E-ISSN: 3108-4192

APSSHS

Academic Publications of Social Sciences and Humanities Studies
2026, Volume 6, Issue 1, Page No: 85-93
Available online at: https://apsshs.com/

Journal of Applied Organizational Systems and Behavior

Behavioral Autonomy, Cognitive Offloading, and Responsibility in Artificial Intelligence–Assisted Work

Siti Rahman[1]*, Ahmad Zaki[1], Nurul Huda[2], Amir Faisal[1], Lim Wei[2]

1. Department of Behavioral Autonomy and AI-Assisted Work, Faculty of Business, University of Malaya, Kuala Lumpur, Malaysia.
2. Department of Cognitive Offloading and Responsibility, Faculty of Management, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.

Abstract

Artificial intelligence is increasingly embedded in organizational tasks that once depended primarily on human judgment, memory, interpretation, and discretion. Yet describing work as AI assisted says little about who retains meaningful authority, which cognitive operations are delegated, or who remains answerable when consequential outcomes follow. This Original Propositional Model Article develops an evidence-informed account of those distinctions. It integrates work-design, agency, delegation, algorithmic-control, judgment, cognitive-offloading, and responsibility scholarship to propose an Assisted-Autonomy Model organized around four linked allocation processes: authority allocation, judgment allocation, cognitive allocation, and responsibility alignment. The model distinguishes assistance from judgment substitution, routine or execution offloading from evaluative-reasoning offloading, and formal accountability from practical evaluative control. It further introduces retained evaluative agency and autonomy–responsibility decoupling as proposed constructs for explaining why workers may remain formally responsible even when their capacity to interrogate, reject, or reformulate AI-generated judgments has narrowed. The framework also separates technological capability from legitimate organizational authority and nominal human presence from consequential oversight. The article argues that autonomy effects are conditional on task verifiability, relative human–AI capability, opacity, workload, organizational control, and meaningful opportunities for intervention. The model and its propositions are not presented as validated causal relationships or universal prescriptions. Their contribution is to specify falsifiable mechanisms, rival explanations, and empirical boundaries for research on AI-assisted work.

Keywords: Behavioral autonomy, Cognitive offloading, Theoretical foundations, Judgment substitution, Theoretical propositions, Safeguards and empirical tests

How to cite this article: Rahman S, Zaki A, Huda N, Faisal A, Wei L. Behavioral Autonomy, Cognitive Offloading, and Responsibility in Artificial Intelligence–Assisted Work. J Appl Organ Syst Behav. 2026;6(1):85-93. https://doi.org/10.51847/qgQp66Ji4d

Received: 20 May 2026; **Revised:** 03 August 2026; **Accepted:** 08 August 2026
Corresponding author: Siti Rahman
E-mail ✉ siti.rahman@um.edu.my

Introduction

Artificial intelligence is increasingly used to recommend, classify, summarize, predict, generate, and prioritize within organizational decisions. Human–AI work therefore cannot be understood through substitution alone. AI may complement human judgment by extending analytic capacity while leaving contextual interpretation and final decision authority with workers [1]. The central problem is that the label “assisted work” can conceal major differences in who frames a problem, evaluates alternatives, rejects system output, and remains answerable for action.

Autonomy offers an entry point, but it must be treated as a work-design property rather than simply as an attitude toward technology. Digital automation and algorithmic systems can reorganize discretion, feedback, task boundaries, and decision

© 2026 The Author(s).
Copyright CC BY-NC-SA 4.0

latitude [2]. Accordingly, this article defines behavioral autonomy as the practical scope to choose, revise, reject, delay, redirect, or escalate action within an AI-assisted task. This differs from nominal authority: a worker may remain officially responsible while having limited substantive influence over the operative judgment.

AI assistance may also be embedded in broader systems of algorithmic control that direct, evaluate, rank, or discipline work [3]. Yet automation and augmentation are not clean opposites; the same system may augment one component of work while automating or constraining another [4]. Recent field evidence likewise shows task-contingent consequences: generative AI can improve performance on some knowledge tasks yet impair it on tasks outside its capability frontier [5]. These findings concern performance rather than autonomy, but they demonstrate why uniform claims about assistance are inadequate.

This Original Propositional Model Article therefore develops the proposed Assisted-Autonomy Model. It connects authority allocation, judgment substitution, cognitive offloading, and responsibility alignment while distinguishing evidence-supported relationships from new theoretical synthesis. The model is intended to generate conditional propositions and empirical tests, not to claim causal validation or universal organizational applicability.

