

The Human-in-the-Loop Paradigm Orchestrating Human Intelligence and Agentic AI for Scalable Customer Experience: A Metrics-Focused Review

Aishwary Bodhale¹; Akanksha Meshram²

¹Hutton School of Business, University of Cumberland, Kentucky, USA

²Assistant Professor, Department of CSE AIML, Oriental Institute of Science & Technology Bhopal

Publication Date: 2026/03/14

Abstract: Human-in-the-Loop (HiL) systems integration into customer experience (CX) schemes is a disruptive strategy in terms of expanding service delivery without sacrificing quality. This is a review article that investigates the paradigm shift of orchestrated human intelligence and agentic AI in the context of CX with particular attention given to the metrics that are needed to assess, optimize, and scale such systems of hybrid forms. By conducting a topical review of literature published in 2015-2024, we realize the fact that the current CX measures do not sufficiently reflect the dynamic interaction between humans and artificial agents. According to the traditional measures of endpoint (CSAT, NPS, CES), it is not possible to measure the quality of orchestration, the effectiveness of handoff and the effectiveness of human-AI cooperation. Our framework of metrics includes five dimensions that are structured as (1) Orchestration Efficiency Metrics, (2) CX Journey Analytics, (3) Agent Performance Indicators, (4) Business Impact Measures, and (5) Learning and Adaptation Metrics. The review provides a synthesis of 87 relevant studies, which uncovered three new trends: the continuous versus discrete approach to measurement, the necessity to track the emotional trajectory, and the necessity to align multi-stakeholder metrics. This is, to the best of our knowledge, one of the first reviews to explicitly conceptualize human, AI orchestration as a quantifiable construct and to systematize measurements in specific Human-in-the-Loop customer experience system metrics. We end with research agenda that focuses on creation of real time orchestration scoring systems, considerations of ethical metrics and industry specific implementation. This publication has a contribution to the academic literature and practice as it offers a systematic method of measuring what is important in human-AI collaborative CX systems.

Keywords: Human-in-the-Loop; Agentic AI; Customer Experience Metrics; Service Orchestration; Human-AI Collaboration; CX Measurement.

How to Cite: Aishwary Bodhale; Akanksha Meshram (2026) The Human-in-the-Loop Paradigm Orchestrating Human Intelligence and Agentic AI for Scalable Customer Experience: A Metrics-Focused Review. *International Journal of Innovative Science and Research Technology*, 11(3), 698-708. <https://doi.org/10.38124/ijisrt/26mar527>

I. INTRODUCTION

➤ *The Evolution of Customer Experience Delivery*

The last ten years have seen a radical change in the world of customer service thanks to the development of artificial intelligence, evolving customer demands, and financial pressures surrounding scalable service provisions. The customer experience that used to be provided mainly by the human-to-human touch is becoming more advanced in modern CX ecosystems where human service operators are being paired with advanced artificial intelligence systems [1]. The result of this integration has become the so-called Human-in-the-Loop paradigm (HiL) paradigm, according to which tasks are strategically distributed between human and artificial intelligence depending on the complexity, context, and capabilities requirements [2].

The potential of HiL systems in CX is huge: theoretically, it will provide scalability and efficiency of automation without losing empathy, creativity, and the ability to understand human beings [3]. Nevertheless, with more organizations currently implementing these hybrid systems, there has arisen a critical question: what can we do to measure how they are performing and the effect they are having on customer experience?

➤ *The Metrics Challenge in Human-AI Orchestration*

The conventional customer experience measures were created on environments where service was dominant by humans. The Net Promoter Score [4], Customer Satisfaction (CSAT) and Customer Effort Score [5] have some inherent shortcomings when used in the context of HiL systems: they are end-of-service measures that capture the customer perception but they tell very little about the dynamics of the

process of human-AI interaction [6]. They do not provide essential answers to such questions: When is a conversation to be transferred? What was the level of smoothness of that handoff? How good is the process of collaborative problem-solving between human and AI agents? And what becomes of this co-operation as time goes on?. Figure 1 compares conventional endpoint-based customer experience measurements with the continuous Human-in-the-Loop orchestration measurement by showing the drawbacks of traditional evaluation models in hybrid service systems.

The inconsistency between the existing measurement habits and the ones required to achieve optimal performance at the HiL level acts as a substantial research problem and a business issue to address [7]. The lack of proper metrics will make organizations fail in improving their human-AI orchestration systematically, which is likely to deprive them of the promised gains of such integrated systems.

