International Journal of
Philosophy, Ethics and Humanities
https://www.ioro.org/ijpeh

Human–AI Teaming in Modern Organizations: A Framework for Collaborative Intelligence

Manikanta R

ORCID: 0009-0005-2576-8731

MBA (Business Analytics & Human Resource Management) Nagarjuna Degree College, Bangalore

Abstract

As artificial intelligence (AI) systems take on increasingly interdependent roles within organizational work-flows, a growing body of scholarship argues that the appropriate metaphor for human-AI interaction is shifting from tool use toward teaming. Yet research on this shift remains split across disconnected literatures: human factors and ergonomics research on human-autonomy teaming, organizational behavior and strategy research on automation and augmentation, and information systems research on algorithmic decision-making and trust. This conceptual paper develops an integrative, organization-level framework for human-AI teaming, synthesizing peer-reviewed literature published between 2020 and 2026 across these streams. The paper argues that collaborative intelligence, the joint performance capability that emerges when humans and AI systems combine complementary strengths under conditions of genuine interdependence, is not an automatic consequence of deploying capable AI systems but an organizational achievement requiring deliberate design across task allocation, trust calibration, and role structure. Drawing on complementarity theory and the automation-augmentation paradox, the paper proposes the Collaborative Intelligence Framework (CIF), which specifies four conditions, task interdependence, calibrated trust, role clarity, and organizational enablement, under which human-AI teams are likely to outperform either humans or AI systems operating alone. The discussion examines how the framework resolves tensions between the human factors and organizational literatures and identifies the managerial conditions under which human-AI teaming succeeds or degrades into either under-reliance or over-reliance on algorithmic recommendations. The paper concludes with practical implications for designing human-AI teams in organizational settings, acknowledges the limitations of a conceptual approach, and proposes a future research agenda centered on empirical validation of the framework across occupational contexts.

Keywords: *human-AI teaming, collaborative intelligence, future of work, human-AI collaboration, complementarity, organizational behavior*

1. Introduction

Artificial intelligence (AI) systems are increasingly embedded within organizational workflows not as passive tools invoked by human operators but as active participants that generate recommendations, flag anomalies, and, in a growing number of applications, take autonomous action within defined boundaries. This shift has prompted a corresponding shift in the scholarly vocabulary used to describe human-AI interaction, from tool use toward teaming, reflecting a growing recognition that AI systems increasingly meet the functional criteria of a teammate: they contribute to a shared task, their behavior affects and is affected by their human counterparts, and the performance of the joint human-AI unit is not reducible to the performance of either party alone (O'Neill et al., 2022).

International Journal of Philosophy, Ethics and Humanities (IJPEH)
Volume 1 | Issue 1 | 2026 | Article ID: 4010-5782-9440 | https://doi.org/10.64823/ijpeh.2601005
IORO Publications · Texas, USA · www.ioro.org

This reframing carries substantive organizational implications. If AI is understood merely as a tool, organizational design questions reduce to questions of adoption and usability. If AI is understood as a teammate, organizational design questions expand to encompass task allocation, role clarity, trust calibration, and the coordination mechanisms that determine whether a human-AI unit outperforms either party working alone, a property termed complementarity in the decision-making literature (Bansal et al., 2021). The stakes of this distinction are considerable: Organization Science research on professional judgment finds that when AI recommendations are treated as authoritative inputs rather than as one voice within a genuinely interdependent process, professionals may either dismiss algorithmic input reflexively or defer to it uncritically, undermining the very complementarity that motivated AI adoption in the first place (Lebovitz et al., 2022).

Despite the growing convergence of human factors research and organizational scholarship on this teaming metaphor, the literatures addressing it remain substantially disconnected. Human factors and ergonomics research has produced a detailed empirical and theoretical account of human-autonomy teaming, specifying the cognitive, behavioral, and design conditions under which humans and autonomous systems function effectively as interdependent units (O'Neill et al., 2022, 2023). Organizational behavior and strategy scholarship, meanwhile, has examined the organizational and institutional conditions under which AI adoption enhances rather than undermines human judgment and organizational learning, largely independent of the human-

autonomy teaming literature's conceptual vocabulary (Raisch & Krakowski, 2021; Balasubramanian et al., 2022; Anthony et al., 2023). A third stream, situated in information systems and AI ethics research, has examined the trust and fairness dynamics that shape whether humans engage productively with algorithmic recommendations (Kordzadeh & Ghasemaghaei, 2022; Glikson & Woolley, 2020). Each stream offers partial insight into collaborative intelligence, but none offers an integrated, organization-level account of the conditions under which human-AI teaming succeeds.