Theoretical foundations

The model begins from a relational conception of agency. Human and technological contributions can become conjoined in organizational activity, meaning that observed action may emerge from an arrangement of interdependent human and technical capacities rather than a single isolated actor [6]. This does not assign moral agency to technology. It instead clarifies why autonomy cannot be inferred merely from whether a person remains visibly present in the workflow.

Delegation provides a second foundation. Agentic information systems can receive delegated activities and return outputs that reshape subsequent human action [7]. Delegation is therefore distinct from simple system use. A worker who consults an AI output and independently evaluates it occupies a different position from one whose task has been structured so that the system determines the available decision path.

At the organizational level, AI can be incorporated into different decision structures that distribute decision roles between humans and systems [8]. Such structures concern authority as well as capability: technical competence does not itself confer legitimate organizational decision rights. Work-design research also indicates that algorithmic management can reshape job demands and resources, including autonomy-related features [9].

These foundations imply that behavioral autonomy depends on the joint configuration of agency, delegation, decision rights, and work design. The model adds a proposed integrative step: these foundations become consequential through how human judgment is allocated, which cognitive processes are offloaded, and whether responsibility remains aligned with practical control. Those connections are theoretically plausible but require separate empirical examination.

Judgment substitution

Judgment substitution is narrower than AI use, reliance, or delegation. This article proposes it as a condition in which AI output materially displaces independent human evaluation because workers no longer meaningfully formulate, interrogate, compare, reject, or revise the judgment before consequential action. Productive human–AI collaboration depends partly on knowing relative human and AI capabilities; experimental evidence shows that inadequate metaknowledge can undermine productive delegation [10]. Heavy reliance can therefore occur without formal removal of judgment, while formal authority can remain even when effective evaluation is weak.

Opacity complicates this distinction. Professionals do not always respond to opaque AI by simply accepting or rejecting it. Medical professionals have been observed to interrogate AI outputs and connect them with contextual knowledge before making critical judgments [11]. This suggests that evaluative agency can survive opacity when expertise, information, time, and procedural latitude support active scrutiny.

Substitution becomes clearer when organizational design narrows such engagement. Research on substitutive decision-making AI shows that employees can lose direct interaction with and influence over decisions that previously formed part of their professional role [12]. The contextual nature of that evidence prevents universal claims, but it illustrates how nominal responsibility can coexist with diminished participation in judgment.

The proposed model therefore distinguishes assistance, deference, and judgment substitution. Assistance informs human evaluation; deference gives AI output substantial voluntary weight while alternatives remain feasible; substitution materially narrows independent evaluation. These states should be identified through observable problem formulation, verification, rejection, override, and decision influence rather than through frequency of AI use alone.

Cognitive offloading

Cognitive offloading is the externalization of cognitive operations to resources outside internal processing. It is not inherently evidence of dependency or diminished competence. Experimental research on reminder use shows that people externalize memory demands partly in response to metacognitive judgments and effort considerations [13]. In AI-assisted work, the

cautious implication is that delegating search, recall, formatting, or routine synthesis may conserve resources, but the organizational consequence depends on what cognition is transferred and what remains actively evaluated.

Metamemory provides a related mechanism. Confidence in internal memory can influence whether information is managed externally [14]. Offloading decisions therefore reflect beliefs about one's own capacity as well as objective task demands. Workplace AI, however, adds organizational pressures, deadlines, defaults, authority relations, and performance expectations that differ from laboratory memory tasks.

Externalization can also change patterns of internal memory use [15]. That finding does not establish long-term deskilling or expertise erosion. The proposed model therefore separates routine or execution offloading from evaluative-reasoning offloading. The former includes memory, search, retrieval, formatting, or procedural operations; the latter transfers interpretation, comparison, inference, or assessment that would otherwise contribute directly to judgment.

This distinction supports the proposed construct retained evaluative agency: the practical capacity to interrogate, verify, reinterpret, reject, or replace AI-generated judgment before consequential action. Routine offloading may free resources for such activity, whereas evaluative-reasoning offloading may weaken it when independent verification also declines. That relationship remains a proposition requiring direct workplace testing.

Table 1 organizes the proposed conceptual architecture and the boundaries needed to distinguish behavioral autonomy, cognitive offloading, judgment allocation, and responsibility in AI-assisted work.

