REVIEW ARTICLE / AUTOMATED HOSPITALITY SERVICES
VOLUME 3 · ISSUE 1 · OCTOBER 2026 | e-ISSN 3064-1780 | DOI: 10.53893/grtss.v3i1.550

From AI Disclosure to Human Handoff: Transparency, Trust, and Accountability in Automated Hospitality Services—A Structured Evidence Review

Dila Ariyogi[1]* · **Ulfan Mulyawan[2]**
Mawar Junita[3]

[1] Department of Tourism, Sekolah Tinggi Pariwisata Mataram, Indonesia

[2] Department of Hotel Management, Sekolah Tinggi Pariwisata Mataram, Indonesia

[3] Department of Travel Management, Sekolah Tinggi Pariwisata Mataram, Mataram, Indonesia

* Corresponding author: Dila Ariyogi · dila.riyogi@gmail.com

RECEIVED 1 September 2026 · REVISED 29 September 2026 · ACCEPTED 13 October 2026

ABSTRACT

Automated hospitality services raise questions about transparency when guests first encounter artificial intelligence (AI) and when systems cannot resolve their requests. This structured evidence review examines AI disclosure, human escalation, trust, and organizational accountability. An access-limited synthesis mapped 19 empirical reports published between 2019 and July 2026, supplemented by relevant reviews and conceptual studies. Diverse methods, incomplete full-text access, and the absence of independent screening precluded meta-analysis and a comprehensive risk-of-bias assessment. The available evidence suggests that disclosure effects depend on service criticality and failure conditions, while recovery evaluations vary with problem severity, agent capabilities, and guest preferences. Responsibility assigned to a robot also differs from that attributed to its service organization. We develop a proposed executable transparency framework linking identity and capability information, accessible human assistance, information continuity, and clearly assigned recovery authority. Five propositions specify how these features could be examined in future research. The framework is conceptual and has not been tested as an integrated model. Because the evidence search and appraisal were incomplete, the findings constitute a bounded synthesis rather than definitive systematic-review conclusions.

KEYWORDS AI disclosure · human handoff · service recovery · trust · hospitality automation

SUGGESTED CITATION Ariyogi, D., Mulyawan, U., & Junita, M. (2026). From AI disclosure to human handoff: Transparency, trust, and accountability in automated hospitality services—A structured evidence review. Global Review of Tourism and Social Sciences, 3(1), 77–90. https://doi.org/10.53893/grtss.v3i1.550

01 Introduction

A hotel chatbot can correctly identify itself as artificial intelligence yet leave a guest unable to reach anyone who can resolve an overcharge. A delivery robot can communicate its failure while an employee lacks permission to replace the missing service. These situations expose a gap between providing information and enabling action. Transparency can tell guests which agent they face, but accountability requires a service organization capable of answering for what happens next. The central problem examined here is how those two functions connect in automated hospitality encounters.

Hospitality automation encompasses conversational agents, robot concierges, delivery systems, and other customer-facing interfaces. Their functions differ: some communicate information, others move goods, and some initiate decisions with consequences for bookings or service access. Foundational service research distinguishes different capabilities of AI and anticipates complementary allocation between technologies and

employees (Huang & Rust, 2018; Wirtz et al., 2018). A strategic marketing perspective likewise treats AI deployment as a choice about the tasks and relationships a firm intends to support (Huang & Rust, 2021). The question is therefore more specific than whether guests accept automation: what happens when an accepted system reaches its limits?

Recent reviews already organize substantial literatures on service robots, conversational agents, and technology-enabled hospitality (Mariani et al., 2023; Mukherjee et al., 2023; Hwang et al., 2026; G. D. Sharma et al., 2026). Zhang and Yuan (2026) specifically review AI service failure and recovery. Consequently, novelty cannot rest on announcing that automated service can fail or that human–robot collaboration matters. This paper contributes a narrower analytical connection between disclosure before or during an encounter, intervention after difficulty emerges, and responsibility for completing recovery. It uses the encounter trajectory as its organizing unit.

That connection has practical significance for service marketing. The guest evaluates the branded service organization through its interfaces and staff, even when technical components are supplied elsewhere. Rifqi (2025) situates AI within hospitality marketing innovation while recognizing privacy and implementation constraints. Related service research separates cognitive and emotional trust and shows why perceived risk can complicate acceptance (Fitriani & Basir, 2025). However, financial-service findings cannot establish effects in hotels; they provide sensitizing distinctions whose applicability must be tested in hospitality settings.

The review asks four questions. RQ1: Under which service conditions does AI identity disclosure support or undermine consumer responses? RQ2: When does human involvement contribute to recovery compared with automated recovery? RQ3: How do trust, empathy, competence, and responsibility attribution connect these responses? RQ4: Which features would make transparency actionable across an automated hospitality encounter? The last question concerns framework development rather than an effect already established by the reviewed evidence.

Three contributions follow. First, the synthesis distinguishes disclosure of identity from disclosure of capability limits and recovery responsibility. Second, it separates trust in an agent from confidence in the organization's capacity to resolve a problem. Third, it specifies observable features of human escalation, including accessibility, information continuity, and decision authority. The evidence does not directly test this complete combination. Accordingly, the proposed framework and propositions are presented as a research agenda supported by partial mechanisms, not a validated intervention.

02 Conceptual boundaries

Transparency as identity, capability, and process information

Identity disclosure identifies an interaction partner as automated. Capability disclosure communicates what the system can do and where its authority ends. Process disclosure explains what occurs when the system cannot complete a request. These are analytically distinct: a message stating that an assistant is AI does not tell the guest whether it can issue a refund or transfer a complaint. The review therefore treats identity as one component of transparency rather than a substitute for the others.

