SANKALPA: International Journal of Sustainability, Leadership & Management (SIJSLAM)
Vol. 2 No. 1 | June-2026| 45-60

From "Human Versus AI" to "Human & AI" - A Psychological Safety and Trust Based Framework for Managing Workforce Transformation in the Age of Artificial Intelligence

Anupama Reddy, Delivery Head, HP India Sales Pvt Ltd., email: anupama.reddy@hp.com, pranum@gmail.com
Dr. Anupama G, Professor & Director (I/C) - CDOE, Dayananda Sagar University, email: dranupama.g@dsu.edu.in

Abstract

Companies are buying artificial intelligence faster than they are getting their people ready for it. Executive enthusiasm about efficiency gains routinely outruns the psychological and organizational conditions employees actually need to work well alongside these tools, and the result is a gap between what leadership announces and what the workforce experiences that most transformation plans never name directly. This paper builds a framework for closing that gap and checks it against the evidence. Drawing on five constructs that keep turning up, separately, across organizational behaviour, human resource, and information systems research psychological safety, organizational trust, institutional memory, human–AI task allocation, and cross generational learning I propose an integrated model, the Human & AI Integration Model (HAIM), and test it against a widely read practitioner text on AI era workforce change. The method is a qualitative integrative review paired with directed thematic analysis, used to see how closely practitioner claims track peer reviewed evidence. They track closely. Psychological safety predicts whether people adopt AI tools at all, though not how much they use them afterward (Choe et al., 2026 , Edmondson, 1999). AI rollouts pursued without attention to trust are associated with higher depression and disengagement among employees (Kim et al., 2025). Knowledge lost when experienced staff leave carries real, delayed costs that rarely show up on the spreadsheet used to justify the departure (Galan, 2023a, 2023b). Firms that deliberately design how work gets split between people and machines, and that build genuine two way mentoring across generations, outperform firms chasing automation for its own sake (Daugherty & Wilson, 2018 , Murphy, 2012). The paper closes with six propositions for future testing and a short diagnostic managers can run before their next AI deployment decision.

Keywords artificial intelligence, psychological safety , organizational trust , institutional memory, human–AI collaboration, workforce transformation, reverse mentoring, change management.

1. Introduction

The spread of artificial intelligence through professional work is not a trend to be monitored anymore. It is a condition managers now operate under. Employer surveys covering more than a thousand global companies project that 22% of today's jobs will be created or displaced by 2030, and that 39% of core workforce skills will change over the same stretch (World Economic Forum [WEF], 2025). Numbers like that describe a labour market in motion. They say almost nothing about what it feels like to be the person inside that motion and that gap, between the structural story and the human one, is what this paper is about. Management research already has a fairly mature vocabulary for AI's economic and technical dimensions. It has a much thinner one for the organizational behaviour side how fear, trust, memory, and cross generational learning determine whether AI adoption builds something real or just looks like progress on a dashboard.

None of this anxiety is new, exactly. Autor (2015) points out that every wave of automation has produced predictions of mass unemployment, and that every prior wave has, on net, created more jobs than it destroyed, even while displacing plenty of people and places along the way. Frey and Osborne's (2017) frequently cited estimate put roughly 47% of U.S. employment at high risk of computerization over the following couple of decades, a figure that anchored a decade of subsequent debate. What is different this time is speed and reach AI is now moving into judgment based cognitive work, not only routine manual tasks (WEF, 2025). That changes the question a manager has to ask. It is no longer just "which tasks can a machine do." It is whether the humans who remain the ones checking, correcting, and taking responsibility for what the machine produces have actually been set up to do that well.

This paper starts from a specific artifact a widely circulated practitioner book on AI era workforce transformation, written for executives, human resource leaders, and employees living through organizational AI adoption. The book proposes a set of terms of its own a "Fear Economy," an "AI Confidence Ladder," "Reverse Mentorship," "Wisdom Augmentation," a "Human Workforce Policy Minimum" that track, with more precision than one might expect, constructs already established in the peer reviewed literature, though usually studied one at a time and by non overlapping author communities. My goal here is not to review or promote that book. It is to use it as a structured artifact for spotting where practice and scholarship already agree, and then to see whether that agreement holds up into something closer to a testable management framework. Toward that end, the paper does three things. It pulls the fragmented research on AI adoption, psychological safety, trust, institutional memory, and human–AI collaboration into one model. It checks how far practitioner level claims actually converge with peer reviewed findings. And it ends with propositions and a diagnostic tool that scholars and practitioners can use to judge whether an organization is genuinely ready for this kind of change, as opposed to merely announcing that it is.