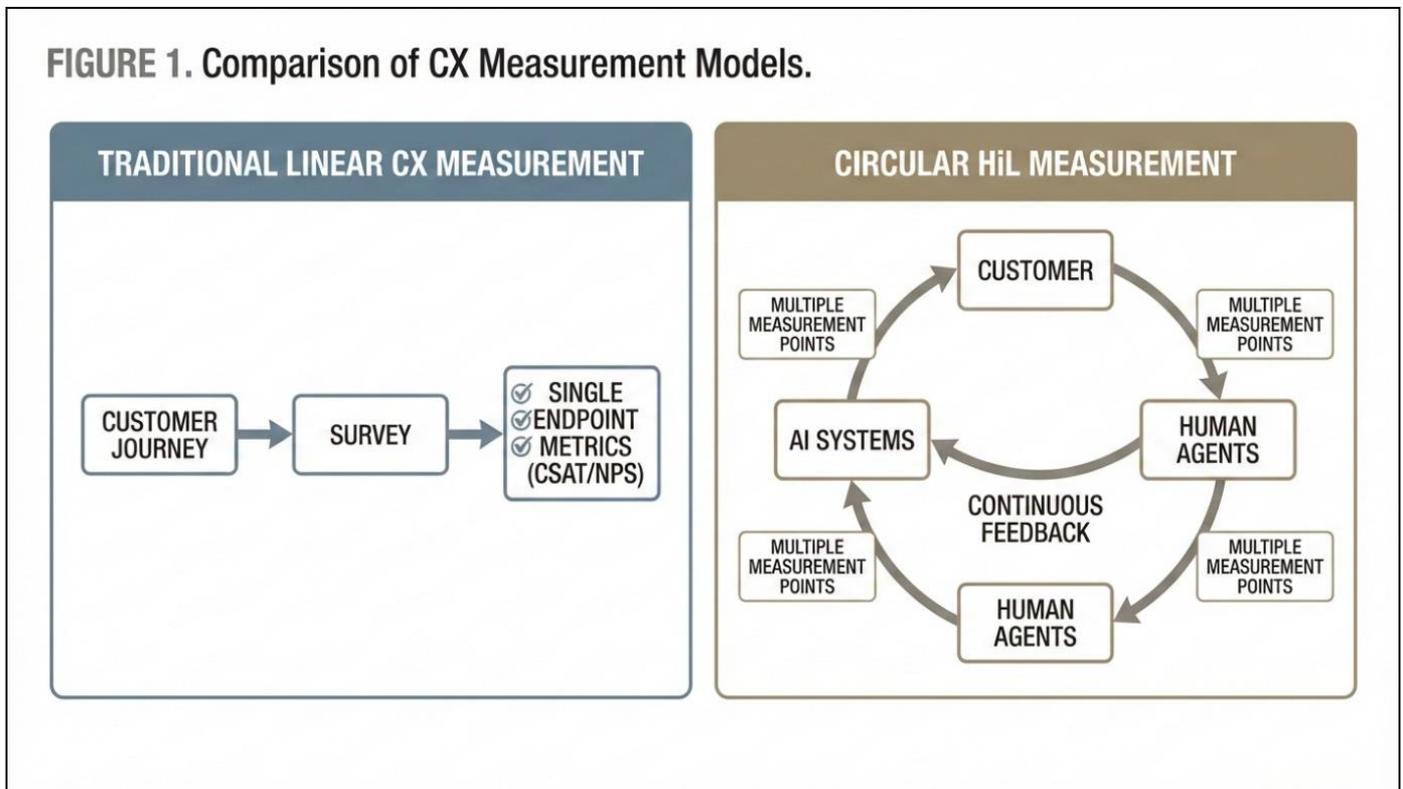


Fig 1 Contrasting Measurement Paradigms: Traditional Endpoint Metrics (Left) Versus Continuous HiL Orchestration Measurement (Right).

➤ *Article Scope and Structure*

This critical gap is closed in this review article, which is a systematic review of available literature on the topic of HiL systems in CX that pays special attention to the methods of measurement and measures frameworks. The four fundamental questions that we base our analysis on are:

- What are the current measures to gauge the performance of HiL systems when applied to CX situations?
- What are the theoretical grounds on which the human-AI collaboration should be measured?
- What are the weaknesses of existing measurement strategies?
- What is a complete metrics system of HiL-CX systems?

We have then proceeded with this introduction, and in Section 2, we will outline our methodology. Section 3 goes over the development of CX metrics and their shortcomings of HiL systems. In section 4, the human-AI collaboration structures are discussed. Section 5 summarizes the results into a suggested metrics framework. Section 6 deals with the

challenge of implementation and the last section, Section 7, deals with the conclusions and future research direction.

➤ *Contributions of This Review*

This review contributes to the literature on customer experience and human-AI collaboration in 3 general ways. The first is that it consolidates the fragmented studies on customer experience measurements, Human-in-the-Loop systems, and agentic AI into one analytical approach that deals with measurement. Second, it redefines the concept of evaluation in hybrid service systems in introducing human-AI orchestration as a first-class measurement construct, beyond traditional endpoint-based measuring CX. Third, it suggests Human-AI Orchestration of Customer Experience (HAIOCX) framework, which suggests a structured and theory-based methodology of measuring, controlling, and enhancing human-AI collaborative customer experience systems.

II. METHODOLOGY

➤ Systematic Review Protocol

The systematic review follows the guidelines of Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) guidelines of systematic literature reviews [8]. The transition to data-driven, learning-based AI systems, which shift the design and measurement needs of Human-in-the-Loop customer experience systems, commences in 2015, when rule-based automation came to an end. We have used three main areas in our search strategy, including: (1) customer experience measurement literature, (2) literature on human-in-the-loop systems, and (3) literature about human-AI collaboration.

• Inclusion Criteria:

- ✓ Peer-reviewed articles (2015-2024) that deal with the topic of HiL systems, CX metrics, or human-AI collaboration.
- ✓ The reports of recognized research companies (Gartner, Forrester, McKinsey) in the industry.
- ✓ Patterns of applied HiL-CX systems.
- ✓ Human-machine teaming or service orchestration theoretical papers.

• Exclusion Criteria:

- ✓ Articles that only dealt with fully automated systems (not human integrated).
- ✓ Artificial intelligence development papers that do not mention the use of CX.
- ✓ Non-English publications
- ✓ Articles whose methodology is not described clearly.

➤ Search Strategy and Sources

We have searched several databases: Scopus, Web of Science, ACM Digital Library, IEEE Xplore and Google

Scholar. The search terms were: [human-in-the-loop OR human-AI collaboration OR human-machine teaming] + [customer experience OR customer service OR CX] + [metrics Or measurement or evaluation or KPIs].

Primary search showed 423 articles. Upon title/abstract screening, 157 articles were left. The review of a full text led to the identification of 87 articles that fit all inclusion criteria. These were accompanied by 23 industry reports and 15 case studies to provide some practical background. In order to enhance the analytical rigor of the review, the triangulation of multiple sources of data was conducted. Besides the 87 peer-reviewed scholarly articles, the synthesis also included 23 industry reports published in 2019-24 by Gartner, Forrester, and McKinsey and 15 documented implementation case studies. The thematic coding was used to analyze the literature, and to achieve greater reliability, several reviewers were involved in the independent coding. Differences were settled by a discussion, thus, achieving high inter-coder agreement. The generated classifications and metric groupings were further narrowed down by expert review by researchers and practitioners in the industry experienced in customer experience management and with experience in human-AI systems.

➤ Analytical Framework

Thematic analysis was done using NVivo software to analyze articles [9]. We have used a hybrid coding method, which consisted of deductive codes using our research questions and emergent codes of the literature. The coding structure of the thematic analysis that was applied to the literature synthesis is presented in Figure 2. Analysis was directed towards: (1) the types and the classification of metrics, (2) the method of measuring, (3) the theoretical basis, (4) the challenge of implementation, and (5) gaps in research.

