INTERNATIONAL JOURNAL OF BUSINESS AND MANAGEMENT (IJBM)
ISSN: 2815-9330 (Online)

VOL. 5, NO. 2 2026 https://iessociety.org/

Human–AI collaboration capability: A new driver of organizational performance

Dr. Naveen Nandal[1], Swati Luthra[2*], Afseer Majeed[3], Dr. Rushi Prasad Sahoo[4] and Ulfah Fajarini[5]

[1]Associate Professor, Delhi Institute of Higher Education (DIHE), Delhi, India

[2]Assistant Professor, Department of Applied Science and Humanities, School of Engineering and Technology, Vivekananda Institute of Professional Studies-Technical Campus, India

[3]Assistant Professor, College of Business, Jumeira University, Dubai, United Arab Emirates

[4]Department of Mathematics and Statistics, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India

[5]Social Education Department, Syarif Hidayatullah State Islamic University Jakarta, Indonesia

Email: naveennandal22@yahoo.com[1]; afseer.majeed@ju.ac.ae[3]; rushiprasad27@gmail.com[4]; ulfah.fajarini@uinjkt.ac.id[5]

*Correspondence: Luthraphd@gmail.com

Abstract

As access to advanced artificial intelligence becomes near-universal among organizations, performance differences among adopters continue to widen, suggesting that the AI asset itself is not the source of competitive advantage. This paper locates the residual explanation in an organizational capability rather than in the technology stock. Human–AI Collaboration Capability (HAICC) is defined as an organization's ability to set up, orchestrate, and continuously recalibrate human–AI work systems such that joint output exceeds what either party could achieve alone with the same resources. Grounded in the resource-based view, dynamic capabilities theory, the complementarity tradition in the economics of organization, and human–automation research, HAICC is conceptualized as a second-order formative construct comprising six dimensions: task decomposition acuity, interpretive competence, calibrated reliance, workflow reconfigurability, joint learning loops, and accountability architecture. Two mechanisms give the framework analytical leverage: the coupling tax, the transaction cost of joint human–AI work, and reliance elasticity, the responsiveness of human reliance to changes in system reliability. The paper links HAICC to operational, market, and financial performance through decision quality, operational agility, and knowledge recombination, and specifies environmental dynamism, task interdependence, data infrastructure maturity, regulatory intensity, and workforce learning orientation as contingencies. A staged validation protocol and candidate item pool are proposed to support future empirical testing. The paper is conceptual: it reports no primary data, and the propositions and candidate items are advanced for subsequent empirical validation rather than

DOI: https://doi.org/10.56879/ijbm.v5i2.131

tested here. The framework reframes the AI productivity problem as a capability challenge rather than a computational one, clarifying the conditions under which human–AI collaboration, rather than computation itself, becomes a scarce and advantage-generating resource.

***Keywords*:** Human AI Collaboration, Organizational Capability, Dynamic Capabilities, Calibrated Reliance, Coupling Tax, Complementarity, Organizational Performance, Artificial Intelligence

JEL Classification: O33, M15, L21

1. Introduction

Two observations sit uneasily together. The first is that access to advanced artificial intelligence has become close to universal among medium and large organizations. Capabilities that were proprietary and expensive a few years ago are now procured through metered interfaces, embedded in productivity suites, or obtained as open-weight models that can be hosted at modest cost. The second observation is that the performance consequences of this access are strikingly uneven. Some organizations report material gains in throughput, decision quality, and product development speed; many others report pilots that do not scale, projects that are quietly retired, and measured productivity that is indistinguishable from the pre-adoption baseline. Experimental and field evidence confirms that generative systems can raise individual output substantially at the task level (Noy & Zhang, 2023; Brynjolfsson et al., 2023), which sharpens rather than resolves the puzzle of uneven organizational returns (Dwivedi et al., 2021). The strategy literature has a standard explanation for uneven returns to a widely available input: the input is not the source of the rent. Resources that are readily purchased on factor markets cannot sustain advantage, because rivals can acquire them on similar terms (Barney, 1986; Dierickx & Cool, 1989). Advantage instead accrues to socially complex, path-dependent, and causally ambiguous organizational arrangements that competitors cannot buy and find difficult to observe or replicate (Barney, 1991; Peteraf, 1993). Applied to the present moment, this logic has a sharp implication. The model is becoming the commodity. What remains scarce is the organizational system through which human judgment and machine inference are joined.

This paper develops that implication into a construct. We propose Human–AI Collaboration Capability (HAICC) as a firm-level capability that is analytically distinct from AI adoption, AI readiness, information technology capability, and analytics capability, and we argue that HAICC is the proximate driver of the performance differences that adoption measures fail to explain. The argument rests on a simple accounting identity. The net value of a joint human–machine work system is the gross augmentation gain minus the coordination cost of joint work. Deployments frequently generate real gross gains at the task level while incurring verification effort, handoff latency, and explanation overhead that consume those gains entirely. We call this coordination cost the *coupling tax*. Capability differences show up less in the size of the gross gain, which is

DOI: https://doi.org/10.56879/ijbm.v5i2.131

largely determined by the model and the task, and more in the size of the tax, which the organization determines.

A second mechanism concerns how reliance is allocated. Human–automation research has long distinguished appropriate reliance from trust as a level (Parasuraman & Riley, 1997; Lee & See, 2004). Organizations that rely uniformly on machine output exhibit automation bias; organizations that discount it uniformly exhibit algorithm aversion (Dietvorst et al., 2015; Logg et al., 2019). Both are failures of the same kind: reliance that does not move with reliability. We therefore define *reliance elasticity* as the responsiveness of reliance to change in actual reliability, and argue that it, rather than the mean level of trust, carries the performance-relevant variance.

The paper makes four contributions. First, it specifies a firm-level construct with explicit boundaries, six dimensions, and stated microfoundations, addressing the requirement that a theoretical contribution identify what a construct is and is not (Whetten, 1989). Secondly, it provides two mechanisms, the coupling tax and reliance elasticity, which transform a general statement about collaboration into statements that can potentially be falsified. Third, it builds the nomological network of 12 propositions as a link between HAICC and performance via three mediators and five contingencies and identifies the conditions under which this relationship should weaken. Fourth, it offers a "stage" process of measurement and a candidate item pool allowing for the operationalizing of the construct without additional conceptualization.

The rest of the story goes on as follows. Section 2 provides an overview of the theory and framework of HAIC and its relationship with neighbouring frameworks. The construct definition and dimensions are provided in Section 3. Section 4 formalizes the two mechanisms. Propositions and a conceptual model are presented in Section 5. The measurement and empirical strategy are described in Section 6. The contributions and implications are discussed in Section 7, limitations and a research agenda in Section 8, and it concludes in Section 9.