This paper develops such an account. As a framework-development contribution, the paper synthesizes literature across these three streams into the Collaborative Intelligence Framework (CIF), specifying the organizational conditions under which human-AI teams achieve genuine complementarity rather than either under-reliance or uncritical over-reliance on algorithmic input. The remainder of the paper specifies the problem the framework addresses, identifies gaps in existing scholarship, reviews the relevant literature, grounds the framework theoretically, presents the framework itself, and discusses its managerial implications.

2. Problem Statement

The problem motivating this paper is that organizations increasingly deploy AI systems into interdependent, team-like roles without applying the design principles that decades of human teaming research indicate are necessary for genuine collaborative performance. Whereas the deployment of conventional software tools requires attention primarily to usability and training, the deployment of AI systems into roles involving judgment, recommendation, or autonomous action introduces team-design considerations, including task allocation, trust calibration, and role clarity, that are frequently absent from organizational AI implementation practice (O'Neill et al., 2023).

This gap manifests in two well-documented and opposing failure modes. The first is algorithmic aversion, in which humans systematically discount or ignore accurate AI recommendations, particularly following observed errors, resulting in organizations failing to capture the performance benefits their AI investment

was intended to produce. The second, and increasingly the more consequential failure mode in professional and managerial contexts, is automation complacency or overreliance, in which humans defer to AI recommen-dations without exercising the critical judgment needed to catch algorithmic errors, a dynamic documented directly in professional judgment contexts such as medical diagnosis, where opacity in AI reasoning led some professionals to disengage their own critical evaluation rather than engage more deeply with it (Lebovitz et al., 2022). Both failure modes represent a breakdown of complementarity: the joint human-AI system performs no better, and in the case of overreliance potentially worse, than either party operating alone (Bansal et al., 2021).

A further dimension of the problem concerns the organizational-level conditions that shape which failure mode, if either, an organization is likely to experience. The automation-augmentation paradox literature argues that automation and augmentation are not cleanly separable organizational choices but interdependent processes that unfold together across time and organizational levels, such that an organization's attempt to purely augment human judgment with AI input often triggers downstream automation pressures, and vice versa (Raisch & Krakowski, 2021). This suggests that human-AI teaming cannot be designed once, at the level of an individual tool deployment, but must be managed as an ongoing organizational process, a requirement that most current organizational AI implementation practice, focused on discrete tool rollouts, is not structured to meet. The problem addressed by this paper is therefore the absence of an integrated framework specifying the organizational conditions under which human-AI teaming achieves genuine, sustained complementarity rather than drifting toward algorithmic aversion or automation complacency.

3. Research Gap

Three gaps in the existing literature motivate this paper. First, the human-autonomy teaming literature, developed substantially within human factors, ergonomics, and cognitive psychology traditions, has produced detailed frameworks specifying the team-level conditions, including interdependence, shared mental models, and trust calibration, required for effective human-AI teaming (O'Neill et al., 2022, 2023), but this literature has been applied primarily to high-stakes, dynamic operational domains such as aviation and defense, with comparatively limited translation into the organizational and managerial contexts, such as professional

services, knowledge work, and administrative decision-making, that characterize the majority of contemporary organizational AI deployment. This paper addresses that gap by translating human-autonomy teaming principles into an organizational management framework.

Second, the organizational behavior and strategy literature on AI and automation has developed a rich account of the automation-augmentation paradox and its implications for organizational learning and professional judgment (Raisch & Krakowski, 2021; Balasubramanian et al., 2022; Lebovitz et al., 2022), but this literature engages only lightly with the specific team-design mechanisms, such as trust calibration and role clarity, iden-tified in the human-autonomy teaming literature as determinants of collaborative success. This paper addresses that gap by incorporating these team-design mechanisms directly into an organization-level framework.