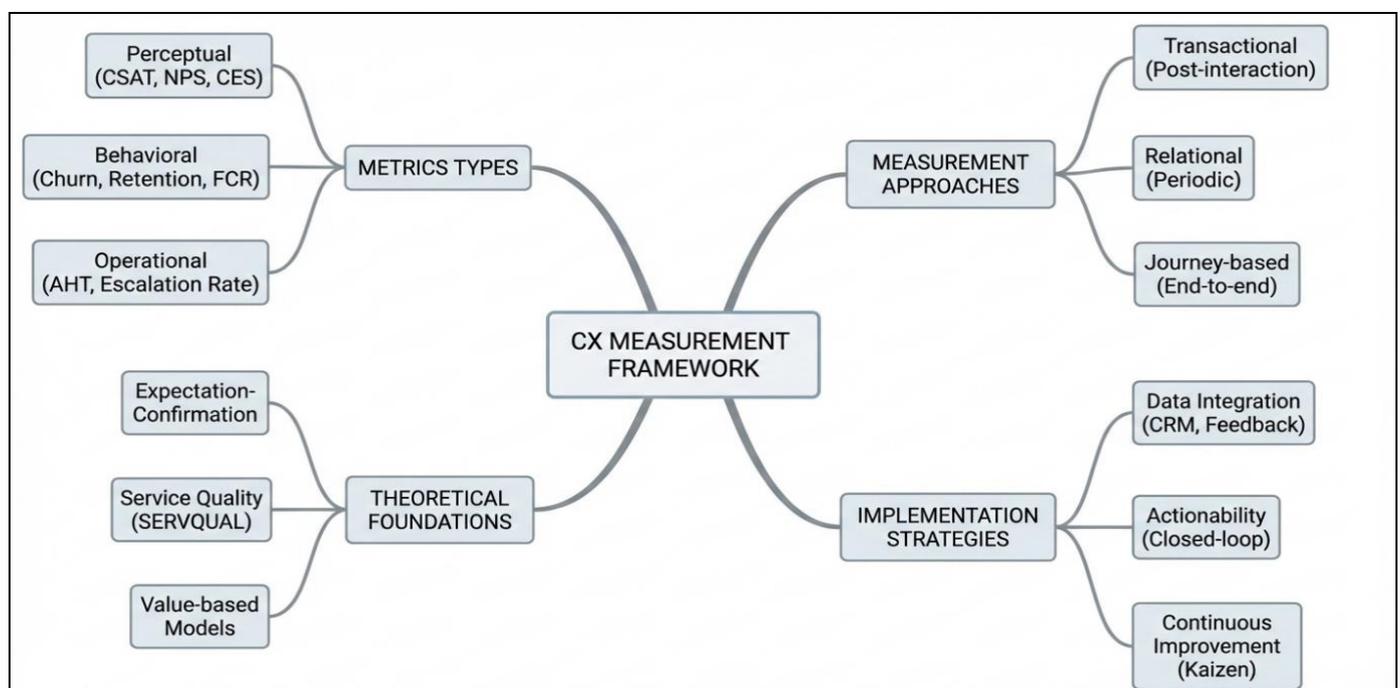


Fig 2 Thematic Analysis Coding Structure Used for Literature Synthesis.

III. THE EVOLUTION AND LIMITATIONS OF CX METRICS

➤ *Historical Development of CX Metrics*

Table 1 Evolution of Dominant CX Metrics.

Era	Dominant Metrics	Primary Focus	Measurement Approach	Key Limitations for HiL Systems
1980s-1990s	Customer Satisfaction (CSAT)	Transaction quality	Post-interaction surveys	Static, doesn't capture process dynamics
2000s	Net Promoter Score (NPS)	Customer loyalty	Likelihood to recommend	Lagging indicator, attribution challenges [4]
2010s	Customer Effort Score (CES)	Service efficiency	Perceived effort required	Over-simplifies complex interactions [5]
2015-Present	Customer Journey Analytics	End-to-end experience	Touchpoint tracking	Often lacks human-AI distinction [10]
Emerging	Emotional Analytics, Real-time Feedback	Momentary experience	Continuous measurement	Integration challenges with HiL systems [11]

Note. Synthesized by the author from seminal and recent literature on customer experience metrics and human-AI collaboration.

Its evolution can be traced as follows: simple transactional metrics (CSAT) to relational ones (NPS), to process measures (CES), and the last step towards the comprehensive journey analytics [10]. Nevertheless, the most sophisticated modern methods will hardly differentiate human-based and AI-based components of services, not to mention quantifying the quality of their orchestration [12]. This development of a transactional to journey-based measurement practices is depicted in Figure 3.

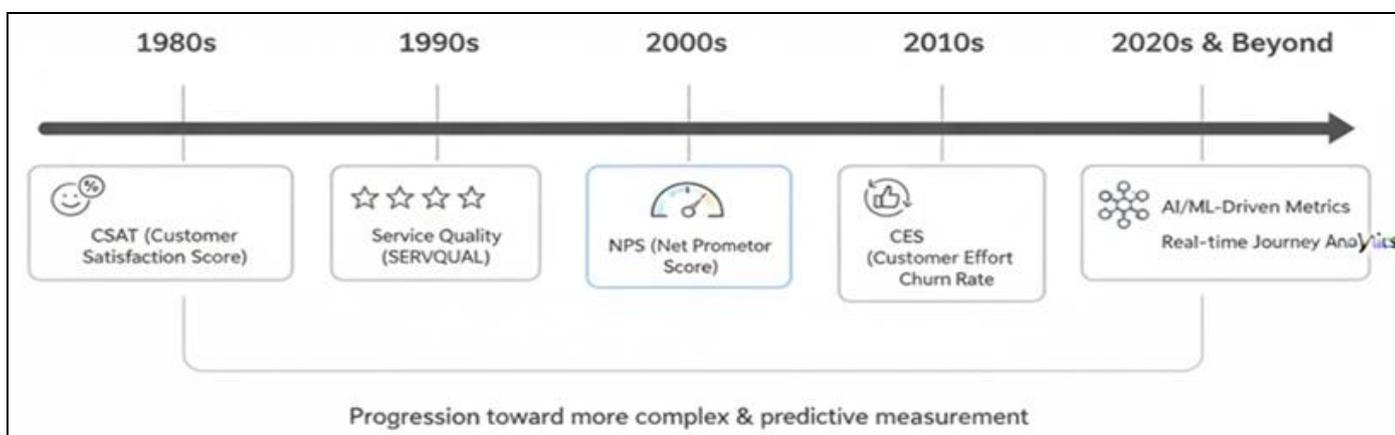


Fig 3 Timeline of CX metric evolution

➤ *Current Industry Practice: A Gap Analysis*

The review of 45 industry reports (2019-2024) shows that 78% of organizations that use HiL systems still rely on the traditional metrics mostly, and only 22% of them develop custom metrics [13, 14]. Issues that have been identified to be common include:

- Attribution Problem: Inability to learn whether the results are due to human intervention, AI performance, or a combination of the two [7].
- Process Opacity: Minimal transparency of the human-AI handoff decision-making and implementation [15] into the black box.
- Temporal Misalignment: Data collection is done quarterly or monthly because the fast learning cycles in AI systems cannot be tracked [3].
- Disconnect between Stakeholders: Various measures of technical teams (AI accuracy), operational teams (handle time), and business units (NPS) [1].