Section 2 reviews six clusters of literature. Section 3 lays out the method. Section 4 presents the findings, including the model itself and six propositions. Section 5 discusses what this means for managers and for theory, and the paper closes with limitations and a conclusion.

2. Literature Review

Six clusters of scholarship mark out the edges of this problem AI adoption and organizational psychology , psychological safety as a precondition for learning and adoption , organizational trust and its erosion during technological change , institutional memory and the economics of losing experience , human–AI collaborative intelligence and task allocation , and intergenerational learning through reverse mentorship. A final subsection asks what these six literatures leave unaddressed once you put them side by side.

2.1 AI Adoption and Organizational Psychology

Research on workplace AI adoption has moved through roughly two phases. The first was economic and task analytic which occupations and tasks were most exposed to automation (Frey & Osborne, 2017), and how anxiety about that exposure has recurred throughout technological history (Autor, 2015). The second, more recent phase, prompted by how visibly and quickly generative AI tools have shown up in ordinary knowledge work, looks instead at what employees actually experience. A bibliometric and thematic review of AI adoption studies from 2015 to 2024 found that AI reliably improves productivity, personalization, and inclusion metrics, but just as reliably raises questions about trust, ethics, and cultural fit, and that well being outcomes depend far more on how AI gets introduced than on whether it does (Frontiers in Artificial Intelligence, 2024). A related systematic review of workers' own accounts of AI driven change, covering 2020 to 2024, found meaningful variation by sector and by implementation context AI adoption is not one uniform event happening the same way everywhere, but something negotiated locally, workplace by workplace (ScienceDirect systematic review, 2025). Raisch and Krakowski (2021) give this tension a name the automation–augmentation paradox. Firms that reach for AI mainly to cut headcount get short term efficiency and stop there, missing the larger, compounding gains available when AI is used to sharpen human judgment instead. That claim lines up with Wilson and Daugherty's (2018) research across 1,500 firms, which found the biggest performance gains show up where people and AI work together not where automation is used chiefly as a cost lever.

2.2 Psychological Safety as a Precondition for AI Engagement

Psychological safety Edmondson's (1999) term for the shared belief that a team is safe for interpersonal risk taking has become one of the load bearing ideas in the AI adoption literature, and for reasonably good empirical reasons. Edmondson's early work found that higher performing teams often report more errors, not fewer, because safety makes it easier to surface a problem early

rather than hide it , she later extended the idea into a fuller account of organizational learning under uncertainty (Edmondson, 2019). Recent work applies this directly to AI. A study of 2,257 employees at a global consulting firm found psychological safety reliably predicted whether people adopted AI tools in the first place, but not how often or how long they kept using them once adoption happened a distinction the authors think practitioners routinely blur when they treat "adoption" and "mature use" as interchangeable (Choe et al., 2026). More troubling, a three wave study of 381 South Korean employees found AI adoption was linked to a drop in psychological safety, which in turn predicted higher depression, with ethical leadership softening the effect (Kim et al., 2025). Put together, these studies back up something the trade literature on AI change management tends to say in looser language psychological safety is not a nice to have during AI transformation. It is measurably tied to whether people adopt the tools, and to whether doing so costs them their mental health.

