Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/



Research Paper Open Access

Responsible Artificial Intelligence in Enterprise Customer Relationship Management Systems: An Accountability Transfer Taxonomy and Research Agenda

Naveen Kumar Govindaraju

Independent Researcher, USA.

Abstract
Responsible AI research can be organized by principle, by regulatory risk tier, by stage of the development lifecycle, or by technique. Each of these lenses is useful, but none indicates the point at which a system displaces the authority of a human decision-maker. This review proposes a different taxonomy that distinguishes, for customer-facing enterprise systems, five loci of accountability transfer: perception, framing, option, execution and commitment. Through perception, the system determines what a human attends to. Through framing, it determines how the human interprets it. Through option, it limits the choice set. Through execution, it performs a reversible action. Through commitment, it performs an action with customer-visible or binding effect. Twenty-four peer-reviewed articles published from 2020 onwards were synthesized against this taxonomy, and coverage across it was uneven. Option, execution and commitment are discussed in some detail, commitment most of all because it carries regulatory obligations. Perception and framing are only weakly discussed, or discussed in work focused on other questions. They are, however, the two loci at which customer relationship management automation operates most often, since triage, prioritization, summarisation and sentiment interpretation are far more common than autonomous action. This asymmetry yields four research gaps and a corresponding research agenda. The main limitation is that the taxonomy is analytically derived from the reviewed corpus rather than empirically validated against practitioners' experience of where authority is lost.

Keywords: Responsible Artificial Intelligence, Transfer of Accountability, Customer Relationship Management, Algorithmic Governance, Human in the Loop, Automation Bias, Explainability, Enterprise System Governance.

1. Introduction

A service representative opens the queue and the top three cases appear, selected by the prioritization model. For each case, a summary, a sentiment label and a recommendation have been generated automatically. The representative reads, judges and decides. By any conventional account this is human decision-making with artificial intelligence support, and the human bears ultimate responsibility. Yet the representative did not select the cases to consider, did not construct the account of what each customer said, and did not assemble the set of available remedies. Considerable authority has moved, and none of it moved at the moment usually described as the decision (Uddandarao, 2026).

The responsible artificial intelligence literature has expanded rapidly and can be organized into four broad categories. Principle-based treatments list fairness, transparency and accountability as obligations to be designed for (Lu et al., 2023; Marzouk et al., 2023). Risk-tier treatments relate legal product classification to compliance obligations (Wagner et al., 2026; Restrepo Amariles & Baquero, 2026). Lifecycle treatments apply controls at successive development stages (Mökander & Floridi, 2023; McGregor, 2023). Technique-oriented treatments develop and test specific methods (Wang & Yin, 2022; Westerstrand, 2025). Bibliometric mapping confirms that the field is growing rapidly and is structurally heterogeneous (Zhang et al., 2025).

None of these organizing schemes answers the question a practising architect must settle: at which point does a named human authority pass to a system, and what control would restore it? Principles bind everywhere and therefore attach nowhere in particular (Vytla, 2024). Risk tiers sort systems rather than points in a workflow, and they place most customer relationship management automation below the threshold at which obligations apply (Vytla, 2025). Lifecycle stages describe when a control might be applied rather than what is lost.

Technique-oriented work examines a mechanism without establishing that this mechanism is what carried the authority away.

This review proposes that the locus of accountability transfer offers a clearer organization of the field, and applies it to twenty-four peer-reviewed publications. The review is timely for a specific reason rather than a general one. The move toward conversational and generative AI has shifted customer-facing automation decisively toward interpretation and summarisation, operations that frame what happens next, at a time when regulatory and academic attention has focused mainly on actions. The gap between what automation is used for and where governance attention is directed is the condition this review addresses.

This review contributes the following:

1. A five-locus taxonomy of accountability transfer in customer-facing enterprise systems: perception, framing, option, execution and commitment (Section 4).
2. A map of how the reviewed corpus covers those loci, and where coverage is absent (Section 5).
3. A comparison of five organizing lenses, stating what each makes visible and what each obscures (Section 6).
4. Four research gaps arising from this asymmetry, and a research agenda to address them (Sections 7 and 8).

2. Background

Enterprise customer relationship management platforms are environments in which automated interpretation, recommendation and action happen continuously across sales, service and marketing work. Scale is the critical point. A platform serving several thousand internal users, and many times that number of customers, applies automated judgements at a volume beyond the ability of any review function to inspect fully. Oversight therefore becomes selective, and the basis of selection becomes an architectural problem. The architectural models describing these environments are typically layered, with artificial intelligence functions distributed across the layers rather than concentrated in one.

Read from the bottom up, such a model places a connected customer data fabric at the base, integration services above it, intelligence services above those, automation functions above that, and finally human decision controls and governance. Responsibility mechanisms such as review queues, approval workflows, audit logs and control monitoring are the focus of most governance-related work, and they attach to the upper layers. The intelligence services layer, which supports scoring, summarisation and sentiment analysis, has no associated responsibility mechanism. It is at this layer, where framing occurs without a paired control, that the asymmetry documented in this review first appears.

The distinction between making a decision and determining what is decided has a long history in organizational theory, and it has re-entered the technical discussion through research on reliance. Decision-support research shows that more decision support can, under some conditions, produce worse decision quality (Ward, 2023). Overreliance research shows that people accept AI output beyond what its accuracy warrants, and that explanation design measurably changes this behaviour (Vasconcelos et al., 2023). Reliance also changes over time: Kahr et al. (2024) show that trust and reliance develop with accumulated experience, so the same reviewer may behave differently with the same system in month twelve than in month one (Gaggar, 2025b).