➤ *Theoretical Foundations for HiL-CX Measurement*

There are four theoretical perspectives on which we view HiL measurement:

- Service-Dominant Logic [16]: Focuses on the value co-creation between the provider and the customer. In HiL, this extends to include value co-production among human, AI, and customer a triple relationship where triple measures are needed [12].
- Distributed Cognition Theory [17]: Takes a cognitive process as distributed between human and artificial agent. According to this view, metrics are recommended to evaluate the system, rather than the individual components of the cognitive system [18, 19].
- Affective Computing [20, 21]: Emphasis is put on the role of emotional aspects in the interaction between a human and a computer. HiL measures should ensure that emotional paths are measured based on service experiences [11].
- Human-AI Teaming Literature [22, 23]: It focuses on such concepts as shared situational awareness, interdependence, and adaptability, which need specialized measurements.

Although every of these theoretical stances provides insightful information about factors involved in Human-in-the-Loop systems, the available methods of measurement are

rather fragmented, and measures are normally translated to the human performance, AI accuracy, or customer outcomes independently. What is striking in the literature of the past is the explicit consideration of orchestration as a construct in its own right, in other words, the attribute of ascription, coordination, and accommodation between human and artificial actors over time. The measures that are synthesized within this review are thus not presented as new measures per se, but rather as theoretical-operationalization based on solid underpinnings in service science, human computer interaction, cognitive psychology and organizational theory. The concepts of satisfaction, effort, autonomy, emotional response, and learning are informed by classic theories and

measurement traditions, and more recent studies of human and artificial intelligence collaboration tell orchestration-specific metrics, including the quality of handoff, and the preservation of the context, and adaptive task allocation. This foundation guarantees the continuation of the previous measurement strategies alongside the extension of these methods to the dynamics of hybrid human-AI service systems. This is the gap that drives the necessity of customer experience orchestration-based measurement framework. These theoretical views are all synthesized in figure 4 and how they collectively inspire orchestration-based measurement in Human-in-the-Loop customer experience systems.



Fig 4 Theoretical foundations of HiL measurement.

IV. HUMAN-AI COLLABORATION FRAMEWORKS IN CX

➤ Typology of HiL Architectures in CX

Table 2 Common HiL Architectures and Their Measurement Implications

Architecture	Description	Primary Use Cases	Measurement Challenges
Sequential Handoff	AI handles initial contact, escalates to human when needed	Customer support, sales qualification	Handoff quality, context preservation [24]
Parallel Processing	Human and AI work simultaneously on different aspects	Complex problem-solving, data analysis	Coordination efficiency, integration quality [25]
Human Oversight	AI executes with human supervision	Content moderation, fraud detection	Oversight effectiveness, alert fatigue [26]
AI Augmentation	Human leads with AI assistance	Personalization, recommendation systems	Augmentation value, cognitive load [3]
Dynamic Orchestration	Continuous reallocation based on real-time assessment	Premium service tiers, crisis management	Orchestration intelligence, adaptability [22]

Note. Synthesized by the author from seminal and recent literature on customer experience metrics and human-AI collaboration.

➤ *Agentic AI Capabilities Relevant to CX*

In CX settings, the capabilities of modern agentic AI systems are generally specific to measurement strategies [25, 27]:

- Independent Task Performance: Capacity to perform specified matters without involving human beings.
- Contextual Understanding: Customer situation recognition, history and emotional condition.
- Predictive Escalation: When intervention of human being is expected.
- Feedback Learning: Behavior Incorporation Human corrections into future behavior.

- Multi-modal Interaction: Fluid interaction between chat and voice, among others.

Every capability implies the following metrics: autonomy success rate, context accuracy score, escalation prediction accuracy, learning rate, and cross-channel consistency [7].

These capabilities are measured to specify what needs to be observable, measurable, and attributable to Human-in-the-Loop customer experience evaluation. The range of agentic AI capabilities applicable to customer experience applications is shown in Figure 5.