2.3 Organizational Trust and the Fear Economy

A related but separate literature treats organizational trust as infrastructure for technological change not sentiment, but something closer to plumbing you only notice when it fails. Trust formation research in AI enabled human resource contexts suggests employees' initial trust in an AI system has less to do with the system's accuracy than with how leadership handles it an experiment with 426 employees found that transparent explanation of the AI system and high trust in leadership each independently raised employees' initial trust and their willingness to adopt, with familiarity and collectivist culture adding further support (Xu et al., 2024). That is a useful reframe. AI trust turns out to be substantially a leadership and communication problem, not a purely technical one. Parallel research on job insecurity shows what happens when that trust never gets built in the first place. Brougham and Haar's (2018) STARA framework Smart Technology, Artificial Intelligence, Robotics, and Algorithms found that employees' awareness of these technologies predicted job insecurity, lower career satisfaction, and higher turnover intent across industries, one of the earliest solid links between AI awareness and negative employee outcomes. Newer studies extend this further AI related job insecurity mediates the relationship between AI awareness and both lower task performance and more deviant workplace behaviour, with career resilience only partly buffering the effect (PMC systematic study, 2025). This literature is, in effect, the empirical backbone of what practitioners describe more vividly as a "Fear Economy" a condition where surface level compliance with AI initiatives covers over anxiety nobody has addressed, and where the behaviours organizations most need during transformation, asking questions, admitting mistakes, experimenting in public, are exactly what fear shuts down.

2.4 Institutional Memory and the Economics of Replacing Experience

A different strand of research examines what happens when organizations respond to AI driven cost pressure by pushing experienced employees out faster. Galan's (2023a, 2023b) two part review of 91 empirical studies on knowledge loss from turnover, what the literature calls KLT, pulls together decades of work under a vocabulary that is almost darkly poetic organizational amnesia, corporate dementia, institutional forgetting. The phenomenon is real and measurable, but hard to see coming, because the knowledge at stake is usually tacit and relational invisible until its absence shows up as a client relationship nobody can rebuild, a quality problem that keeps recurring, or a strategic mistake a veteran would have caught. This body of work backs up a claim that sounds almost too tidy to be true until you look at the numbers behind it the cheapest employee is rarely the least expensive one. Swapping a senior, costly employee for a junior hire plus an AI tool can look correct on a static spreadsheet and still be badly wrong once you account for relationship rebuilding costs, the elevated error rate during the gap, and the slow erosion of judgment that used to prevent expensive mistakes before they happened.

2.5 Human–AI Collaborative Intelligence and Task Allocation

Against the automation first story sits a well established counter argument. Wilson and Daugherty's (2018) research, developed at length in their subsequent book (Daugherty & Wilson, 2018), found that companies automating mainly to shrink headcount got short term gains and then plateaued, while the largest and most durable improvements showed up where humans and AI complemented one another people training the models, explaining their outputs to stakeholders, staying accountable for decisions the AI helped make. Information systems scholars have since sharpened this distinction automation suits repetitive, well structured tasks with little need for human involvement, while augmentation, an actual collaboration between people and AI, fits complex, ambiguous, judgment heavy work better (Information Systems Frontiers review, 2025). The practical upshot is that thoughtful transformation requires an explicit sorting exercise which tasks are structured enough to hand to AI outright, which need human context and ethical judgment, and which need a deliberately designed handoff between the two.

2.6 Intergenerational Learning and Reverse Mentorship

Last, there is a literature on how organizations move capability in both directions across generations during fast technological change. Reverse mentoring pairing a technically fluent junior employee with an experienced senior one in a genuinely two way teaching relationship goes back at least to General Electric's late 1990s practice of having executives learn the internet from junior staff, and has been studied since as a mechanism for cross generational learning and leadership development (Murphy, 2012). Murphy's model identifies structured matching, real reciprocity, and organizational backing as the conditions that make reverse mentoring work rather than feel awkward or breed resentment, a caveat later studies confirm, noting that senior resistance and

unclear reciprocity are the most common ways these programs fail. Read alongside the psychological safety and trust literatures above, reverse mentorship looks less like a nice HR program and more like a structural lever it builds AI fluency in experienced staff, gives real standing to junior employees whose skills would otherwise go underused, and creates exactly the kind of visible, senior modelled vulnerability that Edmondson (1999, 2019) argues drives psychological safety at scale.