2. Theoretical Background

2.1 Resources, Capabilities, and the Limits of the Artificial Intelligence Asset

According to the resource-based view, the resources which create advantage are valuable, rare, imperfectly imitable, and imperfectly substitutable (Barney, 1991). The condition of imitability is the one that is binding here. If a resource is competitively supplied, its rent will be exhausted by the price, and there will be no residual rent (Barney, 1986: p.51). The important difference, as Dierickx and Cool (1989) pointed out, is between asset stocks and asset flows: the latter must be built up over time and are subject to time-compression diseconomies. General-purpose AI systems are nearing the stock market. The processes that a company implements in determining which questions it will ask a model, how it will audit the model's responses, who may edit it, and how the flow will be re-established once the model's results alter the organization's capacity. The point is further accentuated in the framework of dynamic capabilities theory. Teece et al. (1997) and

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Teece (2007) refer to these as capacities: the ability to notice opportunity, commit resources to capitalize on that opportunity, and alter the organisation's structure as circumstances evolve. Eisenhardt and Martin (2000) note that in high-velocity settings such capabilities take the form of simple, experiential rules rather than elaborate routines. Winter (2003) distinguishes ordinary from dynamic capability by whether the routine changes the organization's operating routines. On this criterion, HAICC is dynamic: its core work is the repeated reconfiguration of the division of labour between people and systems as the systems' capability frontier shifts, which under current conditions shifts on a timescale of months. Microfoundations matter because capabilities reside in individual skills, in social interaction and in structure simultaneously (Felin et al., 2015); a construct that stops at the organizational level cannot explain why two firms with identical technology stacks diverge.

2.2 Complementarity and the Intangible-Investment Lag

The economics of organization supplies the second base. Milgrom and Roberts (1990) showed that modern manufacturing technologies deliver returns only in combination with matching changes in strategy and organization, so that partial adoption can be worse than none. Brynjolfsson and Hitt (2000) documented the same pattern for information technology: measured returns depend on complementary organizational investment. Brynjolfsson et al. (2021) formalized the resulting time path as a productivity J-curve, in which unmeasured intangible investment depresses measured productivity before the complementary capital matures and returns appear. Autor (2015) added the task-level logic that automation of some subtasks raises the value of the human subtasks that remain complementary to them. Taken together, these results predict exactly what is observed: a period in which adoption is broad, measured returns are weak, and dispersion is wide, because organizations differ in the intangible complementary asset they are building. The contribution available to organization theory is to say what that asset consists of. HAICC is our answer.

2.3 From Interaction to Capability

Research on joint human–machine performance has existed for decades, indicating that automation does not eliminate human labor, misuse and disuse is systematic, and trust is a tuning knob that can be calibrated according to the perceived reliability and has measurable lags (Parasuraman & Riley, 1997; Lee & See, 2004; Schaefer et al., 2016). The present research builds on these findings by examining opaque, probabilistic, and generative systems in which the cues for calibration are less reliable: output is fluent, or confidence signals are unreliable (Glikson & Woolley, 2020). An emerging field of organization and information systems research has started to detail the corresponding collective adaptations. What they reveal is that occupational groups reinterpret and domesticate the analytical system, rather than follow it (Faraj et al., 2018; Anthony, 2021); that brokers emerge, who translate between the output of the model and the domain practice (Waardenburg et al., 2022); that ground truth is itself constructed and contested (Lebovitz et al., 2021); that opacity provokes reactivity instead of compliance (Rahman, 2021); and that new

DOI: https://doi.org/10.56879/ijbm.v5i2.131

configurations of the human-in-the-loop are assembled locally and unevenly (Grønsund & Aanestad, 2020). These adaptations extend a longer-standing observation that information technology both automates and informates work, redistributing knowledge, skill and authority within the organization (Zuboff, 1988), and that technology in use is enacted in practice rather than given by the artefact (Orlikowski, 2000); recent syntheses accordingly argue that emerging technologies are best theorized relationally, through the practices they reconfigure (von Krogh, 2018; Bailey et al., 2022). Raisch and Krakowski (2021) say that it makes little sense to choose automation or augmentation independently of each other, since the one without the other is not effective. Murray et al. (2021) theorize conjoined agency, distinguishing configurations in which human and machine contributions are separable from those in which they are not. Fügener et al. (2021) show that the benefits of working with a machine partner depend on how the task is delegated, and that gains realized at the individual level need not aggregate to the collective. Puranam (2021) frames the allocation of decision rights between people and algorithms as an organization design problem. Vaccaro et al. (2024) report meta-analytic evidence that human–machine combinations frequently perform worse than the better of the two components acting alone, which is difficult to reconcile with any account in which collaboration is intrinsically beneficial but straightforward to reconcile with a coupling-tax account.

What this literature has not supplied is a firm-level capability construct with dimensions, measurement, and a stated link to performance. Studies of AI capability at the firm level exist (Mikalef & Gupta, 2021; Enholm et al., 2022) but define capability chiefly as a bundle of tangible, human, and intangible resources, which reproduces the resource-inventory logic that the commoditization of models undermines. Krakowski et al. (2023) show that the sources of advantage shift as AI capability improves, but stop short of specifying the collaborative capability itself. Choudhury et al. (2020) demonstrate empirically that machine learning and human capital are complements, which is a necessary premise of our argument rather than a substitute for it.

2.4 Positioning Against Adjacent Constructs

Table 1 positions HAICC against the constructs with which it is most likely to be confused. The comparison is on three grounds: the primary locus of the construct, the way each treats the technology, and the limitation that leaves the present question unanswered. The comparison set spans AI readiness, information technology capability (Bharadwaj, 2000), absorptive capacity (Cohen & Levinthal, 1990; Zahra & George, 2002), analytics capability, human–automation teaming, and algorithmic management.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Table 1. Positioning of Human–AI Collaboration Capability against Adjacent Constructs

Construct	Primary locus	Technology treated as	Residual limitation
AI readiness or maturity	Preparedness prior to deployment	Object to be acquired	Ex ante; silent on how joint work performs after deployment
IT capability	Infrastructure, technical and managerial IT skills	Configurable asset base	Formulated for deterministic systems with stable specifications
Absorptive capacity	Acquisition and assimilation of external knowledge	Knowledge source	Concerns flows between organizations, not within human–machine dyads
Analytics capability	Data, tools, analytical talent	Input to decisions	Treats the use of output as unproblematic
Human–automation teaming	Operator, task and interface	Teammate	Individual and task level; no firm-level rent argument
Algorithmic management	Control and coordination of labour	Instrument of control	Directed at power and resistance, not performance capability
HAICC (this paper)	Repeatable routines coupling judgment to inference	Non-deterministic collaborator whose frontier moves	

Note. Constructs are characterized at the level of their dominant usage; individual studies vary.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

3. Conceptualizing Human–AI Collaboration Capability

3.1 Definition and Construct Boundaries

We define Human–AI Collaboration Capability as an organization's demonstrated and repeatable ability to configure, coordinate, and continuously recalibrate joint human–machine work systems such that joint output exceeds what either humans or systems achieve independently under equivalent resource constraints. Four elements of the definition carry weight. Demonstrated and repeatable. HAICC is inferred from realized patterns of joint work, not from stated intentions, declared strategies or procurement decisions. This distinguishes it from readiness constructs and makes it, in principle, observable in process data. Configure, coordinate and recalibrate. The capability has a static component, the quality of the current division of labour, and a dynamic component, the rate at which that division is revised as system capability changes. The dynamic component is what makes HAICC a dynamic capability in Winter's (2003) sense.

Joint work systems. The unit is the work system, comprising tasks, people, systems, information flows and controls, consistent with the sociotechnical tradition (Bostrom & Heinen, 1977; Sarker et al., 2019). HAICC is not an attribute of a model, an individual, or an interface.