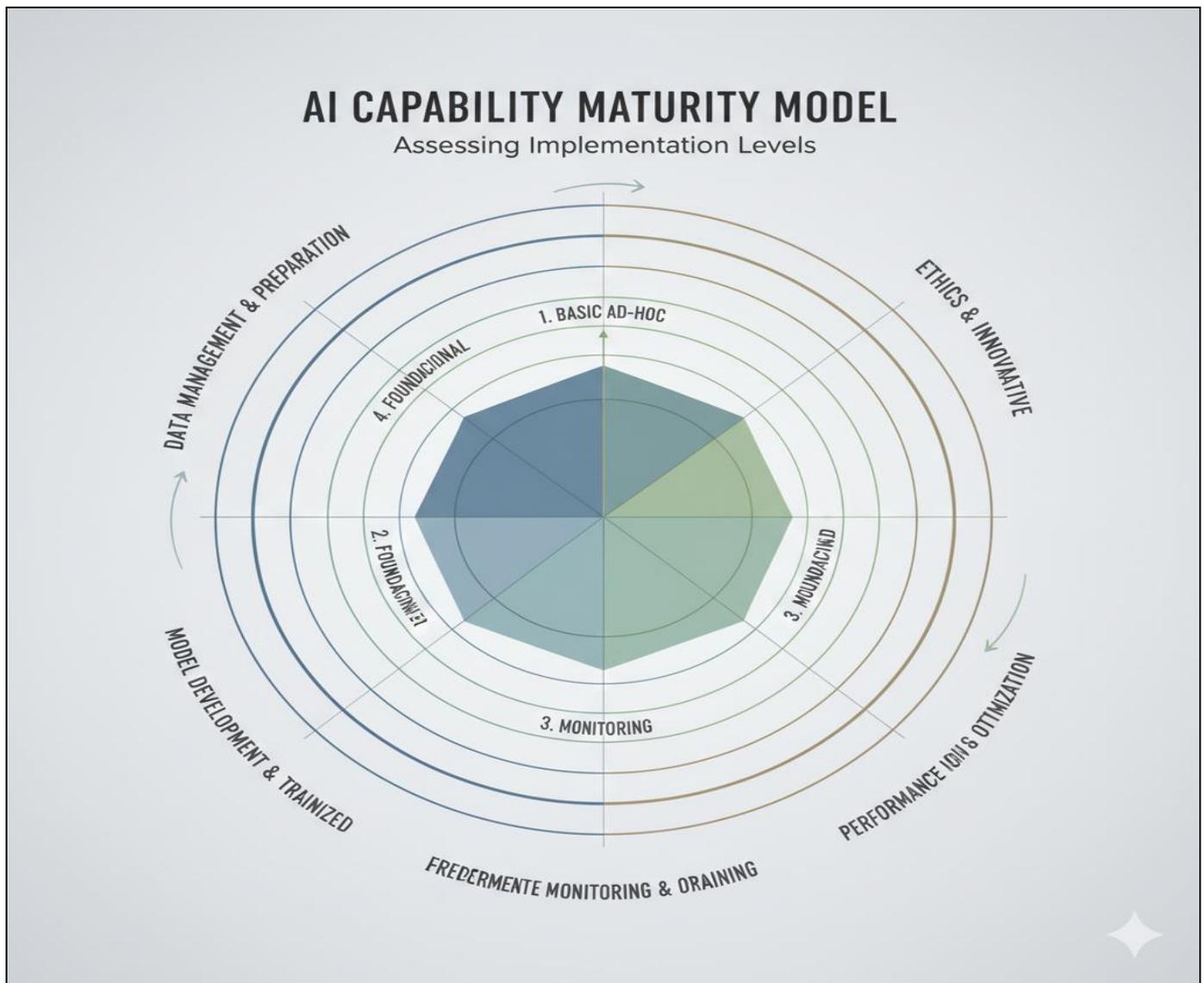


Fig 5 Agentic AI Capabilities Spectrum

➤ *Human Roles in HiL-CX Systems*

Human-involvement aspects of the HiL systems include various functions, which demand various performance indicators [15, 2]:

- Exception Handlers: Working with AI incompetent cases.
- Quality Auditors: Detection and correction of the AI outputs.
- Trainers: Feedback to enhance the performance of AI.

- Orchestrators: These are in charge of the entire human-AI workflow.
- Empathy Providers: How to manage emotionally charged situations.

This distinction in roles implies role-specific measures, instead of generic measures of a human agent [26].

V. PROPOSED METRICS FRAMEWORK FOR HIL-CX SYSTEMS

It is necessary to explain the essence of the contribution that the suggested framework can offer before presenting it. The HAIOCX framework does not aim at substituting the established customer experience, operation, or AI performance measures nor does it purport novelty in the indicators. The value it adds is that it brings human orchestration of AI to the dimension of primary measurement and merges the current and new metrics into a coherent framework to capture the dynamics of hybrid service delivery instead of fixed points.

We suggest the Human-AI Orchestration of Customer Experience Metrics Framework, which advances in five

dimensions and its metrics, based on our literature review. The entire HAIOCX framework is given in Figure 6.

In the case of a pair of target ranges or indicative thresholds having been given, target ranges are to be understood as analytic as opposed to prescriptive thresholds. They capture ranges recorded in previous empirical studies, in the industry, and recorded instances of implementation in the literature reviewed. The reason behind having such ranges is to facilitate comparative evaluation and diagnostic argumentation as opposed to having similar performance expectations across diverse service settings.

➤ *Dimension 1: Orchestration Efficiency Metrics*

These metrics assess the technical execution of human-AI collaboration:

Table 3 Orchestration Efficiency Metrics

Metric	Definition	Measurement Method	Target Range	Key References
Handoff Success Rate	Percentage of transfers completed without information loss	System logging + customer survey	>95%	[24, 22]
Context Preservation Score	Degree to which customer context is maintained during transitions	Comparative analysis of pre/post-handoff context	>90%	[7]
Transfer Time	Time required for complete agent-to-agent transfer	System timestamps	<15 seconds	[25]
Orchestration Decision Accuracy	Percentage of appropriate human/AI task allocations	Expert review of allocation decisions	>85%	[26]
System Uptime	Percentage of time HiL system is fully operational	System monitoring	>99.5%	[3]

Note. Synthesized by the author from seminal and recent literature on customer experience metrics and human-AI collaboration.

Illustrative example: Orchestration efficiency measures are especially important in contact center settings, in cases of AI-to-human escalation. As an example, low success rate in handoff or low context preservation score can suggest that customers have to repeat their issues post-escalation, which directly negatively affects the perceived quality of services even with good scores on endpoint satisfaction.

Even though a high orchestration efficiency is a desired condition, excessive ambition to optimize the speed of handoff can be dangerous due to a possible risk of escalation or loss of customer confidence.

➤ *Dimension 2: Customer Experience Journey Metrics*

These metrics capture the customer's experience through the HiL-mediated service journey:

Table 4 CX Journey Metrics

Metric	Definition	Measurement Method	Theoretical Basis	Key References
Emotional Trajectory Index	Pattern of emotional states throughout interaction	Sentiment analysis across interaction segments	Affective computing	[11, 21]
Cognitive Load Score	Perceived mental effort required from customer	Post-interaction survey + interaction analysis	Cognitive load theory	[28]
Agency Perception	Customer's sense of control and understanding	Validated questionnaire items	Self-determination theory	[29]
Channel Consistency	Uniformity of experience across interaction channels	Cross-channel experience mapping	Service design principles	[10]
Resolution Confidence	Customer certainty that issue is fully resolved	Post-resolution survey	Uncertainty reduction theory	[30]

Note. Synthesized by the author from seminal and recent literature on customer experience metrics and human-AI collaboration.

Illustrative example: In the domain of financial services dispute resolution, emotional trajectory tracking could be used to identify whether customer frustration before human intervention following the introduction of AI reduces or exacerbates the emotional reaction. Negative emotional paths even when problems are resolved in the end may indicate failure of orchestration which cannot be measured by traditional post interaction measures.