Third, existing conceptual work at the intersection of these literatures, such as system-view accounts of AI's role in the future of work (Anthony et al., 2023) and multilevel reviews of AI in organizations (Bankins et al., 2024), has usefully mapped the breadth of relevant organizational behavior research but has stopped short of specifying an actionable, testable framework identifying the specific conditions under which human-AI teams

achieve complementarity. This paper addresses that gap by proposing the Collaborative Intelligence Framework (CIF), a structured model specifying four conditions for effective human-AI teaming that can guide both future empirical research and organizational design practice.

4. Objectives of the Study

This paper pursues four objectives. The first is to synthesize literature from human-autonomy teaming, organi-zational behavior, and information systems research, published between 2020 and 2026, into an integrated account of the conditions under which human-AI teams achieve genuine collaborative performance. The second is to characterize the two principal failure modes, algorithmic aversion and automation complacency, through which human-AI teaming breaks down, and to identify their organizational antecedents. The third is to develop the Collaborative Intelligence Framework (CIF), specifying the conditions required for effective human-AI teaming at the task, dyadic, and organizational levels. The fourth is to translate this framework into managerial guidance for designing and sustaining human-AI teams in organizational settings, while identifying the framework's limitations and the empirical research needed to validate it.

5. Literature Review

This section reviews the literature underpinning the proposed framework across four clusters: the conceptual foundations of human-autonomy teaming; the automation-augmentation paradox in organizational and management scholarship; trust and complementarity in human-AI decision-making; and organizational-level accounts of AI's role in the future of work.

5.1. Conceptual Foundations of Human-Autonomy Teaming

The human-autonomy teaming literature, consolidated through a series of influential reviews, establishes the criteria that distinguish genuine human-AI teaming from conventional human-computer interaction or tool use. O'Neill et al. (2022), synthesizing the empirical human-autonomy teaming literature, identify interdependence, the intelligent agent's status as a quasi-independent entity with meaningful decision-making capacity, and the presence of one or more humans and one or more intelligent agents jointly pursuing a shared task as the defining features of human-AI teaming. This framing deliberately excludes simple automation, in which an AI system executes a predetermined function without meaningful interdependence with a human counterpart, from the teaming category, a distinction with direct organizational design implications, since the governance and trust mechanisms appropriate to simple automation differ substantially from those required for genuine teaming.

Building on this foundational review, O'Neill et al. (2023) argue that human-autonomy teaming research requires a guiding, team-based framework rather than continued reliance on frameworks developed for human-only teams or for simple human-automation interaction, on the grounds that AI teammates introduce distinctive coordination challenges, including asymmetric transparency between human and AI reasoning processes and

rapidly evolving AI capability, that classical team theory does not fully anticipate. This literature's emphasis on interdependence as a defining, rather than incidental, feature of human-AI teaming provides a key building block for the framework developed later in this paper, which treats the degree of genuine task interdependence as a necessary precondition for collaborative intelligence to emerge at all.

5.2. The Automation-Augmentation Paradox

Organizational and management scholarship has approached the question of human-AI collaboration from a largely independent angle, focused on the strategic and organizational-level consequences of automating versus augmenting human tasks with AI. Raisch and Krakowski (2021) argue that automation and augmentation, though often presented as alternative strategic choices, are in practice interdependent processes that unfold together across time and organizational levels, such that an organizational choice to augment human decision-making with AI input frequently generates downstream pressure toward further automation, and that this paradoxical tension cannot be resolved through a one-time strategic choice but must be actively managed on an ongoing basis. This finding has direct implications for the sustainability of human-AI teaming arrangements: an organization that successfully establishes a collaborative, augmentation-oriented human-AI arrangement should not assume that arrangement will remain stable without deliberate ongoing management.