Regulatory concerns appear in compliance research on how firms interpret and implement high-risk obligations (Wagner et al., 2026) and, more generally, on how they navigate regulatory complexity (Restrepo Amariles & Baquero, 2026). Fairness engineering has also begun to engage with these obligations (Westerstrand, 2025). This body of work is large and consequential, but it is usually organized according to risk classification, which would place a customer service assistant that summarizes conversations and manages queues outside the regulated category entirely (Gaggar, 2025a).

3. Methodology

The search and screening process is set out here so that readers can assess and replicate the resulting corpus.

3.1 Databases and search strategy

The search was run on Crossref bibliographic metadata, filtered by publisher member identifier to target established journal venues of the Institute of Electrical and Electronics Engineers, Elsevier, Springer Nature, Wiley, the Association for Computing Machinery and Emerald. The filter was a methodological decision, made because unfiltered searches returned a large proportion of records from venues with no editorial process.

Contamination by such records is a known problem in this subject area, where responsible artificial intelligence attracts both serious scholarship and a large volume of low-barrier publication.

Search strings combined responsible-computing terms with application or mechanism terms. The responsible-computing terms were responsible artificial intelligence, artificial intelligence governance, algorithmic accountability, algorithmic fairness, explainability, human oversight and automated decision making. The application terms included customer service, enterprise, organizational adoption, decision support, managerial decision making, trust and regulatory compliance, often in pairwise combination.

3.2 Inclusion and exclusion criteria

Records were included if they were journal articles or named association proceedings published from 2020 onwards, had a resolvable digital object identifier, and discussed responsible AI in relation to human decision authority. Books, vendor documentation, technical reports, agency web publications and unrefereed preprints were excluded, as were records with no editorial process, records that discussed model architecture without addressing governance or oversight, and records whose thematic contribution was matched or exceeded by a stronger included record. In borderline cases, studies of consumer trust were included only when the trust was placed in the automated decision rather than in the brand, and studies of firm behaviour under regulation were included while general legal commentary on the regulatory instrument was excluded.

3.3 Screening flow

Figure 1 shows the progression of records through each phase of selection: 210 records were identified, 52 were retained after title and abstract screening, 28 were assessed for full eligibility, and 24 were included. Four potentially eligible records were excluded because of thematic redundancy with a more informative included record.

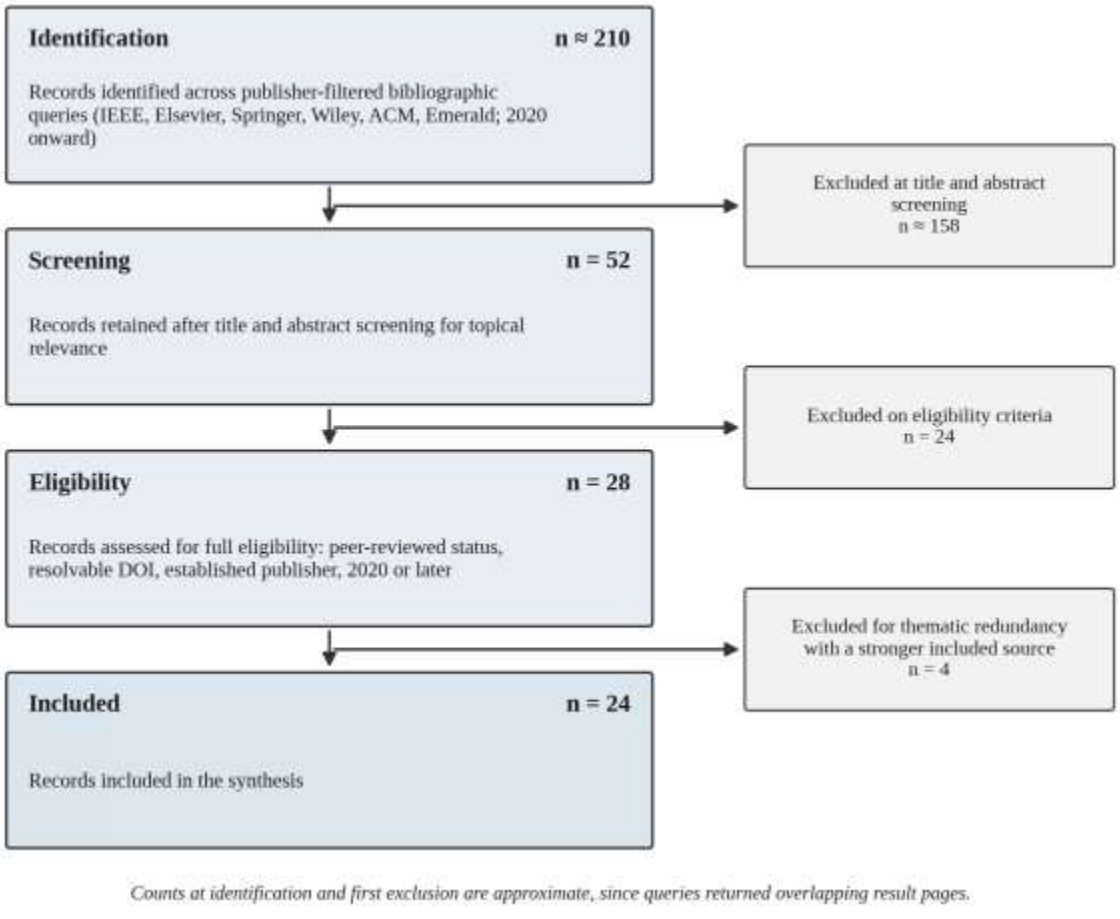


Figure 1. Record identification, screening and inclusion flow
Source: Author's own analysis.