Notably, more detailed experiential measures are what should be supplementary not substitutive to more basic ones since too fine a grained measurement might blur actionable information.

➤ *Dimension 3: Agent Performance Metrics*

These metrics assess both human and AI agent performance within the collaborative system:

Table 5 Agent Performance Metrics

Metric Type	Human-Specific Metrics	AI-Specific Metrics	Collaborative Metrics	Key References
Efficiency	Cases handled per hour	Conversations per compute unit	Total system throughput	[15]
Quality	First-contact resolution rate	Intent recognition accuracy	Joint problem-solving success	[2]
Learning	Skill acquisition rate	Model improvement from feedback	System adaptation speed	[25]
Adaptability	Range of handled case types	Performance across domains	Dynamic role adjustment	[22]

Note. Synthesized by the author from seminal and recent literature on customer experience metrics and human–AI collaboration.

Illustrative example: In triage in the healthcare sector or insurance claims processing, collaboration indicators like the success of joint problem-solving and the rate of system adaptation can be used to show how well human expertise is being transferred into the AI-enhanced workflows and can support the quality of services and their scalability to the long term.

➤ *Dimension 4: Business Impact Metrics 5.4.*

These measures can relate the performance of the HiL to the organizational results [12, 31]:

- Cost Per Resolved Query: Human and AI cost total/successful resolutions.
- HiL ROI: (Value of outcomes - System costs)/ System costs.
- Scalability Efficiency: Percentage change in capacity/percentage change in cost.
- Impact of Customer Lifetime Value: Change in CLV due to HiL implementation.
- Employee Retention Correlation: Correlation between the use of the HiL system and retention of the agents.

It is difficult to attribute business performance to Human-in-the-Loop configurations only, especially in the context of a setting in which multiple transformation initiatives operate simultaneously.

➤ *Metric D5: Learning and Adaptation.*

These indicators follow system improvement as time goes on [3, 23]:

- Human-to-AI Knowledge Transfer Rate: The rate of human knowledge automation into capability.
- Reduction Trajectory: Rate of the reduction in errors in the human and AI components.
- Orchestration Optimization Rate: Enhancement of the decisions related to the allocation of tasks with time.
- New Scenario Management Ability: Deepening of successfully resolved case types.

Although the dimensions suggested are analytically different to each other, in reality, they are highly interdependent and enhancement in one aspect might bring about trade-offs in another.

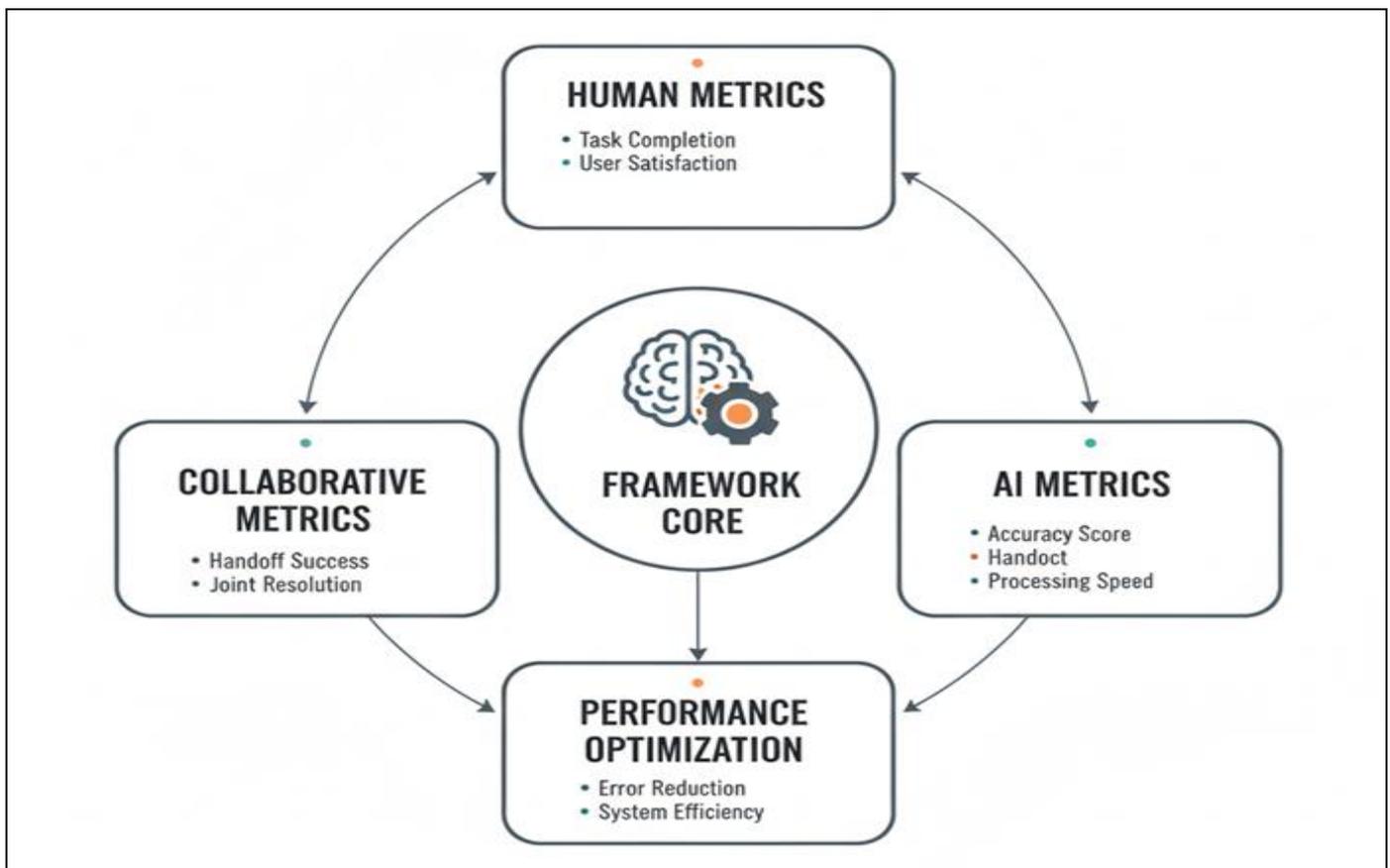


Fig 6 Complete Human-AI Orchestration for Customer Experience (HAIOCX) Framework Visualization

VI. IMPLEMENTATION CONSIDERATIONS AND CHALLENGES

➤ *Measurement Infrastructure Requirements*

To apply the HAIOCX framework, one will need the following capabilities [1, 7]:

- Unified Data Layer: Customer feedback, business data and system logs integration.
- Real-time Analytics: Capacity to calculate measurements in a very brief period.
- Multi-modal Data collection: text, voice, sentiment and behavioral data combination.
- Longitudinal Tracking: Tracking measures between customer journeys and time.
- Benchmarking Capability: Relatedness with industry standards and historical performance.

