

Journal Pre-proof

How to manage AI and human input in customer feedback systems without eroding empathy and trust

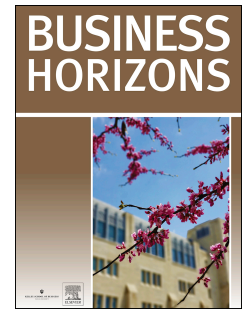
Bodicia Washington, Research Associate, Omar Merlo, Associate Professor of Marketing Strategy, Valentina Prados Spitaleri, Andreas Eisingerich, Professor of Marketing

PII: S0007-6813(26)00078-9

DOI: <https://doi.org/10.1016/j.bushor.2026.05.012>

Reference: BUSHOR 2151

To appear in: *Business Horizons*



Please cite this article as: Washington B., Merlo O., Prados Spitaleri V. & Eisingerich A., How to manage AI and human input in customer feedback systems without eroding empathy and trust, *Business Horizons*, <https://doi.org/10.1016/j.bushor.2026.05.012>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Kelley School of Business, Indiana University. Published by Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

How to manage AI and human input in customer feedback systems without eroding empathy and trust

Bodicia Washington

Research Associate
Imperial Business School, Imperial College
London SW7 2AZ, United Kingdom
bodiciawashington@outlook.com

Omar Merlo*

Associate Professor of Marketing Strategy
Imperial Business School, Imperial College, London
London SW7 2AZ, United Kingdom
o.merlo@imperial.ac.uk

Valentina Prados Spitaleri

Research Assistant
Imperial Business School, Imperial College, London
London SW7 2AZ, United Kingdom
valentina.prados-spitaleri25@imperial.ac.uk

Andreas Eisingerich

Professor of Marketing
Imperial Business School, Imperial College, London
London SW7 2AZ, United Kingdom
a.eisingerich@imperial.ac.uk

* Contact author

How to manage AI and human input in customer feedback systems without eroding empathy and trust

Abstract

Artificial intelligence (AI) is increasingly embedded in customer feedback systems, promising speed, scale, and predictive insight. Yet many organizations struggle to integrate AI without eroding empathy, trust, and relational quality. This paper examines how managers balance AI-driven efficiency with human judgment in customer feedback management. Drawing on twenty semi-structured interviews across government, healthcare, marketing, finance, and technology sectors, we surface five interconnected findings: AI's operational power as a high-volume extraction engine; its structural limitation of nuance blindness in contextual interpretation; the counterintuitive finding that human involvement introduces its own distortions and is not a reliable corrective; AI's underutilized role as a governance stabilizer that depoliticizes decisions and legitimizes authority; and AI's upstream influence in reshaping the generative conditions of feedback itself, altering what people are willing to say and to whom. Building on these insights, we develop a structured hybrid playbook outlining how managers can design, govern, and audit human–AI feedback systems. The playbook offers actionable guidance on task allocation, escalation protocols, accountability roles, and governance mechanisms that allow firms to scale intelligence without sacrificing empathy. The findings position AI not as a replacement for human judgment, but as a complementary engine for service improvement when deliberately orchestrated.

KEYWORDS: Artificial intelligence; Customer feedback; Human–AI collaboration; Hybrid systems; Service management

1. Introduction

Artificial intelligence (AI) is transforming how organizations capture, interpret, and act on customer feedback. Once limited to automating surveys or analyzing satisfaction data, AI is now embedded across the customer journey, from service interactions to predictive insights and decision support (Ameen et al. 2021; Mattingly et al., 2025; Samson et al., 2022). This rapid diffusion presents managers with a central question: how can AI be integrated into feedback systems without eroding the empathy, trust, and contextual understanding that define high-quality service relationships?

AI-driven feedback systems promise efficiency, scalability, and analytic precision. They can process vast amounts of textual and behavioral data in real time, uncovering patterns that would escape human analysis. Yet over-automation can damage relational value. When firms such as Klarna replaced customer-service staff with AI and later reversed course, with the CEO publicly acknowledging the move had “gone too far” and announcing a recruitment drive so customers could always reach a human (Bloomberg, 2025), it became clear that technical capability alone does not guarantee perceived quality or customer trust. The pattern is not isolated: Gartner (2025) predicts that 50% of organizations planning to reduce customer-service headcount through AI will abandon those plans by 2027, citing concerns about service quality, complaint volume, and customer experience. The challenge is therefore not whether to use AI, but how to balance machine accuracy with human empathy in ways that enhance service improvement.

Although research on AI in service contexts is developing, it remains fragmented. Extant research highlights AI's operational advantages in speed and consistency (e.g., Cubric, 2020; Jarrahi et al., 2023) and the human strengths of judgment, emotional intelligence, and contextual reasoning (e.g., Fürst et al., 2025; Puntoni et al., 2021). What this literature lacks is operational guidance: most frameworks treat AI and human input as substitutes rather than as roles to be deliberately allocated within a single feedback loop. Hence, three specific research gaps remain: first, the assumption that human involvement reliably corrects AI's limitations has not been tested against practitioners' own accounts of how human judgment might fail; second, AI's role in shaping internal decision-making, beyond analysis, has been largely unexamined; and third, AI's effect on the upstream conditions of feedback generation has been overlooked: existing research treats feedback as an input AI processes, rather than as a process AI shapes.

This paper addresses these gaps by asking: how do managers design and govern hybrid human-AI feedback systems in practice, and what conditions determine when AI input, human input, or a

combination of the two creates value? Drawing on established frameworks in service quality (Parasuraman et al., 1988) and service innovation (Berry et al., 2020), and on twenty semi-structured interviews with professionals across government, healthcare, finance, marketing, and technology sectors, we derive cross-sector boundary conditions that shape when AI or human input is most effective, and we translate these conditions into a five-rule managerial playbook for hybrid system design.

Our findings make three key contributions. Empirically, they provide cross-sector evidence that complicates the conventional efficiency versus empathy binary: not only does AI exhibit structural contextual limitations, but human input introduces its own distortions, and AI performs governance and generative functions that are widely underrecognized in practice. Conceptually, the findings advance a more differentiated complementarity model that identifies the conditions in which AI and human judgment each create and destroy value, moving beyond the general claim that the two are complementary to explain when, how, and why the boundary between them should be actively managed. Managerially, we offer a hybrid playbook comprising five design rules grounded in practitioner accounts, providing actionable guidance on task allocation, escalation protocols, governance stabilization, and continuous auditing. The next section reviews the evolving literature on AI's role in service decision-making and its implications for customer insight, empathy, and organizational learning.

2. The case for AI: Efficiency, prediction, and the new logic of service decisions

AI offers a strategic advantage in the customer feedback loop, delivering scale, objectivity, and efficiency that often complement or exceed human capabilities. Recent research suggests AI is no longer confined to a purely functional role: rather than serving only as a support tool, it is increasingly framed as an autonomous decision-making agent (Mortati & Viana Mundstock Freitas, 2025). This is a shift in decision authority that raises new managerial questions about where human oversight remains essential, making AI's comparative advantage in efficiency and prediction central to how organizations balance technological and human decision inputs.

AI's value is most evident in high-volume, structured tasks, where it improves both speed and consistency, enabling large-scale feedback analysis while reducing human error. Cubric (2020) identifies increased productivity, error reduction, improved forecasting, and cognitive decision support as core drivers of AI adoption. Extending this logic, Katsiuba et al. (2026) show that AI's

contribution to customer feedback management evolves through a multi-phase process: modular tools such as automated sentiment analysis and difficulty scoring enable organizations to detect issues and opportunities that may be overlooked in human-authored text, strengthening sensing and prioritization capabilities. These gains translate into tangible strategic benefits by allowing firms to automate knowledge-intensive processes at near-zero marginal cost, thereby enhancing returns on R&D investments (Mattingly et al., 2025).

Beyond automation, AI can shift organizations from reactive to proactive service logics. In the customer feedback loop, real-time automation and prediction enable firms to anticipate dissatisfaction rather than merely respond to it (Samson et al., 2022), addressing a persistent challenge of data underutilization, with an estimated 60–70% of collected data left unused (Jarrahi et al., 2023). By optimizing workflows and automating repetitive tasks, AI augments worker capacity, reduces fatigue, and supports higher-quality service delivery (Bankins et al., 2024). These predictive capabilities position AI as a driver of continuous service improvement, while also revealing the limits of purely data-driven enhancement in emotionally sensitive service contexts.

In summary, AI is a powerful resource for service decision-making, especially effective in structured, data-driven contexts. Yet its potential is fully realized only when set against an honest account of where human judgment remains essential, and where that judgment itself may fall short.