Conversational agents and embodied robots also require separate interpretation. The artificial identity of a physical robot may be visually apparent, while the artificial identity of a text interface may be ambiguous. Studies manipulating robot appearance or humanness are not automatically studies of explicit disclosure. Likewise, a comparison between human and robot recovery agents is not necessarily an evaluation of a working handoff from one to the other. These distinctions prevent adjacent literatures from being mistaken for direct tests of the same intervention.

Trust and neighboring outcomes

Trust concerns expectations about an agent or organization under uncertainty. Acceptance, satisfaction, forgiveness, reuse intention, and retention intention are related outcomes but are not interchangeable. A guest might forgive an error yet choose human service next time. Conversely, a guest might continue using a convenient system without believing that its recommendations are reliable. The synthesis retains the outcomes reported by each source rather than recoding every favorable response as trust.

Functional and relational evaluations can coexist. Song et al. (2024) connect functional expectations, social presence, trust, and rapport with hotel robot acceptance. In a different service domain, Wusko (2026) links security, convenience, trust, and satisfaction and identifies digital literacy as a boundary condition. These studies suggest useful distinctions between system characteristics and user capabilities; they do not demonstrate that a disclosure message causes durable confidence in a hospitality firm. That distinction becomes particularly relevant when guests are unfamiliar with a service interface.

Human presence, collaboration, and handoff

Human presence means that employees are available or visible. Collaboration means that humans and automated systems jointly participate in service delivery. Handoff means that responsibility for handling a request moves from one agent to another. An effective handoff therefore requires more than the presence of a staff member: someone must receive the problem, obtain enough information, and be able to act. This operational definition is developed for the present framework; it should not be attributed to every source that studies human involvement.

Wong et al. (2023) examine employee presence in smart-hotel acceptance, while Nguyen et al. (2025) examine configurations of human–robot participation in recovery. Both inform the argument, but neither is treated here as evidence that a particular transfer queue, transcript-sharing mechanism, or response-time standard is effective. Majeed et al. (2024) offer conceptual insights into virtual service agents in recovery, reinforcing the need to specify the intervention rather than assume that all human–AI arrangements are equivalent.

Responsibility attribution and organizational accountability

Responsibility attribution concerns whom customers perceive as causing or controlling a failure. Accountability concerns arrangements through which an organization explains decisions, accepts an obligation to respond, and provides a remedy. Psychological blame is not a legal determination. Sharma and Mattila (2026) explore ethical and organizational questions surrounding service robot responsibilities; their conceptual discussion informs the distinction but does not establish jurisdiction-specific liabilities.

The distinction also separates employees from firms. Assigning an employee to receive an escalated complaint does not establish that the employee caused the failure. Malau et al. (2026) associate governance quality, role clarity, trust, and engagement in cross-organizational public-sector work. Its relevance here is conceptual: clear responsibilities may help coordinate multiple actors. It is not hospitality evidence about robot failures. The present framework therefore assigns recovery ownership at organizational level while distinguishing the employee who performs corrective work from the actors responsible for system design and resource allocation.

03 Review methods and evidential status

Review design and reporting boundary

This version reports a rapid, structured synthesis of accessible empirical reports and develops a focused conceptual framework. Literature-review guidance emphasizes alignment between the research question, search procedure, selection, and synthesis (Snyder, 2019; Paul et al., 2021). PRISMA 2020 and PRISMA-S provide reporting standards for selection and search documentation (Page et al., 2021; Rethlefsen et al., 2021). They do not make an incomplete search comprehensive. No Scopus or Web of Science search, preregistration, independent dual screening, or complete PRISMA assessment is claimed here.

Searches were undertaken on 10 July 2026 using public scholarly web discovery, publisher sites, institutional repositories, and Crossref metadata. Crossref was used for discovery and bibliographic checking, not as an equivalent substitute for citation databases. Its relevance-ranked searches can return records that do not contain every query term. The empirical time boundary was 1 January 2019 to 10 July 2026. Older foundational sources were used for concepts and were not counted as contemporary empirical reports. Online publication and issue publication dates were distinguished where records differed.

Discovery and eligibility

Four Crossref queries were attempted, each requesting the first 25 relevance-ranked journal-article records within the date window: “hospitality service robot failure recovery”; “hotel chatbot disclosure trust”; “hospitality human robot collaboration recovery”; and “hospitality robot responsibility trust privacy.” Three queries returned 75 records, including 74 distinct DOI records. The first query failed with a rate-limit response and is retained as a failed attempt in the search archive. These retrieval numbers are not an estimate of all eligible literature. The uncompleted query and relevance truncation limit coverage.

Additional web searches used combinations of hospitality, hotel, chatbot, service robot, disclosure, trust, human intervention, responsibility, and service recovery. Reference lists of recent reviews were inspected to locate specific studies, and exact-title searches were used to verify relevant publications. Supplementary searching was adaptive rather than a fixed exhaustive database search. The source archive records retrieved pages and metadata; results from those searches were not converted into fictitious database identification counts.

The focal evidence map comprises 19 empirical reports for which a primary publisher abstract, author-affiliated institutional record, or full text supported extraction of a relevant result. Eligible focal topics were identity disclosure, customer trust or rapport in automated hospitality, comparisons of recovery agents, human involvement in recovery, and responsibility attribution or recovery work after automated failures. Hospitality studies were prioritized. Three explicitly identified adjacent-service reports were retained to clarify disclosure and communication mechanisms: Luo et al. (2019), Mozafari et al. (2022), and Haupt et al. (2023). They are not represented as hotel field evidence.

The 19 reports form a bounded, verified evidence map, not a claim that all potentially eligible records in the retrieval archive have completed full-text screening. Reviews, conceptual papers, methodological guidance, and contextual studies inform interpretation but are not counted as empirical reports in that map. Records whose relevance or full-text methods require further assessment remain unresolved in the supporting register rather than being described as exclusions for methodological quality. This separation permits substantive synthesis while preserving the incompleteness of the systematic-review stage.