Exceeds either party independently. The comparison standard is the better of the two components acting alone, not the pre-adoption baseline. This is a deliberately demanding benchmark, and it is the benchmark against which many deployments fail (Vaccaro et al., 2024).

Three boundary conditions apply. HAICC is defined for work in which machine inference is probabilistic and its errors are consequential; it has little purchase on fully deterministic automation, where the design problem is engineering rather than collaboration. It is defined at the organizational level, although it aggregates from individual and group microfoundations. Moreover, it is defined relative to a technological frontier: the same routines that constitute high HAICC at one capability level may constitute low HAICC at another, which is why reconfigurability is a dimension rather than an antecedent.

3.2 Dimension 1: Task Decomposition Acuity

The first dimension is the ability to partition work into subtasks and to assign each subtask to the party holding comparative advantage, including assignment to joint execution where neither party dominates. Agrawal et al. (2019) frame the underlying economics as a separation between prediction, which machines increasingly supply cheaply, and judgment, which specifies the payoffs over predicted states. Organizations differ markedly in whether they perform this separation at all. The common failure modes are wholesale: either an entire role is automated, importing errors on the subtasks where machine performance is weakest, or an entire role is protected, forgoing gains on the subtasks where machine performance is strongest. Acuity also has a temporal aspect, since a partition that was correct at one capability level becomes incorrect when the frontier moves. Hence, the capability includes periodic re-scoping rather than one-time design.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

3.3 Dimension 2: Interpretive Competence

The second dimension is the collective capacity to read, contextualize, and interrogate machine output: to distinguish output that is grounded in evidence from output that is merely fluent, to recognize when a case lies outside the distribution on which the system performs reliably, and to identify the assumptions embedded in a model's framing of the problem. Interpretive competence is collective rather than individual. It resides in domain expertise, in a working level of statistical literacy, and in routines that make critique a normal part of the workflow rather than an act of dissent. Lebovitz et al. (2021) show that where the reference standard is itself uncertain, practitioners cannot straightforwardly evaluate model output, and Waardenburg et al. (2022) show that translation work is often performed informally by individuals whose role is unrecognized. Organizations with high interpretive competence institutionalize that work.

3.4 Dimension 3: Calibrated Reliance

The third dimension is the extent to which reliance tracks reliability at the level of the individual task rather than the level of the system. High calibrated reliance means the organization accepts machine output where it is dependable and overrides it where it is not, and revises those judgments as evidence accumulates. Low calibrated reliance takes two symmetric forms: uniform acceptance, which imports machine error directly into outcomes, and uniform rejection, which forfeits the gains and additionally wastes the cost of the system. Both are inelastic. Section 4.1 formalizes the elasticity notion. The microfoundations are informational and social: reliance can only track reliability where outcome feedback is available and attributable, where interfaces make error patterns visible rather than concealing them behind uniform confidence, and where overriding the system carries no career penalty, which is a matter of psychological safety (Edmondson, 1999).

3.5 Dimension 4: Workflow Reconfigurability

The fourth dimension is the speed and cost at which roles, handoffs, controls, and interfaces can be redesigned when system capability changes. Reconfigurability is structural rather than attitudinal. It is supported by modular process design, by the ability to run low-cost pilots without extensive authorization, by slack that permits redesign to proceed alongside operations, and by the absence of deep hard-coding of process assumptions into surrounding systems. Its characteristic failure is a process frozen around a superseded model generation: controls designed for a system that hallucinated frequently, retained after the failure profile has changed, so that the organization pays verification costs calibrated to a risk it no longer faces.

3.6 Dimension 5: Joint Learning Loops

The fifth dimension is the bi-directional learning. On one side, the human workers get involved with the system and improve it systematically: corrections are not thrown away but captured, evaluation sets are updated when the system changes its case mix, retrieval corpora and prompt assets are curated as organizational assets and not personal artefacts. The opposite, the system

DOI: https://doi.org/10.56879/ijbm.v5i2.131

grows and does not subtract from the human competence. The other direction is obvious and is the one area which is often overlooked, and the one that poses the most significant long-run risk. Through this mechanism, where normal cases are absorbed and practitioners work primarily with the exceptions, the experiential foundation for expert judgment becomes reduced, and the organization's capacity to supervise the system gets weaker as it becomes more dependent on the system. We refer to this as the "deskilling spiral." Most importantly, it is a failure that happens very slowly and will not be reflected in near-term performance measures, so it is more important to design against it. The mechanism that supports the concept is that of the transactive memory (Argote & Ren, 2012), which means that joint performance depends on the accurate knowledge of what or who knows what, and that knowledge should not be forgotten.

3.7 Dimension 6: Accountability Architecture

The sixth dimension is the distribution of decision rights, the transparency of decision making, and the accountability of results; in other words, reliance that is defensible. When it is not known who is responsible for a decision made using a machine, two costs come into play. There are two types of risk - unmanaged risk and managed risk. The latter is not as widely publicized, but it is nonetheless important and is also defensive non-use: people do not want to depend on the systems over which they are responsible if they are challenged as to their output, so the organisation pays for the capability which is not used. The accountability architecture has a facilitating, not limiting role, then. It contains documented decision rights, as well as the ability to recreate the rationale for a decision after the event with audit trails, and thresholds on escalation that turn ambiguous decision-making into a routine and not a 'logistical work-out'. Kellogg et al. (2020) and Rahman (2021) demonstrate the resistance created by opacity in these arrangements, and Berente et al. (2021) demonstrate the design problem in the context of management of AI.

3.8 Specification as a Second-Order Formative Construct

The six dimensions are treated as forming rather than reflecting HAICC. Three considerations support formative specification. The dimensions are conceptually distinct rather than interchangeable manifestations of a common cause; dropping one alters the meaning of the construct, which is the diagnostic test for formative indicators (Diamantopoulos & Winklhofer, 2001; MacKenzie et al., 2011). Correlation among them is expected but not required, and organizations plausibly exhibit uneven profiles, for example, strong interpretive competence with weak reconfigurability. Moreover, the causal direction runs from dimension to capability: an improvement in accountability architecture increases HAICC rather than being a symptom of it. Formative specification has two consequences for empirical work. Internal consistency statistics are inappropriate as evidence of validity, and content validity must be established by argument and expert judgment rather than by factor structure. Section 6 sets out the resulting protocol. Table 2 summarizes the six dimensions with their microfoundations and characteristic failure modes.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Table 2. Dimensions of Human–AI Collaboration Capability

Dimension	Definition	Microfoundations	Characteristic failure
1. Task decomposition acuity	Partitioning work and assigning subtasks by comparative advantage; re-partitioning as the frontier moves.	Task analysis skill; workflow mapping routines; scheduled re-scoping reviews	Wholesale automation or wholesale retention of a role
2. Interpretive competence	Reading, contextualizing and interrogating output, including detection of ungrounded output and distribution shift	Domain expertise; statistical literacy; institutionalized critique; recognized translation roles.	Fluency mistaken for validity
3. Calibrated reliance	Adjusting reliance to reliability at task level; high reliance elasticity	Attributable outcome feedback; error-visible interfaces; psychological safety to override	Automation bias; algorithm aversion
4. Workflow reconfigurability	Speed and cost of redesigning roles, handoffs and controls as capability changes	Modular process design; low-cost piloting; redesign slack	Process frozen around a superseded model generation
5. Joint learning loops	Bidirectional improvement: humans improve the system; the system builds rather than erodes human skill	Correction capture; refreshed evaluation sets; curated prompt and retrieval assets; deliberate practice	Deskilling spiral; stale evaluation sets

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Dimension	Definition	Microfoundations	Characteristic failure
6. Accountability architecture	Decision rights, traceability and answerability that make reliance defensible	Documented decision rights; audit trails; escalation thresholds	Responsibility gap; defensive non-use

Note. Dimensions are specified as formative indicators of HAICC; see Section 3.8.