3. The human advantage: Judgment, empathy, and the limits of AI

Despite clear gains in efficiency, scale, and objectivity, AI adoption across service sectors remains cautious. This reflects a managerial paradox in which strong economic potential is constrained by relational and perceptual barriers. In automated sales calls, disclosing a chatbot's identity before interaction reduces purchase rates by 79.7% (Luo et al., 2019). Crucially, this resistance is perceptual rather than technical: customers judge AI agents as less knowledgeable and empathic even when their objective performance matches that of high-performing humans (Luo et al., 2019). Social factors further amplify resistance, including fears related to job security, understanding, and distrust (Cubric, 2020), as well as algorithmic aversion among highly expert employees driven by heightened accountability concerns or perceived judgmental superiority (Bankins et al., 2024, p. 164). On the customer side, repetitive or ineffective AI responses intensify frustration, reduce trust, and increase service abandonment (Yeh et al., 2025). Such failures, whether technical or perceptual, can trigger strong negative emotions, including anger and resentment (Guo et al., 2025). This

tension between technical competence and relational resistance reveals a persistent gap: while AI excels at process optimization, it often lacks the emotional and relational capabilities required to effectively close the service loop.

The managerial imperative is to recognize how AI's expanding competencies, spanning logic-mathematical, social, and processing-speed intelligence, enable increasingly complex service tasks, including high-risk functions such as service recovery (Ameen et al., 2025). Effective service improvement depends on discerning when machine competence is sufficient and when human warmth is indispensable. From a managerial perspective, this requires a clear division between strategic and tactical responsibilities. Humans must retain control over strategic domains such as employee motivation, value proposition design, and the communication of empathy and accountability, all of which rely on emotional intelligence and contextual sensitivity (Hermann & Puntoni, 2025). AI is better suited to tactical and operational tasks, including repetitive marketing activities and low-risk service requests. However, successful recovery cannot be determined by managerial logic alone; outcomes ultimately depend on how customers perceive fairness, authenticity, and emotional sincerity in AI-mediated interactions.

Empirical research clarifies the conditions under which customers prefer humans or AI. Fürst et al. (2025) show that service recovery preferences depend on three factors: the locus of failure (customer versus firm responsibility), the type of symbolic recovery (explanation versus apology), and the type of utilitarian recovery (monetary versus functional redress). Customers tend to favor AI when failures are self-attributed and when objective, utilitarian outcomes such as monetary compensation are desired. By contrast, when the firm is at fault and customers seek symbolic redress or emotional reassurance, they prefer human agents capable of conveying genuine empathy, particularly during process failures that undermine social resources (Huang & Lo, 2025; Lee, 2024). Generative AI complicates this distinction by increasingly simulating emotional expression and conversational warmth, yet practical limitations persist, including weak multilingual performance, poor interpretation of nonverbal cues, and inconsistencies that frustrate customers and undermine trust (Rana et al., 2025; Wen & Laporte, 2025).

This evolution introduces a paradox: as AI becomes more human-like, customers evaluate it according to human relational norms. Hao et al. (2025) show that attempts by large language models to express empathy in emotionally charged interactions can trigger identity-contingent trust violations. Consumers remain concerned about privacy and AI-driven tools (Herriger et al., 2025; Selim et al., 2024) and place high value on authenticity (Duffek et al., 2025), genuine storytelling

(Eisingerich et al., 2023; Tam et al., 2026), and transparency (Merlo et al., 2023, 2024). Ding et al. (2025) further identify a “collaboration trap,” whereby increased AI involvement erodes profitability when consumers are highly sensitive to service agent identity.

At the same time, well-designed AI-driven empathy can generate positive relational outcomes. Empathetic AI perceived as sufficiently human-like can function as a social model, eliciting empathy and encouraging prosocial behavior, particularly when supported by human cues such as names or avatars (Shi et al., 2025). However, heightened anthropomorphism also raises risk by elevating service recovery expectations and reducing forgiveness during process failures (Huang & Lo, 2025). The managerial question therefore shifts from whether AI can express empathy to identifying the conditions under which artificial empathy builds trust rather than undermines it. Overall, the literature suggests that human judgment remains essential for resolving emotionally complex service failures, while elements of relational intelligence can be embedded in carefully designed AI systems.

Yet the human advantage cannot be assumed to be always reliable. A growing body of research shows that professional human judgment is subject not only to bias but also to noise: unwanted variability in judgments of the same case by different people, or by the same person at different times (Kahneman et al., 2021). Noise may be driven by seemingly irrelevant factors such as mood, fatigue, or the order in which information is encountered, and it can produce inconsistency that is largely invisible to organizations but consequential in its effects. Crucially, Kahneman et al. (2021) emphasize that mechanical rules are often more consistent than human judgment, a finding that directly challenges the assumption that expert human involvement constitutes a reliable correction to AI's limitations.

This conclusion is echoed by Dietvorst, Simmons, and Massey (2015), who show that people abandon algorithmically superior evidence after observing a single algorithmic error, even when the algorithm continues to outperform human judgment. This is a phenomenon they term algorithm aversion. Paradoxically, the reverse occurs among non-experts: laypeople adhere more strongly to algorithmic advice than to human advice, a tendency termed algorithm appreciation (Logg et al., 2019). Experienced professionals, by contrast, rely less on algorithmic input than laypeople do, to the detriment of their accuracy. Together, these findings suggest that the relationship between human judgment and AI output is very context-dependent, shaped by expertise, institutional culture, and the visibility of algorithmic error. The human advantage in service contexts is real, but

conditional. And in certain conditions, human involvement may introduce distortions that structured AI evaluation does not.

4. Hybrid human-AI systems: Augmentation and trust

Extant research points to four key principles. First, complementarity refers to the principle that AI and human inputs each contribute capabilities the other lacks, e.g., speed and pattern detection on one side, contextual judgment and relational accountability on the other (Paschen et al., 2020). Second, augmentation refers to the design choice to enhance human work with AI rather than substitute for it, maintaining human authority over a final decision (Jarrahi, 2018). Third, hybrid systems are operational architectures that allocate specific tasks to AI, human, or combined effort according to specific rules of escalation and oversight. Fourth, orchestration is the managerial activity of coordinating these allocations over time, sequencing, monitoring, and recalibrating them as contexts change. With these distinctions in place, research increasingly shows that effective service improvement depends not on replacing humans with AI, but on strategic augmentation (Jarrahi, 2018). Collaboration models with substantial AI involvement, including low human/high tech and high human/high tech configurations, generate stronger outcomes than full automation (Le et al., 2025), and emerging frameworks define their core function as the coordinated production of decisions through human-AI teaming (Mortati & Viana Mundstock Freitas, 2025; Simón et al., 2024).

Adoption typically happens in stages, along dimensions of automation, augmentation, and data richness, and managers need tools such as decision canvases and cross-organizational data pooling to navigate the trade-offs at each stage (Holmström & Magnusson, 2026; Bagchi & Sharma, 2026). Several frameworks operationalize how AI and humans should work together. For example, Blaurock et al. (2025) build collaborative intelligence around four design principles: transparency, process control, outcome control, and reciprocal strength enhancement. Ameen et al. (2025) propose matching AI capabilities to the cognitive demands of specific service stages. Trust is a separate design constraint: delaying AI disclosure until users have experienced its competence reduces the disclosure penalty (Luo et al., 2019). But trust alone is not sufficient. AI remains constrained by structured data and historical dependence, limiting intuition, creativity, and moral judgment (Jarrahi, 2018; Porkodi & Cedro, 2025), so its outputs are best treated as input to a decision, not as instruction.

This complementarity reflects the distinction between information and knowledge: AI generates information, while humans interpret context, ethics, and significance to produce actionable knowledge (Paschen et al., 2020). Operationally, AI manages analytical complexity while humans address uncertainty and ambiguity, with human interpretation remaining essential regardless of the level of automation (Jarrahi, 2018).

As AI embeds more deeply in service systems, governance and explainability become critical. Opaque outputs undermine trust and accountability, creating the explainability problem (Chui et al., 2018), necessitating transparency, interpretability, and oversight (Hermann & Puntoni, 2025; Pillai, 2024). Explainable AI transforms black-box models into glass-box systems that improve acceptance of unfavorable outcomes and reduce employee alienation (Sahoh & Choksuriwong, 2023; Blaurock et al., 2025). Ethical governance is equally essential to prevent ethical inflexibility and align AI-assisted decisions with organizational and societal values (Cubric, 2020; Krotov, 2025).