Every record was checked against its bibliographic authority and then checked again independently, with the first check treated as unverified. This double-verification step is not yet standard practice, and it is reported here because it materially affected the corpus: the second pass identified a truncated title and an omitted co-author that the first pass had missed. This suggests that single-pass verification may no longer be sufficient at this scale.

3.4 Synthesis approach

Records are assigned to a locus according to the claims they make about it, and are organized thematically by locus rather than by journal or method. A record that makes claims about several loci appears under each, which is why the counts in Section 5 exceed the size of the corpus. Uneven coverage was not anticipated: the taxonomy was built to organize the synthesis, and the unevenness emerged from that grouping rather than motivating it. It is therefore presented as an empirical observation rather than a premise.

4. A Taxonomy of Accountability Transfer

The five loci are ordered by their position in the decision sequence, not by their potential impact, and a deployed capability may operate at several of them. Figure 2 positions the loci along the decision sequence and indicates, for each, what is handed to the system, what remains with the human, and the architect's guiding question. Its coverage bars anticipate the finding reported in Section 5.

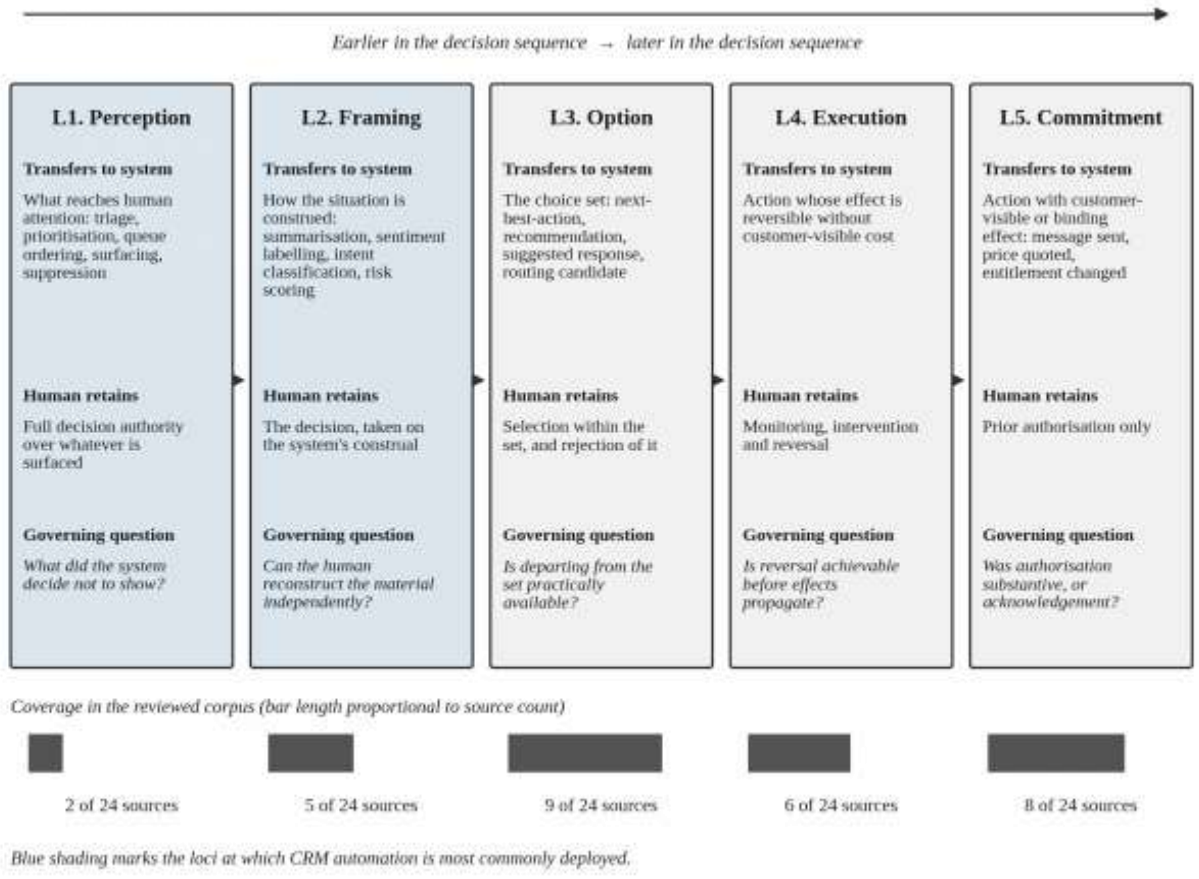


Figure 2. Five loci of accountability transfer in customer-facing enterprise systems, ordered along the decision sequence, with corpus coverage
Source: Author's own analysis.

Perception transfer occurs when the system decides what reaches human attention, as in queue ordering, case triage, alert suppression and lead prioritization, while the human retains discretion over what is shown. This

is precisely why the locus escapes scrutiny: no decision has been taken away. What has been taken away is the agenda. A case that is never surfaced is never decided badly, because it is never decided at all.