2.7 Synthesis and Research Gap

Read separately, each of these six literatures tells only part of the story. Adoption research explains what is changing. Psychological safety explains who engages with the change. Trust explains why engagement collapses. Institutional memory explains what gets lost when transformation moves carelessly. Collaborative intelligence explains how tasks should be redistributed. Reverse mentorship explains a mechanism for moving capability across generations. What is mostly missing from the scholarly record is a model that treats these as parts of one system rather than six separate conversations that rarely cite one another. Practitioner writing, including the book this paper draws on, has already started stitching these together informally, with frameworks a Fear Economy, a Confidence Ladder, a Reverse Mentorship Matrix that map onto the peer reviewed literature above more precisely than one might expect. That convergence, largely undocumented in the journals, motivates the questions driving this study

- RQ1 How far do practitioner level constructs of AI era workforce transformation actually converge with peer reviewed findings across the psychological safety, trust, institutional memory, and collaborative intelligence literatures?
- RQ2 Can these constructs be combined into one internally coherent management framework capable of generating testable propositions?
- RQ3 What diagnostic indicators can practitioners use to gauge whether an organization is ready for psychologically safe, trust preserving AI adoption?

3. Methodology

3.1 Research Design

I use a qualitative, interpretive integrative review here (Torraco, 2005), a design suited to questions that call for pulling a scattered, multi disciplinary literature into a new framework rather than testing one hypothesis against one new dataset. Integrative reviews differ from systematic reviews in purpose a systematic review tries to catalogue and, ideally, statistically pool findings on a narrowly defined question, while an integrative review tries to critically combine literatures that are conceptually related but methodologically different, in service of building new theory

(Torraco, 2005 , Snyder, 2019). That fit matters here because the six literatures reviewed in Section 2 use different methods entirely large sample surveys, experiments, systematic reviews, bibliometrics, qualitative case work and have rarely been theoretically combined. A quantitative meta analysis was not feasible, or honestly appropriate, at this stage of theory development.

3.2 Data Sources and Sampling

Two sources of data feed this analysis. The first is a purposive set of peer reviewed and high credibility gray literature sources systematic reviews, empirical studies, major institutional reports located through structured searches of Google Scholar, PubMed, ScienceDirect, Emerald Insight, Frontiers, and the World Economic Forum's publication archive, using combinations of terms including "artificial intelligence," "workplace," "psychological safety," "trust," "knowledge loss," "institutional memory," "reverse mentoring," and "human–AI collaboration." Sources were kept if they were peer reviewed, published in or after 2015 (with a small number of earlier foundational pieces retained for grounding, notably Edmondson, 1999), and spoke directly to one of the six clusters. Sixty nine sources were screened , twenty made it into the final citation list, chosen for relevance, methodological clarity, and how directly they bore on the framework being built.

The second source is a single, purposively chosen practitioner text a contemporary field guide on AI era workforce transformation written for executive, human resource, and employee audiences, organized around sixteen chapters and a set of named frameworks (the Fear Economy, the AI Confidence Ladder, the Reverse Mentorship Matrix, the Human Workforce Policy Minimum). I do not treat this text as an empirical data source in its own right. I treat it the way one might treat an unusually rich set of field notes, or a single information dense case (Patton, 2002) a place to extract practitioner level constructs and then test whether they hold up against the peer reviewed literature. Choosing one information rich case for the depth it offers, rather than for statistical representativeness, follows standard qualitative case logic in management research (Patton, 2002 , Eisenhardt, 1989).

3.3 Analytical Procedure

Both sources were coded using directed thematic analysis (Hsieh & Shannon, 2005). The initial coding frame came deductively from the six clusters in Section 2. I then went chapter by chapter through the practitioner text, mapping each named construct, the Fear Economy, say, or the AI Confidence Ladder, onto its closest peer reviewed counterpart. Where a construct had no obvious scholarly analogue, it was kept as an emergent code and flagged for discussion, either as a possible contribution or as something needing empirical testing later. Constructs appearing independently across three or more sources, in either data stream, became core dimensions of the resulting framework , constructs appearing in only one or two sources were kept as secondary or illustrative

themes. That threshold follows standard practice for establishing thematic saturation in qualitative synthesis (Braun & Clarke, 2006).

3.4 Trustworthiness and Limitations of the Method

A few steps were taken to keep this analysis honest. Every core construct in the final framework had to be backed by at least one peer reviewed source independent of the practitioner text, so the model is not simply a restatement of one author's opinions with citations attached after the fact. A running log of screening and coding decisions was kept to make the process auditable. I also went looking, deliberately, for literature that would undercut the "automation first is risky" argument, a negative case check in the usual qualitative sense, and where counter evidence turned up, it is reported in Section 4 rather than quietly dropped. The main limitation of a design like this is that it produces no new primary data and cannot, by itself, establish causation between the constructs , Section 6 treats the propositions that follow explicitly as hypotheses for future research rather than as settled findings.