Collectively, the literature highlights a clear trade-off: AI delivers predictive efficiency and analytical competence, while humans provide trust, empathy, and moral judgment essential for relational service delivery (Fürst et al., 2025). Service robot research reinforces these conclusions, showing that human–robot pairing compensates for functional and emotional limitations (Rana et al., 2025; Blut et al., 2021), that message congruence improves effectiveness (Nguyen et al., 2025; Yeh et al., 2025), and that urgency requires dynamic shifts from empathic to utilitarian responses (Guo et al., 2025; Zhu et al., 2018). Emerging guidance emphasizes embedding process and outcome control to strengthen responsibility and brand authenticity (Blaurock et al., 2025; Liu et al., 2025), continuous learning for system evolution (Simón et al., 2024), investment in employee learning and development to address generative AI implementation challenges (Alhusban et al., 2025), and human contextualization of generative AI outputs (Porkodi & Cedro, 2025).

An emerging stream of organizational research highlights a further dimension of AI's role in hybrid systems: its capacity to function not merely as an analytical tool, but as a structural governance mechanism. Algorithmic systems are increasingly used to direct, evaluate, and discipline workers in ways that remove human decision-making from control structures (Kellogg et al., 2020). Algorithmic control is more encompassing, instantaneous, and disintermediating than human-led oversight. These are properties that reduce opportunities for appeals to human empathy or for rule exceptions to be granted (Kellogg et al., 2020). While Kellogg et al. frame this as a concern for worker autonomy, the same structural properties have a managerial counterpart: in contexts where human authority hierarchies distort the flow of information: suppressing dissent, rewarding

sycophancy, or filtering feedback through political relationships. AI's capacity to bypass those hierarchies and provide a reference point perceived as neutral becomes a governance resource. This positions AI not only as a tool for processing customer feedback, but as a mechanism for stabilizing internal decision-making in politically constrained organizations.

A final dimension concerns AI's influence on the upstream conditions of feedback generation itself. Research in computer-mediated communication and human-computer interaction consistently finds that the medium through which feedback is solicited shapes the honesty and completeness of what is disclosed. For instance, Joinson (2001) demonstrates that computer-mediated communication produces higher levels of spontaneous self-disclosure than face-to-face interaction, driven by reduced public self-awareness and the sense of visual anonymity that digital interfaces provide. Lucas et al. (2014) extend this logic to automated interview contexts, showing that participants interacting with a virtual human they believed to be computer-operated reported lower fear of self-disclosure and lower impression management than those who believed a human was present. This mechanism was explained by the absence of perceived judgment, allowing for more candid responses. These findings raise a consequential question for customer feedback system design: if the medium of feedback collection systematically shapes what people are willing to say, organizations relying exclusively on human-mediated channels may be capturing a structurally filtered version of customer sentiment. Recent research confirms that this pattern extends to conversational AI specifically: customers and users disclose differently to chatbot interlocutors than to humans, in ways shaped by perceived anonymity, judgment, and interface design (Croes et al., 2024; Papneja & Yadav, 2024).

In sum, the managerial challenge is no longer deploying AI efficiently, but orchestrating human-AI relationships effectively. This study therefore examines how managers assess relative strengths, determine leadership and intervention conditions, and govern hybrid feedback systems to enhance service outcomes.

5. Methodology

This study adopts an exploratory qualitative design using semi-structured, in-depth interviews to examine how managers perceive and enact human-AI complementarities in customer feedback systems. The semi-structured format combined a common discussion guide (covering current feedback architecture, decision points where AI or human input was prioritized, observed failure

modes, governance practices, and trust calibration) with open-ended probes that allowed unanticipated patterns to surface. A qualitative approach was chosen because complementarity involves tacit judgment, accountability, and relational risk that are difficult to capture through standardized research methods.

We used purposive maximum-variation sampling to recruit twenty professionals with direct strategic or operational responsibility for customer feedback systems across government, finance, healthcare, technology, and marketing. This sectoral spread enabled cross-context comparison of how task allocation, trust, and governance are interpreted under different operational, ethical, and resource constraints (e.g., regulated public-sector settings versus agile technology firms). Saturation was assessed continuously rather than declared at a fixed sample size: we tracked the rate at which new first-order codes emerged after each interview, with no new codes added after the seventeenth interview and the final three confirming thematic stability (Guest, Bunce, & Johnson, 2006). Participant profiles are summarized in Table 1.

[Insert Table 1 About Here]

Interviews were conducted online between June and July 2025 and lasted 20–25 minutes. The interview window was a function of practitioner availability: participants held senior or operational roles in active organizations and accepted on the condition that interviews would not exceed half an hour. We addressed this constraint analytically by tightening the discussion guide around the study's focal constructs and by recruiting a relatively large sample ($n = 20$) to compensate for shorter individual interviews. The study followed the Market Research Society Code of Conduct (Market Research Society, 2023), which sets out professional standards for participant consent, confidentiality, data handling, and the responsible representation of findings. All participants provided informed consent; interviews were audio-recorded and transcribed verbatim. Identifying details (i.e., participants' names, employer names, distinctive project descriptors, and any client identifiers raised in conversation) were redacted at transcription, and participants are referred to throughout the manuscript by anonymized identifiers (P1–P20) coupled only with their broad role and sector.

Analysis followed a three-cycle thematic process (Saldaña, 2021). The first cycle, open coding, stayed close to participants' own language and produced 184 first-order codes (e.g., “manual verification,” “emotive nuance,” “operational speed”). The second cycle, axial coding, consolidated these into 26 categories organized around recurrent relationships, including how perceived machine

accuracy shaped required levels of human oversight. The third cycle integrated a deductive lens drawn from established constructs (i.e., complementarity, task allocation, trust calibration, disclosure expectations, and governance) mapping the inductive categories onto these dimensions to derive the five cross-cutting themes reported in the findings. To support coding consistency, two members of the research team reviewed the codebook and a subset of transcripts together, discussing coding decisions and resolving disagreements through team discussion, with the codebook updated accordingly.

To support trustworthiness, we maintained an audit trail recording each coding decision and codebook revision so that the analytical path from raw data to final theme can be retraced (Lincoln & Guba, 1985). The iterative team review described above further supported consistency in code application across the dataset. Triangulation across sectors was used as an analytical rule: no theme is reported unless grounded in accounts from at least three participants across at least two industries. The following section presents the findings, structured around five interconnected themes that together inform the hybrid playbook developed in the discussion.

6. Findings

Interviews reveal a more complex and counterintuitive picture of human-AI collaboration than the prevailing narrative suggests. While AI's operational strengths in speed and scale are widely acknowledged, participants also described limitations: not only does AI exhibit contextual deficits, but human involvement itself is not uniformly superior. Across sectors, participants identified AI as a stabilizing and legitimizing force in organizational decision-making and described how AI reshapes the conditions under which feedback is produced. These findings offer a basis for hybrid system design that moves beyond the efficiency-empathy framing and toward a more precise logic of task allocation, governance and feedback architecture.

6.1. High-volume extraction as an operational engine

Across sectors, participants emphasized AI-driven feedback systems' strength in handling high-volume tasks at speeds and scales unattainable through human labor alone. A government participant (P8) noted that manually transcribing large volumes of verbal feedback "would take days of effort for a member of staff," whereas AI automates collation and summarization, directly augmenting capacity and reducing fatigue (Jarrahi et al., 2023; Bankins et al., 2024). Similarly, a

marketing consultant (P1) observed that AI “can cover a lot more ground than one individual,” framing its value primarily in terms of efficiency and effectiveness.

These accounts align with prior research highlighting AI’s role in structured automation, cognitive support, and real-time analytics (Cubric, 2020; Mattingly et al., 2025; Samson et al., 2022). Participants consistently described AI as a “first-draft engine,” capable of templating, summarizing, and filtering data prior to human review. A sales representative (P12) explained that AI-generated customer summaries “save us time [and] resources,” accelerating expertise rather than replacing it by freeing attention for higher-order decisions. Critically, AI’s operational value is not confined to speed alone, it also improves the completeness of what is captured. An account brand director (P18) observed that in multi-speaker meetings, AI “fills in the blanks” by recording content that “human ears fail to hear”, particularly when multiple participants speak simultaneously. This distinction matters: AI does not merely process the feedback record faster, it produces a more complete one. Yet this operational power also reveals a critical limitation.