➤ *Organizational Alignment Challenges*

To achieve the implementation of advanced Human-in-the-Loop metrics, several organizational barriers will have to be overcome in terms of the metric ownership, interpretation, and utilization. [15, 3]

- Siloed Metrics Ownership: The various departments tend to have disparate metrics which are not integrated.
- Opposition to New Measurement: Usage of Familiar Measures in spite of constraints.
- Skill Gaps: Non-existence of analytical skills of advanced HiL metrics.
- Privacy Issues: Weighing the needs of measuring and protecting customer data.
- Change Management: The shift to process-oriented measurement End to End.

➤ *Ethical Considerations in HiL Measurement*

In addition to technical and organizational issues, the formulation and application of Human-in-the-Loop metrics bring up significant issues of ethics which directly influence the manner in which performance is assessed and optimized. The development of metrics will have to deal with ethical aspects [26, 32]:

- Transparency: definite information as to what is being measured and why.
- Fairness: Making sure that metrics do not provide perverse incentives and biased evaluation of agents.
- Privacy Protection: The reduction of obtrusive data gathering and preservation of measuring validity.
- Human Dignity: Meaningful human work as opposed to a blind maximization towards automation.
- Customer Autonomy: Customer preferences in human vs. AI interaction.

Technically robust metrics can still have undesirable effects in that they implicitly place more emphasis on automation than on meaningful human input.

➤ *Limitations of This Review*

There are a number of limitations with this review. To start with, the proposed HAIOCX framework is purely conceptual and has not been confirmed empirically. This is an intentional design decision in line with the aims of a systematic review that is designed to synthesize and organize the current knowledge and not hypotheses. Empirical validation in the various industries and Human-in-the-Loop architectures is thus a significant research direction in future studies.

Second, the results of this review are subject to nature to the size and quality of the literature at hand. Although systematic searches in various academic databases and in the industry sources were conducted, the relevant studies might have been left out because of publication bias, restrictions of databases coverage, or the exclusion of non-English articles.

Third, customer experience situations differ greatly in terms of industry and organizational cultures and regulatory environments. Consequently, not every measure presented in the HAIOCX framework will be applicable in the same environment equally. The paradigm is thus to be understood as a loose guideline and not a prescriptive measure rule.

Besides, certain of suggested metrics presuppose the presence of the supportive technological infrastructure, including real-time analytics, detailed system logs, and integrated customer data platforms. Companies that are less digitally mature might need to adapt or implement the framework at a slow pace. Moreover, much of the evaluated literature represents the service setting in Western organizational settings, which implies that cross-cultural validation and situational normalization of future studies are necessary.

VII. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

➤ *Key Findings Summary*

There are a number of important insights that have been identified in this review:

- Measurement Gap: The existing CX measures are ineffective when assessing the quality of processes in HiL systems, which are endpoint-based [6, 12].
- Multi-dimensional Requirement: Achieving the effective HiL measurement needs measures in such dimensions as technical, experiential, agent, business, and learning [22].
- Dynamic Nature: HiL systems must have metrics that reflect the system adaptation and performance improvement with time, not a static performance [3].
- The Imperative of Integration The measures of isolated human or AI components lack the collaborative aspect that is critical [2].
- Practical Issues: There are technical, organizational, and ethical barriers to implementation that have to be overcome [26].

➤ *Theoretical Contributions*

The main theoretical contribution of the review is the conceptualization of human-AI orchestration as a specific and quantifiable phenomenon in the customer experience system that goes beyond the scope of the current service, IS, and HCI measurement systems that still rely on the independent actors or results.

The work makes contributions to various areas of theoretical work:

- **Service Science:** Applies the measurement frameworks of services to human-AI systems [16]
- **Human-Computer Interaction:** It gives measures to assess cooperative interfaces and workflows [18]
- normally identifies measurement methods that enable ethical human-AI integration [32].
- **Organizational Learning:** Relates measurement of the system to the processes of continuous improvement [33].

➤ *Practical Implications*

In the case of practitioners, it is recommended in this review [1, 7]:

- **Begin with Diagnosis:** Evaluate existing gaps in measurement with new measures to be put in action.
- **Phase Implementation:** Start with critical orchestration measures and then widen to full measurements.
- **Cross-functional Teams:** CX, IT, analytics, and HR are involved in the development of metrics.
- **Iterative Refinement:** Consider metrics to be dynamic and not static.
- **Clear Communication:** State new metrics clearly to all the stakeholders.

➤ *Future Research Agenda*

The areas of priority research are:

- **Validation Studies:** Empirical experimentation of suggested metrics in the various industries and in architecture of hi-L.
- **Real-time Orchestration Scores:** Real-time adjustment of human-AI allocation with the development of composite indices.
- **Sector Adoptions:** General frameworks Fit to Sector (healthcare, finance, retail).
- **Longitudinal Impact Studies:** Evidence on the relationship of HiL metrics and business long-term performance.
- **Cross-cultural Issues:** The study of the variation in HiL measurement in different cultures.
- **Development of Ethical Framework:** Development of responsible guidelines on measurement of HiL.
- **Integration with AI Governance:** Bridging the gap between metrics and AI ethics and governance.

➤ *Final Remarks*

The Human-in-the-Loop paradigm is a major shift in the sphere of providing customer experience, and such an approach has the potential to use both AI scalability and the

human intelligence capacity to deliver and adjust to situations [3, 22]. This possibility demands equally developed measurement methods that represent the peculiarities of human-AI interaction. The model suggested in this review can be used by researchers and practitioners who want to create such approaches to have a beginning point. Since HiL systems are becoming more advanced and popular, it will be necessary to have proper metrics in place that would guarantee that these systems would not just provide efficiency but also an improved customer experience and a meaningful human work.

ACKNOWLEDGMENT

The authors would like to acknowledge the valuable academic environment and research support provided by the Hutton School of Business, University of the Columbians, Kentucky, USA, and the Department of Computer Science and Engineering (AIML), Oriental Institute of Science & Technology, Bhopal, India, which facilitated the development of this study.