4. Findings and Analysis

4.1 Convergence Between Practitioner Constructs and the Peer Reviewed Literature

The thematic analysis found substantial overlap between the practitioner text's named frameworks and independent findings already sitting in the peer reviewed literature. Table 1 lays this out across six constructs.

Practitioner Construct	Core Claim	Corroborating Peer Reviewed Evidence	Convergence Strength
Fear Economy	Undisclosed job security anxiety produces surface compliance and suppresses experimentation with AI tools.	Brougham & Haar (2018) , AI related job insecurity mediation studies (PMC, 2025)	Strong
AI Confidence Ladder	Literacy and confidence are distinct constructs , most programs stop at literacy.	Choe et al. (2026) psychological safety predicts adoption, not usage depth	Strong
Collapse of Trust	Trust functions as operating infrastructure , its loss undermines AI rollout regardless of tool quality.	Xu et al. (2024) , Kim et al. (2025)	Strong
Death of Institutional Memory	Removing experienced staff before AI	Galan (2023a, 2023b)	Strong

	deployment destroys undocumented, tacit organizational knowledge.		
Collaborative Intelligence	Automation first strategies yield only short term gains , augmentation yields durable performance improvement.	Wilson & Daugherty (2018) , Raisch & Krakowski (2021)	Strong
Reverse Mentorship	Structured bidirectional teaching between junior and senior employees builds both AI fluency and psychological safety.	Murphy (2012)	Moderate–Strong

4.2 The Human & AI Integration Model (HAIM)

Pulling together the constructs from Table 1, I propose the Human & AI Integration Model, or HAIM five interdependent dimensions that, together, appear to determine whether AI adoption builds real capability or produces compliance that merely looks good in a slide deck. The five are Psychological Safety, the base condition without which people will not experiment, disclose AI errors, or ask honest questions (Edmondson, 1999, 2019 , Choe et al., 2026) , Trust Architecture, the visible match between what an organization says and what it does during rollout (Xu et al., 2024 , Kim et al., 2025) , Institutional Memory Preservation, deliberately mapping tacit, experience based knowledge before any AI driven role redesign happens (Galan, 2023a, 2023b) , Collaborative Task Allocation, explicitly sorting tasks by structure and judgment intensity to decide who, or what, should own them (Wilson & Daugherty, 2018 , Raisch & Krakowski, 2021) , and Structured Intergenerational Learning, reciprocal reverse mentorship pairings that build AI fluency in experienced staff and organizational standing in junior staff at the same time (Murphy, 2012).

One thing stood out in the analysis these five dimensions do not function as independent levers you can pull in any order. They form something closer to a sequence. Trust Architecture and Psychological Safety act as upstream conditions without them, everything downstream, task redesign, mentorship programs, whatever it might be, tends to get experienced as performance rather than substance. That tracks with the practitioner observation that "a workforce can be technologically trained and still psychologically broken," and with Kim and colleagues' (2025) finding that AI adoption pursued without attention to safety correlates with more depression, not more capability. Institutional Memory Preservation, meanwhile, has to come before Collaborative Task Allocation makes much sense at all an organization cannot rationally decide which tasks require human judgment if it has not first figured out what judgment its experienced people

actually hold (Galan, 2023a). And Structured Intergenerational Learning turns out to be the mechanism that operationalizes all four of the others at once, because a reverse mentorship pairing needs psychological safety to function, is itself a trust signal, creates a natural setting for capturing institutional memory, and implements collaborative task allocation at the level of one person's actual work.