Complementary research on organizational learning finds that substituting human decision-making with machine learning can, under certain conditions, erode the tacit organizational knowledge and judgment that would otherwise accumulate through human decision-making practice, even when the machine learning system's predictions are statistically accurate in the short term (Balasubramanian et al., 2022). This literature suggests that the apparent efficiency gains of automating decisions that could instead be handled through human-AI collaboration may carry longer-term organizational learning costs that are not visible in short-term performance metrics, a consideration directly relevant to the framework's treatment of role clarity, discussed below, since decisions about which tasks to automate versus which to structure as genuine human-AI collaboration have consequences extending beyond immediate task performance.

5.3. Trust, Opacity, and Complementarity in Human-AI Decision-Making

A substantial literature has examined the trust dynamics that determine whether humans engage productively with AI-generated recommendations. Glikson and Woolley (2020), reviewing the empirical literature on human trust in AI, distinguish cognitive trust, based on perceived reliability and competence, from emotional trust, based on the perceived benevolence and predictability of the AI system, arguing that both dimensions shape whether humans appropriately calibrate their reliance on AI input. Insufficient trust produces algorithmic aversion, in which accurate AI recommendations are discounted; excessive or miscalibrated trust produces automation complacency, in which flawed recommendations are accepted without adequate scrutiny.

Direct evidence of this latter failure mode emerges from research on professional judgment under conditions of AI-generated opacity. Lebovitz et al. (2022), examining how medical professionals engage with AI diagnostic tools, find that when AI systems produce recommendations without adequately transparent reasoning, profes-sionals do not uniformly respond with heightened scrutiny; some instead disengage from independent critical judgment, effectively ceding decision authority to a system whose reasoning they cannot fully evaluate. This finding directly parallels the concerns raised in the algorithmic bias literature regarding insufficient human oversight of AI-based decisions in other domains (Kordzadeh & Ghasemaghaei, 2022), and it underscores that trust calibration is not merely a psychological nicety but a functional precondition for the complementarity that justifies human-AI teaming in the first place.

Complementarity itself has been examined directly in experimental decision-making research. Bansal et al. (2021) demonstrate that optimizing an AI system purely for standalone accuracy does not necessarily produce the best teammate for a human-AI decision-making unit, since a highly accurate but unpredictable AI system can undermine the human's ability to form an accurate mental model of when to trust and when to

override its recommendations, reducing joint team performance even as standalone AI accuracy increases. This finding has significant implications for how organizations should evaluate AI systems intended for team-based deployment: accuracy benchmarks assessed in isolation are an incomplete, and potentially misleading, basis for evaluating

a system's suitability as a human teammate.

5.4. Organizational Perspectives on AI's Role in the Future of Work

A fourth literature examines human-AI collaboration from a broader organizational and future-of-work perspective. Anthony et al. (2023), taking a systems view of AI's role in organizations, distinguish several distinct pathways through which humans come to relate to AI, ranging from AI as a passive tool through AI functioning as a more genuine teammate, arguing that the pathway an organization follows shapes the broader system of work practices, professional identity, and organizational structure that develops around the AI system, not merely the immediate task outcomes of any single interaction. Bankins et al. (2024), synthesizing organizational behavior research on AI across multiple levels of analysis, similarly find that workers' perceptions of AI capability, and their consequent willingness to collaborate with or resist a given AI system, are shaped by organizational and institutional factors extending well beyond the technical performance of the AI system itself, including perceptions of whether the AI threatens valued human skills or organizational status.

These organizational-level findings reinforce a theme recurring throughout this review: technical AI capability, whether measured through standalone accuracy or explainability, is necessary but insufficient for effective human-AI teaming. The organizational and psychological context within which a human-AI team operates, including trust calibration, perceived role security, and the broader system of work practice surrounding the AI system, substantially shapes whether collaborative intelligence emerges in practice.

6. Theoretical Foundation

This paper grounds the proposed framework in two complementary theoretical perspectives: complementarity theory, as developed within the human-AI decision-making literature, and paradox theory, as applied to the automation-augmentation tension in organizational scholarship.