The authors also acknowledge the contributions of the broader research community whose scholarly works on human-AI collaboration, customer experience analytics, and service orchestration informed the literature synthesis presented in this review. Their foundational research provided critical insights that helped shape the conceptual framework proposed in this paper.

REFERENCES

- [1]. T. H. Davenport, A. Guha, D. Grewal, and T. Bressgott, "How artificial intelligence will change the future of marketing," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 24-42, 2020.
- [2]. K. Holstein, J. W. Vaughan, H. Daumé III, M. Dudík, and H. Wallach, "Improving fairness in machine learning systems: What do industry practitioners need?" in *Proc. 2019 CHI Conf. Human Factors in Computing Systems*, pp. 1-16, 2019.
- [3]. A. Rai, P. Constantinides, and S. Sarker, "Editorial case for integrative digital ensembles: Adopting complementary multi-entity perspectives," *MIS Quarterly*, vol. 43, no. 4, pp. iii-xviii, 2019.
- [4]. F. F. Reichheld, "The one number you need to grow," *Harvard Business Review*, vol. 81, no. 12, pp. 46-55, 2003.
- [5]. M. Dixon, K. Freeman, and N. Toman, "Stop trying to delight your customers," *Harvard Business Review*, vol. 88, no. 7/8, pp. 116-122, 2010.
- [6]. B. Larivière, D. Bowen, T. W. Andreassen, W. Kunz, N. J. Sirianni, C. Voss, N. V. Wunderlich, and A. De Keyser, "Service encounter 2.0: An investigation into the roles of technology, employees and customers," *Journal of Business Research*, vol. 79, pp. 238-246, 2017.
- [7]. R. Knote, A. Janson, M. Söllner, and J. M. Leimeister, "Value co-creation in smart services: A

- functional affordances perspective on smart personal assistants,” *Journal of the Association for Information Systems*, vol. 22, no. 2, pp. 418-458, 2021.
- [8]. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, and D. Moher, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, p. n71, 2021.
- [9]. V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.
- [10]. K. N. Lemon and P. C. Verhoef, “Understanding customer experience throughout the customer journey,” *Journal of Marketing*, vol. 80, no. 6, pp. 69-96, 2016.
- [11]. A. De Keyser, K. Verleye, K. N. Lemon, T. L. Keiningham, and P. Klaus, “Moving the customer experience field forward: Introducing the touchpoints, context, qualities (TCQ) nomenclature,” *Journal of Service Research*, vol. 23, no. 4, pp. 433-455, 2020.
- [12]. V. Kumar, B. Rajan, R. Venkatesan, and J. Lecinski, “Understanding the role of artificial intelligence in personalized engagement marketing,” *California Management Review*, vol. 63, no. 4, pp. 135-155, 2021.
- [13]. Gartner, “Market guide for conversational AI platforms,” Gartner Research, 2023.
- [14]. Forrester Research, “The state of AI in customer service,” Forrester Research Report, 2022.
- [15]. A. Baird and L. M. Maruping, “The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts,” *MIS Quarterly*, vol. 45, no. 1, pp. 315-342, 2021.
- [16]. S. L. Vargo and R. F. Lusch, “Institutions and axioms: An extension and update of service-dominant logic,” *Journal of the Academy of Marketing Science*, vol. 44, no. 1, pp. 5-23, 2016.
- [17]. E. Hutchins, *Cognition in the wild*. Cambridge, MA: MIT Press, 1995.
- [18]. J. Zhang and D. A. Norman, “Representations in distributed cognitive tasks,” *Cognitive Science*, vol. 18, no. 1, pp. 87-122, 1994.
- [19]. P. Wright, B. Fields, and M. Harrison, “Analyzing human-computer interaction as distributed cognition: The resources model,” *Human-Computer Interaction*, vol. 15, no. 1, pp. 1-41, 2000.
- [20]. R. W. Picard, *Affective computing*. Cambridge, MA: MIT Press, 1997.
- [21]. R. W. Picard, “Affective computing: Challenges,” *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55-64, 2003.
- [22]. Seeber, E. Bittner, R. O. Briggs, T. de Vreede, G. J. de Vreede, A. Elkins, and M. Söllner, “Machines as teammates: A research agenda on AI in team collaboration,” *Information & Management*, vol. 57, no. 2, p. 103174, 2020.
- [23]. N. J. McNeese, M. Demir, N. J. Cooke, and C. Myers, “Teaming with a synthetic teammate: Insights into human-autonomy teaming,” *Human Factors*, vol. 63, no. 5, pp. 690-706, 2021.
- [24]. E. Bittner, S. Oeste-Reiß, and J. M. Leimeister, “Where is the bot in our team? Toward a taxonomy of design option for hybrid collaboration in teams,” in *Proc. 52nd Hawaii Int. Conf. System Sciences*, 2019.
- [25]. D. Dellermann, A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel, “The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems,” in *Proc. 52nd Hawaii Int. Conf. System Sciences*, 2019.
- [26]. B. Green and Y. Chen, “The principles and limits of algorithm-in-the-loop decision making,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, CSCW, pp. 1-24, 2019.
- [27]. A. Maedche, C. Legner, A. Benlian, B. Berger, H. Gimpel, T. Hess, and M. Söllner, “AI-based digital assistants,” *Business & Information Systems Engineering*, vol. 61, no. 4, pp. 535-544, 2019.
- [28]. J. Sweller, “Cognitive load theory,” in *Psychology of Learning and Motivation*, vol. 55. New York: Academic Press, 2011, pp. 37-76.
- [29]. R. M. Ryan and E. L. Deci, “Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being,” *American Psychologist*, vol. 55, no. 1, pp. 68-78, 2000.
- [30]. C. R. Berger and R. J. Calabrese, “Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication,” *Human Communication Research*, vol. 1, no. 2, pp. 99-112, 1975.
- [31]. R. T. Rust and M. H. Huang, “The service revolution and the transformation of marketing science,” *Marketing Science*, vol. 33, no. 2, pp. 206-221, 2014.
- [32]. B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, “The ethics of algorithms: Mapping the debate,” *Big Data & Society*, vol. 3, no. 2, pp. 1-21, 2016.
- [33]. C. Argyris and D. A. Schön, *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley, 1978.