6.2. The contextual deficit and nuance blindness

Building on the well-documented observation that AI systems struggle with sarcasm, tone, and culturally embedded meaning (Fürst et al., 2025; Puntoni et al., 2021; Rana et al., 2025), we use the term nuance blindness as a diagnostic shorthand. The term is not intended to introduce a new phenomenon but to give managers a concept they can act on at the level of escalation logic; its contribution is operational rather than conceptual, mapping a known limitation onto the specific trigger conditions developed in the playbook (Rule 2). AI-generated output often lacks the contextual sensitivity required for high-quality decision-making, and participants repeatedly emphasized the need to validate, recalibrate, or reinterpret AI output before acting on it. For instance, a music director (P3) noted that AI tools “don’t get the context, the cultural context, the language nuances,” while a healthcare data lead (P2) described how AI’s reliance on past data risks producing “a misleading or even harmful direction of action” in clinical contexts. A front-end developer (P17) similarly explained that AI struggles to weigh the strategic implications of feedback: “AI may not understand if there’s a competitor doing the same thing slightly different... and how that affects us”.

These findings echo prior research showing that while AI excels at objectivity and pattern recognition, it struggles with tacit knowledge and emotionally embedded cues (Fürst et al., 2025; Puntoni et al., 2021; Rana et al., 2025). For example, an account brand director (P18) noted that AI

could accurately transcribe feedback while missing sarcastic intent that fundamentally altered meaning. The same participant noted that AI systems routinely misclassify feedback from elderly customers who conflate the star-rating scale with a schoolroom reward system, giving one star alongside an enthusiastic written review. In each case, AI correctly captured the literal data but fundamentally misread its meaning.

Nuance blindness is further compounded in multilingual and cross-cultural settings. As P18 noted, when client meetings shift mid-conversation between languages, AI transcription systems produce records that are technically complete but interpretively incoherent, while a human listener with the appropriate linguistic range would follow the shift without disruption. A finance manager (P4) added a high-stakes variant of this problem: AI reading a client negotiation recommended a one-to-two week deadline extension, failing to detect from the tone and phrasing of the client's message that any pushback "will react negatively", a judgment a human reader would make immediately from contextual cues. When errors of this kind are scaled across large feedback volumes, as one participant (P7) cautioned, "we really cannot trust AI" without structured human oversight. When errors scale, reputational and relational consequences scale with them, reinforcing evidence that GenAI-assisted feedback systems require explicit quality criteria and human supervision to ensure alignment and mitigate risk (Katsiuba et al., 2026). In short, AI identifies patterns; humans interpret meaning. However, this does not inherently establish human involvement as superior, it introduces a more complex set of trade-offs.

6.3. Human involvement: Sources of distortion and interpretive value

A recurring and counterintuitive theme across interviews was that human involvement in feedback processing does not automatically produce better outcomes. While participants acknowledged AI's contextual limitations, they were equally candid about the ways in which human judgement introduces its own forms of distortion: bias, mood effects, inconsistency, and motivated interpretation. Ultimately, these distortions can undermine the quality of feedback handling just as readily as nuance blindness in AI.

The most common source of human variability was emotional state. As a music director (P3) observed, "people can have a bad day...and because they're in an annoyed mood they'll find a way to complain," highlighting how the production and interpretation of feedback are both shaped by fluctuating emotional conditions that AI does not share. This variability extended to how feedback was processed internally. An IT support analyst (P6) described a case where a customer calling

someone a “liar” was “interpreted multiple ways” by different human reviewers, some treating it as a personal attack, others as institutional criticism, triggering unnecessary escalation and further contact. The divergent interpretations ultimately produced more friction than resolution.

Several participants also pointed to more systematic forms of human bias in feedback handling. A sales representative (P12) noted that the accuracy and usefulness of human interpretation was contingent on the reviewer’s expertise, lived experience, and, critically, timing: when customer satisfaction surveys represent either the first or last task of a working day, reviewers are less likely to chase up anomalies, flag concerns, or read every detail carefully. Human judgment, in other words, is not a stable resource; it fluctuates with cognitive load, attention, and situational context in ways that are difficult to control for at scale. An account brand director (P18) added a further dimension, describing how colleagues conducting client insight sessions while unwell or emotionally preoccupied would omit data “consciously or subconsciously”, and would sometimes mis-attribute client answers to the wrong questions entirely when clients responded out of sequence, an error AI does not make.

Participants also described how human reviewers can selectively interpret or soften feedback in ways that distort its original meaning. A front-end developer (P17) noted: “if someone said your phone calls are absolutely horrific... and they’ve got to feed that back to their manager, they may be like “this person is a bit upset”, whereas AI “would be like ‘they just said that this is the worst person they’ve ever met’”, preserving the intensity of the original signal. Similarly, P18 noted that human interpretation of meeting content was shaped by “what they hear and what they didn’t hear, how much do they actually like the client”, introducing relational and affective filters that AI does not apply. When feedback is routed through individuals who have a personal or professional stake in its reception, the signal is at risk of distortion before it reaches decision-makers.

In contrast, several participants described AI as offering a form of consistency and resistance to distortion that human reviewers could not reliably match. A business and relationship manager (P9) observed that “humans can miss stuff... humans can misinterpret information... if there was something negative there, they could just overlook it,” whereas AI is less likely to omit or soften critical signals. A healthcare data lead (P2) went further, noting that AI systems trained on large datasets can recognize patterns, including sarcasm and linguistic variation, that “a human reviewer might miss out on,” particularly in cross-cultural or multilingual contexts where human expertise is unevenly distributed.

Yet this does not render human involvement dispensable—it clarifies where human involvement is irreplaceable. A clear boundary emerged across interviews: while AI can structure, evaluate, and surface information at scale, it cannot persuade, negotiate, or translate insight into organizational action. An associate delivery manager (P11) emphasized that although AI generates quantitative insights, “it can’t talk... it can’t persuade or back up” recommendations to leadership. A marketing consultant (P1) noted that while AI offers a “holistic approach” that individual humans may lack in scope, interpretation and advocacy remain essentially human responsibilities. A digital delivery manager (P14) similarly cautioned against “missing any motive”, stressing that humans interpret contextual and political significance in ways that AI cannot reliably replicate.

These accounts reinforce a principle of complementarity: AI processes and optimizes information, while humans convert it into actionable knowledge. This interpretive role performs three functions that AI cannot substitute for: contextualizing output, identifying blind spots, and ensuring ethical and strategic alignment (Paschen et al., 2020; Jarrahi et al., 2018). Complementarity thus arises through role differentiation rather than substitution: AI generates; humans interpret, advocate, and act. The question managers should therefore ask is not whether to involve humans, but when and how human involvement is most likely to add interpretive and persuasive value rather than introduce noise.

This reframing also introduces a second managerial layer: trust calibration. Participants engaged more confidently with AI when they trusted its outputs would not expose them to reputational or relational risk. AI generates efficiently at scale, but its outputs gain relevance only through human contextual refinement (Csaszar et al., 2024; Doshi et al., 2025; Porkodi & Cedro, 2025). Importantly, interpretation is not purely cognitive but relational; and this relational dimension shapes whether AI's stabilizing potential can be realised in practice, a dynamic explored in the following section.

6.4. The neutral arbiter: AI as a stabilizing and legitimizing force in organizational decision-making

Participants also described a distinct internal role for AI as a neutral arbiter. Beyond its role as an analytical tool, participants described a more structurally significant function for AI: its capacity to stabilize decision-making processes, reduce the influence of hierarchy and organizational politics, and provide a reference point perceived as more objective and harder to dismiss than individual

human judgment. This governance function operated across two distinct registers: internal political dynamics and external accountability with clients.

The most vivid illustration of AI's internal legitimizing function came from a growth marketing executive (P10) who described working under a senior leader whose feedback "really wasn't" of the quality the leader believed it to be. Rather than challenging this directly, a path constrained by hierarchy and interpersonal risk, P10 would input the leader's content into an AI system, ask it to "evaluate it and give constructive feedback," and then use the AI's assessment to demonstrate that "what you're giving me is really not good". Here, AI functioned not as an efficiency tool but as a neutral third party whose evaluations carried legitimacy precisely because they were perceived as free from personal agenda or political motive.

This pattern of using AI to bypass or neutralise political dynamics was echoed across other accounts and sectors. When teams were divided or discussions became circular, participants described turning to AI to cut through the ambiguity. Both a growth marketing executive (P10) and a deal advisor (P20) independently reached for the same framing: when internal disagreement reaches an impasse, "it's better to just ask a neutral AI chatbot", using AI not to generate new information, but to provide a shared reference point for consensus. The convergence of this language across participants in different industries suggests that AI's neutrality is being recognized as a functional governance property, not merely a technical one.

AI also performed a parallel legitimizing role in external relationships with clients. An account brand director (P18) described using AI-generated meeting summaries as an objective record to resolve disputes about what had been agreed: "I pull out the summary, I send it over to them and be like, listen, this did not happen... this is not human notes. So at no point can you say that our organization omitted that data". Where human meeting notes are vulnerable to selective recording and contested recollection, AI-generated records carry a different kind of authority, one grounded in their perceived impartiality rather than the credibility of any individual who produced them. This function extends AI's governance role beyond internal politics into the management of external accountability.