Table 1. Conceptual Architecture of Behavioral Autonomy, Cognitive Offloading, and Responsibility in Artificial Intelligence–Assisted Work

Construct or component	Working definition	Central theoretical problem	Distinction from adjacent concepts	Role in the proposed model	Key uncertainty	Conceptual boundary
AI-assisted work	Work in which AI contributes information, recommendations, generated content, classifications, predictions, or executable actions while human participation remains part of the task arrangement	The label "assistance" conceals substantial variation in who controls interpretation and action	Differs from both fully human work and complete task automation	Defines the technological work context within which autonomy is allocated	The degree of human influence may vary across tasks and decision stages	AI participation alone does not establish augmentation, substitution, or autonomy loss
Behavioral autonomy	Practical capacity to choose, revise, reject, delay, redirect, or escalate action during task execution	Formal discretion may exist without consequential influence over the operative decision	Differs from perceived freedom, job satisfaction, technology acceptance, and nominal authority	Primary human-agency outcome of the model	Formal and enacted autonomy may diverge	Autonomy must be assessed through actual decision possibilities rather than job titles or policy statements
Authority allocation	Distribution of permission to recommend, decide, initiate, approve, or execute actions between human and AI actors	Technological capability can be mistaken for legitimate decision authority	Differs from technical capability and task execution	First allocation dimension in the proposed Assisted-Autonomy Model	Formal authority may differ from authority enacted in practice	Delegating activity does not automatically transfer accountability
Judgment allocation	Distribution of interpretive and evaluative work between the human and the AI system	Human presence may conceal displacement of substantive evaluation	Differs from simple exposure to AI advice	Determines whether evaluation remains human-led, shared, or AI-dominant	Degree of meaningful independent evaluation may be difficult to observe	Reliance does not necessarily constitute judgment substitution
Judgment substitution	**Proposed:** a condition in which AI output materially displaces independent human evaluation before consequential action	Workers may retain nominal final authority while their opportunity to formulate or contest judgment narrows	Differs from voluntary deference and routine decision support	Proposed high-substitution state within judgment allocation	Requires behavioral rather than attitudinal identification	Frequent AI use or agreement with AI output is insufficient evidence of substitution
Cognitive offloading	Externalization of cognitive operations to technological resources rather than completing them entirely through internal processing	Offloading is often interpreted too quickly as dependency or reduced competence	Differs from incapacity, automation, and long-term deskilling	Specifies how cognitive activity is redistributed during assisted work	Consequences depend on what type of cognition is externalized	Offloading does not itself demonstrate deterioration of expertise

Routine/execution offloading	**Proposed:** externalization of memory, search, formatting, transcription, retrieval, or procedural operations	Cognitive effort may be reduced without reducing meaningful evaluation	Differs from offloading interpretation or judgment	May release resources for higher-order evaluation	Whether released resources are actually redirected to evaluation remains uncertain	Efficiency gains do not prove increased autonomy
Evaluative-reasoning offloading	**Proposed:** transfer of interpretation, comparison, inference, or assessment that would otherwise form part of human judgment	Cognitive support may become substantive judgment displacement	Differs from routine assistance	Proposed pathway through which cognitive allocation may affect evaluative agency	Effects likely depend on verification opportunities and task characteristics	Externalized reasoning does not necessarily eliminate meaningful human review
Retained evaluative agency	**Proposed:** practical capacity to interrogate, verify, reinterpret, reject, or replace AI-generated judgment	Nominal autonomy may persist after substantive evaluative influence has weakened	Differs from final-signoff authority and subjective perceptions of control	Proposed mediating state linking judgment and cognitive allocation to enacted autonomy	No established operational measure currently defines the construct	Must be demonstrated through observable opportunities and behavior
Responsibility alignment	**Proposed:** correspondence among authority, information access, evaluative involvement, intervention capability, traceability, and answerability	Responsibility can remain attached to a human whose substantive control has diminished	Differs from responsibility attribution considered in isolation	Final alignment dimension in the proposed model	Organizational and individual understandings of responsibility may diverge	Alignment cannot be inferred merely from formal accountability assignment
Autonomy–responsibility decoupling	**Proposed:** condition in which formal human answerability remains while meaningful control, information, or evaluative influence has shifted elsewhere	Accountability may persist after substantive agency has been redistributed	Differs from the general claim that AI necessarily creates a responsibility gap	Proposed risk state of the Assisted-Autonomy Model	Its prevalence, causes, and consequences require direct empirical testing	The construct does not imply that every AI-assisted decision generates a responsibility gap

Note. Constructs explicitly marked Proposed represent original conceptual developments or distinctions advanced in this article. The table defines analytical categories rather than validated measures, causal relationships, organizational prescriptions, or demonstrated outcomes.