Figure 1. The Human & AI Integration Model (HAIM)

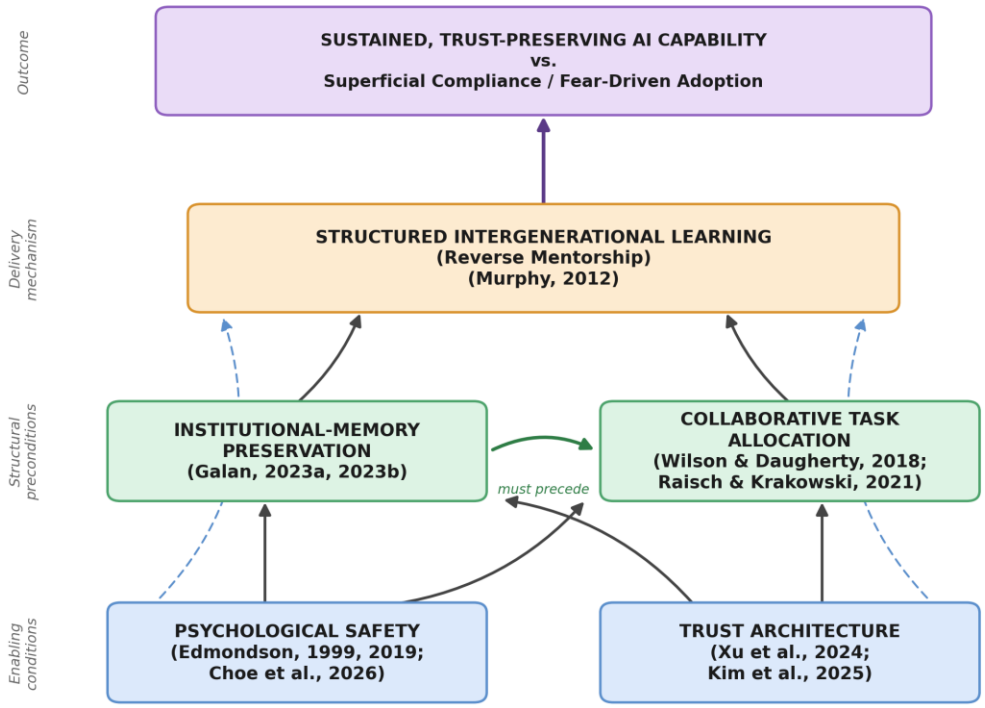

Figure 1. The Human & AI Integration Model (HAIM), showing the proposed causal ordering from enabling conditions through to sustained AI enabled organizational capability.

Figure 1 lays this ordering out visually. Psychological Safety and Trust Architecture sit at the base as enabling conditions , Institutional Memory Preservation and Collaborative Task Allocation sit above them as structural preconditions, with memory preservation feeding directly into task allocation, not the other way around , Structured Intergenerational Learning sits above both as the mechanism that puts all four into practice at once , and the whole structure feeds into the outcome that actually matters to a management team deciding whether its AI investment is working sustained, trust preserving capability on one side, or fear driven, superficial compliance on the other.

4.3 Research Propositions

Six propositions follow from this synthesis, offered here as a research agenda rather than as findings already established

- P1 Psychological safety will predict initial AI tool adoption but not sustained usage intensity, replicating and extending Choe et al. (2026) across industry and national contexts.
- P2 Ethical, transparent leadership will moderate the relationship between AI adoption intensity and employee well being, such that adoption under strong ethical leadership will not predict increased depression or disengagement, while adoption under weak ethical leadership will (extending Kim et al., 2025).
- P3 Organizations that run a structured "wisdom audit," mapping tacit institutional knowledge before AI driven role redesign, will show lower post restructuring error rates and less client relationship attrition than organizations that skip this step.
- P4 Organizations that explicitly classify tasks by structure and judgment intensity before deploying AI will realize larger, more durable productivity gains than organizations pursuing automation first, headcount reduction strategies, consistent with Wilson and Daugherty (2018).
- P5 Structured, reciprocal reverse mentorship programs will be associated with both senior employees' progress up the AI confidence ladder and junior employees' perceived organizational standing, more so than unstructured or one directional digital skills training.
- P6 The five HAIM dimensions will show incremental predictive validity for successful AI transformation outcomes sustained usage, retained institutional knowledge, workforce trust metrics beyond what any single dimension predicts alone.