Complementarity theory holds that a human-AI team achieves genuine collaborative value only when the joint performance of the team exceeds the performance of either the human or the AI system operating independently, a condition that depends not merely on the individual capabilities of each party but on the structure of their interaction, including how tasks are allocated between them and how effectively the human party can calibrate reliance on AI input (Bansal et al., 2021). This theoretical lens directly informs the framework's emphasis on task interdependence and trust calibration as necessary, rather than incidental, conditions for collaborative intelligence: a human-AI arrangement lacking genuine interdependence reduces to simple tool use, in which complementarity is not a meaningful construct, while a human-AI arrangement lacking calibrated trust produces either aversion or complacency, both of which undermine joint performance regardless of each party's independent capability.

Paradox theory, as applied to the automation-augmentation tension by Raisch and Krakowski (2021), provides a complementary account of why human-AI teaming arrangements are inherently unstable absent deliberate organizational management. Paradox theory holds that certain organizational tensions, including the tension between automating and augmenting a given task, cannot be permanently resolved through a single strategic choice but must instead be actively and continuously managed, since engagement with one

side of the tension tends to generate pressure toward the other over time. Applied to human-AI teaming, this perspective suggests that an organization's initial success in establishing a genuinely collaborative, augmentation-oriented arrangement between humans and AI does not guarantee the persistence of that arrangement; organizational, competitive, or efficiency pressures may progressively erode the human role within the arrangement unless the organization actively sustains the conditions that support genuine teaming.

Together, these two theoretical perspectives support the framework's central claim: that collaborative intelligence is not a fixed property that emerges automatically from deploying a sufficiently capable AI system, but an achieved and actively sustained organizational state, requiring deliberate attention to task design, trust calibration, and ongoing management of the automation-augmentation tension. This framing distinguishes the proposed framework from more technology-centered accounts of human-AI collaboration, which tend to treat

collaborative outcomes as primarily a function of AI system design rather than organizational design.

7. Conceptual Framework

Building on the literature review and theoretical foundation, this paper proposes the Collaborative Intelligence Framework (CIF), which specifies four conditions under which human-AI teams are likely to achieve genuine complementarity: task interdependence, calibrated trust, role clarity, and organizational enablement. The framework further specifies the two principal failure modes, algorithmic aversion and automation complacency, that result when these conditions are inadequately met, and positions collaborative intelligence as the joint outcome that emerges only when all four conditions are simultaneously satisfied.

The first condition, task interdependence, requires that the task in question genuinely require joint contribution from both human and AI parties, rather than being fully decomposable into an AI-only subtask and a human-only subtask executed in sequence without meaningful interaction. Drawing on the definitional criteria established in the human-autonomy teaming literature (O'Neill et al., 2022), the framework treats interdependence as a threshold condition: tasks lacking genuine interdependence should be organizationally classified as automation rather than teaming, since applying team-design principles, such as trust calibration mechanisms, to tasks that do not require them introduces unnecessary organizational complexity without a corresponding complementarity benefit.

The second condition, calibrated trust, requires that the human party's reliance on AI input be proportionate to the AI system's actual, task-specific reliability, avoiding both algorithmic aversion, in which accurate recom-mendations are systematically discounted, and automation complacency, in which flawed recommendations are accepted without scrutiny (Glikson & Woolley, 2020; Lebovitz et al., 2022). The framework specifies that calibrated trust depends substantially on the transparency of AI reasoning and the human party's opportunity to develop an accurate mental model of the AI system's performance boundaries, consistent with evidence that AI systems optimized purely for standalone accuracy, without regard to predictability, can undermine rather than support this calibration process (Bansal et al., 2021).

The third condition, role clarity, requires that the division of decision authority and responsibility between the human and AI parties be explicit and organizationally legitimated, rather than left to emerge informally through practice. Drawing on evidence that ambiguous AI opacity can prompt professionals to disengage their own judgment rather than exercise it more critically (Lebovitz et al., 2022), the framework specifies that

role clarity should address not only what each party contributes to the task but explicitly when and how the human party is expected to exercise override authority, ensuring that human oversight is a defined organizational expectation rather than an implicit and easily eroded norm.