Responsibility gaps

Responsibility becomes theoretically difficult when the person held answerable for an AI-assisted outcome is not the actor or configuration with meaningful control over the judgment that produced it. Responsibility-gap scholarship distinguishes multiple ways in which AI may complicate the relationship between control and answerability [16]. For organizational analysis, the implication is not that AI automatically dissolves responsibility, but that responsibility allocation should be examined alongside authority, knowledge, and intervention capability.

The novelty and severity of responsibility gaps are contested. Critical analysis argues that some claims about unprecedented AI responsibility gaps overstate what is conceptually new because established responsibility practices can sometimes accommodate technologically mediated action [17]. Apparent gaps may instead arise from poor role design, fragmented decision rights, or unclear accountability.

Responsibility attribution is also relational. Work connecting explainability and responsibility emphasizes relations among actors, technologies, information, and opportunities to understand or intervene [18]. The model therefore introduces autonomy–responsibility decoupling as a proposed organizational condition in which a worker remains formally answerable while meaningful evaluative control, information access, or practical intervention capability has shifted elsewhere.

Responsibility alignment is correspondingly proposed as the fit among decision authority, evaluative involvement, information access, intervention capability, traceability, and answerability. Neither construct is validated. Their purpose is to make misalignment observable and testable while distinguishing personal perceptions of responsibility from formal policy and enacted decision practice.

Proposed assisted-autonomy model

Organizational research shows why the preceding domains must be integrated at the level of practice. Knowledge workers can develop routines for validating analytical technologies, yet repeated practices may also contribute to black boxing algorithmic outputs [19]. Autonomy is therefore not determined only when a system is introduced; it may be enacted through recurring practices of checking, contesting, accepting, and routinizing output.

Figure 1 presents the original conceptual synthesis for assisted-autonomy model, showing how the article's principal constructs, mechanisms, uncertainties, and decision boundaries are connected.