Framing transfer occurs when the system supplies the interpretation on which the human then decides, as in summarisation, sentiment labelling, intent classification and risk scoring. The human decides, but on a construal they did not build and generally cannot audit in the time available. Explainability research is the closest related area, with evaluation frameworks for managerial comprehension and trust (Love et al., 2024) and for high-stakes financial decision-making (Ioannou et al., 2026), both of which address how a conclusion was reached. Whether the construal was sound is a different question, and explainability does not reach it.

Option transfer occurs when the system constrains the choice set, through next-best-action recommendations, suggested responses or routing candidates. This locus is studied in depth. One study compared the effects of different explanation types on acceptance (Wang & Yin, 2022), and an experimental evaluation showed that explanations can reduce overreliance rather than merely increase confidence (Vasconcelos et al., 2023). Longitudinal work shows reliance calibrating through repeated interaction (Kahr et al., 2024), and presentation detail strongly influences consumer acceptance of recommendations (Kim et al., 2021).

Execution transfer occurs when an agent acts in a way that can be reversed without visible cost to the customer, which Kshetri (2025) frames as a question of identity, permission and oversight. This locus is addressed at programme scale in impact assessment practice (McGregor, 2023), and tooling is emerging to support the documentation burden (Bogucka et al., 2024). The real issue is reversibility. A change is reversible only if it can be undone within the window before its side effects propagate to other systems or to the customer; a record update that has already triggered three downstream operations is not reversible merely because the field can be set back to its former value.

Commitment transfer occurs when actions create externally visible and binding effects: a message is sent, a price is quoted, an entitlement is changed. Reversal is not possible, because the customer has already been told. Most work here focuses on regulatory consequences, although studies of firms have examined how accountability is constructed around opaque systems in practice (Lioliou et al., 2026). Evidence from industrial service recovery shows what failure at this locus costs: automated recovery can erode the trust and engagement it was deployed to protect (Naidoo, 2026).

4.1 A single capability across several loci

The loci are analytically distinct but operationally entangled, so a capability cannot be governed by assigning it to a single locus. Consider a service assistant, a type of agent now widely deployed. It orders the incoming queue, operating at perception. It summarizes the interaction history, operating at framing. It proposes a resolution, operating at option. It updates the case record automatically, operating at execution. It drafts the reply to the customer and sends it after approval, operating at commitment (Konatham et al., 2025).

One capability therefore performs five transfers, and a governance process will almost certainly classify it by its most visible function. If the classification is commitment, the controls built will be authorisation controls, and the four upstream transfers will be governed by an approval step that occurs after all of them have already happened. If the classification is assistance, because a human approves the final message, the capability may receive no substantive control at all. Neither outcome reflects negligence; both follow from a classification scheme that lacks a vocabulary for the distinction.

5. Thematic Review and Coverage

The distribution of the corpus across the taxonomy is the central finding of the review. It is uneven, and inversely so: coverage is thinnest at the loci where customer relationship management automation is most common. Table 1 summarizes the distribution.

Table 1. Distribution of the reviewed corpus over the transfer loci

Locus	Reviewed sources addressing it	Depth of treatment	Dominant framing in the literature	Coverage assessment
L1. Perception	2 of 24	Incidental	Treated as a data-pipeline or ranking concern rather than an accountability question.	Substantially under-addressed. No reviewed source treats attention allocation as a transfer of responsibility.
L2. Framing	5 of 24	Partial	Approached through explainability, which asks how a conclusion was reached rather than whether the construal was sound.	Under-addressed. Explainability research is adjacent but answers a different question.
L3. Option	9 of 24	Substantial	Well covered through reliance, overreliance and recommendation-acceptance research.	Adequately addressed, though largely in laboratory rather than enterprise settings.
L4. Execution	6 of 24	Substantial	Covered through agentic governance, impact assessment and auditing literature.	Adequately addressed at the level of principle; thinner on operational mechanism.
L5. Commitment	8 of 24	Substantial	Heavily covered through regulatory compliance, high-risk classification and accountability construction.	Best addressed of the five, driven by regulatory attention.

Source: Author's own analysis.

Of the twenty-four sources, two discuss perception, and neither does so in accountability terms; both treat it as a ranking or pipeline concern. The closest existing work concerns automated data cleaning, which has been shown to harm fairness in machine learning-based decision making, demonstrating that upstream processing choices carry downstream normative consequences (Guha et al., 2024). The argument carries over to triage, although the source does not make it.

Five studies address framing, and all do so through explainability. There are evaluation frameworks for users' comprehension of and trust in AI outcomes (Love et al., 2024), comparisons of explanation effects (Wang & Yin, 2022), analyses of explainability in financial decision-making (Ioannou et al., 2026), and studies of transparency as a driver of managerial trust (Cao et al., 2026). None considers whether the output's construal of the underlying situation was adequate. A summary can be perfectly explainable and still omit the decisive sentence. The distinction is not pedantic: an explanation accounts for the route from input to output, and is therefore silent about material that never entered the input.

The corpus concentrates at the action end of the taxonomy. Option, execution and commitment together account for twenty-three of the thirty locus assignments in Table 1, against seven for perception and framing. Regulatory compliance research sits at commitment (Wagner et al., 2026; Restrepo Amariles & Baquero, 2026), fairness engineering at option and commitment (Westerstrand, 2025), and auditing and assessment practice at execution and commitment (Mökander & Floridi, 2023; McGregor, 2023). Practice-oriented syntheses of responsible artificial intelligence building blocks (Marzouk et al., 2023) and design pattern collections (Lu et al., 2023) also cluster toward the action end.