AI was also described as a mechanism for aggregating and depersonalising expertise. P10 explicitly problematized reliance on singular human knowledge, noting that an AI system could "aggregate [input] from almost everyone" and "attach weights" to different dimensions of a problem, reframing AI not as a replacement for human expertise, but as a means of synthesising it at scale into a more

robust and representative form. A sales representative (P12) captured the structural tension this addresses: human input is filtered through social dynamics, such that “if you have the right person saying important things, you’re going to act upon it... however, if you don’t completely trust or respect their experience, you’re going to be sceptical”. AI output operates under a different logic of legitimacy, evaluated based on perceived objectivity rather than the social standing of its source. In hierarchical organizations where information flow is shaped by power relations, this distinction is managerially significant: AI can surface insights that would be dismissed or diluted if attributed to a junior or less credible human source.

Collectively, these accounts suggest that AI’s value in organizational life is not confined to efficiency or analytical accuracy, but extends to the legitimation of decisions and the reduction of political distortion in knowledge flows. Managers who recognize this role can deploy AI not merely as a back-end processing tool but as an active governance mechanism, one that depoliticizes judgment, reduces dependence on hierarchical authority, and provides a shared and contestable basis for organizational action.

6.5. AI reshapes feedback generation

Across interviews, participants described AI not only as a tool for analyzing feedback more efficiently, but as a mechanism that alters what people are willing to say in the first place. Participants noted how the presence of AI, as either the recipient or the medium of feedback, appeared to change the tone, candor, and content of the input itself. They reported observing qualitatively different information from the same individuals when feedback was AI-mediated rather than human-mediated, an observation consistent with recent experimental evidence that disclosure patterns to conversational AI differ systematically from disclosure to human interlocutors (Croes et al., 2024; Papneja & Yadav, 2024). This upstream effect aligns with a longstanding literature showing that the very act of anticipating evaluation reshapes what consumers report (Ofir & Simonson, 2001), and with more recent work indicating that perceived interlocutor type (human vs. AI) influences candor and willingness to disclose in conversational settings (Croes et al., 2024). Together, this evidence suggests that the channel through which feedback is solicited does not merely transmit data, but also shapes it.

The clearest articulation of this came from a music director (P3), who reflected on why customers might share more direct input when interacting with an AI system: “when you’re talking to a robot or AI... you can just say whatever you want... with no repercussions. And it’s just pure data”. In

human-to-human feedback settings, social friction (the desire to avoid seeming impolite, rude, or confrontational) can lead people to filter or soften what they say. Participants in our study perceived AI as reducing this inhibition, describing customers as more likely to express dissatisfaction directly when interacting with an automated channel. While our data captures practitioners' observations of customer behavior rather than direct measurement of it, the pattern is broadly consistent with experimental findings that anonymity and reduced perceived judgment in computer-mediated settings facilitate more candid disclosure (Croes et al., 2024; Papneja & Yadav, 2024). This observation was corroborated from a different angle by an IT support analyst (P6), who noted that anonymized feedback mechanisms produce more candid responses because respondents are not "scared of conflict" in the way they would be when giving feedback directly to a known individual. These accounts suggest that organizations relying exclusively on human-gathered feedback may be systematically receiving a more palatable, but less accurate, version of customer sentiment.

This mechanism also operates within organizations, reshaping how feedback is communicated and received internally. An account brand director (P18) described a meaningful shift in the character of internal communication when AI is present as the record-keeper: colleagues and clients alike behave differently knowing that an objective, uneditable transcript is being produced. Rather than the selective and relationally filtered notes a human colleague might produce, AI generates a record that is resistant to retrospective contestation, "if you set 73, it's 73. It's not 80". Participants described this as changing not just what is recorded, but what colleagues and clients are willing to say and commit to in the first place. An associate delivery manager (P11) described a complementary workaround that reveals the same underlying dynamic: when forwarding feedback to stakeholders, they deliberately send it without any accompanying text, specifically to prevent their own interpretive framing from influencing how the data is received. AI formalizes and scales this instinct, removing the human intermediary whose framing would otherwise color the message.

Similarly, several participants described how AI-mediated interfaces expand who participates in feedback processes at all. A sales representative (P12) noted that for individuals "not comfortable speaking over the phone", chatbot interfaces provide an accessible alternative, one that simultaneously generates structured, traceable data including "trigger words" and recurring queries that would not otherwise be captured. A front-end developer (P17) observed that chatbot interfaces lower the threshold for low-stakes feedback, a customer who wants to report that a council website cannot find bin collection times is more likely to do so via a text-based AI interface than by calling a service desk. This is more than an operational efficiency: by lowering the social and interpersonal barriers to feedback provision, AI widens the pool of voices captured, potentially shifting the

demographic and experiential composition of the data organizations use to make service decisions. This accessibility function operates with particular significance in public sector contexts, where government participants noted that digital services must accommodate citizens who may be uncomfortable with telephone-based or face-to-face contact, and where the range of users is broader and more varied than in commercial settings.

Taken together, these practitioner accounts suggest that AI's influence on feedback systems begins earlier than is typically recognized. Rather than simply processing the feedback that humans generate and collect, AI appears to reshape the generative conditions of feedback: who provides it, under what social conditions, with what degree of candor, and through what channels. While direct measurement of these effects on customer behavior is beyond the scope of the present study, the convergence of practitioner observations across sectors, and their alignment with the disclosure literature cited above, suggests this is a managerially consequential dimension warranting further empirical attention. This influence has consequences for the quality, diversity, and representativeness of the data organizations use to make service decisions. For managers, this raises a design question that is rarely posed: not just how to integrate AI into feedback analysis, but how to deliberately configure the conditions under which feedback is gathered to capture more honest and complete information.

6.6. Sector variation: Where the efficiency–empathy boundary sits

A consistent pattern of sector-specific variation emerged in where participants drew the boundary between AI-appropriate and human-essential tasks. Rather than a universal threshold, the efficiency–empathy boundary appears to shift depending on the nature of the service relationship, the vulnerability of the user population, and the reputational or relational consequences of miscalibration.

In healthcare, the boundary is drawn at emotional sensitivity rather than task complexity. A healthcare data lead (P2) articulated a principled distinction: where objectivity, scale, and pattern recognition dominate (e.g., data analysis, code review, system architecture) AI is the preferred tool. Where emotional intelligence and deep contextual nuance are clinically required (e.g., therapy, sensitive care decisions, patient feedback) “human feedback is essential and will take the priority” (P2). Their organization's current 20% AI, 80% human split reflects deliberate institutional caution: not a technical limitation, but a governance decision driven by the stakes of miscalibration in

emotionally sensitive contexts. The boundary in healthcare is therefore not about whether AI can process the data, but about whether the consequences of nuance blindness are acceptable.

In government, the boundary is structurally mandated rather than discretionally drawn. An associate delivery manager (P11) framed this directly: “AI isn’t filling out our forms or our digital services. So we need a humane aspect to it”, positioning human involvement not as a corrective to AI limitation but as a democratic obligation to citizens. Government participants consistently described a culture in which quantitative scores dominate upward reporting because “the number was really easy to report on” (P8), while qualitative insight embedded in citizen narratives requires labor-intensive human processing that resource constraints routinely deprioritize. Across this cluster, AI adoption is lowest and human oversight most entrenched, driven by data sensitivity, citizen vulnerability, and accountability expectations that commercial contexts do not face in the same form. Where AI is used, it is in a tightly bounded supporting role: automating statistical summaries while humans retain interpretive and relational authority.

In commercial and B2B professional services contexts, the boundary is drawn at relationship value and reputational risk rather than emotional sensitivity. A sales representative (P12) described how the same numerical score carries different urgency depending on the strategic significance of the account; a four-out-of-five from a local client treated very differently from the same score from a high-value one. AI cannot apply this contextual weighting; humans must. A finance manager (P4) and a strategy and pricing analyst (P7) both described the client-facing layer as non-negotiable human territory: in high-value B2B relationships, automated interaction is “severely irritating” (P7), and AI misreading negotiation tone carries direct commercial consequences (P4). By contrast, internal analytics, forecasting, and content generation in commercial contexts are more readily AI-led, with participants in marketing roles (P1, P10) expressing strong preference for AI as the primary system. The efficiency–empathy boundary in commercial settings therefore operates less as a fixed threshold and more as a tiered logic: AI leads at volume and internal analysis; humans lead wherever the client relationship or brand reputation is directly at stake.