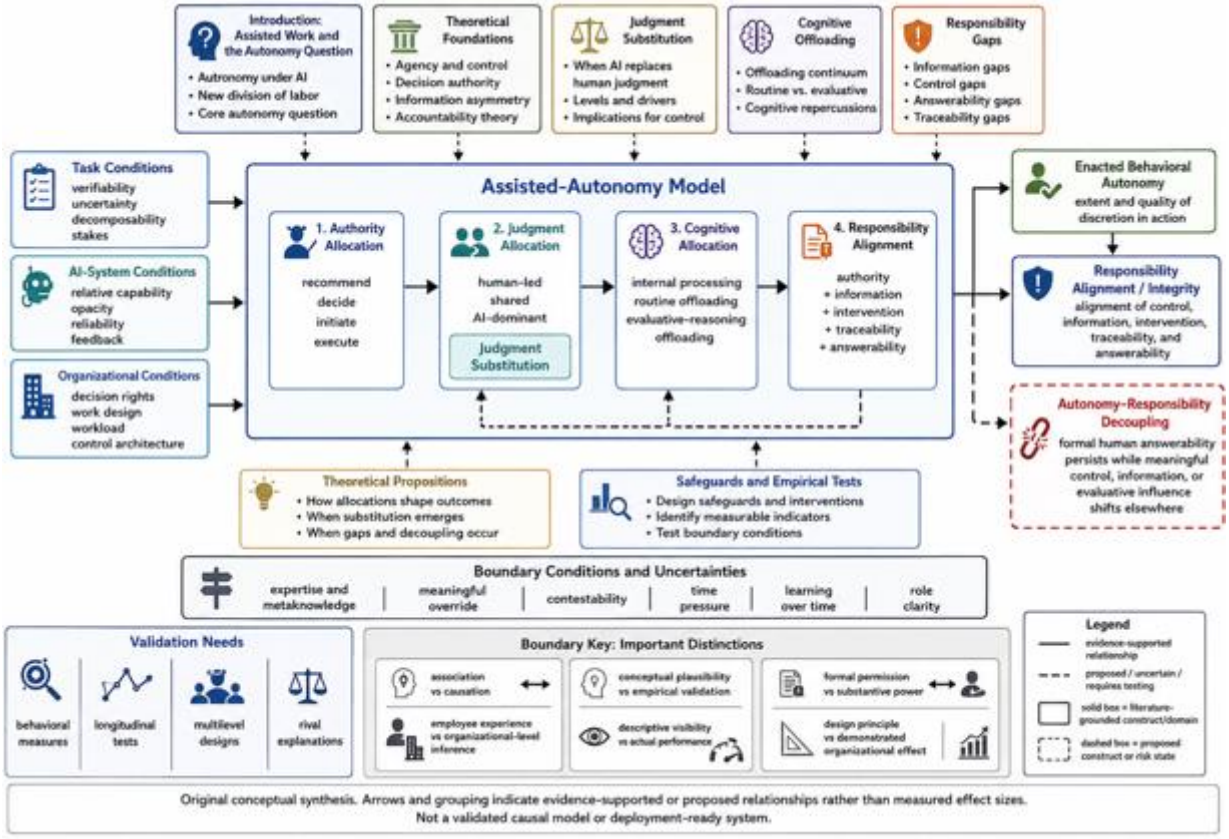

Figure 1. Assisted-Autonomy Model

The model begins with authority allocation: what an AI system may recommend, decide, initiate, or execute. It then considers judgment allocation, asking whether workers independently evaluate outputs or whether system judgments materially substitute for their own. Cognitive allocation identifies what is internally processed and what is offloaded. Responsibility alignment then evaluates whether answerability remains proportionate to authority, information, evaluative involvement, and intervention capability. These are proposed analytic dimensions, not a validated temporal sequence.

Algorithm implementation can alter established regimes of knowing and whose expertise becomes consequential [20]. Algorithmic coordination may also combine facilitation with extensive control, as platform-work research illustrates [21]. Trust adds another layer: empirical research shows that trust in AI varies with system, task, and human characteristics and should not be treated as equivalent to calibrated behavioral reliance [22].

The model consequently proposes task verifiability, relative human–AI capability, expertise and metaknowledge, opacity, workload, feedback, and meaningful override as boundary conditions. Retained evaluative agency connects judgment and cognitive allocation to enacted autonomy. When workers can interrogate and revise AI output, offloading need not imply reduced autonomy. When evaluative reasoning is displaced while formal answerability remains, autonomy–responsibility decoupling becomes more plausible.

The allocations may also be recursive. Repeated acceptance, detected error, or successful contestation may change later confidence, delegation, and checking. Because the selected evidence does not establish a common longitudinal trajectory, these feedback processes remain testing priorities rather than claims of inevitable dependence or skill loss.

Table 2 organizes the proposed mechanisms, boundary conditions, rival explanations, and validation requirements of the Assisted-Autonomy Model.

Table 2. Proposed Mechanisms, Boundary Conditions, and Testable Implications of the Assisted-Autonomy Model

Proposed mechanism or relationship	Triggering condition	Proposed process	Key boundary conditions	Expected theoretical implication	Plausible rival explanation	Empirical test required
AI assistance → differentiated authority allocation	AI becomes capable of recommending, initiating, or executing consequential task components	Decision rights are redistributed between worker and system	Task stakes; organizational policy; override rights; system capability	Behavioral autonomy should depend on the location of substantive decision authority rather than AI use alone	Apparent authority change reflects pre-existing managerial hierarchy	Map formal and enacted decision rights across comparable AI-assisted and non-assisted tasks