Extraction, access, and appraisal

Extraction recorded DOI, bibliographic details, service context, reported design, focal comparison, outcome, boundary conditions, access level, and the inference supported by the accessible report. DOI-based checking reduced duplicate citation of the same work through multiple websites. Bibliographic discrepancies were resolved against current publisher records where possible and retained in the audit notes where records had changed. Missing sample sizes, effect estimates, confidence intervals, and procedural details were recorded as unavailable; they were not reconstructed from related papers.

The workflow was AI-assisted and involved one analytical pass followed by consistency checks, without independent human second-reviewer screening. Full texts were accessible for selected studies, including Huang and Lo (2025), Haupt et al. (2023), and the author-hosted report by Mozafari et al. (2022). Other focal reports were coded at abstract or institutional-summary level. Access level is reported separately from study quality: an unavailable full text is not proof of a weak study, and a detailed abstract does not establish low risk of bias.

A complete design-specific risk-of-bias appraisal was not feasible across the map. Instead, the synthesis distinguishes experimentally manipulated comparisons from cross-sectional associations and flags uncertainty about samples, measures, procedures, and statistical estimates. No numeric quality scores or certainty ratings are assigned. Abstract-reported findings are treated as provisional until full-text checks establish manipulation validity, exclusions, measurement properties, and the correspondence between statistical estimates and reported conclusions.

Synthesis procedure

Reports were grouped by disclosure, relational and functional trust, recovery-agent matching, responsibility attribution, and employee recovery work. Within each group, the comparison and its moderators were considered before interpreting the reported direction. Convergent mechanisms were then linked across encounter stages to develop propositions. This is an interpretive synthesis, not a frequency

count of significant effects. Shared constructs across studies are not assumed to use identical measures, and study-level mediation estimates are not treated as proof of a common causal sequence.

Meta-analysis was not undertaken because the verified extraction did not establish comparable effect estimates, common outcomes, or independence of contrasts across the corpus. Pooling forgiveness, purchase, satisfaction, and trust would obscure substantive differences. The evidence table provides traceability, while the framework distinguishes source-supported findings from propositions introduced by the review. Completing the failed query, broader database discovery, full-text eligibility decisions, and independent appraisal is necessary for a comprehensive SLR claim.

04 Evidence synthesis

Disclosure has context-dependent consequences

The literature does not support a universal expectation that identifying AI improves or reduces consumer responses. Luo et al. (2019) report a purchase penalty associated with prior chatbot disclosure in a sales setting. That finding concerns conversational selling and cannot be converted into a hotel cancellation or loyalty effect. Mozafari et al. (2022) examine service criticality and failure: disclosure can reduce trust-related retention responses in critical service settings but can become beneficial when the chatbot fails. Together, these studies shift attention from the disclosure label alone to expectations about capability and the encounter outcome.

Recent hospitality research adds a different behavioral consequence. Li et al. (2026) report that AI identity disclosure can shape mind perception, cognitive trust, and information distortion in recovery interactions. The accessible report supports this mechanism as a study-specific finding; it does not establish its prevalence across all hotels. Nonetheless, it broadens the problem beyond whether a customer likes the interface: disclosure may influence the quality of information on which recovery depends.

The implication is that transparency affects at least two service processes. It can alter expectations about the agent and it can alter how guests communicate with the firm. A credible service design must therefore evaluate both customer understanding and the information needed to solve the problem. An efficient conversation that leaves a guest confused about authority is not necessarily successful. Similarly, apparent acceptance that depends on guests mistaking an automated agent for a human does not demonstrate informed confidence.

Social qualities help, but outcomes depend on the task

Functional usefulness and social qualities are both relevant to automated hospitality. Song et al. (2024) report relationships involving performance and effort expectations, social presence, trust, rapport, and hotel robot acceptance. Song et al. (2026) examine trust under privacy concerns in hospitality scenarios. These are associational findings; they do not establish that adding a human-like voice or disclosure banner will cause trust to rise.

Privacy also complicates design choices. Jia et al. (2024) examine social robot privacy concerns in hospitality, and Xie and Lei (2022) investigate nonlinear responses to anthropomorphism from a privacy-calculus perspective. Their contributions caution against interpreting human likeness as a uniformly beneficial feature. The present review does not infer an optimal appearance level or privacy policy from these papers. It instead treats perceived warmth, capability, and information exposure as potentially different evaluations of the same system.

This separation matters during failure. A guest may welcome a friendly interface while doubting its ability to correct a booking. Another may prefer an impersonal robot for a routine request while seeking an employee for an embarrassing complaint. Frameworks about unintended robot consequences recognize that customer and employee outcomes extend beyond adoption metrics (Heirati et al., 2025). Service marketing should therefore evaluate what the interface promises against what the organization can actually deliver.

Recovery effectiveness depends on agent–problem fit

Hu et al. (2021) compare human recovery with robot-generated text and voice. Human recovery is reported as more sincere and satisfying, especially for customers with a stronger need for human interaction;

differences between the two robot modes are not identified. However, this should not be generalized into a rule that humans always recover better. Liu and Wang (2025) report that recovery effectiveness varies with failure type and whether the recovery agent conveys warmth or competence. The relevant comparison is a match between the problem and the corrective response.

Huang and Lo (2025) distinguish process from outcome failures in hotel scenarios. Greater agent humanness increases dissatisfaction through forgiveness and recovery expectations in process failures, while comparable differences are not identified for outcome failures. Their experimental comparison concerns agents associated with failure, not a full operational transfer protocol. It nevertheless demonstrates why the type of service breakdown needs to be retained in synthesis.

Communication content also matters. Haupt et al. (2023) find that solution-oriented and empathy-seeking chatbot messages elicit different competence and warmth evaluations, with preference depending on failure attribution and recurrence. Because this is adjacent-service evidence, its transfer to hotel complaint handling is a proposition for testing. It supports preserving the distinction between addressing a technical problem and acknowledging a relational concern rather than assuming one generic apology serves both purposes.