The fourth condition, organizational enablement, requires that the broader organizational context, including governance structure, professional identity considerations, and the ongoing management of the automation-augmentation tension, actively support the sustained operation of the human-AI team rather than passively permitting it. Drawing on the paradox theory foundation discussed above, this condition specifies that organizations must treat human-AI teaming arrangements as requiring continuous management attention, including periodic reassessment of task allocation and monitoring for organizational or competitive pressures that might erode the human role over time (Raisch & Krakowski, 2021), and must attend to workers' broader perceptions of AI's relationship to valued human skills and professional identity (Bankins et al., 2024).

The framework depicts these four conditions as jointly necessary rather than independently sufficient: strong performance on any three conditions does not compensate for failure on the fourth. A team with strong task interdependence, calibrated trust, and role clarity operating within an organization that fails to actively manage the automation-augmentation tension may see its collaborative arrangement gradually erode into pure automation over time, even without any explicit organizational decision to that effect. Conversely, strong organizational enablement without genuine task interdependence produces a well-supported but functionally unnecessary team structure applied to a task that did not require it. Collaborative intelligence, in this framework, is therefore modeled as the emergent property of an organizational system in which all four conditions are simultaneously and continuously satisfied, rather than a fixed outcome of AI system capability alone.

8. Discussion

The Collaborative Intelligence Framework developed in this paper carries several implications that extend the reviewed literature. First, by treating task interdependence as a threshold condition rather than a matter of degree, the framework offers organizations a practical diagnostic for distinguishing tasks appropriate for team-design investment from tasks better classified, and governed, as straightforward automation. This distinction addresses a common practical confusion in organizational AI deployment, in which team-oriented language, such as describing an AI tool as a colleague's assistant, is applied to systems that in practice function as simple automation, creating a mismatch between the governance mechanisms an organization applies and the mechanisms the underlying human-AI arrangement actually requires.

Second, the framework's treatment of calibrated trust as dependent on the interaction between transparency and predictability, rather than on accuracy alone, has direct implications for how organizations should evaluate AI systems intended for collaborative deployment. The finding that a highly accurate but unpredictable AI system can degrade joint team performance relative to a more predictable, if marginally less accurate, system (Bansal et al., 2021) suggests that procurement and evaluation processes for collaborative AI systems should incorporate predictability and explainability alongside standalone accuracy benchmarks, a practice not yet standard in most organizational AI evaluation processes.

Third, the framework's role clarity condition addresses a specific mechanism through which automation complacency emerges: not simply insufficient training or awareness on the part of human team members, but the absence of an explicit organizational specification of when human override is expected. The finding that professionals faced with AI opacity may disengage their independent judgment rather than heighten it (Lebovitz et al., 2022) suggests that organizations cannot rely on individual professional diligence alone to

sustain appropriate human oversight; oversight expectations must be organizationally specified and reinforced as an explicit role requirement, not left as an implicit professional norm vulnerable to erosion under time pressure or repeated positive experience with AI recommendations.

Fourth, the framework's organizational enablement condition, grounded in paradox theory, implies that human-AI teaming arrangements require ongoing organizational maintenance in a manner that more conventional software deployments do not. This has structural implications for how organizations govern human-AI teaming initiatives: rather than treating a collaborative AI deployment as a completed project once initial trust calibration and role clarity have been established, the framework suggests that organizations should assign ongoing governance responsibility for monitoring whether the automation-augmentation balance within a given human-AI arrangement is drifting toward automation, potentially eroding the complementarity the arrangement was designed to achieve.

Finally, the framework helps reconcile an apparent tension between the human-autonomy teaming literature's optimism about the potential of genuine human-AI teaming and the organizational behavior literature's more cautious findings regarding automation's tendency to erode human judgment and organizational learning over time (Balasubramanian et al., 2022). The framework suggests that this tension reflects two different organizational trajectories rather than a genuine theoretical contradiction: where all four conditions of the framework are actively maintained, genuine collaborative intelligence is achievable and sustainable; where organizational enablement in particular is neglected, the more pessimistic trajectory documented in the organizational learning literature becomes the more likely outcome, even in arrangements that initially exhibited strong task interdependence, trust calibration, and role clarity.