Research on the customer side includes the trust effects of AI-mediated communication (Hennighausen et al., 2025), transparency and bias mitigation in service contexts (Ehsan et al., 2026), and the tension between

personalization and privacy (Chen et al., 2022). Together, these studies examine the customer's perspective after the system has acted. Compared with this outcome-focused work, research on the operational side of the system is sparse.

5.1 The operating model and where it does not reach

A tiering common in enterprise customer relationship management practice grades human oversight in four levels: assistive, supervised, conditional autonomy and restricted autonomy. The tiers are defined by what the system does and how far the human operator is involved, which aligns them with execution and commitment, loci four and five. Table 2 shows which loci are actually operating at each tier, and what each tier's control presupposes about the reviewing human.

Table 2. Conventional oversight tiers read against the accountability transfer loci

Conventional tier	Typical CRM capabilities	Loci actually operating	Control the tier assigns	What that control presupposes
Assistive	Conversation summarisation, reply drafting, knowledge-article suggestion, account insight panels.	L1 perception and L2 framing	Reviewer edits or rejects each output; sources and outputs are logged.	Time and access to check the construal against source material, which is the capacity framing transfer reduces.
Supervised	Queue routing, lead prioritisation, renewal-risk flagging, triggering of journey steps.	L1, L2 and L3	Reviewer approves flagged exceptions and sets thresholds.	That problems surface as exceptions; a suppressed item generates no exception to approve.
Conditional autonomy	Reminder scheduling, record enrichment, internal status changes, routing that can be undone.	L4, with L1 where routing is involved	Aggregate monitoring, bounded permissions and a rollback path.	That reversal is achievable within the window before effects propagate.
Restricted autonomy	Customer messages on sensitive matters, discount and pricing exceptions, regulated determinations, significant account changes.	L5, preceded by L1 to L3	A named approver for each action, with audit and escalation records.	That authorisation remains substantive, although the upstream transfers have already occurred.

Source: Author's own analysis.

Conventional tiering treats assistive automation as the low-risk case, requiring only traceability and logging. Read against the taxonomy, it is the tier at which perception and framing transfer occur. The control assigned to it, human review and edit, presupposes the very capacity that the transfer has reduced. Consider what review means here: a representative cannot check a summary against sources they have no time to read, and the tier model classifies actions by their reversibility rather than by the authority they absorb.

The field as a whole shows a similar pattern in its bibliometric profile (Zhang et al., 2025). Growth has concentrated where regulation and visible harm concentrate, which is a reasonable response at the level of the domain, but one that has left the quiet loci unexamined.

5.2 What the customer-side evidence adds

A further stream of the corpus concerns what customers experience, and read against the taxonomy it yields a finding that these studies suggest but do not state. In personalisation, trust and reactance mediate the tension between tailoring and privacy concern (Chen et al., 2022). The level of numerical detail with which a recommendation is presented influences acceptance independently of its quality (Kim et al., 2021). AI-

mediated communication has measurable effects on customer trust (Hennighausen et al., 2025), and service transparency and bias mitigation shape trust in ways that depend on the organization's ethical culture (Ehsan et al., 2026).

Each of these studies measures an effect at the point of customer contact, that is, at commitment. What produced the effect, however, usually occurred upstream. The recommendation whose presentation shifted acceptance was constructed at the option locus, and the message that provoked reactance drew on a profile assembled at perception. The customer-side literature is thus measuring the output of a pipeline while attributing the result to its final stage. Evidence that automated service recovery can erode rather than repair engagement (Naidoo, 2026) is most plausibly read as an upstream failure surfacing at the only point where anyone was looking.

This is consistent with managerial research on trust, which identifies ethics, transparency and explainability as drivers of AI trustworthiness in managerial decision-making (Cao et al., 2026); these attributes, however, concern what a system displays, not what it withholds. A manager can trust a system appropriately on everything it shows them and still be badly served by it.

6. Comparative Analysis of Organizing Lenses

The proposed taxonomy is one way of organizing the field, and it is not presented as superior to the alternatives on every dimension. Table 3 shows what each lens makes visible, and what each obscures, in a customer-facing setting.

Table 3. Comparison of organizing lenses for responsible artificial intelligence

Organising lens	Representative basis in the reviewed literature	What it makes visible	What it obscures in a customer-facing setting
Principle-based (fairness, transparency, accountability)	Responsible artificial intelligence practice landscapes and design pattern collections	A shared vocabulary that transfers across domains and supports policy drafting.	Where in a workflow each principle actually binds; principles are stated without an attachment point.
Risk-tier based (regulatory classification)	European regulatory compliance research and high-risk obligation studies	Legal obligation, enforcement exposure and documentation duty.	Everything below the high-risk threshold, which is where most customer relationship management automation sits.
Lifecycle-stage based (design, training, deployment, monitoring)	Impact assessment programmes and ethics-based auditing practice	When in a development process a control should be applied.	Which human loses which form of authority at each stage, and to what.
Technique-based (explainability, fairness metrics, uncertainty)	Explanation-effects research and fairness engineering studies	Concrete, measurable interventions with established evaluation methods.	Whether the technique addresses the authority that actually moved, or a different one.
Accountability-transfer based (this review)	Synthesis across the reviewed corpus	The specific human authority displaced at each point, and therefore the control that would restore it.	Offers no compliance mapping; requires translation before it satisfies a regulator.

Source: Author's own analysis.

