwer & van Doorn, 2023).

The Relevance–Intrusiveness Paradox. Greater relevance does not necessarily translate into more positive experiences (Liu-Thompkins et al., 2022). Instead, highly tailored content may simultaneously signal surveillance or excessive familiarity, thereby triggering perceptions of intrusiveness and irritation (Ranieri et al., 2024). The paradox is that the signal of relevance, content that fits precisely, is the same signal as the mark of surveillance, and customers cannot always distinguish between the two.

Together, these three paradoxes reveal that irritation in hyper-personalized environments is not a failure of execution but a structural consequence of the technology's design logic. Each paradox identifies a condition under which the features most responsible for personalization's value are simultaneously those most responsible for its harm, a relationship that linear models of personalization value are structurally unable to capture and that firms cannot address through service recovery alone.

Boundary Conditions

The proposed framework is most applicable under specific contextual conditions. It is particularly salient in AI-dominant, high-personalization environments, such as recommender systems, algorithmic advertising, and chatbot-driven service interactions, where AI substantially shapes the customer journey and interaction dynamics (Castillo et al., 2021). A systematic review of AI in frontline service encounters confirms that these environments generate the widest range of customer emotional responses, precisely because the stakes of misalignment in terms of perceived competence, empathy, and relevance are highest when AI is the primary customer-facing agent (George et al., 2025). The mechanism is further amplified in hedonic and identity-relevant domains, including fashion, entertainment, and social media, where personalization failures are more likely to be interpreted as misrecognition or violations of personal boundaries (Puntoni et al., 2021).

The framework is also especially relevant among digitally native segments, who simultaneously expect seamless personalization and exhibit heightened sensitivity to data privacy, transparency, and fairness (Hardcastle et al., 2025), and in emerging and digitally transforming markets, where rapid AI adoption intersects with evolving norms around autonomy and data use, as evidenced by research showing divergent AI quality expectations across service contexts in developing economies (Nguyen et al., 2022).

Conversely, the proposed relationships may be weaker in low-involvement, utilitarian contexts; hybrid human–AI service settings; and among customers with low AI literacy or limited privacy concerns, where personalization is less salient and affective responses are less intense (Zaman et al., 2025). The framework is also less applicable in early-stage AI adoption contexts where customer expectations for personalization have not yet been elevated to the level required to produce disconfirmation-driven irritation. These boundary conditions carry direct methodological implications: empirical tests of the framework should prioritize high-personalization, digital-native, hedonic contexts and should measure affect longitudinally rather than through single post-encounter surveys (Aslam, 2023).

Future Research Directions

The proposed framework opens five specific avenues for empirical investigation, each addressing a distinct methodological or substantive gap.

First, longitudinal research designs are needed to capture the temporal accumulation of irritation and identify the threshold at which sub-threshold frictions cross into behaviorally consequential territory. This is not merely a design preference but a theoretical requirement: if irritation is cumulative rather than episodic, as the framework claims, then cross-sectional and single-encounter measurement strategies are structurally unsuited to testing it (Ameen et al., 2021). Panel designs that track affect, tolerance, and behavioral intention across multiple interactions with the same AI system are the appropriate methodological response.

Second, experimental studies using scenario-based or vignette designs can establish causal directionality by systematically manipulating AI response irrelevance and perceived intrusiveness across conditions. Such designs can isolate the mediating role of irritation and test the boundary conditions proposed in §4.3, including the moderating effects of AI literacy and privacy concern (Pavone et al., 2023). Factorial designs crossing antecedent type (irrelevance vs. intrusiveness) with accumulation level (single vs. repeated exposure) would be particularly informative.

Third, behavioral and digital trace data, including clickstreams, chatbot logs, dwell times, and interaction histories, offer opportunities to infer irritation dynamically from behavioral signatures without relying on retrospective self-report. Research demonstrating that AI-triggered discomfort manifests in observable compensatory behaviors supports the feasibility of this approach (Mende et al., 2019), as does evidence that emotion patterns in service interactions can be inferred from interaction data with sufficient precision to predict service outcomes (Bromuri et al., 2020). Developing behavioral proxies for irritation accumulation would address a core limitation of survey-based research on sub-threshold affect.

Fourth, future work should examine how AI-induced irritation interacts with broader consumer decision-making processes. AI increasingly mediates not only service encounters but also product evaluation, comparison, and choice (Jain et al., 2024), meaning that irritation accumulated in one domain may transfer to, or be amplified by, AI-mediated encounters in adjacent domains. Cross-channel irritation transfer is a theoretically underspecified phenomenon with significant practical implications for firms operating integrated AI ecosystems.