9. Managerial and Practical Implications

The Collaborative Intelligence Framework developed in this paper carries several practical implications for organizations designing human-AI teams. First, before investing in team-design mechanisms such as trust calibration training or explicit override protocols, organizations should assess whether a given AI deployment genuinely satisfies the task interdependence threshold, reserving team-design investment for tasks that require joint human-AI contribution and applying simpler automation governance to tasks that do not.

Second, organizations evaluating AI systems intended for collaborative deployment should incorporate

predictability and explainability into procurement and evaluation criteria alongside standalone accuracy metrics, given evidence that highly accurate but unpredictable systems can degrade rather than enhance joint team performance (Bansal et al., 2021). Vendor demonstrations and internal validation processes focused solely on accuracy benchmarks provide an incomplete basis for assessing a system's suitability as a human teammate.

Third, organizations should explicitly define, document, and periodically reinforce override protocols specifying when human team members are expected to exercise independent judgment rather than defer to AI recommendations, rather than relying on general training or professional norms to sustain appropriate oversight. This is particularly important in contexts involving AI opacity, where evidence suggests professionals may otherwise disengage critical judgment under the pressure of unclear reasoning (Lebovitz et al., 2022).

Fourth, organizations should assign explicit, ongoing governance responsibility for monitoring the automation-augmentation balance within established human-AI teaming arrangements, treating this as a continuous organizational management function rather than a one-time deployment decision. This

governance function should specifically monitor for gradual erosion of human role scope, decision authority, or engagement over time, consistent with the paradox theory expectation that such erosion is likely absent active management (Raisch & Krakowski, 2021).

Fifth, organizations should attend explicitly to workers' perceptions of how a given AI deployment relates to valued human skills and professional identity, since these perceptions shape willingness to engage collaboratively with AI systems independent of the AI system's technical performance (Bankins et al., 2024). Communication and change management around AI deployment should address these identity and status concerns directly rather than focusing solely on technical capability and expected efficiency gains.

Finally, HR and organizational design functions should treat human-AI team composition, specifically the allocation of decision authority and the design of interdependence between human and AI contributions, as a deliberate organizational design decision requiring the same rigor traditionally applied to human team design, rather than as a byproduct of technical system configuration determined primarily by IT or data science functions.

10. Limitations

This paper is subject to several limitations inherent in its conceptual, framework-development design. First, because the Collaborative Intelligence Framework is developed through literature synthesis rather than empirical data collection, its four proposed conditions have not been operationalized into measurable constructs or tested for their relative predictive weight in determining human-AI team performance across organizational contexts.

Second, much of the empirical foundation for the framework's trust calibration and role clarity conditions derives from research conducted in specific occupational contexts, including professional judgment domains such as medical diagnosis (Lebovitz et al., 2022) and experimental decision-making tasks (Bansal et al., 2021), which may not generalize uniformly to the full range of organizational contexts in which human-AI teaming is being pursued, including lower-stakes administrative and creative work.

Third, the framework's organizational enablement condition, grounded in paradox theory, offers a qualitative account of why human-AI teaming arrangements require ongoing management attention but does not specify precise indicators or thresholds by which organizations could detect early-stage drift toward automation complacency or erosion of human role scope before significant organizational consequences occur.

Fourth, the framework treats its four conditions as jointly necessary based on theoretical synthesis rather than empirical demonstration; the possibility that certain conditions substitute for or compensate for weaknesses in others, under specific organizational or task circumstances, remains an open empirical question that this conceptual paper cannot resolve.

11. Future Scope

Several avenues for future research follow from the limitations identified above. First, the four conditions of the Collaborative Intelligence Framework should be operationalized into validated measures and tested empirically across a range of human-AI team contexts, to assess their relative contribution to joint team performance and to test the framework's central proposition that all four conditions are jointly necessary for collaborative intelligence to emerge.

Second, comparative research across occupational and organizational contexts, extending beyond the profes-sional judgment and experimental decision-making contexts that dominate the current literature, would help clarify the boundary conditions of the framework, particularly its applicability to lower-stakes administrative, creative, and service-oriented human-AI teaming arrangements.