Two entries require comment. The risk-tier lens has the strongest claim to practical authority, since it maps directly onto enforceable obligation. It is also the lens under which most customer relationship management

automation disappears, since a queue-prioritization model in a service centre would not be classified as high-risk, and nothing in the classification asks what authority the model absorbs. The accountability-transfer lens, by contrast, identifies the authority that has been displaced, but it offers no compliance mapping and must be translated before it can be presented to a regulator of high-risk systems.

The lenses are therefore not in competition. The one proposed here is useful for determining which controls a capability should carry while it is being designed; it does not show that an existing capability complies. An organization needs both, and the literature has long provided the latter (Tokas, 2026).

6.1 Mapping the taxonomy onto governance practice

Lifecycle-stage governance of the kind shown in Table 4 remains as relevant as ever. The taxonomy does not replace it, but it adds an accountability-transfer question to each stage, most importantly at design (Table 4, fourth column). Before deciding where to place the autonomy boundary, an organization must first decide whose authority the capability will absorb. A framing capability may call for different design-stage artifacts than a commitment capability, yet current practice assumes that a single risk-classification exercise serves both.

Table 4. Governance lifecycle stages, their evidence artifacts and the accountability-transfer question each should settle

Lifecycle stage	Architecture focus	Evidence artifact	Accountability-transfer question	Risk reduced
Design	Classification of the use case, purpose limitation for data, placement of the autonomy boundary.	Use-case assessment and decision-control plan.	Which loci will the capability occupy, and whose authority does each absorb?	Automation misaligned with purpose; accountability left unassigned.
Data readiness	Quality, lineage, identity resolution, consent and minimisation.	Data provenance register and readiness assessment.	What will perception-locus filtering be permitted to exclude, and who can see what was excluded?	Recommendations built on unreliable data; privacy exposure.
Model enablement	Validation, versioning, prompt control and evaluation sets.	Model documentation, prompt register and validation evidence.	How will construal adequacy, and not only explainability, be evaluated?	Opaque behaviour; results that cannot be reproduced.
Workflow deployment	Review queues, escalation routes, action logging and rollback design.	Oversight playbook and action audit trail.	Does the reviewer have the time and access that the assigned control presupposes?	Unchecked execution; automation bias.
Operations	Drift monitoring, feedback capture, incident review and governance cadence.	Monitoring record and remediation log.	Is review still functioning, as distinct from merely being performed?	Undetected degradation; stalled improvement.

Source: Author's own analysis.

7. Research Gaps

The coverage analysis highlights four gaps, stated here as specific absences rather than as a general call for more work.

The first is measurement at the perception locus. No reviewed source provides a means of measuring what an attention-allocation system excluded, or with what consequence. The practical difficulty is genuine, since checking the counterfactual for an unsurfaced item requires the outcome of a case that was never worked,

which is unobservable by construction, and this may be a larger methodological barrier than is recognized. The gap is nonetheless consequential: suppression is among the most pervasive and least examined automated operations in customer service (Varanasi, 2025).

The second is construal adequacy, which is distinct from the more mature question of explainability. Explainability concerns how the model produced its output (Wang & Yin, 2022; Love et al., 2024). No reviewed source offers a way to assess whether a summary or classification faithfully reflects the situation it claims to describe, in the terms a reviewer would need.

The third is the decay of review capacity over time. Reliance research indicates that acceptance calibrates with experience (Kahr et al., 2024) and that more decision support can reduce performance (Ward, 2023). However, no reviewed study measures whether a production review queue is still genuinely reviewing a year after an automation reaches high accuracy, which is precisely the condition under which enterprise oversight matters most (Tokas, 2026a).

The fourth is the gap between assessment artifacts and operative control. Impact assessment (McGregor, 2023; Bogucka et al., 2024) and ethics-based auditing (Mökander & Floridi, 2023) both produce documentation. Neither line of work reports whether the documented controls remained in force after the assessment, which is the only question that determines whether the assessment made a lasting difference. A documented but unenforced control looks identical on paper to an enforced one (Uddandarao et al., 2025).

The four gaps share a structure worth naming, because it explains why they will not close one at a time. Each is an absence, not an event. In each case something did not happen: the case was not surfaced, the detail was not summarized, the review was not genuinely performed, the control was not actually exercised. Existing methods are mostly geared toward artifacts and incidents, which exist and can be studied. Absences create no artifact and no incident, so the instruments that detect them must be designed for that purpose rather than adapted from infrastructures built to observe what is present (Uddandarao, 2026).

8. Future Research Directions

Three directions follow from the gaps, ordered from least to most difficult, together with a fourth that cuts across them.

The first, addressing the decay of review capacity, is the most immediately feasible. Override rate, review dwell time and the catch rate for deliberately seeded errors could be monitored in production without further theoretical work. Over a longer horizon, a longitudinal cohort study of review queues across enterprises would show whether the laboratory finding on reliance calibration (Kahr et al., 2024) holds at organizational scale, and over what timescale.

The second, construal adequacy evaluation, is less developed but has a clear starting point: sampling generated summaries against their source documents and comparing the omitted material with decision-relevant content. This is closer to the evaluation of generated summaries than to explainability, so existing summary-evaluation frameworks offer a base. What must change is the metric rather than the summary format. Summary-evaluation metrics normally measure coverage of salient content as a general reader would prioritize it, whereas construal adequacy requires coverage of the content that would have led the handler to a different decision. The two targets diverge most where it matters most (Gaggar, 2026).