4. Two Mechanisms

A capability construct earns its place only if it generates claims that could be wrong. This section formalizes two mechanisms that do so. Notation is introduced for precision; the symbols denote quantities to be estimated and no numerical values are asserted here.

4.1 Reliance Elasticity

Let A denote the actual reliability of a system on a defined class of tasks, measured as the proportion of outputs that would be judged correct by a defensible reference standard, and let R denote the reliance rate, the proportion of outputs accepted without material human amendment. Reliance elasticity is the proportional responsiveness of the second to the first:

$$\varepsilon_R = (\partial R / R) \div (\partial A / A) \quad (1)$$

Three regions are of interest. Where ε_R is near zero, reliance does not respond to reliability. This is the inelastic case, and it subsumes both automation bias, where R is high irrespective of A, and algorithm aversion, where R is low irrespective of A. Where ε_R is positive and bounded, reliance tracks reliability, and the organization captures gains where the system is dependable while containing error where it is not. Where ε_R is large and unbounded, reliance overreacts to reliability signals, producing the pattern Dietvorst et al. (2015) document at the individual level, in which a single observed error triggers disproportionate abandonment.

The construct has two properties that matter for theory. First, it separates the level of trust from its responsiveness, and it is the responsiveness that should carry performance-relevant variance: an organization with moderate mean reliance that is well allocated across tasks should outperform one with high mean reliance that is not. This is a testable divergence from trust-level accounts. Second, it is estimable in practice wherever override decisions and outcomes are both logged, which is increasingly common, so the mechanism is not merely conceptual.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

4.2 The Coupling Tax

Let G denote the gross gain from joint work relative to the better single-party benchmark, and let C denote the coordination cost incurred in achieving it. The net contribution is:

$$\Delta\Pi = G - C, \quad where\ C = C_v + C_h + C_e \qquad (2)$$

The three components are verification cost C_v, the effort of checking output before acting on it; handoff cost C_h, the latency and information loss at transitions between human and machine steps; and explanation cost C_e, the effort of rendering machine-assisted decisions communicable to those entitled to an account, including regulators, clients, and colleagues. The coupling tax is the sum. It is a transaction cost of the ordinary kind, arising because joint production requires coordination that single-party production does not.

The argument is that G is determined largely by the model and the task, and is therefore approximately common across organizations with comparable access. In contrast, C is determined by organizational arrangements and varies widely. This yields a specific and falsifiable explanation of null results: where a deployment shows no net effect, the prediction is that gross task-level gains are present and measurable but are offset by verification, handoff, and explanation costs. The alternative explanations, that the technology does not work or that adoption is too shallow, make different predictions about what a task-level audit would find.

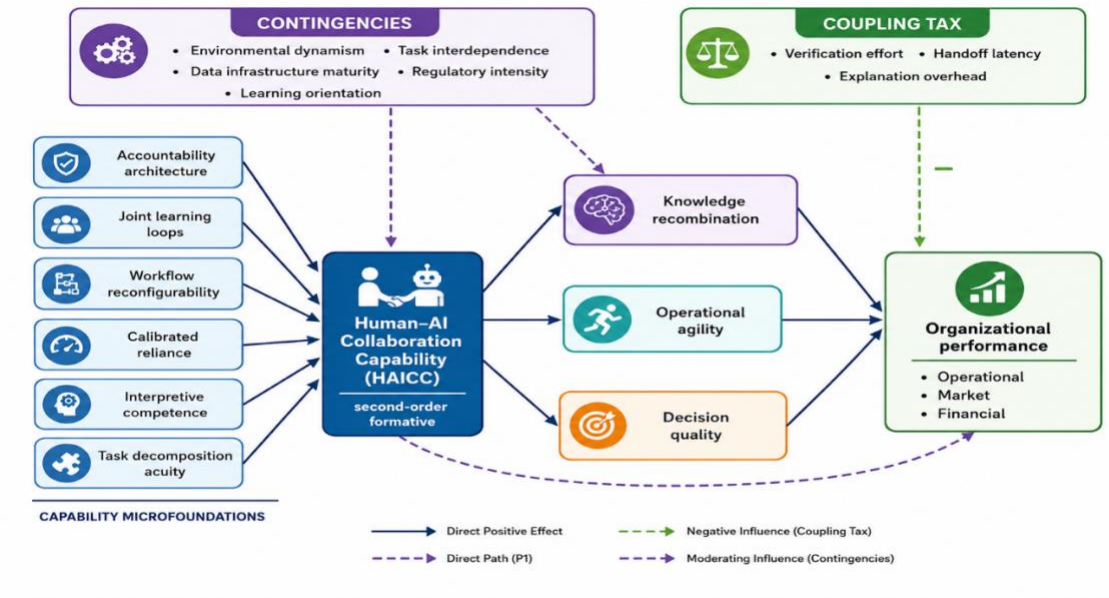
Each dimension of HAICC maps onto a component of the tax. Task decomposition acuity reduces C_h by minimizing the number of transitions and placing them where information loss is lowest. Interpretive competence reduces C_v by directing scrutiny towards cases where error is likely rather than distributing it uniformly. Calibrated reliance reduces C_v for the same reason, from the reliance side. Reconfigurability prevents the accumulation of controls calibrated to superseded risk profiles. Learning loops reduce C_v over time by raising A. Accountability architecture reduces C_e by making the basis of decisions reconstructable as a by-product of the workflow rather than as a subsequent reconstruction effort. The mapping is not decorative: it implies that the dimensions should be jointly rather than separately necessary, and that partial improvement may not show a net effect, which is the Milgrom and Roberts (1990) complementarity result restated at the level of the work system.

5. Conceptual Model and Propositions

Figure 1 presents the conceptual model. The six dimensions form HAICC; HAICC affects

Figure 1. Conceptual model of Human–AI Collaboration Capability and organizational performance

DOI: https://doi.org/10.56879/ijbm.v5i2.131



organizational performance directly and through three mediating mechanisms; the coupling tax attenuates the transmission; and five contingencies condition the strength of the relationships.

5.1 The Direct Relationship

The first claim is the base claim. If HAICC is the complementary intangible asset whose absence explains weak measured returns, then it should predict performance after the technology stock is controlled for. The technology stock should predict performance only conditionally.