Fifth, cross-cultural and multi-level research is needed to examine how institutional contexts, cultural norms, and regulatory environments moderate irritation thresholds and the conditions under which value co-destruction processes unfold (Liu-Thompkins et al., 2022). Cultures that differ in privacy orientation, power distance, and AI trust are likely to exhibit meaningfully different irritation accumulation rates and tipping points, a prediction the present framework generates but cannot, as a conceptual study, empirically verify. Multi-level designs that nest individual customer trajectories within firm-level AI design choices and country-level regulatory regimes would substantially advance this line of inquiry (Gouveia & Santos, 2025). More broadly, as consumer-AI experience emerges as a distinct research domain spanning emotional, relational, and behavioral dimensions across retail and service contexts, future theoretical frameworks must be equipped to capture both the positive and negative experiential trajectories that AI-mediated service generates, of which the irritation construct proposed here represents one foundational component (Noble & Mende, 2023).

Conclusion

Repeated AI-driven hyper-personalization failures rarely provoke immediate complaints, yet their cumulative effect can gradually erode trust, increase switching intention, and encourage customer disengagement. This subtle process has remained underexplored because existing AI-enabled customer experience literature has largely conceptualized negative affect as an episodic response to discrete service failures rather than as a cumulative emotional mechanism.

This study contributes to both theory and practice by conceptualizing customer irritation as the central affective mechanism linking AI-driven hyper-personalization failure to value co-destruction in AI-mediated customer experiences. By integrating Expectation–Disconfirmation Theory, Psychological Reactance Theory, and Affective Events Theory, the proposed framework identifies AI response irrelevance and perceived intrusiveness as key antecedents of customer irritation and advances five research propositions, six boundary-condition moderators, and three structural paradoxes that provide a foundation for future empirical research. Managerially, the findings suggest that organizations should manage not only major AI service failures but also the accumulation of everyday interactional frictions embedded in personalization systems.

This study has several limitations. As a conceptual paper, the proposed framework has not yet been empirically validated, and its proposed relationships have not been examined across different industries or AI-mediated service contexts. Future research is encouraged to test the proposed framework using longitudinal, experimental, or cross-cultural designs to further validate and extend the theoretical model.

References

Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. *Computers in Human Behavior*, *114*(September 2020), 106548. https://doi.org/10.1016/j.chb.2020.106548

Aslam, U. (2023). Understanding the usability of retail fashion brand chatbots: Evidence from customer expectations and experiences. *Journal of Retailing and Consumer Services*, *74*, 103377. https://doi.org/10.1016/j.jretconser.2023.103377

Becker, L., & Jaakkola, E. (2020). Customer experience: fundamental premises and implications for research. *Journal of the Academy of Marketing Science*, *48*, 630–648. https://doi.org/10.1007/s11747-019-00718-x

Belanche, D., Belk, R. W., Casaló, L. V, & Flavián, C. (2024). The dark side of artificial intelligence in services. *The Service Industries Journal*, *44*(3–4), 149–172. https://doi.org/10.1080/02642069.2024.2305451

Bilal, M., Zhang, Y., Cai, S., Akram, U., & Halibas, A. (2024). Artificial intelligence is the magic wand making customer-centric a reality! An investigation into the relationship between consumer purchase intention and consumer engagement through affective attachment. *Journal of Retailing and Consumer Services*, *77*, 103674. https://doi.org/10.1016/j.jretconser.2023.103674

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y

Bromuri, S., Henkel, A. P., Iren, D., & Urovi, V. (2020). Using AI to predict service agent stress from emotion patterns in service interactions. *Journal of Service Management*, *32*(4), 581–611. https://doi.org/10.1108/JOSM-06-2019-0163

Canhoto, A., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, *63*, 183–193. https://doi.org/10.1016/j.bushor.2019.11.003

Castillo, D., Canhoto, A. I., & Said, E. (2021). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. *The Service Industries Journal*, *41*(13–14), 900–925. https://doi.org/10.1080/02642069.2020.1787993

Chen, C. (2024). How consumers respond to service failures caused by algorithmic mistakes: The role of algorithmic interpretability. *Journal of Business Research*, *176*, 114610. https://doi.org/10.1016/j.jbusres.2024.114610

Chen, Q., Lu, Y., Gong, Y., & Xiong, J. (2023). Can AI chatbots help retain customers? Impact of AI service quality on customer loyalty. *Internet Research*, *33*(6), 2205–2243. https://doi.org/10.1108/INTR-09-2021-0686

Chen, Y., & Prentice, C. (2025). Integrating Artificial Intelligence and Customer Experience. *Australasian Marketing Journal*, *33*(2), 141–153. https://doi.org/10.1177/14413582241252904