The third, evaluation at the perception locus, is both the hardest and the most valuable. It would require randomised holdouts of items from automated suppression and the collection of the resulting outcomes. This raises ethical questions in customer-facing settings, and the design would need to fall within the evolving governance apparatus described in the auditing literature (Mökander & Floridi, 2023). The uncomfortable symmetry is that establishing whether suppression harms customers requires leaving some cases unsuppressed. Whether that is acceptable depends on the harm profile of the domain, and it is a question for a research ethics committee rather than an architect (Tokas, 2026b).

A fourth direction cuts across the other three. The taxonomy was derived from the literature and has not been tested against practitioners' experience. Interviews with service and sales professionals about where they experience authority as having left them would show whether the five loci hold or whether a different set emerges (Bobba, 2025).

9. Conclusion

Responsible artificial intelligence in customer relationship management is well studied where harm is visible and little studied where it is hidden. Regulation, auditing and design guidance concentrate downstream, on

actions, while triage, summarisation and interpretation operate upstream. Governance of the upstream loci also presumes a reviewing human who can trace the workings of the automation under review, which is precisely the capacity these transfers erode.

The five-locus taxonomy locates that mismatch, but it is not in itself a governance framework. Design-time decisions about which authority a capability will absorb can nevertheless change what gets built. The taxonomy's limitations are that it offers no compliance mapping, that it was derived from literature rather than practice, and that the coverage counts in Section 5 rest on a single reviewer's coding of the corpus, which other reviewers might have coded differently.

What the analysis does establish is an asymmetry that is hard to attribute to anything other than the visibility of harm. The loci that generate an identifiable incident attract governance. The loci of absence, the case not surfaced and the sentence not summarized, generate no incident to investigate and accordingly attract none. Closing the gap does not require new principles; it requires measuring things that currently leave no trace.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of Interest: The author declares no conflict of interest.

Data Availability: No primary datasets were generated or analysed. The full corpus synthesised in this review is listed in the references and is publicly retrievable via the digital object identifiers provided.

Ethics Statement: This work involved no human participants, no animal subjects and no personal data. Ethical approval was not required.

References

1. Bobba, C. T. (2025). Advancements in cosmetic dentistry. Innovation, 1, 2.
2. Bogucka, E., Constantinides, M., Šćepanović, S., & Quercia, D. (2024). AI design: A responsible artificial intelligence framework for prefilling impact assessment reports. IEEE Internet Computing, 28(5), 37-45. https://doi.org/10.1109/MIC.2024.3451351
3. Cao, Y., Duan, Y., & Edwards, J. S. (2026). AI trustworthiness in managerial decision-making: Ethics, transparency, and explainability as key drivers. Journal of Business Ethics. https://doi.org/10.1007/s10551-026-06400-x
4. Chen, Q., Sun, Y., & Liu, X. (2022). Balancing web personalization and consumer privacy concerns: Mechanisms of consumer trust and reactance. Journal of Consumer Behaviour, 21(3), 572-582. https://doi.org/10.1002/cb.1947
5. Ehsan, U., Hussain, S., Song, Y., & Shaukat, K. (2026). AI transparency and bias mitigation in customer service: How ethical culture shapes customer trust. International Journal of Ethics and Systems, 1-25. https://doi.org/10.1108/IJOES-03-2026-0197
6. Gaggar, H. (2025). Architecting scalable enterprise AI agent infrastructure for distributed intelligent systems. International Journal of Advances in Signal and Image Sciences, 193–208. https://doi.org/10.29284/da9w0117
7. Gaggar, H. (2025). Policy-driven access control for secure deployment of autonomous AI agents in enterprise environments. International Journal of Computational and Experimental Science and Engineering, 11(4). https://doi.org/10.22399/ijcesen.5497
8. Gaggar, H. (2026). Protocol-preserving context trimming for agentic workflows: Benefits, failure regimes, and budget guardrails. arXiv. https://doi.org/10.48550/arXiv.2609.16461
9. Guha, S., Khan, F. A., Stoyanovich, J., & Schelter, S. (2024). Automated data cleaning can hurt fairness in machine learning-based decision making. IEEE Transactions on Knowledge and Data Engineering, 36(12), 7368-7379. https://doi.org/10.1109/TKDE.2024.3365524
10. Hennighausen, C., Yarza Navarro-Schär, L., & Eller, E. (2025). Artificial intelligence-mediated communication in e-commerce: Implications for customer trust. International Journal of Consumer Studies, 49(5). https://doi.org/10.1111/ijcs.70111
11. Ioannou, A., Tavalaei, M. M., Vatn, D. M. K., & Mikalef, P. (2026). The role of explainability in artificial intelligence-driven financial decision making. Information Systems Frontiers. https://doi.org/10.1007/s10796-026-10727-1
12. Kahr, P. K., Rooks, G., Willemsen, M. C., & Snijders, C. C. P. (2024). Understanding trust and reliance development in AI advice: Assessing model accuracy, model explanations, and experiences from previous