Human involvement is conditional rather than automatically superior

Nguyen et al. (2025) examine recovery configurations, apology styles, and consumption emotions. Their reported results differentiate economic recovery when robots lead from social recovery when humans participate. Liu et al. (2026) provide a more explicit severity comparison in unmanned smart hotels: robot-only apologies can outperform collaboration in low-severity failures, while collaboration benefits high-severity failures, with additional variation by regulatory focus. These findings challenge a simple instruction to add a human to every automated interaction.

Ho et al. (2020) compare recovery by staff, robots, and fellow customers. Firm-provided instrumental recovery is evaluated more favorably than purely informational recovery under the reported conditions. Huang et al. (2026) examine recovery effort in restaurants and report diminishing additional benefit where blame is low and rapport is strong. Across these papers, the capacity to provide an appropriate remedy appears more useful analytically than the mere presence of a human recovery agent.

The synthesis therefore distinguishes optional human access from compulsory intervention. A guest should have an intelligible route to help, but intervention can be triggered by severity, repeated failure, uncertainty, guest preference, or a request outside system authority. Those triggers are proposed operational criteria, not validated thresholds. None of the summarized comparisons establishes a universal maximum waiting time, an optimal number of failed bot responses, or a standard proportion of encounters that should transfer to employees.

Accountability can move toward the firm after robot failure

Leo and Huh (2020) report that customers assign less responsibility to robots than humans for equivalent failures, while assigning more responsibility to the firm when a robot fails. This is a psychological attribution result, not a statement of legal liability. Liu and Wang (2026) further show that robot responsibility attribution varies with failure type, relationship norms, and perceptions of warmth or competence. The findings suggest that firm-level evaluation cannot be inferred simply from how much blame an individual robot attracts.

These mechanisms create a service-governance problem. A message that a robot made a mistake may clarify the immediate cause without convincing the guest that anyone owns the remedy. An effective response must connect the failure explanation to an identifiable service obligation. The review proposes organizational recovery ownership to describe that connection. It is distinct from apologetic language and distinct from blame assigned to the employee receiving the complaint.

Employees bear the costs of failed automation

Liu et al. (2025) identify illegitimate tasks, customer mistreatment, and robotic instability as recovery stressors. Their mixed-method study links emotional-labor pathways with recovery work engagement and considers collaborative climate. The implication is that human fallback cannot be assumed to be a costless safety net. It is work requiring preparation, resources, and clear responsibilities.

Wang et al. (2025) review AI and hotel employee service performance, reinforcing the relevance of training and employee outcomes. Dasmadi et al. (2026) connect dynamic capability and service quality with resilience in small accommodation businesses. Their study does not examine AI handoff, but it grounds the contextual argument that service technology must be interpreted through organizational capabilities. A small hotel cannot obtain reliable recovery simply by advertising human availability if employees lack time, information, or authority to act.

Evidence coverage and the missing integrated test

Table 1 summarizes the 19 mapped empirical reports. The map contains direct hospitality studies and three designated adjacent-service studies. It should not be read as a sample of all published studies, a ranking of article quality, or evidence that every reported effect has been independently replicated. Access limitations are part of the map, not hidden editorial conditions.

Table 1. Mapped empirical reports and access levels

Source	Context and design	Reported finding	Type
Li et al. (2026)	Hospitality recovery; empirical study	Disclosure is linked to mind perception, cognitive trust, and information distortion. Detailed procedures require full-text verification.	A
Liu et al. (2026)	Smart hotels; four experiments	Robot-only apology favors low-severity failures; collaboration favors high-severity failures. Regulatory focus further conditions responses.	A
Liu and Du (2025)	Hospitality; experiments	Proactive recovery benefits forgiveness under reported robot-induced failure conditions; locus and privacy-sensitive context require attention.	A
Huang and Lo (2025)	Hotels; 3 × 2 experiment	Greater agent humanness increases dissatisfaction in process failures through forgiveness and recovery expectations; comparable outcome-failure differences are absent.	F
Liu and Wang (2025)	Robot-induced failures; three studies	Warm agents suit process failures and competent agents suit outcome failures; human recovery also performs favorably under reported comparisons.	A
Liu and Wang (2026)	Hospitality; four experiments and meta-analysis	Robot responsibility attribution varies with failure type, relationship norms, and warmth or competence; controllability is a reported mediator.	A
Nguyen et al. (2025)	Recovery; scenario experiment, n = 311	Economic apologies fit robot-led recovery; social apologies fit human participation. Outcomes include comfort and continuance-related intentions.	A
Hu et al. (2021)	Hospitality; recovery-agent comparison	Human apologies are perceived as more sincere and satisfying, especially with high need for human interaction; robot text–voice differences are absent.	A
Lv et al. (2022)	Hospitality; AI empathic recovery	Empathic recovery responses are examined in relation to continued use. Exact estimates and design details are not extracted here.	A
Ho et al. (2020)	Service failure; three experiments	Firm-provided recovery is favored over fellow-customer recovery; instrumental recovery can outperform informational recovery.	A
Song et al. (2026)	Hospitality; survey, n = 576	Privacy concerns and perceived usefulness condition trust-related associations; survey mediation does not establish causal effects.	F
Song et al. (2024)	Hotels; survey, n = 456	Functional expectations and social presence relate to trust and rapport; relational evaluations relate to robot acceptance.	A
Huang et al. (2026)	Restaurants; field and online experiments	Recovery effort can improve forgiveness and satisfaction; additional benefit diminishes under low blame and stronger rapport.	A
Leo and Huh (2020)	Services; three vignette experiments	Robots receive less blame than humans for equivalent failures, while firms receive more responsibility when robots fail.	A
Liu et al. (2025)	Hospitality workers; interviews, n = 25; survey, n = 435	Robot failure stressors connect to emotional labor and recovery engagement; collaborative climate is a boundary condition.	A
Luo et al. (2019)	Adjacent sales service; experiment	Prior chatbot disclosure reduces purchases in the studied setting; the result is not a hospitality loyalty estimate.	A
Haupt et al. (2023)	Adjacent chatbot service; three experiments	Solution and empathy-seeking messages elicit competence and warmth differently; attribution and recurrence condition preference.	F
Wong et al. (2023)	Smart hotels; survey	Employee presence moderates acceptance-related pathways. Presence is not a measure of completed handoff.	A
Mozafari et al. (2022)	Adjacent service; experiments, n = 201 and 197	Disclosure can harm trust-related retention under high criticality and benefit responses after chatbot failure; interaction strength warrants careful interpretation.	F