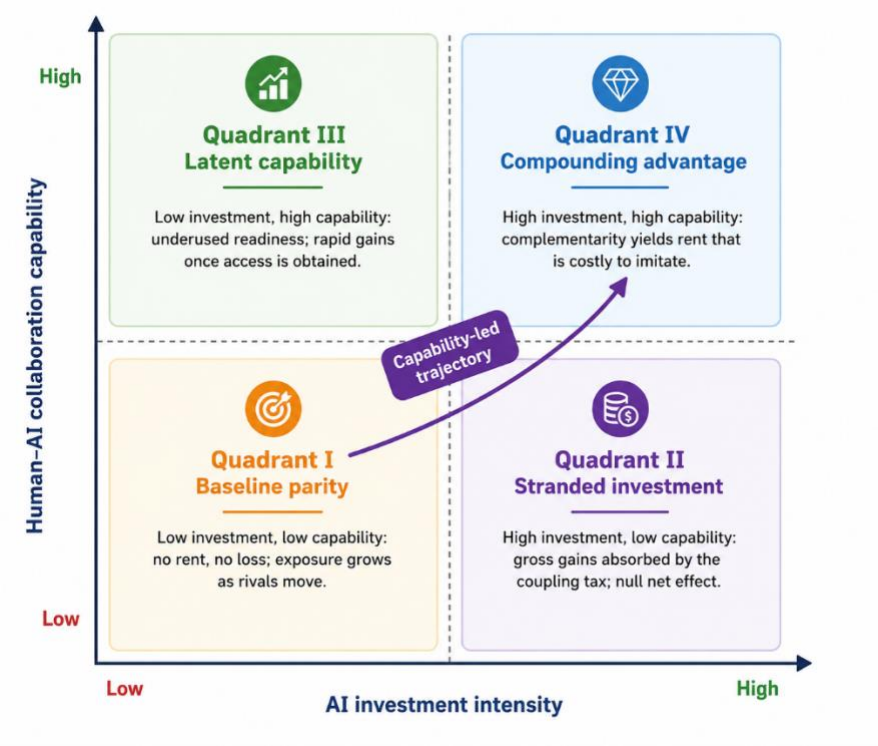
Proposition 1. HAICC is positively associated with organizational performance, controlling for AI investment intensity, firm size, industry, and prior performance.

Proposition 2. The association between AI investment intensity and organizational performance is conditional on HAICC; at low levels of HAICC, the association does not differ from zero, and it becomes positive as HAICC increases.

Proposition 2 is the sharper of the two. It predicts a specific pattern of cross-sectional heterogeneity that a main-effects specification would render as a weak average and dismiss as noise. Figure 2 renders the implied configuration as four archetypes.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Figure 2. Archetypes Generated By The Interaction Of Investment Intensity And Collaboration Capability



Quadrant II is the case of interest. High expenditure combined with low capability yields stranded investment: gross gains exist at task level but are consumed by the coupling tax, and the organization concludes, incorrectly, that the technology does not work. Quadrant III, low expenditure with high capability, is a latent position from which gains are realized rapidly once access is obtained, and it is a plausible description of organizations with strong process discipline that have adopted late. Quadrant I is exposed rather than safe, since its exposure grows as rivals move. Only Quadrant IV yields rent that resists imitation, because it depends on the accumulated flow rather than the purchasable stock.

5.2 Mediating Mechanisms

Three mediators are proposed, chosen because they correspond to the three ways joint work can create value: better decisions, faster response, and new combinations.

Decision quality. Where reliance is well calibrated, and interpretation is competent, the error rate of decisions taken with machine assistance falls below the rate achievable by either party alone, and the variance of decision quality across cases and decision-makers narrows. Variance reduction may matter more than mean improvement in regulated and safety-relevant settings.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Operational agility. Reconfigurability and decomposition acuity shorten the time between the recognition of a change in conditions and the corresponding adjustment of the work system, which is the operational content of the sensing and seizing capacities described by Teece (2007).

Knowledge recombination. Systems that make distant literatures, prior cases, and adjacent domains cheaply accessible expand the recombinant search space, which is the classical source of innovative output. The gain is realized only where interpretive competence permits the organization to distinguish productive combinations from plausible-sounding ones, and where learning loops retain what is found. March's (1991) exploration–exploitation tension applies: cheap search raises the risk of the failure that Levinthal and March (1993) called the myopia of learning, in which the organization explores only where search is easy.

Proposition 3. Decision quality partially mediates the relationship between HAICC and organizational performance.

Proposition 4. Operational agility partially mediates the relationship between HAICC and organizational performance.

Proposition 5. Knowledge recombination capacity partially mediates the relationship between HAICC and organizational performance.

Proposition 6. HAICC is negatively associated with the coupling tax, and the coupling tax mediates the negative component of the total effect of AI investment intensity on performance.

Proposition 7. Within the calibrated reliance dimension, reliance elasticity accounts for the performance-relevant variance; the mean reliance level does not predict performance once elasticity is included.

5.3 Contingencies

Five moderators are proposed. Each specifies a condition under which the return to HAICC should be higher or lower, and each therefore also specifies a condition under which the theory should fail.

Environmental dynamism. Where conditions change rapidly, the dynamic component of HAICC dominates, because a well-designed but static division of labour decays. Where conditions are stable, a one-time good configuration substitutes for the capability, and the return to reconfigurability falls.

Task interdependence. The coupling tax rises with the density of interdependence among tasks, because interdependence multiplies handoffs (Thompson, 1967). Capability that reduces the tax is therefore worth more where interdependence is high. In loosely coupled work, the tax is small, and HAICC has less to contribute.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Data infrastructure maturity. Reliable integration, lineage and access are preconditions for calibration and learning, since neither is possible without attributable feedback. Maturity is therefore a positive moderator, but with diminishing returns: beyond the threshold at which feedback is attributable, further infrastructure investment does not raise the return to HAICC. This is the necessary-but-not-sufficient structure, and it predicts a concave rather than linear interaction.

Regulatory intensity. Where decisions must be justified to external authorities, explanation cost is high, and the accountability architecture dimension carries a larger share of the total return. In contrast, the return to unconstrained automation falls.

Workforce learning orientation. Where the workforce treats capability change as an occasion for learning rather than for defence, joint learning loops function; where it does not, feedback is withheld and the loops break, which attenuates the relationship regardless of the technical arrangements.

Proposition 8. Environmental dynamism strengthens the relationship between HAICC and organizational performance, chiefly through workflow reconfigurability.

Proposition 9. Task interdependence strengthens the relationship between HAICC and organizational performance, because the coupling tax it reduces is larger.

Proposition 10. Data infrastructure maturity moderates the relationship positively with diminishing returns, consistent with a necessary but not sufficient condition.

Proposition 11. Regulatory intensity strengthens the contribution of accountability architecture to HAICC and weakens the contribution of unconstrained automation.

5.4 Path Dependence and Divergence

A final proposition concerns the trajectory rather than the level. Each dimension of HAICC accumulates from experience with prior configurations. Decomposition acuity improves with each partition attempted and evaluated; calibration requires a history of outcomes; reconfigurability improves as modular structures are built and reused. This is the classical structure of routine-based capability accumulation (Nelson & Winter, 1982) with time-compression diseconomies (Dierickx & Cool, 1989). *Proposition 12.* HAICC is path dependent: prior-period HAICC positively predicts current-period growth in HAICC, producing divergence in performance among organizations with comparable technology access over time. Proposition 12 has an uncomfortable implication for practice and policy. If capability compounds and technology does not, the distribution of returns to AI should widen rather than converge, and late movers cannot close the gap by increasing expenditure alone. Table 3 summarizes the twelve propositions, their focal relationships and the evidence each would require.