Delegation → judgment reallocation	Workers or organizations transfer task components to AI	Human evaluation is retained, shared, deferred, or displaced depending on the delegation arrangement	Relative human–AI capability; expertise; task verifiability; time pressure	Delegation should affect autonomy primarily when it changes substantive evaluative influence	Delegation simply reflects rational specialization with no autonomy consequence	Manipulate delegation scope while measuring independent evaluation, rejection, and correction behavior
AI judgment dominance → judgment substitution	AI output becomes the default or practically unavoidable basis for action	Independent problem formulation, comparison, or evaluation becomes narrower	Override capability; alternative information; system opacity; organizational expectations	Greater substitution should be associated with lower retained evaluative agency	Workers accept AI because its judgment is objectively superior	Compare reliance under equal AI accuracy while experimentally varying contestability and alternative-generation opportunities
Routine offloading → cognitive resource reallocation	AI performs memory, search, retrieval, formatting, or procedural operations	Internal cognitive demands decrease while evaluative cognition may remain active	Workload; expertise; task complexity; use of released cognitive capacity	Routine offloading may preserve or enhance evaluative agency when released resources support higher-order judgment	Reduced effort merely reflects convenience without changing evaluation	Measure both cognitive load and subsequent verification, alternative generation, and decision revision
Evaluative-reasoning offloading → reduced retained evaluative agency	AI performs interpretation, comparison, inference, or recommendation formation	Workers engage less deeply in constructing or testing the underlying judgment	Task verifiability; explanation quality; expertise; error detectability	Extensive evaluative offloading may weaken substantive autonomy when independent checking also declines	Reduced checking reflects high justified confidence in superior AI performance	Hold AI performance constant while varying requirements for independent reasoning and verification
Opacity → interrogation or black boxing	AI output is difficult to inspect or reconstruct	Workers develop checking routines or progressively normalize unexamined system output	Expertise; feedback; error visibility; organizational learning climate	Opacity should not have a uniform autonomy effect; workplace practices determine whether evaluative agency is preserved	Observed checking practices originate from occupational norms unrelated to AI	Longitudinally observe how verification practices evolve after AI adoption
Algorithmic coordination → constrained enacted autonomy	AI becomes embedded in task assignment, evaluation, scheduling, or workflow control	Technology structures available choices and permissible responses	Employment arrangement; managerial oversight; worker contestability	AI can simultaneously facilitate coordination and narrow behavioral discretion	Reduced autonomy results from conventional managerial control rather than algorithmic mediation	Compare matched work settings differing in algorithmic versus human control architecture
Knowledge-authority redistribution → changing evaluative influence	AI-generated knowledge becomes institutionally privileged	What counts as credible expertise or acceptable evidence shifts	Professional expertise; institutional norms; system performance; organizational sponsorship	Human expertise may become more or less consequential independently of formal job authority	Authority shifts reflect broader restructuring rather than AI	Track changes in whose judgments alter decisions before and after AI integration
Responsibility–control misalignment → autonomy–responsibility decoupling	Human answerability remains while authority, information, or intervention capability declines	Responsibility is retained despite reduced practical control over the decision process	Traceability; escalation options; role clarity; decision stakes	Responsibility tensions should intensify where answerability exceeds meaningful agency	Ambiguity reflects poorly specified roles rather than AI involvement	Independently measure formal accountability, actual intervention capacity, and retrospective answerability

Nominal human oversight → potentially symbolic control	A human remains formally "in the loop"	Oversight exists procedurally but may lack information, time, authority, or meaningful override capability	Workload; interface design; automation bias; organizational incentives	Human presence should protect autonomy only when oversight is consequential	Human presence itself increases scrutiny sufficiently to prevent problematic reliance	Experimentally vary nominal presence versus substantive intervention capability
Trust → calibrated or inappropriate reliance	Workers develop positive expectations regarding AI performance	Trust influences acceptance of recommendations and willingness to verify them	AI reliability; task uncertainty; expertise; feedback quality	Trust should affect autonomy through behavior rather than through attitudes alone	High reliance simply reflects objectively superior AI accuracy	Measure trust and behavioral reliance separately under manipulated AI reliability
Repeated AI-assisted decisions → recursive allocation changes	Workers repeatedly interact with AI across similar tasks	Prior outcomes shape confidence, delegation, checking, and future cognitive allocation	Feedback visibility; learning opportunities; error experience; organizational routines	Authority and cognitive allocation may evolve over time rather than remain fixed	Changes reflect ordinary learning unrelated to AI	Use longitudinal or repeated-task designs tracking reliance, verification, skill, and decision influence

Note. Relationships in this table are proposed theoretical mechanisms unless explicitly described as established background conditions in the manuscript. Expected implications indicate directions for empirical examination, not demonstrated causal effects. Rival explanations are included to make the model falsifiable and to prevent technological explanations from being preferred automatically over organizational, task, or individual alternatives.

Theoretical propositions

The propositions translate the model into conditional expectations rather than universal claims. Experimental research shows that people can weight algorithmic advice more heavily than human advice, although expertise conditions this tendency [23]. Reliance also varies with task perception: algorithm aversion is stronger when tasks are perceived as subjective than when they are perceived as objective [24]. More broadly, evidence synthesizing human–AI combinations indicates that complementarity is heterogeneous rather than guaranteed [25]. Reliance on algorithmic judgment and the value of human–AI combination therefore vary with task and decision conditions rather than following a universal pattern of aversion, appreciation, or complementarity [23–25].

Proposition 1. Moving from execution support toward AI substitution of evaluative judgment is proposed to reduce enacted behavioral autonomy when workers lack meaningful opportunities to interrogate, override, or reformulate AI output; task verifiability and practical intervention capability are proposed moderators.

Proposition 2. Cognitive offloading is proposed to have different autonomy implications depending on what is offloaded. Routine memory or execution offloading may preserve resources for evaluation, whereas evaluative-reasoning offloading is proposed to reduce retained evaluative agency when independent verification is also weakened.