Note. F = full text accessed; A = publisher abstract or author-affiliated institutional summary. Access is not a quality rating. Results reported only at abstract level require full-text verification. Three reports concern adjacent services rather than hospitality.

None of the reports in the extracted map establishes the integrated effect of identity disclosure, capability disclosure, a functioning transfer pathway, and organizational recovery ownership across the full hospitality encounter. This is a bounded observation about the verified map, not an assertion that no such study exists anywhere. The absence of that integrated test motivates the framework below, while the limitations of discovery prevent a definitive priority claim.

05 An executable transparency framework

Linking information to the ability to obtain a remedy

We define executable transparency as information about services that enables a guest to understand the automated interaction and take a feasible next step when the system reaches its limits. Four elements operationalize the concept: intelligible identity and capability information; an accessible route to human help; continuity of relevant information across the transfer; and an identifiable actor authorized to complete recovery. This is a conceptual configuration of service-process features, not a new validated psychometric scale.

The framework separates agent evaluation from organizational assurance. Identity and capability information can influence expectations about the automated agent. Escalation availability can influence confidence that the organization will respond when the agent cannot. Successful transfer and an adequate remedy can then influence recovery satisfaction and subsequent behavior. Severity, task type, privacy sensitivity, customer preference, and staff resources can modify those relationships. Figure 1 organizes these proposed connections without treating them as already estimated paths.

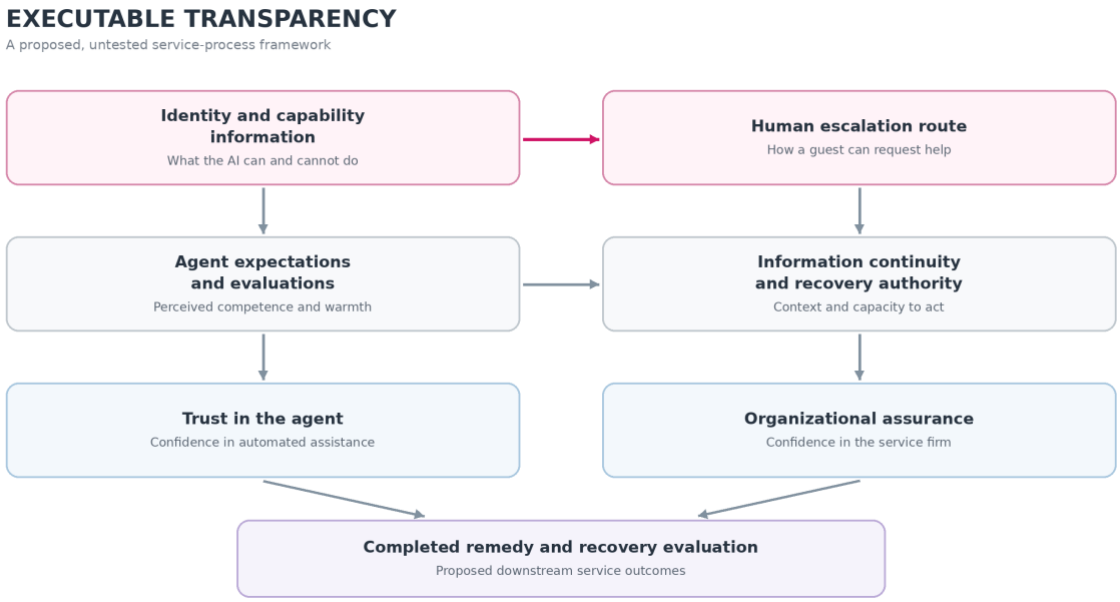

Figure 1. Proposed executable transparency framework

Note. Agent trust and organizational assurance are distinct proposed evaluations. The framework links information, escalation capacity, and recovery ownership; it is not an estimated structural model.

Proposition development

P1. When an automated hospitality system encounters a task beyond its capabilities, disclosure that also explains a feasible recovery route will produce greater perceived organizational assurance than identity disclosure alone. This proposition combines the context dependence of disclosure with the importance of recovery provision. It does not assume that richer disclosure increases liking of the automated agent itself. A guest may judge the system less capable while judging the organization more responsible.

P2. The contribution of human intervention to recovery evaluation will be stronger when relational injury, failure severity, or guest preference exceeds what the automated recovery response can address. Existing agent comparisons provide partial support for matching, but the proposition concerns intervention during a continuing encounter. Human availability should therefore be tested separately from mandatory intervention and from successful problem resolution.

P3. The positive association between offered human access and organizational assurance will weaken when the transfer requires substantial repetition or leads to an employee without decision authority. Information continuity and authority are proposed boundary conditions. Their effects are not directly

established by the mapped agent-comparison studies. Testing them would determine whether the promised fallback is an operational capability or simply a persuasive message.

P4. After robot-induced failure, identifiable organizational recovery ownership will improve responsibility clarity and recovery evaluation relative to responses that identify only the technical agent. This proposition follows the distinction between robot blame and firm responsibility. It predicts an effect of ownership communication and corrective capacity, not a transfer of liability. A study should measure perceived responsibility clarity independently from forgiveness and satisfaction.