4.4 A Managerial Diagnostic The HAIM Readiness Audit

To make the framework usable rather than merely descriptive, the analysis yields a seven item diagnostic that management teams can run before a major AI deployment decision, drawn directly from convergent themes in the literature and the practitioner artifact. Have we mapped what our most experienced employees know that exists nowhere in a documented system? Have we calculated the full, multi year cost of replacing experienced roles, including relationship rebuilding, error rate, and knowledge erosion costs, rather than only the immediate payroll saving? Do we have a genuine, resourced transition and reskilling pathway for employees in affected roles, at a scale consistent with what the Future of Jobs Report 2025 recommends (WEF, 2025)? Have we communicated honestly, including about what remains uncertain, on how AI will affect roles

and evaluation? Is there a safe, formal channel for employees to raise concerns about AI systems without career risk? Are we measuring confidence, trust, and knowledge retention, or only tool adoption and usage frequency numbers? And have we built at least one structured, reciprocal reverse mentorship mechanism connecting AI fluent and experience rich employees? Organizations answering no to three or more of these should expect, on the evidence reviewed here, a meaningfully higher risk of the fear driven, trust eroding adoption pattern documented throughout this paper.

4.5 Boundary Conditions and Disconfirming Evidence

Consistent with the negative case procedure described in Section 3.4, the analysis also turned up boundary conditions that qualify the framework rather than support it unconditionally. Choe et al.'s (2026) finding that psychological safety predicts adoption but not usage intensity suggests safety alone is not sufficient for deep organizational AI capability , sustained usage seems to depend on other factors too, task relevance, workload capacity, perceived output quality, that sit outside HAIM's immediate scope and deserve separate study. Sectoral differences are substantial as well employees in research oriented or creative roles report lower AI related threat perception than employees in routine, transaction heavy roles like hospitality and retail (PMC systematic study, 2025), which implies the five HAIM dimensions probably need different weighting by occupation rather than a single uniform application across an organization. And national or cultural context appears to moderate trust formation specifically Xu et al.'s (2024) study, run in a Chinese organizational setting, found organizational collectivism strengthened the path from leadership trust to AI trust, a moderator that may not travel unchanged into more individualist contexts and that future cross cultural work should test directly rather than assume.

5. Discussion and Implications

5.1 Theoretical Implications

This paper contributes to management theory in three ways. First, it shows that constructs usually studied in separate literatures organizational behaviour’s psychological safety, human resource management's trust formation, knowledge management's institutional memory loss, information systems' human–AI task allocation are empirically and causally tangled up in a single organizational phenomenon AI era workforce transformation. Treating them as one system, the way HAIM does, gives a fuller account of why some organizations end up with genuine AI enabled capability while others end up with tool adoption and not much else. Second, the analysis extends Raisch and Krakowski's (2021) automation–augmentation paradox by naming the organizational behaviour mechanisms, psychological safety and trust specifically, through which augmentation oriented strategies turn into durable performance gains, rather than treating augmentation as a

purely strategic choice divorced from how people experience it. Third, cross checking a contemporary practitioner framework against peer reviewed evidence offers a template other researchers could use to translate fast moving practitioner discourse into something closer to cumulative scholarly knowledge, a translation problem that matters more now than usual, given how far ahead of the journals AI management practice is currently running.

5.2 Managerial Implications

For practicing managers, the main implication is about sequence, not additional cost. Nothing in the evidence reviewed here argues for adopting AI more slowly in some absolute sense , the competitive and labour market pressures WEF (2025) documents make delay itself a real risk. What it argues for is sequencing build trust oriented communication and psychologically safe leadership behaviour before or alongside large scale AI deployment, not after it , map institutional knowledge before redesigning roles, not after workforce reductions have already happened , and make task allocation decisions explicitly, through a structured classification exercise, rather than implicitly, on the basis of whichever roles look cheapest to automate on a spreadsheet that stops at this quarter. Human resource and learning and development teams in particular might want to reconsider measuring AI maturity purely by tool adoption rates , the evidence here (Choe et al., 2026) suggests adoption and confident, sustained, high quality use are genuinely different outcomes and need different interventions.