Choi, Sungwoo, Mattila, Anna S, & Bolton, Lisa E. (2021). To Err Is Human(-oid): How Do Consumers React to Robot Service Failure and Recovery? *Journal of Service Research*, *24*(3), 354–371. https://doi.org/10.1177/1094670520978798

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, *117*, 587–595. https://doi.org/10.1016/j.jbusres.2018.10.004

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

Gao, Y., & Liu, H. (2022). Artificial intelligence-enabled personalization in interactive marketing: a customer journey perspective. *Journal of Research in Interactive Marketing*, *17*(5), 663–680. https://doi.org/10.1108/JRIM-01-2022-0023

Garvey, Aaron M, Kim, TaeWoo, & Duhachek, Adam. (2023). Bad News? Send an AI. Good News? Send a Human. *Journal of Marketing*, *87*(1), 10–25. https://doi.org/10.1177/00222429211066972

George, S. R., C, M., & Edward, M. (2025). Artificial Intelligence in Frontline Service Encounters: A

Systematic Review and Research Agenda. *International Journal of Consumer Studies*, *49*(3), e70048. https://doi.org/10.1111/ijcs.70048

Gouveia, J., & Santos, S. (2025). Rethinking the Customer Journey: Impact of AI for Consumers and Businesses—A Systematic Literature Review and Research Agenda. *International Journal of Consumer Studies*, *49*(6), e70136. https://doi.org/https://doi.org/10.1111/ijcs.70136

Guha, A., Grewal, D., Kopalle, P. K., Haenlein, M., Schneider, M. J., Jung, H., Moustafa, R., Hegde, D. R., & Hawkins, G. (2021). How artificial intelligence will affect the future of retailing. *Journal of Retailing*, *97*(1), 28–41. https://doi.org/10.1016/j.jretai.2021.01.005

Hardcastle, K., Vorster, L., & Brown, D. (2025). Understanding Customer Responses to AI-Driven Personalized Journeys: Impacts on the Customer Experience. *Journal of Advertising*, *54*, 176–195. https://doi.org/10.1080/00913367.2025.2460985

Hermann, E., & Puntoni, S. (2024). Artificial intelligence and consumer behavior: From predictive to generative AI. *Journal of Business Research*, *180*, 114720. https://doi.org/10.1016/j.jbusres.2024.114720

Hollebeek, Linda D, Sprott, David E, & Brady, Michael K. (2021). Rise of the Machines? Customer Engagement in Automated Service Interactions. *Journal of Service Research*, *24*(1), 3–8. https://doi.org/10.1177/1094670520975110

Holthöwer, J., & van Doorn, J. (2023). Robots do not judge: service robots can alleviate embarrassment in service encounters. *Journal of the Academy of Marketing Science*, *51*(4), 767–784. https://doi.org/10.1007/s11747-022-00862-x

Hoyer, W., Kroschke, M., Schmitt, B., Kraume, K., & Shankar, V. (2020). Transforming the Customer Experience through New Technologies. *Journal of Interactive Marketing*, *51*, 57–71. https://doi.org/10.1016/j.intmar.2020.04.001

Huang, Ming-Hui, & Rust, Roland T. (2024). The Caring Machine: Feeling AI for Customer Care. *Journal of Marketing*, *88*(5), 1–23. https://doi.org/10.1177/00222429231224748

Huang, B., & Philp, M. (2021). When AI-based services fail: examining the effect of the self-AI connection on willingness to share negative word-of-mouth after service failures. *The Service Industries Journal*, *41*(13–14), 877–899. https://doi.org/10.1080/02642069.2020.1748014

Jaakkola, E. (2020). Designing conceptual articles: four approaches. *AMS Review*, *10*(1), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Jain, V., Wadhwani, K., & Eastman, J. K. (2024). Artificial intelligence consumer behavior: A hybrid review and research agenda. *Journal of Consumer Behaviour*, *23*(2), 676–697. https://doi.org/https://doi.org/10.1002/cb.2233

Kuanr, A., Moharana, T. R., Yan, M., Pradhan, D., & El-Manstrly, D. (2025). When chatbots cause trouble: how agent type and conversation style shape consumer responses to service failures. *European Journal of Marketing*, *59*(12), 2788–2839. https://doi.org/10.1108/EJM-08-2023-0663

Li, Bingqing, Lai, Edward Yuhang, & Wang, Xin (Shane). (2026). From Tools to Agents: Meta-Analytic Insights into Human Acceptance of AI. *Journal of Marketing*, *90*(3), 13–33. https://doi.org/10.1177/00222429251355266