Proposition 3. Human–AI complementarity is proposed to depend partly on accurate metaknowledge regarding relative capability. Miscalibration is expected to increase misdelegation and inappropriate reliance rather than producing a uniform tendency toward either algorithm aversion or appreciation.

Proposition 4. Autonomy–responsibility decoupling is proposed to become more likely when formal human answerability remains high while actual decision authority, relevant information, or intervention capability shifts away from the accountable human.

Proposition 5. Human oversight is proposed to reduce autonomy–responsibility decoupling only when it is consequential: the human has sufficient information, time, authority, and practical capacity to alter the decision.

Proposition 6. The consequences of AI assistance are proposed to vary across task characteristics and relative human–AI capabilities; neither automation nor augmentation is expected to be uniformly autonomy-preserving or performance-enhancing.

These propositions can be rejected or revised by evidence contradicting their mechanisms. That requirement prevents unfavorable AI outcomes from being treated automatically as confirmation of reduced autonomy.

Safeguards and empirical tests

Safeguards should be treated as testable design conditions rather than implementation prescriptions. Experimental evidence indicates that assurance that humans remain in the decision loop can influence trust in AI-assisted decisions [26]. Human presence, however, does not establish meaningful oversight if the person lacks information, time, authority, or a practical route to reject an output.

Explanation likewise requires contextual evaluation. Experiments in personnel selection show that reactions to human versus algorithmic decision makers can depend on explanations and characteristics of the selection test [27]. Explanation should

therefore be distinguished from substantive power: receiving a reason for an automated judgment does not necessarily create the capacity to contest or replace it. Behavioral research also distinguishes expressed trust from actual reliance and shows that overreliance can occur when scrutiny would be warranted [28]. Formal human involvement, explanations, and expressed trust should therefore not be assumed to produce calibrated reliance without contextual and behavioral validation [26–28].

Strong tests of the model should manipulate AI authority, reliability, override capacity, information access, and task verifiability while measuring actual rejection, verification, correction, and decision influence. Longitudinal studies should examine whether repeated offloading changes evaluative practice rather than inferring long-term effects from short tasks. Multilevel designs should separate individual reliance from organizational autonomy and compare formal governance policies with enacted work practice.

Measurement should also distinguish perceptions from consequential behavior. Behavioral autonomy requires observable choice or revision opportunities; retained evaluative agency requires evidence of independent checking and alternative formulation; responsibility alignment requires mapping answerability against actual authority and intervention capability. Such designs could refine, delimit, or reject the proposed model.

Conclusion

AI-assisted work creates an autonomy problem not because assistance necessarily removes human agency, but because the distribution of agency can become difficult to observe. The proposed Assisted-Autonomy Model separates authority allocation, judgment allocation, cognitive allocation, and responsibility alignment while introducing retained evaluative agency and autonomy–responsibility decoupling as proposed constructs.

Its central implication is analytical. Cognitive offloading can conserve resources or relocate consequential reasoning; delegation can support complementarity or become judgment substitution; human oversight can be consequential or merely formal. These possibilities depend on task, system, worker, and organizational conditions.

The framework therefore directs attention away from symbolic labels such as “augmented” or “human in the loop” and toward enacted decision arrangements: which judgments are contestable, which cognitive operations remain active, and whether the actor held answerable could meaningfully have acted otherwise. The model does not establish a validated causal sequence, universal safeguard, or deployment rule. Its contribution is to specify relationships and boundaries that empirical research can distinguish, test, refine, or reject.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Jarrahi MH. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Bus Horiz. 2018;61(4):577-86. doi:10.1016/j.bushor.2018.03.007
2. Parker SK, Grote G. Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. Appl Psychol. 2022;71(4):1171-204. doi:10.1111/apps.12241
3. Kellogg KC, Valentine MA, Christin A. Algorithms at work: The new contested terrain of control. Acad Manag Ann. 2020;14(1):366-410. doi:10.5465/annals.2018.0174
4. Raisch S, Krakowski S. Artificial intelligence and management: The automation–augmentation paradox. Acad Manag Rev. 2021;46(1):192-210. doi:10.5465/amr.2018.0072
5. Dell’Acqua F, McFowland E III, Mollick E, Lifshitz H, Kellogg KC, Rajendran S, et al. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organ Sci. 2026;37(2):403-23. doi:10.1287/orsc.2025.21838
6. Murray A, Rhymer J, Sirmon DG. Humans and technology: Forms of conjoined agency in organizations. Acad Manag Rev. 2021;46(3):552-71. doi:10.5465/amr.2019.0186
7. Baird A, Maruping LM. The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. MIS Q. 2021;45(1):315-41. doi:10.25300/MISQ/2021/15882
8. Shrestha YR, Ben-Menahem SM, von Krogh G. Organizational decision-making structures in the age of artificial intelligence. Calif Manage Rev. 2019;61(4):66-83. doi:10.1177/0008125619862257