interactions. ACM Transactions on Interactive Intelligent Systems, 14(4), 1-30. https://doi.org/10.1145/3686164
13. Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. Psychology & Marketing, 38(7), 1140-1155. https://doi.org/10.1002/mar.21498
14. Konatham, M. R., Uddandarao, D. P., Vadlamani, R. K., & Konatham, S. K. (2025). Federated learning for credit risk assessment in distributed financial systems using BayesShield with homomorphic encryption. In 2025 International Conference on Computing Technologies & Data Communication (ICCTDC) (pp. 1–6). IEEE. https://doi.org/10.1109/ICCTDC64446.2025.11158863
15. Kshetri, N. (2025). Governing agentic AI: Security, identity, and oversight in the age of autonomous intelligent systems. Computer, 58(8), 123-129. https://doi.org/10.1109/MC.2025.3572173
16. Lioliou, E., Seitanidi, M. M., & Stadtler, L. (2026). Governance under algorithmic opacity: How financial firms construct accountability and control around artificial intelligence in risk disclosures. Information Systems Frontiers. https://doi.org/10.1007/s10796-026-10762-y
17. Love, P. E. D., Matthews, J., Fang, W., Porter, S., Luo, H., & Ding, L. (2024). Learning to comprehend and trust artificial intelligence outcomes: A conceptual explainable AI evaluation framework. IEEE Engineering Management Review, 52(1), 230-247. https://doi.org/10.1109/EMR.2023.3342200
18. Lu, Q., Zhu, L., Xu, X., & Whittle, J. (2023). Responsible-AI-by-design: A pattern collection for designing responsible artificial intelligence systems. IEEE Software, 40(3), 63-71. https://doi.org/10.1109/MS.2022.3233582
19. Marzouk, R., Zitoun, S., Belghith, O., & Skhiri, S. (2023). The building blocks of a responsible artificial intelligence practice: An outlook on the current landscape. IEEE Intelligent Systems, 38(6), 9-18. https://doi.org/10.1109/MIS.2023.3320438
20. McGregor, S. (2023). A scaled multiyear responsible artificial intelligence impact assessment. Computer, 56(8), 20-27. https://doi.org/10.1109/MC.2022.3231551
21. Mökander, J., & Floridi, L. (2023). Operationalising artificial intelligence governance through ethics-based auditing: An industry case study. AI and Ethics, 3(2), 451-468. https://doi.org/10.1007/s43681-022-00171-7
22. Naidoo, V. (2026). The dark side of artificial intelligence-enabled industrial service recovery: The erosion of trust and customer engagement. Industrial Marketing Management, 135, 205-223. https://doi.org/10.1016/j.indmarman.2026.05.004
23. Restrepo Amariles, D., & Baquero, P. M. (2026). Compliance under the EU AI Act: How firms navigate regulatory complexity in the digital landscape. Computer Law & Security Review, 62, 106390. https://doi.org/10.1016/j.clsr.2026.106390
24. Tokas, D. (2026). Distributed big data frameworks and high-scale service design: Proven engineering patterns across cloud-native deployments. International Journal of Computational and Experimental Science and Engineering, 12(1). https://doi.org/10.22399/ijcesen.5086
25. Tokas, D. (2026). Interoperability standards in modern workflow orchestration systems. Journal of Information Systems Engineering and Management, 11(2s), 646–654. https://jisem-journal.com/index.php/journal/article/view/14459
26. Uddandarao, D. P., Konatham, M. R., & Vadlamani, R. K. (2025). Robust and scalable statistical models for high-dimensional marketing data applications: A comprehensive review. In 2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) (pp. 1–8). IEEE. https://doi.org/10.1109/ICERECT65215.2025.11377070
27. Uddandarao, D. P. (2026). Statistics and the science of causal economics: A new paradigm for data science. Deep Science Publishing.
28. Uddandarao, D. P. (2026). The reality of an AI implementation. Communications of the ACM, 69(8), 8–9. https://doi.org/10.1145/3821443
29. Varanasi, B. S. S. (2025). AI-enabled continuous authority-to-operate (ATO) frameworks for information systems. Enterprise Development and Microfinance, 35(2), 748–757. http://www.papjournals.com/index.php/edm/article/view/1201
30. Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on artificial intelligence systems during decision-making. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW1), 1-38. https://doi.org/10.1145/3579605

31. Vytla, S. K. (2024). POLICY-ML: Policy-as-code enforcement for compliance, auditability, and trust across data and machine learning pipelines. International Journal of Computational and Experimental Science and Engineering, 10(4). https://doi.org/10.22399/ijcesen.5393
32. Vytla, S. K. (2025). DevOps-GRID: DevOps-driven reliability and digital-twin operations for intelligent power systems. 57(3), 3407–3416.
33. Wagner, M., Song, H., Borg, M., Engström, E., & Lysek, K. (2026). AI Act high-risk AI compliance challenge and industry impact: A multiple case study. Information and Software Technology, 194, 108067. https://doi.org/10.1016/j.infsof.2026.108067
34. Wang, X., & Yin, M. (2022). Effects of explanations in artificial intelligence-assisted decision making: Principles and comparisons. ACM Transactions on Interactive Intelligent Systems, 12(4), 1-36. https://doi.org/10.1145/3519266
35. Ward, P. (2023). Choice, uncertainty, and decision superiority: Is less artificial intelligence-enabled decision support more? IEEE Transactions on Human-Machine Systems, 53(4), 781-791. https://doi.org/10.1109/THMS.2023.3279036
36. Westerstrand, S. (2025). Fairness in artificial intelligence systems development: EU AI Act compliance and beyond. Information and Software Technology, 187, 107864. https://doi.org/10.1016/j.infsof.2025.107864
37. Zhang, Y., Wu, M., Zhang, G., & Lu, J. (2025). Responsible artificial intelligence: Portraits with intelligent bibliometrics. IEEE Transactions on Artificial Intelligence, 6(8), 1977-1990. https://doi.org/10.1109/TAI.2024.3510474