Together, these patterns suggest that hybrid system design cannot be approached as a universal configuration. The appropriate AI-to-human ratio, the trigger conditions for escalation, and the scope of human interpretive authority all require calibration to the specific sector context — its user population, its accountability obligations, and the relational and reputational consequences of getting the boundary wrong. The playbook rules proposed in the discussion are intended as a design

framework rather than a fixed prescription: the architecture is consistent across sectors, but the thresholds, timelines, and task allocations must be set with sector-specific logic in mind.

7. Discussion: A hybrid playbook

Figure 1 integrates the study's five core findings into a structured hybrid human–AI feedback loop. Stage 1 captures AI's generative advantage in speed, scale, and analytical breadth. Stage 2 represents the human interpretive layer, where contextual judgment, empathy, and relational calibration convert output into actionable insight. Stage 3 emphasizes orchestration, where governance design and iterative refinement align analytical capability with strategic intent. Service improvement does not arise from any single stage, but from the disciplined coordination of all three within a system of continuous organizational learning.

[Insert Figure 1 About Here]

7.1. Rule 1: Institutionalize a first-draft mandate (AI leads)

AI should be the formal generator for high-volume, structured tasks (e.g., intake, collation, anomaly detection, first-pass summarization) but no AI-generated summary should reach decision-makers without a human interpretation pass. Implementation requires three explicit decisions: who is the named reviewer, on what timeline (daily for high-volume customer-facing contexts; monthly for lower-volume internal systems), and before which decision point the reviewed output must be available. A sales representative (P12) described AI generating live satisfaction scores with daily alerts for sub-threshold interactions, while human managers scan regional data for contextual anomalies the score alone misses. An associate delivery manager (P11) described a parallel design in government: one designated reviewer reads all qualitative comments after automated processing, before any output is routed onward. Naming a specific individual rather than any available staff member is what creates accountability for the interpretive pass.

7.2. **Rule 2: Implement sentiment-based escalation protocols (humans lead)** Organizations must define in advance the specific conditions that trigger mandatory human escalation. Four trigger conditions emerged consistently. First, a numerical threshold: any score below a defined level triggers an automatic alert to a named reviewer (P12: below five-out-of-five). Second, categorical terms; For example, words such as “complaint,” “cancel,” “refund,” or “legal” route

immediately to humans regardless of sentiment score (P9). Third, rating-comment mismatches: a one-star rating with positive text (or vice versa) is flagged for review (P18). Fourth, account-tier rules: the same score carries different urgency depending on the strategic value of the relationship (P12). Trigger-word lists should be reviewed at least quarterly for linguistic drift. An account brand director (P18) described how “wicked,” coded negatively by the system, is used positively by younger customers, producing false positives. Each trigger should carry a defined response window: same-day for commercial customer-facing escalations; within the weekly or monthly review cycle for internal contexts.

7.3. Rule 3: Formalize the strategic spokesman role

A recurring failure mode in hybrid systems is mechanical passivity: the insights accumulate but produce no action, because no one is explicitly responsible for converting AI output into organizational movement. AI cannot persuade, negotiate, or absorb the relational risk of advocating for a position; the strategic spokesman fills this gap. This role echoes what organizational theorists have long described as the boundary spanner or knowledge broker (Tushman & Scanlan, 1981; Carlile, 2004); our contribution is not to identify a new role but to specify how it functions in AI-mediated feedback systems, where the spokesman now carries responsibility for converting AI-derived insight, rather than human-collected insight, into organizational action. The role requires institutional credibility as well as analytical competence. For example, a sales representative (P12) observed that whether action follows depends on whether decision-makers trust the person presenting the data. Organizations should designate this role explicitly, attach it to someone with sufficient seniority to access decision-making forums, and embed AI-derived insight into governance calendars rather than leaving it to discretionary reporting.

7.4. Rule 4: Use AI as a governance stabilizer

Hybrid systems frequently stall due to organizational politics rather than technical limitations. AI can function as a structural corrective, providing an evaluation perceived as more objective and harder to contest on personal grounds than any individual's advocacy. A growth marketing executive (P10) described inputting a senior leader's work into AI for evaluation precisely because the AI's assessment could not be dismissed as a junior colleague's opinion. P10 and a deal advisor (P20) independently described turning to AI when team discussions reached impasse, not to generate new information but to provide a shared neutral reference point. An account brand

director (P18) used AI-generated meeting records to resolve client disputes about what had been agreed. Three governance applications follow: structured assumption-testing before major decisions; parallel scenario evaluation when teams are divided; and AI as the default meeting recorder, producing a record that resists retrospective contestation. Transparency about whether AI is acting as a generator of options or an evaluator of positions is essential.

7.5. Rule 5: Audit the loop continuously

Hybrid effectiveness must be actively maintained. Language shifts, priorities evolve, and risks accumulate in automated channels without triggering visible failure until the consequences become costly. An account brand director (P18) described keyword lists requiring regular manual updating as language evolves faster than AI training cycles. An associate delivery manager (P16) identified a governance risk that had gone unaddressed: only one person held the license for the analytics system, making the entire pipeline a single point of failure. A business and relationship manager (P9) described a loop-closure gap in which feedback reached the right channels but “we don't know whether they do anything with it”. Here, the audit trail ended at distribution, not implementation. Continuous auditing requires four recurring activities: a monthly sample review of AI-flagged and AI-binned cases (including items the system chose not to escalate); a quarterly threshold and keyword review for linguistic and demographic drift; a governance resilience check ensuring no critical function sits with a single individual; and a disclosure audit verifying that customers interacting with automation know they are doing so.

7.6. The rules

These five rules are sequential control points within a single system. Failure at any stage destabilizes the loop: without a validated first-draft mandate, nuance blindness propagates into decisions; without calibrated escalation triggers, relational damage accumulates below the threshold of detection; without a strategic spokesman, insight stalls; without governance stabilization, organizational politics distort data before it reaches decision-makers; without continuous auditing, all of the above drift back into failure. Table 2 operationalises these rules into execution checkpoints.

[Insert Table 2 About Here]

Taken together, the playbook forms a closed-loop architecture: AI generation expands analytical surface area; triage filters relational risk; the spokesman converts insight into action; pressure-testing stabilizes governance; and audits recalibrate the system. The objective is not merely to use AI more efficiently, but to institutionalize a repeatable hybrid control system that scales intelligence while preserving trust.

8. Conclusion

This study set out to address a timely and important managerial question: how can organizations integrate AI into customer feedback systems without undermining trust, empathy, and contextual judgment? Drawing on cross-sector interviews, we moved beyond the binary logic of automation versus human touch, and surfaced five findings with direct managerial implications.

First, AI is most effective as a high-volume operational engine, but its outputs require mandatory human interpretation before informing decisions. Second, nuance blindness is a structural limitation: AI systematically tends to misread emotional tone, sarcasm, and cross-cultural cues, with errors that scale if left unchecked. Third, human involvement is not a reliable correction: mood effects, motivated interpretation, and inconsistency mean the managerial question is when and how to involve humans, not whether to. Fourth, AI plays a governance role that is widely underutilized: as a neutral arbiter, it can depoliticize decisions, pressure-test assumptions, and produce records that resist retrospective contestation. Fifth, AI reshapes the generative conditions of feedback itself: people communicate more candidly through AI-mediated channels, so organizations relying exclusively on human-gathered feedback may receive a systematically filtered version of customer sentiment.

Five managerial actions follow. We advise managers to designate AI as the mandatory first-pass processor with a named human reviewer. There is a need to define explicit escalation triggers (numerical thresholds, categorical terms, rating-comment mismatches, and account-tier rules) before going live. We also recommend appointing a strategic spokesman with the seniority to convert insight into organizational movement. And AI should be used deliberately as a governance stabilizer to cut through political distortion. Finally, organizations should audit the system continuously, treating language drift, single-license risks, and loop-closure gaps as expected failure modes rather than exceptions. Competitive advantage will accrue not to firms with the most

advanced models, but to those that govern human-AI collaboration with precision, preserving legitimacy while scaling intelligence.

Declaration: An AI-based tool was used solely for proofreading and minor language corrections (e.g., grammar and typographical errors).