Table 3. Summary of Propositions

DOI: https://doi.org/10.56879/ijbm.v5i2.131

No.	Proposition
P1	HAICC is positively associated with organizational performance, controlling for AI investment intensity and firm resources.
P2	The association between AI investment intensity and performance is conditional on HAICC; at low HAICC it does not differ from zero.
P3	Decision quality partially mediates the HAICC–performance relationship.
P4	Operational agility partially mediates the HAICC–performance relationship.
P5	Knowledge recombination capacity partially mediates the HAICC–performance relationship.
P6	HAICC is negatively associated with the coupling tax, which mediates the negative component of the investment–performance relationship.
P7	Reliance elasticity, not mean reliance level, carries the performance-relevant variance in calibrated reliance.
P8	Environmental dynamism strengthens the HAICC–performance relationship, chiefly through reconfigurability.
P9	Task interdependence strengthens the HAICC–performance relationship.
P10	Data infrastructure maturity moderates positively with diminishing returns (necessary but not sufficient).
P11	Regulatory intensity raises the return to accountability architecture and lowers the return to unconstrained automation.
P12	HAICC is path dependent, producing performance divergence among organizations with comparable technology access.

Note. HAICC = Human–AI Collaboration Capability.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

6. Measurement and Empirical Strategy

6.1 Specification of the Estimating Relationship

For a cross-sectional test of Propositions 1 and 2, the estimating relationship may be written as:

$$Perf_i = \beta_0 + \beta_1 HAICC_i + \beta_2 AI_i + \beta_3 (HAICC_i \times AI_i) + \gamma' X_i + u_i \quad (3)$$

where Perf denotes organizational performance, AI denotes investment intensity, X is a vector of controls including size, industry, prior performance and data infrastructure maturity, and u is a disturbance term. Proposition 1 predicts $\beta_1 > 0$; Proposition 2 predicts $\beta_3 > 0$ with β_2 not distinguishable from zero at low HAICC. The coefficients are notation for quantities to be estimated; no values are asserted, and none should be inferred from this paper. Mediation in Propositions 3 to 5 requires the indirect paths to be estimated jointly rather than through sequential regressions, and Proposition 12 requires panel data with at least two capability observations.

6.2 A Staged Validation Protocol

Because HAICC is specified as second-order formative, validation cannot rest on factor structure. We propose five stages.

- *Stage 1. Content specification.* Establish the conceptual domain and the census of dimensions by structured review and expert panel, following MacKenzie et al. (2011). Because indicators are formative, the criterion is coverage of the domain, and the omission of a dimension is a specification error rather than a loss of reliability.
- *Stage 2. Item generation and expert sorting.* Generate candidate items per dimension (Appendix 1) and test assignment through sorting by informants independent of the item authors, retaining items assigned to the intended dimension by a substantial majority.
- *Stage 3. Pilot administration and diagnostics.* Administer to a pilot sample, examine indicator collinearity by variance inflation factor, and assess indicator weights and their significance rather than loadings. Redundancy analysis against a global single-item measure provides convergent validity evidence for formative constructs (Hair et al., 2019).
- *Stage 4. Nomological and criterion testing.* Estimate the model of Figure 1 on an independent sample, with performance obtained from archival sources where possible to avoid common-method inflation (Podsakoff et al., 2003), and with informants for capability items drawn from a different function than informants for outcome items.
- *Stage 5. Objective corroboration.* Corroborate the calibrated reliance dimension against behavioural traces, since override and outcome logs permit direct estimation of reliance elasticity as defined in Equation 1. Agreement between survey-based and trace-based estimates is the strongest available validity evidence and is the stage most often omitted.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

6.3 Operationalizing Mediators, Moderators and Outcomes

Table 4 sets out conceptual roles and candidate operationalizations. Where established scales exist they should be adapted rather than replaced, both to conserve construct clarity and to permit comparison with prior work.

Table 4. Mediators, Moderators and Outcomes: Roles and Candidate Operationalizations

Construct	Conceptual role	Candidate operationalization
Decision quality	Mediator: accuracy and consistency of assisted decisions	Error and reversal rates against a defensible reference standard; dispersion of outcomes across decision-makers
Operational agility	Mediator: speed of adjustment to changed conditions	Elapsed time from detection to workflow change; adapted agility scales
Knowledge recombination	Mediator: breadth of productive search	New-combination counts in outputs; share of output drawing on previously unused sources
Coupling tax	Suppressor: coordination cost of joint work	Verification time per case; count and latency of handoffs; effort per explanation request
Reliance elasticity	Sub-mechanism within calibrated reliance	Elasticity of override rate to task-class reliability from override and outcome logs (Equation 1)
Environmental dynamism	Moderator	Established dynamism scales; volatility of industry demand or price indices
Task interdependence	Moderator	Handoff density in process maps; adapted interdependence scales

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Construct	Conceptual role	Candidate operationalization
Data infrastructure maturity	Moderator (concave)	Integration, lineage and access coverage; feedback attributability
Regulatory intensity	Moderator	Count of binding disclosure or justification obligations; sector coding
Workforce learning orientation	Moderator	Established learning-orientation scales; participation in structured capability development
Organizational performance	Outcome	Operational (throughput, cycle time, quality), market (growth, retention) and financial (margin, return on assets) indicators from archival sources where available

Note. Candidate operationalizations are proposals for empirical work, not measures validated in this paper.

6.4 Research Designs

Different propositions call for different designs, and no single study can address the set. Table 5 maps propositions to designs and identifies the principal threat to inference in each case.

Table 5. Research Agenda: Propositions, Designs and Principal Inferential Threats

Propositions	Suitable design	Principal threat to inference
P1, P2	Cross-sectional survey with archival performance and interaction terms	Reverse causality: high performers may invest in capability. Requires lagged capability measures or instrumentation.
P3–P5	Multi-informant survey with mediation estimated jointly	Common-method inflation; mediator–outcome endogeneity.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Propositions	Suitable design	Principal threat to inference
P6	Process study or field measurement of verification, handoff, and explanation cost	Measurement burden; selective recording of costs that are normally invisible.
P7	Archival analysis of override and outcome logs within a single organization	External validity; log completeness; reliability of the reference standard.
P8–P11	Cross-industry survey stratified by dynamism and regulatory intensity	Unobserved industry heterogeneity confounded with the moderators.
P12	Panel design with at least two capability observations, or a longitudinal comparative case study	Attrition; capability measurement equivalence across periods.
Whole model	Field experiment varying one dimension, for example a controlled change in override authority	Feasibility and ethics of manipulating decision rights in live operations.

Note. Designs are indicative. Combining a within-organization trace study for P7 with a cross-organizational survey for P1 and P2 addresses the sharpest threats at acceptable cost.

7. Discussion

7.1 Theoretical Contributions

The paper contributes to three conversations. To the resource-based and dynamic-capabilities literature, it identifies a capability whose theoretical interest derives from a change in the imitability of a complementary asset. The classical account of information technology value allowed the technology itself to be partly differentiating, since large systems were configured idiosyncratically and integration was costly (Bharadwaj, 2000). General-purpose AI inverts this: the technology approaches a competitively supplied input while the organizational complement

DOI: https://doi.org/10.56879/ijbm.v5i2.131

remains socially complex. The paper thus describes a case of capability-side differentiation under technological convergence. It suggests that this configuration will recur whenever a general-purpose technology is commoditized faster than the organizational practices required to use it.