Third, longitudinal research is needed to examine the organizational enablement condition directly, tracking specific human-AI teaming arrangements over time to identify observable early indicators of drift toward automation complacency or erosion of human role scope, extending the largely theoretical treatment of this dynamic in the current paradox theory literature.

Fourth, future research should investigate whether the framework's four conditions interact multiplicatively, as the framework's jointly-necessary formulation implies, or whether certain conditions can partially substitute for weaknesses in others under specific circumstances, a question with direct implications for how organizations should prioritize limited implementation resources across the four conditions.

Finally, future work should examine how the framework applies to emerging forms of human-AI teaming involving multiple AI systems or multiple human team members simultaneously, extending the framework's current dyadic focus toward the more complex, multi-agent team structures increasingly characteristic of organizational AI deployment.

12. Conclusion

This paper has developed the Collaborative Intelligence Framework (CIF), an integrative, organization-level model synthesizing literature from human-autonomy teaming, organizational behavior, and information systems research published between 2020 and 2026. The review found that existing scholarship relevant to human-AI teaming remains distributed across disciplinary traditions, human factors and ergonomics research on team-level design conditions, organizational and strategy research on the automation-augmentation paradox, and information systems research on trust and complementarity, without an integrated account of the organizational conditions under which collaborative intelligence emerges.

Grounded in complementarity theory and paradox theory, the proposed framework specifies four jointly necessary conditions, task interdependence, calibrated trust, role clarity, and organizational enablement, and positions collaborative intelligence as an achieved and actively sustained organizational state rather than an automatic consequence of AI capability. The framework is intended to provide scholars with an integrative structure for future empirical research on human-AI teaming, and to provide organizational leaders with a diagnostic vocabulary for designing and governing human-AI teams more deliberately than current practice typically allows.

As AI systems continue to take on more interdependent organizational roles, the central argument of this paper is that the performance benefits of human-AI collaboration are not guaranteed by AI capability alone. Organizations that treat human-AI teaming as an ongoing design and governance responsibility, rather than a one-time technology deployment, are better positioned to achieve genuine collaborative intelligence and to avoid the algorithmic aversion and automation complacency failure modes that otherwise undermine the value of AI investment.

References

[1] Anthony, C., Bechky, B. A., & Fayard, A. L. (2023). "Collaborating" with AI: Taking a system view to explore the future of work. Organization Science, 34(5), 1672–1694. https://doi.org/10.1287/orsc.2022.1651

[2] Balasubramanian, N., Ye, Y., & Xu, M. (2022). Substituting human decision-making with machine learning: Implications for organizational learning. Academy of Management Review, 47(3), 448–465.

[3] Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. Journal of Organizational Behavior, 45(2), 159–182. https://doi.org/10.1002/job.2735

[4] Bansal, G., Nushi, B., Kamar, E., Horvitz, E., & Weld, D. S. (2021). Is the most accurate AI the best teammate? Optimizing AI for teamwork. Proceedings of the AAAI Conference on Artificial Intelligence, 35(13), 11405–11414.

[5] Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660.

[6] Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions.

[7] European Journal of Information Systems, 31(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

[8] Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. Organization Science, 33(1), 126–148.

[9] O'Neill, T. A., Flathmann, C., McNeese, N. J., & Salas, E. (2023). Human-autonomy teaming: Need for a guiding team-based framework? Computers in Human Behavior, 146, Article 107762. https://doi.org/10.1016/j.chb.2023.107762

[10] O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. Human Factors, 64(5), 904–938. https://doi.org/10.1177/0018720820960865

[11] Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox.

[12] Academy of Management Review, 46(1), 192–210. https://doi.org/10.5465/amr.2018.0072

[13] Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. Knowledge-Based Systems, 263, Article 110273. https://doi.org/10.1016/j.knosys.2023.110273

[14] Tanantong, T., & Wongras, P. (2024). A UTAUT-based framework for analyzing users' intention to adopt ar-tificial intelligence in human resource recruitment: A case study of Thailand. Systems, 12(1), Article 28. https://doi.org/10.3390/systems12010028