Journal Pre-proof

References

- Alhusban, M. I., Khatatbeh, I. N., & Alshurafat, H. (2025). Exploring the influence, implications and challenges of integrating generative artificial intelligence into organizational learning and development. *Competitiveness Review*. <https://doi.org/10.1108/CR-06-2024-0121>
- Ameen, N., Pagani, M., Pantano, E., Cheah, J.-H., Tarba, S., & Xia, S. (2025). The Rise of Human–Machine Collaboration: Managers’ Perceptions of Leveraging Artificial Intelligence for Enhanced B2B Service Recovery. *British Journal of Management*, 36(1), 91–109. <https://doi.org/10.1111/1467-8551.12829>
- Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. *Computers in Human Behavior*, 114, 106548. <https://doi.org/10.1016/j.chb.2020.106548>
- Bagchi, S. N., & Sharma, R. (2026). Managerial decision-making and AI: A decision canvas approach. *Business Horizons*, 69(1), 55–66. <https://doi.org/10.1016/j.bushor.2024.12.001>
- Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*. <https://doi.org/10.1002/job.2735>
- Berry, L. L., Danaher, T. S., Aksoy, L., & Keiningham, T. L. (2020). Service safety in the era of COVID-19. *Journal of Service Research*, 23(4), 391–395. <https://doi.org/10.1177/1094670520944608>
- Blaurock, M., Büttgen, M., & Schepers, J. (2025). Designing Collaborative Intelligence Systems for Employee-AI Service Co-Production. *Journal of Service Research*, 28(4), 544–562. <https://doi.org/10.1177/10946705241238751>
- Bloomberg. (2025, May 8). *Klarna turns from AI to real person customer service*. <https://www.bloomberg.com/news/articles/2025-05-08/klarna-turns-from-ai-to-real-person-customer-service>
- Blut, M., Wang, C., Wunderlich, 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>
- Carlile, P. R. (2004). Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science*, 15(5), 555–568. <https://doi.org/10.1287/orsc.1040.0094>

Chui, M., Manyika, J., & Miremadi, M. (2018). What AI can and can't do (yet) for your business. *McKinsey Quarterly*. <https://www.mckinsey.com/featured-insights/artificial-intelligence/what-ai-can-and-cant-do-yet-for-your-business>

Croes, E. A. J., Antheunis, M. L., van der Lee, C., & de Wit, J. M. S. (2024). Digital confessions: The willingness to disclose intimate information to a chatbot and its impact on emotional well-being. *Interacting with Computers*, 36(5), 279–292. <https://doi.org/10.1093/iwc/iwae016>

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors. *Strategy Science*, 9(4), 322–345. <https://doi.org/10.1287/stsc.2024.0190>

Cubric, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. *Technology in Society*, 62, 101257. <https://doi.org/10.1016/j.techsoc.2020.101257>

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>

Ding, Z., Zhang, Y., Sun, J., Goh, M., & Yang, Z. (2025). Harmonizing human touch and AI precision in customer service. *Journal of Service Research*. Advance online publication. <https://doi.org/10.1177/10946705251384692>

Doshi, A. R., Bell, J. J., Mirzayev, E., & Vanneste, B. S. (2025). Generative artificial intelligence and evaluating strategic decisions. *Strategic Management Journal*, 46(3), 583–610. <https://doi.org/10.1002/smj.3677>

Duffek, B., Eisingerich, A. B., Merlo, O., & Lee, G. (2025). Authenticity in influencer marketing: how can influencers and brands work together to build and maintain influencer authenticity? *Journal of Marketing*, 89(5), 21-46. <https://doi.org/10.1177/00222429251319786>

Eisingerich, A. B., MacInnis, D. J., & Park, C. W. (2023). Do CSR Efforts that Focus on Helping the Environment Influence Brand Purchase More than Other Forms of CSR?. *Journal of Business Research*, 168, 114240. <https://doi.org/10.1016/j.jbusres.2023.114240>

Fürst, A., Trißler, L., Friedrich, R., & Wirtz, J. (2025). Service Recovery by AI or Human Agents: Do Failure and Strategy Context Matter? *Journal of Service Management*. <https://doi.org/10.1108/JOSM-04-2024-0190>

Gartner. (2025, June 10). *Gartner predicts 50% of organizations will abandon plans to reduce customer service workforce due to AI* [Press release]. <https://www.gartner.com/en/newsroom/press->

[releases/2025-06-10-gartner-predicts-50-percent-of-organizations-will-abandon-plans-to-reduce-customer-service-workforce-due-to-ai](#)

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.
<https://doi.org/10.1177/1525822X05279903>

Guo, Y., Xu, L., & Wang, C. (2025). Exploring the effect of empathic response and its boundaries in artificial intelligence service recovery. *Journal of Retailing and Consumer Services*, 82, 104065.
<https://doi.org/10.1016/j.jretconser.2024.104065>

Hao, X., Dong, D., Zhang, Y., & Demir, E. (2025). When customers know it's AI: Experimental comparison of human and LLM-Based Communication in service recovery. *Journal of Marketing Communications*, Advance online publication. <https://doi.org/10.1080/13527266.2025.2540376>

Hermann, E., & Puntoni, S. (2025). Generative AI in Marketing and Principles for Ethical Design and Deployment. *Journal of Public Policy & Marketing*, 44(3), 332–349.
<https://doi.org/10.1177/07439156241309874>

Herriger, C., Merlo, O., Eisingerich, A. B., & Arigayota, A. R. (2025). Context-Contingent Privacy Concerns and Exploration of the Privacy Paradox in the Age of AI, Augmented Reality, Big Data, and the Internet of Things: Systematic Review. *JMIR*, 27, e71951. <https://doi.org/10.2196/71951>

Holmström, J., & Magnusson, J. (2026). Navigating the organizational AI journey: The AI transformation framework. *Business Horizons*, 69(1), 89–100.
<https://doi.org/10.1016/j.bushor.2025.01.002>

Huang, Z., & Lo, A. (2025). Human vs. robot service provider agents in service failures: Comparing customer dissatisfaction and the mediating role of forgiveness and service recovery expectation. *Information Technology & Tourism*, 27(2), 417–448. <https://doi.org/10.1007/s40558-025-00314-6>

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586.
<https://doi.org/10.1016/j.bushor.2018.03.007>

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons*, 66(1), 87–99.
<https://doi.org/10.1016/j.bushor.2022.03.002>

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, 31(2), 177–192. <https://doi.org/10.1002/ejsp.36>

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. William Collins.

Katsiuba, D., Dolata, M., Crowston, K., & Schwabe, G. (2026). Improving customer feedback management with GenAI: An action design research study. In Proceedings of the 59th Hawaii International Conference on System Sciences (pp. 5304–5313). <https://hdl.handle.net/10125/112033>

Krotov, V. (2025). Beyond AI automation: Redesigning organizations for human-AI synergy. *Business Horizons*. <https://doi.org/10.1016/j.bushor.2025.06.007>

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410. <https://doi.org/10.5465/annals.2018.0174>

Le, K. B. Q., Sajtos, L., Kunz, W. H., & Fernandez, K. V. (2025). The Future of Work: Understanding the Effectiveness of Collaboration Between Human and Digital Employees in Service. *Journal of Service Research*, 28(1), 186–205. <https://doi.org/10.1177/10946705241229419>

Lee, W.-J. (2024). Service Failure by Human, Service Recovery by AI Chatbot: The Impact of Justice, AI Efficacy on Recovery Effort. *Virtual Economics*, 7, 48–63. [https://doi.org/10.34021/ve.2024.07.04\(3\)](https://doi.org/10.34021/ve.2024.07.04(3))

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Liu, Y., Duffek, B., & Eisingerich, A. B. (2025). How Can Businesses Enhance Customer Resistance to Brand Switching? *Journal of the Academy of Marketing Science*. Advance online publication. <https://doi.org/10.1007/s11747-025-01115-3>

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94–100. <https://doi.org/10.1016/j.chb.2014.04.043>

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science*, 38(6), 937–947. <https://doi.org/10.1287/mksc.2019.1192>

Market Research Society. (2023). MRS Code of Conduct. Market Research Society. <https://www.mrs.org.uk/standards/code-of-conduct>