To the complementarity literature, it specifies the content of the intangible complement whose accumulation the productivity J-curve describes (Brynjolfsson et al., 2021). That literature establishes the existence and timing of unmeasured complementary investment without stating what is being built. The six dimensions constitute a candidate specification, and the coupling tax supplies a mechanism through which the complement operates on measured output.

To the growing literature on AI in organizations, it supplies a firm-level construct that connects micro-level findings on algorithmic work to performance outcomes. The findings that ground truth is constructed (Lebovitz et al., 2021), that brokers mediate (Waardenburg et al., 2022), that occupational groups reinterpret systems (Anthony, 2021), and that reliance is systematically miscalibrated (Dietvorst et al., 2015; Logg et al., 2019) are individually well established but do not aggregate into a statement about the firm. HAICC provides the aggregation, and the coupling tax explains why the meta-analytic result that human–machine combinations often underperform the better component (Vaccaro et al., 2024) is consistent with, rather than contrary to, the claim that collaboration can be a source of advantage. The paper also complements accounts of automation and augmentation as interdependent (Raisch & Krakowski, 2021) and of conjoined agency (Murray et al., 2021) by specifying what an organization must be able to do to manage that interdependence.

7.2 Implications for Practice

Four implications follow directly. First, measure the capability, not the deployment. Counts of use cases, licences or model calls are inputs; the diagnostic questions concern whether the partition of work between people and systems is deliberate and revisited, whether reliance varies with reliability, and how long a workflow redesign takes. Second, treat the coupling tax as a managed cost. Verification effort, handoff latency, and explanation overhead are usually unmeasured and therefore unmanaged. Where a deployment shows no benefit, the first diagnostic step is to establish whether gross gains are being consumed by these costs rather than to conclude that the technology is ineffective.

Third, protect the experiential base. If routine cases are absorbed and practitioners see only exceptions, expertise thins as dependence deepens. The organization loses its capacity to supervise the system at the moment it most needs it. Deliberate exposure to routine work, retention of independent judgment before consulting the system, and periodic unassisted calibration exercises are the available countermeasures. Fourth, invest asymmetrically in decision rights. Clarity about who may rely on what, and who answers for the result, resolves both the risk problem and the defensive non-use problem, and it is usually cheaper than the technical alternatives.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

7.3 Implications for Policy

Two implications concern policy. Where returns depend on an accumulated organizational capability rather than on technology access, policy instruments that address access alone, such as subsidized compute or model availability, will produce dispersed and disappointing results, and Quadrant II of Figure 2 will be populated at public expense. Instruments directed at capability, including process redesign support, evaluation infrastructure and skills provision for interpretive competence, address the binding constraint. Second, the path dependence of Proposition 12 implies widening rather than converging outcomes among adopters, with particular consequences for smaller organizations that lack the slack that reconfiguration requires. This is a distributional matter that access-oriented policy will not reach.

8. Limitations and Future Research

The paper is conceptual, and its propositions are untested. It provides a definition and a dimensional structure, but not evidence, and these definitions and dimensional structure are provisional until the validation sequence, described in Section 6.2, has been carried out once.

The following are four limitations that must be highlighted. First, the construct is defined in terms of a moving technological frontier, which poses a measurement-equivalence problem for the study of process over time. What may be measured as high capability at one capability level may be measured differently at a second level of capability two years later. Using relative instead of absolute practices will reduce but not eliminate the difficulty. Second, the six dimensions are suggested by theory, and it is possible that the census is not exhaustive; under formative specification, every estimate is influenced by the specification error of the omitted dimension, leading to a greater weight accorded to the expert-panel stage compared to a reflective construct. Third, the cost of the coupling tax is not one that any organization would typically track, and initial measurement will rely on purpose-built instruments that will be developed in the field, which may lead to what is easy to measure being used instead of what is important. Fourth, the framework is broad. With six dimensions, two mechanisms, three mediators, five contingencies and twelve propositions, full empirical validation of the complete nomological network is unlikely to be achievable in a single study. The research agenda should therefore be staged. We suggest that the first wave of testing prioritise Propositions 1, 2, 6 and 7, which together establish the direct effect, the conditional value of AI investment, the mediating role of the coupling tax, and the elasticity refinement of calibrated reliance; the remaining mediation and contingency propositions (3 to 5 and 8 to 11) and the path-dependence Proposition 12 require larger multi-industry samples and longitudinal designs.

Four directions seem to be the most fruitful ones. In several contexts, it is now possible to make estimates of reliance elasticity from override and outcome logs, and such estimates would test Proposition 7 directly with objective data. Longitudinal work on the deskilling spiral is scarce and consequential, since the failure is slow and invisible in near-term metrics. Cross-national

DOI: https://doi.org/10.56879/ijbm.v5i2.131

comparative work would establish whether the returns to accountability architecture vary with regulatory regime as Proposition 11 implies. Finally, the framework should be tested against settings in which the machine component is not language-based, including forecasting and scheduling applications, to establish whether the dimensions generalize beyond the generative case in which they were formulated.

9. Conclusion

The organizations that will benefit most from artificial intelligence are not, on this account, those that buy the most capable systems. They are those that become skilled at the specific and unglamorous work of dividing labour between people and machines, checking machine output where checking pays, changing the arrangement when the machine changes, and being clear about who answers for the result. That work is difficult to observe, slow to accumulate and hard to copy, which is precisely why it can support advantage while the systems themselves cannot. This paper has named that competence, specified its dimensions, supplied two mechanisms through which it operates, and set out how it might be measured and tested. The central empirical claim is falsifiable and, we suggest, testable with data that many organizations already hold: that capability rather than expenditure conditions the return to artificial intelligence, and that the difference is visible in the coordination costs of joint work. If the claim survives, the practical conclusion is that the scarce asset is collaboration rather than computation.

Acknowledgement

None apply.

Declarations

Funding. No funding was received.

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

Data availability. This is a conceptual paper. No datasets were generated or analysed. The candidate measurement items reported in Appendix 1 are provided in full in this article.