P5. Staff capability and workload resources will condition whether human escalation produces a completed remedy and sustainable recovery performance. The employee recovery literature motivates this proposition, but the joint interaction with disclosure remains untested. Organizational assurance should not be measured exclusively through guest perceptions; it should be examined alongside actual resolution and employee outcomes.

Falsifiability and competing explanations

The framework would be weakened if identity-only disclosure performed as well as actionable process disclosure after controlling for information length and service quality. It would also be weakened if continuity and staff authority did not explain outcomes once severity and remedy adequacy were controlled. Competing explanations include simple reassurance, novelty, information overload, and the reputational effect of an apparently caring hotel. These alternatives should be tested rather than folded into the framework after the fact.

The conceptual contribution therefore concerns the organization of observable interventions. It avoids assuming that every transparency effect is mediated by trust, or that every response to automation reflects algorithm aversion. Identity disclosure, privacy explanations, and descriptions of service limits may activate different concerns. A strong test would measure them separately and include actual transfer performance, rather than rely on a single favorable-attitude scale.

06 Discussion and implications

Theoretical implications

First, the encounter perspective joins literatures commonly organized around different outcomes and agents. Disclosure research concerns what consumers know about the interface; recovery research concerns corrective responses; accountability research concerns attributed and organizational responsibility. The framework connects them through the guest's ability to proceed when automation fails. This integration complements existing broad reviews without claiming to supersede their more comprehensive coverage.

Second, the distinction between agent trust and organizational assurance helps interpret apparently conflicting results. A disclosure may reduce expectations of empathy from an AI agent while improving the intelligibility of the firm's service arrangement. Conversely, a friendly robot may attract trust while an inaccessible complaint pathway weakens confidence in the hotel. These are conceptual possibilities derived from the synthesis, not demonstrated two-level effects. They provide a basis for more precise measurement in future hospitality research.

Third, the framework makes the handoff itself visible. Human presence and collaborative apology are insufficient proxies for whether a request was transferred successfully. This matters because completed recovery requires a connection between communication and action. Measuring availability, continuity, authority, and resolution can reveal where an otherwise reassuring interface fails. It also links customer-facing transparency with the employee work needed to fulfill the promise.

Fourth, the framework treats guest and employee outcomes jointly. A policy that protects customer satisfaction by repeatedly shifting unresolvable work to staff may undermine recovery capacity over time. Conceptual work on automated service consequences and hospitality robot responsibilities supports examining organizational arrangements alongside interface design (Heirati et al., 2025; Sharma & Mattila, 2026). However, the review does not establish a causal workforce effect of any specific disclosure policy.

Service design and marketing implications

Hotels should evaluate disclosure in the language of the actual service promise. A hypothetical message could state: "I am the hotel's automated assistant. I can provide booking information, but a team member handles billing disputes. You can request a team member here." The value of that message depends on a working transfer route and accurate capability limits. It is an illustrative design proposal, not wording shown by the reviewed studies to outperform alternatives.

Recovery policies should distinguish routine difficulties from cases involving significant inconvenience, relational injury, or requests outside the system's authority. Automation can potentially support simple corrections, while staff participation may be appropriate for more consequential problems or guests who request it. The reviewed evidence does not justify a fixed severity score or automatic transfer quota. Managers should establish and evaluate rules appropriate to their service context and resources.

Information continuity should be selective. Staff need enough context to avoid making guests repeat the problem, but continuity does not justify transferring every item of personal data. The privacy-sensitive design problem is to convey what is necessary for the remedy and clarify how the information will be used. Privacy research motivates attention to that trade-off without establishing that more information sharing automatically raises trust.

Managers should also identify who can authorize corrective action. A concierge who receives an escalated complaint but cannot amend a booking may provide emotional reassurance without completing recovery. The proposed framework distinguishes those functions. Responsibility for system limitations should remain traceable within the organization, while staff receive the resources and discretion required for their assigned recovery tasks.

Table 2 turns the proposed framework into an operational comparison. The measures are suggested for evaluation; they are not effect sizes extracted from the review. For small hospitality enterprises, a modest and dependable escalation pathway may be more feasible than extensive interface personalization. The service quality and capability findings discussed by Dasmadi et al. (2026) provide context for that judgment, while deployment-specific evidence remains necessary.

Table 2. Proposed service-process features and evaluation measures

Feature	Operational question	Suggested measure
Identity and capability	Does the guest understand the agent and its limits?	Comprehension; ability to identify tasks outside system authority
Human access	Can the guest obtain human help when needed?	Request-to-transfer completion; unresolved escalation requests
Information continuity	Does the receiving employee understand the problem?	Repetition burden; missing-context incidents
Recovery ownership	Can an identified actor authorize a remedy?	Responsibility clarity; remedy completion; repeated contacts
Employee resources	Can staff sustain the promised recovery service?	Recovery workload; role clarity; perceived decision authority

Note. These are proposed operational measures for testing the framework, not observed effects or validated thresholds.

Research agenda

A manageable first test would randomly assign guests or participants to identity-only versus identity-plus-capability-and-escalation disclosure, crossed with accessible versus obstructed human escalation. The simulated task should keep the initial AI response and failure constant. Outcomes should separately measure comprehension, agent trust, organizational assurance, recovery satisfaction, and actual choice to seek assistance. Disclosure and transfer functionality should be manipulated independently to distinguish information from service capacity.

A second test would hold human availability constant while manipulating whether the receiving employee has relevant context and authority to provide a remedy. This design would test P3 and separate the comfort of human presence from successful transfer. Repetition burden and remedy adequacy should be measured,

while identity cues and information length are controlled. Ethical scenario experiments can establish mechanisms before a field evaluation introduces service inconvenience to guests.