- Merlo, O., Eisingerich, A. B., Gillingwater, R., & Cao, J. J. (2023). Exploring the changing role of brand archetypes in customer-brand relationships: Why try to be a hero when your brand can be more? *Business Horizons*, 66(5), 615–629. <https://doi.org/10.1016/j.bushor.2022.11.001>
- Merlo, O., Eisingerich, A. B., & Hoyer, W. D. (2024). Immunizing Customers Against Negative Brand-Related Information. *Journal of the Academy of Marketing Science*, 52(1), 140-163. <https://doi.org/10.1007/s11747-023-00929-3>
- Mortati, M., & Viana Mundstock Freitas, G. (2025). AI in Service Design: A New Framework for Hybrid Human–AI Service Encounters. *Journal of Service Research*, 29(1) <https://doi.org/10.1177/10946705251344387>
- Nguyen, H. N., Nguyen, N. T., & Hancer, M. (2025). Human-robot collaboration in service recovery: Examining apology styles, comfort emotions, and customer retention. *International Journal of Hospitality Management*, 126, 104028. <https://doi.org/10.1016/j.ijhm.2024.104028>
- Ofir, C., & Simonson, I. (2001). In search of negative customer feedback: The effect of expecting to evaluate on satisfaction evaluations. *Journal of Marketing Research*, 38(2), 170–182. <https://doi.org/10.1509/jmkr.38.2.170.18841>
- Papneja, H., & Yadav, N. (2024). Self-disclosure to conversational AI: A literature review, emergent framework, and directions for future research. *Personal and Ubiquitous Computing*, 29, 119-151. <https://doi.org/10.1007/s00779-024-01823-7>
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.
- Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, 63(3), 403–414. <https://doi.org/10.1016/j.bushor.2020.01.003>
- Porkodi, S., & Cedro, T. L. (2025). The Ethical Role of Generative Artificial Intelligence in Modern HR Decision-Making: A Systematic Literature Review. *European Journal of Business and Management Research*, 10(1). <https://doi.org/10.24018/ejbmr.2025.10.1.2535>
- Pillai, V. (2024). Enhancing transparency and understanding in AI decision-making processes. *Iconic Research and Engineering Journals*, 8(1), 168-172.
- Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, 85(1), 131–151. <https://doi.org/10.1177/0022242920953847>

Rana, N. P., Begum, N., Faisal, M. N., & Mishra, A. (2025). Customer experiences with service robots in hotels: A review and research agenda. *Journal of Hospitality Marketing & Management*, 34(2), 145–174. <https://doi.org/10.1080/19368623.2024.2403640>

Sahoh, B., & Choksuriwong, A. (2023). The role of explainable Artificial Intelligence in high-stakes decision-making systems: A systematic review. *Journal of Ambient Intelligence and Humanized Computing*, 14(6), 7827–7843. <https://doi.org/10.1007/s12652-023-04594-w>

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Sage Publications.

Samson, F., Timilehin, O., & Emmanuel, F. V. (2022). AI-driven decision-making in customer service: A study of Pega's real-time analytics. *International Journal of Applied Management and Business*, 6(11), 8.

Selim, R., Basu, A., Anto, A., Foscht, T., & Eisingerich, A. B. (2024). Effects of Large Language Model-Based Offerings on the Well-Being of Students: Qualitative Study. *JMIR Formative Research*, 8, e64081. <https://doi.org/10.2196/64081>

Shaunn Mattingly, E., Kroes, J. R., & Manikas, A. S. (2025). AI as an emerging trend for managing employee efficiency in the retail and services industries. *Business Horizons*. <https://doi.org/10.1016/j.bushor.2025.06.005>

Shi, H., Yi, B. (Miranda), Ma, B., & Teng, L. (2025). Empathetic AI Encounters: Pathways to Prosocial Behavior. *Journal of Service Research*. <https://doi.org/10.1177/10946705251380568>

Simón, C., Revilla, E., & Sáenz, M. J. (2024). Integrating AI in organizations for value creation through Human-AI teaming: A dynamic-capabilities approach. *Journal of Business Research*, 182, 114783. <https://doi.org/10.1016/j.jbusres.2024.114783>

Tam, W. Merlo, O., Eisingerich, A. B., & Hoyer, W. D. (2026). Building Brand Immunity: How to Create Resilient Customer Relationships in Turbulent Times. *Business Horizons*. Advance online publication. <https://doi.org/10.1016/j.bushor.2026.02.001>

Tushman, M. L., & Scanlan, T. J. (1981). Boundary spanning individuals: Their role in information transfer and their antecedents. *Academy of Management Journal*, 24(2), 289–305. <https://doi.org/10.5465/255842>

Wen, Y., & Laporte, S. (2025). Experiential Narratives in Marketing: A Comparison of Generative AI and Human Content. *Journal of Public Policy & Marketing*, 44(3), 392–410. <https://doi.org/10.1177/07439156241297973>

Yeh, G. Y.-H., Chiu, C.-M. and Hsu, J. (2025) 'When AI Apologises: The impact of recovery strategy and tone on user response. In *PACIS 2025 Proceedings*, (pp. 1–9). <https://aisel.aisnet.org/pacis2025/hci/hci/9>

Zhu, M., Yang, Y. and Hsee, C.K. (2018) 'The mere urgency effect', *Journal of Consumer Research*, 45(3), pp. 673–690. <https://doi.org/10.1093/jcr/ucy008>

Table 1: Participant Profile Summary

ID	Professional Title / Role	Sector	Seniority Level	AI Involvement
P1	Marketing Consultant	Marketing	Senior	Strategic
P2	Healthcare Data Lead	Healthcare	Senior	Strategic / Copilot
P3	Music Director	Music / Entertainment	Senior	Operational (Administrative)
P4	Finance Manager	Finance	Mid-Senior	Strategic / Operational
P5	Global Sportswear Manager	Retail / Sportswear	Senior	Strategic (Hypothetical)
P6	IT Support Analyst	Technology	Mid-Level	Operational
P7	Strategy & Pricing Analyst	Logistics	Mid-Level	Operational (Restricted)
P8	Government Digital Manager	Government	Senior	Strategic (hypothetical)
P9	Business & Relationship Manager	Public Sector / Professional Services	Mid-Senior	Operational (Analytical)
P10	Growth Marketing Executive	Marketing / Logistics	Mid-Level	Strategic / Operational
P11	Associate Delivery Manager	Public Sector/ Government	Mid-Level	Operational (Analytical)
P12	Sales Representative	Repair Services / Appliances	Mid-Level	Operational
P13	Front End Developer	Technology / Public Sector	Mid-Senior	Hypothetical
P14	Digital Delivery Manager	Government	Senior	Operational / Exploratory
P15	Front End Developer	Technology	Mid-Level	Operational (Copilot)
P16	Associate Delivery Manager	Public Sector/ Government	Mid-Senior	Analytical / Operational
P17	Front End Developer	Technology / Public Sector	Mid-Level	Hypothetical
P18	Account Brand Director	Marketing / Branding	Senior	Operational
P19	Digital Delivery Manager	Public Sector/ Government	Senior	Analytical / Operational
P20	Deal Advisor	Logistics/ Consulting	Senior	Strategic / Operational

Table 2: The Hybrid Playbook – Execution Blueprint for Managers

Rule	Operational instruction	Named accountability	Review cadence	Risk if ignored
1. First-draft mandate	AI processes all structured feedback. Named human reviewer performs mandatory interpretive pass before output reaches decision-makers. No AI summary is treated as final.	One named reviewer per feedback channel. Escalation owner in each team.	Daily (commercial). Monthly (internal / government).	Mechanical passivity; nuance blindness; strategic drift caused by unchallenged automation.
2. Sentiment-based escalation	Four triggers defined in advance: (1) score below threshold; (2) categorical terms — ‘complaint’, ‘cancel’, ‘refund’, ‘legal’; (3) rating-comment mismatches; (4) account-tier rules.	Triage owner for routing. Keyword list owner for quarterly updates.	Escalated items same-day (commercial). Weekly / monthly (public sector). Keywords reviewed quarterly for linguistic drift.	Over-automation of sensitive cases; false positives divert attention from genuine signals.
3. Strategic spokesman	Named individual contextualizes AI output for senior forums, advocates for action, and confirms loop closure. Chosen for organizational credibility, not only analytical skill.	Named spokesman with seniority to access decision-making forums.	AI insights presented on defined governance cycle. Loop closure confirmed after each action cycle.	Insight accumulates without action. Feedback loop never closes.
4. Governance stabilizer	Use AI for assumption-testing before major decisions; parallel scenario comparison when teams are divided; default meeting recorder producing objective, contestable records. Disclose whether AI is acting as generator or evaluator.	Decision owner specifies AI governance role per forum. Transparency obligation on all participants.	Before major decision points. Meeting records produced in real time.	Political distortion embedded in strategy. Contested records. Junior insights suppressed.
5. Continuous audit	Monthly: sample review of flagged and binned cases including non-escalated items. Quarterly: threshold and keyword review for drift. Governance resilience check: no critical function with a single license-holder. Disclosure audit: customers know when interacting with automation	Audit owner for monthly review and quarterly assessment. Resilience owner ensures no single point of failure.	Monthly sample review. Quarterly threshold review. Annual full governance audit.	Silent system drift; customer trust erodes when automation is undisclosed; loop closure never confirmed.

Figure 1: The Structured Hybrid Human-AI Feedback Loop