References

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. Information Economics and Policy, 47, 1–6. https://doi.org/10.1016/j.infoecopol.2019.05.001

Anthony, C. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. Administrative Science Quarterly, 66(4), 1173–1212. https://doi.org/10.1177/00018392211016755

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Argote, L., & Ren, Y. (2012). Transactive memory systems: A microfoundation of dynamic capabilities. Journal of Management Studies, 49(8), 1375–1382. https://doi.org/10.1111/j.1467-6486.2012.01077.x

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Bailey, D. E., Faraj, S., Hinds, P. J., Leonardi, P. M., & von Krogh, G. (2022). We are all theorists of technology now: A relational perspective on emerging technology and organizing. Organization Science, 33(1), 1–18. https://doi.org/10.1287/orsc.2021.1562

Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. Management Science, 32(10), 1231–1241. https://doi.org/10.1287/mnsc.32.10.1231

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120. https://doi.org/10.1177/014920639101700108

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. MIS Quarterly, 45(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. MIS Quarterly, 24(1), 169–196. https://doi.org/10.2307/3250983

Bostrom, R. P., & Heinen, J. S. (1977). MIS problems and failures: A socio-technical perspective. Part I: The causes. MIS Quarterly, 1(3), 17–32. https://doi.org/10.2307/248710

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. Journal of Economic Perspectives, 14(4), 23–48. https://doi.org/10.1257/jep.14.4.23

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work (NBER Working Paper No. 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. American Economic Journal: Macroeconomics, 13(1), 333–372. https://doi.org/10.1257/mac.20180386

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. Strategic Management Journal, 41(8), 1381–1411. https://doi.org/10.1002/smj.3152

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128–152. https://doi.org/10.2307/2393553

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. Journal of Marketing Research, 38(2), 269–277. https://doi.org/10.1509/jmkr.38.2.269.18845

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. Management Science, 35(12), 1504–1511. https://doi.org/10.1287/mnsc.35.12.1504

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., … Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. Administrative Science Quarterly, 44(2), 350–383. https://doi.org/10.2307/2666999

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? Strategic Management Journal, 21(10–11), 1105–1121. https://doi.org/10.1002/1097-0266(200010/11)21:10/11<1105::AID-SMJ133>3.0.CO;2-E

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. Information Systems Frontiers, 24(5), 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. Information and Organization, 28(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Felin, T., Foss, N. J., & Ployhart, R. E. (2015). The microfoundations movement in strategy and organization theory. Academy of Management Annals, 9(1), 575–632. https://doi.org/10.5465/19416520.2015.1007651

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. MIS Quarterly, 45(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. Journal of Strategic Information Systems, 29(2), 101614. https://doi.org/10.1016/j.jsis.2020.101614

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. Academy of Management Annals, 14(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. Strategic Management Journal, 44(6), 1425–1452. https://doi.org/10.1002/smj.3387

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. MIS Quarterly, 45(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. Strategic Management Journal, 14(S2), 95–112. https://doi.org/10.1002/smj.4250141009

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. MIS Quarterly, 35(2), 293–334. https://doi.org/10.2307/23044045

March, J. G. (1991). Exploration and exploitation in organizational learning. Organization Science, 2(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. Information & Management, 58(3), 103434. https://doi.org/10.1016/j.im.2021.103434

Milgrom, P., & Roberts, J. (1990). The economics of modern manufacturing: Technology, strategy, and organization. American Economic Review, 80(3), 511–528.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. Academy of Management Review, 46(3), 552–571. https://doi.org/10.5465/amr.2019.0186

Nelson, R. R., & Winter, S. G. (1982). An evolutionary theory of economic change. Harvard University Press.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. Organization Science, 11(4), 404–428. https://doi.org/10.1287/orsc.11.4.404.14600

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. Strategic Management Journal, 14(3), 179–191. https://doi.org/10.1002/smj.4250140303

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Puranam, P. (2021). Human–AI collaborative decision-making as an organization design problem. Journal of Organization Design, 10(2), 75–80. https://doi.org/10.1007/s41469-021-00095-2

Rahman, H. A. (2021). The invisible cage: Workers' reactivity to opaque algorithmic evaluations. Administrative Science Quarterly, 66(4), 945–988. https://doi.org/10.1177/00018392211010118

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review, 46(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. MIS Quarterly, 43(3), 695–719. https://doi.org/10.25300/MISQ/2019/13747

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding

DOI: https://doi.org/10.56879/ijbm.v5i2.131

autonomy in future systems. Human Factors, 58(3), 377–400. https://doi.org/10.1177/0018720816634228

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 28(13), 1319–1350. https://doi.org/10.1002/smj.640

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. Strategic Management Journal, 18(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

Thompson, J. D. (1967). Organizations in action: Social science bases of administrative theory. McGraw-Hill.

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour, 8(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. Academy of Management Discoveries, 4(4), 404–409. https://doi.org/10.5465/amd.2018.0084

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. Organization Science, 33(1), 59–82. https://doi.org/10.1287/orsc.2021.1544

Whetten, D. A. (1989). What constitutes a theoretical contribution? Academy of Management Review, 14(4), 490–495. https://doi.org/10.5465/amr.1989.4308371

Winter, S. G. (2003). Understanding dynamic capabilities. Strategic Management Journal, 24(10), 991–995. https://doi.org/10.1002/smj.318

Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. Academy of Management Review, 27(2), 185–203. https://doi.org/10.5465/amr.2002.6587995

Zuboff, S. (1988). In the age of the smart machine: The future of work and power. Basic Books.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Appendix 1. Candidate Measurement Items

Items are proposed for Stage 2 of the validation protocol described in Section 6.2 and have not been empirically validated. A seven-point agreement format is suggested, with the organization or the focal work system as the referent and with informants drawn from both operational and managerial levels. Under formative specification, items index distinct dimensions and should not be averaged into a single composite without estimating indicator weights.

Table A1. Candidate Items by Dimension

Dimension	Code	Candidate item
Task decomposition acuity	TDA1	We can state, for our main workflows, which steps are performed by people, which by systems, and which jointly.
	TDA2	Decisions about which steps to delegate to systems are made deliberately rather than by default.
	TDA3	We revisit the division of work between people and systems when system capability changes.
	TDA4	We can identify the steps in our workflows where machine performance is weakest.
Interpretive competence	IC1	Our staff can tell when a system's output is confident but not well founded.
	IC2	Our staff recognize cases that fall outside the conditions under which our systems perform reliably.
	IC3	Questioning system output is a normal part of our work rather than an exception.
	IC4	People who translate between system output and domain practice hold recognized roles.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Dimension	Code	Candidate item
Calibrated reliance	CR1	How much we rely on system output differs across task types according to how well the system performs on them.
	CR2	We change how much we rely on a system when evidence about its accuracy changes.
	CR3	Staff can override system output without adverse consequences for themselves.
	CR4	We receive feedback on the accuracy of decisions taken with system support.
Workflow reconfigurability	WR1	We can change a workflow that involves system support within weeks rather than quarters.
	WR2	Small changes to how people and systems share work can be trialled without extensive authorization.
	WR3	Our processes are designed so that a step can be reassigned without redesigning the whole process.
	WR4	We remove controls that are no longer justified by the current failure profile of our systems.
Joint learning loops	JL1	Corrections made by staff to system output are captured and used to improve the system.
	JL2	We refresh the cases against which we evaluate system performance as our work changes.
	JL3	Prompts, reference materials and retrieval sources are maintained as shared assets rather than personal ones.

DOI: https://doi.org/10.56879/ijbm.v5i2.131

Dimension	Code	Candidate item
	JL4	We take deliberate steps to ensure staff retain the skills that systems now perform routinely.
Accountability architecture	AA1	It is documented which decisions may be delegated to systems and which require human confirmation.
	AA2	We can reconstruct the basis of a decision taken with system support after the fact.
	AA3	It is clear who is answerable for the outcome of a decision taken with system support.
	AA4	Staff are able to justify decisions taken with system support to those entitled to an explanation.

Note. Codes are for reference during instrument development. Items are stated positively for clarity; reverse-worded variants should be developed and tested to reduce acquiescence bias.

DOI: https://doi.org/10.56879/ijbm.v5i2.131