Field research should then examine repeated encounters and employee outcomes. Useful records would include transfer completion, unresolved requests, time to remedy, repeated contacts, and recovery workload. A longitudinal design could distinguish novelty-driven reactions from durable confidence. Cross-cultural and multilingual research is needed because guest expectations, technological familiarity, and norms of human interaction can vary. Findings from scenarios and adjacent services should guide these tests without being treated as evidence of universal deployment performance.

Limitations

This review has substantial coverage and appraisal limits. Discovery used public web sources and three completed, relevance-truncated Crossref searches rather than a comprehensive multi-database strategy. One planned metadata search failed. The bounded evidence map does not complete full-text eligibility decisions for every candidate in the discovery archive. Consequently, it cannot estimate study prevalence or support a definitive assertion that a specific mechanism has never been investigated.

Many mapped findings were extracted from abstracts or institutional summaries. Full-text methods, detailed estimates, publication-bias indicators, and complete risk-of-bias judgments could not be verified across the set. The synthesis therefore supports interpretation of reported mechanisms, not pooled causal conclusions or a graded certainty statement. Access may preferentially expose papers with more discoverable metadata, and an independent second reviewer has not assessed selection and coding.

The empirical evidence includes heterogeneous technologies, settings, and outcomes. Three adjacent-service reports are included for transferability analysis, and some direct hospitality evidence is scenario-based or cross-sectional. Human presence, collaboration, and completed handoff are not interchangeable. The proposed process features are therefore hypotheses requiring dedicated tests. Completing the systematic-review search and appraisal stages remains a prerequisite to presenting this manuscript as a comprehensive, publication-ready SLR.

07 Conclusion

Automated hospitality transparency becomes consequential when guests need to understand both the interaction partner and the path to help. The accessible evidence indicates that disclosure consequences vary by service conditions, that recovery agents should be interpreted in relation to the problem, and that responsibility can attach differently to an automated agent and its service firm. These patterns motivate a proposed executable transparency framework connecting identity and capability information, accessible escalation, information continuity, and organizational recovery ownership.

The framework contributes specific propositions and operational measures rather than a universal recommendation to maximize disclosure or human intervention. Its central claim is testable: transparency should be evaluated partly through what it enables guests and employees to do when automation reaches its limits. The current access-limited evidence map supports development of that argument; comprehensive discovery, full-text appraisal, independent screening, and integrated empirical tests are still required.

08 Declarations

Competing interests. The authors declare no competing interests.

Author contribution. Dila Ariyogi—conceptualization, methodology, writing (original draft), and project administration; Ulfan Mulyawan—literature investigation, evidence organization, and writing (review and editing); Mawar Junita—evidence synthesis, bibliographic verification, and writing (review and editing).

Funding. The authors received no external funding for this work.

Data availability. Not Applicable for this type of study.

Use of artificial intelligence. No AI was used.

09 References

Dasmadi, Hartono, S., & Ghalih, M. (2026). Dynamic capability and service quality: Linking entrepreneurial ecosystem and leadership to MSME resilience in small hotels. *Global Review of Tourism and Social Sciences, 2*(3), 272–291. https://doi.org/10.53893/grtss.v2i3.518

Fitriani, N., & Basir, I. (2025). Understanding user acceptance of AI-powered financial advisory: A dual-process model integrating trust, satisfaction, and perceived risk. *Global Review of Tourism and Social Sciences, 1*(3), 225–239. https://doi.org/10.53893/grtss.v1i3.402

Haupt, M., Rozumowski, A., Freidank, J., & Haas, A. (2023). Seeking empathy or suggesting a solution? Effects of chatbot messages on service failure recovery. *Electronic Markets, 33*(1), Article 56. https://doi.org/10.1007/s12525-023-00673-0

Heirati, N., Pitardi, V., Wirtz, J., Jayawardhena, C., Kunz, W., & Paluch, S. (2025). Unintended consequences of service robots – Recent progress and future research directions. *Journal of Business Research, 194*, Article 115366. https://doi.org/10.1016/j.jbusres.2025.115366

Ho, T. H., Tojib, D., & Tsarenko, Y. (2020). Human staff vs. service robot vs. fellow customer: Does it matter who helps your customer following a service failure incident? *International Journal of Hospitality Management, 87*, Article 102501. https://doi.org/10.1016/j.ijhm.2020.102501

Hu, Y., Min, H., & Su, N. (2021). How sincere is an apology? Recovery satisfaction in a robot service failure context. *Journal of Hospitality & Tourism Research, 45*(6), 1022–1043. https://doi.org/10.1177/10963480211011533

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

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

Huang, Y.-S., Kao, W.-K., Tao, C.-W., & Ko, S. (2026). Is compensation necessary to recover robotic service failures? The role of attribution and rapport in restaurants. *International Journal of Contemporary Hospitality Management, 38*(2), 409–426. https://doi.org/10.1108/IJCHM-02-2025-0218

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

Hwang, J., Bai, B., & Erdem, M. (2026). Service robot technologies as social actors: A systematic review. *Journal of Hospitality and Tourism Technology*. Advance online publication. https://doi.org/10.1108/JHTT-01-2025-0097

Jia, S., Chi, O. H., & Lu, L. (2024). Social robot privacy concern (SRPC): Rethinking privacy concerns within the hospitality domain. *International Journal of Hospitality Management, 122*, Article 103853. https://doi.org/10.1016/j.ijhm.2024.103853

Leo, X., & Huh, Y. E. (2020). Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms. *Computers in Human Behavior, 113*, Article 106520. https://doi.org/10.1016/j.chb.2020.106520

Li, J., Han, M., & Yuan, Q. (2026). When transparency backfires: How AI identity disclosure fuels consumer information distortion in service recovery. *Journal of Hospitality Marketing & Management, 35*(5), 751–777. https://doi.org/10.1080/19368623.2026.2623405