9. Parent-Rocheleau X, Parker SK. Algorithms as work designers: How algorithmic management influences the design of jobs. Hum Resour Manag Rev. 2022;32(3):100838. doi:10.1016/j.hrmr.2021.100838
10. Fügener A, Grahl J, Gupta A, Ketter W. Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. Inf Syst Res. 2022;33(2):678-96. doi:10.1287/isre.2021.1079
11. Lebovitz S, Lifshitz-Assaf H, Levina N. To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. Organ Sci. 2022;33(1):126-48. doi:10.1287/orsc.2021.1549
12. Strich F, Mayer AS, Fiedler M. What do I do in a world of artificial intelligence? Investigating the impact of substitutive decision-making AI systems on employees' professional role identity. J Assoc Inf Syst. 2021;22(2):304-24. doi:10.17705/1jais.00663
13. Gilbert SJ, Bird A, Carpenter JM, Fleming SM, Sachdeva C, Tsai PC. Optimal use of reminders: Metacognition, effort, and cognitive offloading. J Exp Psychol Gen. 2020;149(3):501-17. doi:10.1037/xge0000652
14. Hu X, Luo L, Fleming SM. A role for metamemory in cognitive offloading. Cognition. 2019;193:104012. doi:10.1016/j.cognition.2019.104012
15. Kelly MO, Risko EF. Offloading memory: Serial position effects. Psychon Bull Rev. 2019;26(4):1347-53. doi:10.3758/s13423-019-01615-8
16. Santoni de Sio F, Mecacci G. Four responsibility gaps with artificial intelligence: Why they matter and how to address them. Philos Technol. 2021;34(4):1057-84. doi:10.1007/s13347-021-00450-x
17. Königs P. Artificial intelligence and responsibility gaps: What is the problem? Ethics Inf Technol. 2022;24:36. doi:10.1007/s10676-022-09643-0
18. Coeckelbergh M. Artificial intelligence, responsibility attribution, and a relational justification of explainability. Sci Eng Ethics. 2020;26(4):2051-68. doi:10.1007/s11948-019-00146-8
19. Anthony C. When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. Adm Sci Q. 2021;66(4):1173-212. doi:10.1177/00018392211016755
20. Pachidi S, Berends H, Faraj S, Huysman M. Make way for the algorithms: Symbolic actions and change in a regime of knowing. Organ Sci. 2021;32(1):18-41. doi:10.1287/orsc.2020.1377
21. Möhlmann M, Zalmanson L, Henfridsson O, Gregory RW. Algorithmic management of work on online labor platforms: When matching meets control. MIS Q. 2021;45(4):1999-2022. doi:10.25300/MISQ/2021/15333
22. Glikson E, Woolley AW. Human trust in artificial intelligence: Review of empirical research. Acad Manag Ann. 2020;14(2):627-60. doi:10.5465/annals.2018.0057
23. Logg JM, Minson JA, Moore DA. Algorithm appreciation: People prefer algorithmic to human judgment. Organ Behav Hum Decis Process. 2019;151:90-103. doi:10.1016/j.obhdp.2018.12.005
24. Castelo N, Bos MW, Lehmann DR. Task-dependent algorithm aversion. J Mark Res. 2019;56(5):809-25. doi:10.1177/0022243719851788
25. Vaccaro M, Almaatouq A, Malone T. When combinations of humans and AI are useful: A systematic review and meta-analysis. Nat Hum Behav. 2024;8(12):2293-303. doi:10.1038/s41562-024-02024-1
26. Aoki N. The importance of the assurance that "humans are still in the decision loop" for public trust in artificial intelligence: Evidence from an online experiment. Comput Human Behav. 2021;114:106572. doi:10.1016/j.chb.2020.106572
27. Wesche JS, Hennig F, Kollhed CS, Quade J, Kluge S, Sonderegger A. People's reactions to decisions by human vs. algorithmic decision-makers: The role of explanations and type of selection tests. Eur J Work Organ Psychol. 2024;33(2):146-57. doi:10.1080/1359432X.2022.2132940
28. Klingbeil A, Grützner C, Schreck P. Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI. Comput Human Behav. 2024;160:108352. doi:10.1016/j.chb.2024.108352