Li, B., Liu, L., Mao, W., Qu, Y., & Chen, Y. (2023). Voice artificial intelligence service failure and customer complaint behavior: The mediation effect of customer emotion. *Electronic Commerce Research and Applications*, *59*, 101261. https://doi.org/https://doi.org/10.1016/j.elerap.2023.101261

Liu-Thompkins, Y., Okazaki, S., & Li, H. (2022). Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *Journal of the Academy of Marketing Science*, *50*(6), 1198–1218. https://doi.org/10.1007/s11747-022-00892-5

MacInnis, Deborah J. (2011). A Framework for Conceptual Contributions in Marketing. *Journal of Marketing*, *75*(4), 136–154. https://doi.org/10.1509/jmkg.75.4.136

Mende, Martin, Scott, Maura L, van Doorn, Jenny, Grewal, Dhruv, & Shanks, Ilana. (2019). Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses. *Journal of Marketing Research*, *56*(4), 535–556. https://doi.org/10.1177/0022243718822827

Mustak, M., Salminen, J., Plé, L., & Wirtz, J. (2021). Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. *Journal of Business Research*, *124*, 389–404. https://doi.org/https://doi.org/10.1016/j.jbusres.2020.10.044

Nguyen, T.-M., Quach, S., & Thaichon, P. (2022). The effect of AI quality on customer experience and brand relationship. *Journal of Consumer Behaviour*, *21*(3), 481–493. https://doi.org/10.1002/cb.1974

Noble, S. M., & Mende, M. (2023). The future of artificial intelligence and robotics in the retail and service sector: Sketching the field of consumer-robot-experiences. *Journal of the Academy of Marketing Science*, *51*(4), 747–756. https://doi.org/10.1007/s11747-023-00948-0

Ozuem, W., Ranfagni, S., Willis, M., Salvietti, G., & Howell, K. (2025). Chatbots, service failure recovery, and online customer experience through lenses of frustration – aggression theory and signaling theory. *Journal of Services Marketing*, *39*(5), 493–512. https://doi.org/10.1108/JSM-10-2024-0525

Pantano, Eleonora, & Scarpi, Daniele. (2022). I, Robot, You, Consumer: Measuring Artificial Intelligence Types and their Effect on Consumers Emotions in Service. *Journal of Service Research*, *25*(4), 583–600. https://doi.org/10.1177/10946705221103538

Pavone, Giulia, Meyer-Waarden, Lars, & Munzel, Andreas. (2023). Rage Against the Machine: Experimental Insights into Customers' Negative Emotional Responses, Attributions of Responsibility, and Coping Strategies in Artificial Intelligence–Based Service Failures. *Journal of Interactive Marketing*, *58*(1), 52–71. https://doi.org/10.1177/10949968221134492

Pentina, I., Xie, T., Hancock, T., & Bailey, A. (2023). Consumer–machine relationships in the age of artificial intelligence: Systematic literature review and research directions. *Psychology & Marketing*, *40*(8), 1593–1614. https://doi.org/10.1002/mar.21853

Prentice, C., & Nguyen, M. (2020). Engaging and retaining customers with AI and employee service. *Journal of Retailing and Consumer Services*, *56*, 102186. https://doi.org/10.1016/j.jretconser.2020.102186

Puntoni, Stefano, Reczek, Rebecca Walker, Giesler, Markus, & Botti, Simona. (2021). Consumers and Artificial Intelligence: An Experiential Perspective. *Journal of Marketing*, *85*(1), 131–151. https://doi.org/10.1177/0022242920953847

Ranieri, A., Di Bernardo, I., & Mele, C. (2024). Serving customers through chatbots: positive and negative effects on customer experience. *Journal of Service Theory and Practice*. https://doi.org/10.1108/jstp-01-2023-0015

Srinivasan, Raji, & Sarial-Abi, Gülen. (2021). When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors. *Journal of Marketing*, *85*(5), 74–91. https://doi.org/10.1177/0022242921997082

Valentini, S., Orsingher, C., & Polyakova, A. (2020). Customers' emotions in service failure and recovery: a meta-analysis. *Marketing Letters*, *31*(2), 199–216. https://doi.org/10.1007/s11002-020-09517-9

Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of Operations Management*, *16*(4), 361–385. https://doi.org/10.1016/S0272-6963(98)00019-9

Zaman, M., Mohamed Jasim, K., Hasan, R., Akter, S., & Vrontis, D. (2025). Understanding customers' intentions to use AI-enabled services in online fashion stores – a longitudinal study. *International Marketing Review*, *42*(4), 604–632. https://doi.org/10.1108/IMR-04-2024-0118