Liu, K., Zhang, X., Li, J., & Wu, H. (2026). Are human employees still needed? The interaction effect between service failure severity and apology agents on customer forgiveness in unmanned smart hotels. *Journal of Consumer Behaviour*. Advance online publication. https://doi.org/10.1002/cb.70226

Liu, X., Zhang, L., Lin, M. S., & Jia, G. (2025). Paying for robotic errors: Exploring the relationship between robot service failure stressors, emotional labor and recovery work engagement. *International Journal of Contemporary Hospitality Management, 37*(6), 2023–2048. https://doi.org/10.1108/IJCHM-08-2024-1188

Liu, Y., & Du, Y. (2025). From failure to forgiveness: Proactive recovery strategies for service robots in hospitality. *Journal of Hospitality and Tourism Management, 63*, 316–328. https://doi.org/10.1016/j.jhtm.2025.05.009

Liu, Y., & Wang, X. (2025). Can service robots recover themselves? The effect of service recovery agents and robot service failure types on customer response. *International Journal of Hospitality Management, 124*, Article 103951. https://doi.org/10.1016/j.ijhm.2024.103951

Liu, Y., & Wang, X. (2026). Service robot failures in hospitality: Understanding customer attribution of responsibility. *Tourism Review, 81*(6), 1594–1611. https://doi.org/10.1108/TR-11-2024-0998

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

Lv, X., Yang, Y., Qin, D., Cao, X., & Xu, H. (2022). Artificial intelligence service recovery: The role of empathic response in hospitality customers' continuous usage intention. *Computers in Human Behavior, 126*, Article 106993. https://doi.org/10.1016/j.chb.2021.106993

Majeed, S., Kim, W. G., & Nimri, R. (2024). Conceptualizing the role of virtual service agents in service failure recovery: Guiding insights. *International Journal of Hospitality Management, 123*, Article 103889. https://doi.org/10.1016/j.ijhm.2024.103889

Malau, G. C., Wang, T.-L., & Fikri, Y. (2026). Collaborative governance and employee engagement in cross sector organizations: Evidence from East Java. *Global Review of Tourism and Social Sciences, 2*(2), 252–271. https://doi.org/10.53893/grtss.v2i2.488

Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research, 161*, Article 113838. https://doi.org/10.1016/j.jbusres.2023.113838

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2022). Trust me, I'm a bot – repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management, 33*(2), 221–245. https://doi.org/10.1108/JOSM-10-2020-0380

Mukherjee, S., Baral, M. M., Nagariya, R., Venkataiah, C., Rao, U. V. A., & Rao, K. S. (2023). Systematic literature review and future research directions for service robots in hospitality and tourism industries. *The Service Industries Journal, 43*(15-16), 1083–1116. https://doi.org/10.1080/02642069.2023.2243592

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*, Article 104028. https://doi.org/10.1016/j.ijhm.2024.104028

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies, 45*(4), O1–O16. https://doi.org/10.1111/ijcs.12695

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., & PRISMA-S Group (2021). PRISMA-S: An extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews, 10*(1), Article 39. https://doi.org/10.1186/s13643-020-01542-z

Rifqi, H. M. (2025). Navigating trends of marketing innovations in the hospitality and tourism industry: A comprehensive review. *Global Review of Tourism and Social Sciences, 1*(3), 182–191. https://doi.org/10.53893/grtss.v1i3.339

Sharma, A., & Mattila, A. (2026). Rights and responsibilities of hospitality service robots. *Journal of Hospitality & Tourism Research, 50*(6), 904–920. https://doi.org/10.1177/10963480251393749

Sharma, G. D., Kharbanda, A., Parihar, J. S., Dawar, G., & Taheri, B. (2026). Navigating the robotics revolution: A review of research on service automation in shaping the future of tourism and hospitality. *International Journal of Contemporary Hospitality Management*. Advance online publication. https://doi.org/10.1108/IJCHM-03-2025-0362

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Song, B., Xu, H., Cai, Z., Huang, H., & Guo, Y. (2026). How to trust hospitality service robots? Privacy concerns in human–robot interaction. *Humanities and Social Sciences Communications, 13*(1), Article 1530. https://doi.org/10.1057/s41599-026-07667-1

Song, X., Gu, H., Ling, X., Ye, W., Li, X., & Zhu, Z. (2024). Understanding trust and rapport in hotel service encounters: Extending the service robot acceptance model. *Journal of Hospitality and Tourism Technology, 15*(5), 842–861. https://doi.org/10.1108/JHTT-12-2023-0428

Wang, T., Aw, E. C.-X., Tan, G. W.-H., Sthapit, E., & Li, X. (2025). Can AI improve hotel service performance? A systematic review using ADO-TCM. *Tourism and Hospitality Research*. Advance online publication. https://doi.org/10.1177/14673584251356817

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

Wong, I. A., Zhang, T., Lin, Z., & Peng, Q. (2023). Hotel AI service: Are employees still needed? *Journal of Hospitality and Tourism Management, 55*, 416–424. https://doi.org/10.1016/j.jhtm.2023.05.005

Wusko, A. U. (2026). Security, trust, and transaction convenience in mobile banking: A moderated mediation model of usage decision and satisfaction. *Global Review of Tourism and Social Sciences, 2*(2), 191–206. https://doi.org/10.53893/grtss.v2i2.443

Xie, L., & Lei, S. (2022). The nonlinear effect of service robot anthropomorphism on customers' usage intention: A privacy calculus perspective. *International Journal of Hospitality Management, 107*, Article 103312. https://doi.org/10.1016/j.ijhm.2022.103312

Zhang, Y., & Yuan, Y. (2026). A systematic literature review of AI-based service failure and service recovery in hospitality and tourism industries using the TCCM framework. *Information Technology & Tourism, 28*, Article 53. https://doi.org/10.1007/s40558-026-00392-0