5.3 Policy Implications

At the policy level, the overlap between the practitioner "Human Workforce Policy Minimum" idea and WEF's (2025) finding that most of the global workforce will need reskilling or upskilling by 2030, with a meaningful share unlikely to receive it, suggests voluntary, firm level good practice is probably not enough on its own at scale. The gap between the 92% of surveyed employers who say they intend to prioritize reskilling (WEF, 2025) and the documented tendency of low trust environments to turn training investment into performance theatre rather than real capability (Kim et al., 2025) points to a measurement and enforcement gap worth policymakers' attention intending to invest in reskilling is not the same thing as creating the psychologically safe conditions under which that investment actually gets absorbed.

6. Limitations and Future Research

This study has four main limitations. As an integrative qualitative synthesis, it produces no new primary quantitative data and cannot establish causal direction among the five HAIM dimensions , the six propositions in Section 4.3 should be read as a research agenda, not as established findings. The practitioner artifact used here, while information rich, is one author's synthesis and may reflect whatever selection effects shaped which examples got included , future work should repeat this

convergence analysis across a larger, more varied set of practitioner texts. The peer reviewed evidence itself is also unevenly distributed across the six literatures psychological safety and trust formation are comparatively well studied with large sample quantitative work, while the intersection of reverse mentorship with AI specific capability building is still thin and would benefit from dedicated longitudinal field studies. And national context, sector, and organizational size all look like plausible moderators of several HAIM relationships but could not be systematically tested within a review of this kind , multi country, multi sector survey research explicitly testing the six propositions, ideally with a longitudinal or time lagged design similar to Kim et al. (2025), is probably the highest priority next step. It would also be worth testing whether the HAIM Readiness Audit in Section 4.4 actually predicts downstream outcomes, AI project success rates, voluntary turnover among experienced staff, workforce reported trust, rather than just organizing them conceptually.

7. Conclusion

Artificial intelligence will keep changing what organizations produce and how they produce it , on that point the economic and labour market literature is fairly settled (Autor, 2015 , Frey & Osborne, 2017 , WEF, 2025). What is still genuinely open, and what this paper has tried to address, is whether that change gets managed in a way that preserves the trust, memory, and judgment organizations depend on, or in a way that discards them under short term cost pressure. The evidence pulled together here, on psychological safety (Edmondson, 1999, 2019 , Choe et al., 2026), organizational trust (Xu et al., 2024 , Kim et al., 2025), institutional memory (Galan, 2023a, 2023b), collaborative intelligence (Wilson & Daugherty, 2018 , Raisch & Krakowski, 2021), and intergenerational learning (Murphy, 2012), converges on a fairly consistent conclusion organizations that treat AI adoption as a narrowly technical or cost accounting exercise tend to underestimate its human costs, while organizations that treat it as a human systems design problem, putting trust and psychological safety ahead of automation and pairing institutional memory capture with deliberate task reallocation, are better positioned to realize AI's larger, more durable benefits. The Human & AI Integration Model proposed here gives management scholars a structured, testable way to investigate that claim further, and gives practicing managers a diagnostic starting point for a transformation that, handled carelessly, risks becoming the most sophisticated replacement engine in organizational history, and, handled well, could instead become one of the more significant expansions of human organizational capability in a generation.

References

Autor, D. H. (2015). The history of technological anxiety and the future of economic growth Is this time different? Journal of Economic Perspectives, 29(3), 3–30. https //doi.org/10.1257/jep.29.3.3

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https //doi.org/10.1191/1478088706qp063oa

Brougham, D., & Haar, J. (2018). Smart technology, artificial intelligence, robotics, and algorithms (STARA) Employees' perceptions of our future workplace. Journal of Management & Organization, 24(2), 239–257. https //doi.org/10.1017/jmo.2016.55

Choe, A., et al. (2026). Safety first Psychological safety as the key to AI transformation [Preprint]. arXiv. https //arxiv.org/abs/2602.23279

Daugherty, P. R., & Wilson, H. J. (2018). Human + machine Reimagining work in the age of AI. Harvard Business Review Press.

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

Edmondson, A. C. (2019). The fearless organization Creating psychological safety in the workplace for learning, innovation, and growth. Wiley.

Eisenhardt, K. M. (1989). Building theories from case study research. Academy of Management Review, 14(4), 532–550. https //doi.org/10.2307/258557

Frey, C. B., & Osborne, M. A. (2017). The future of employment How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280. https //doi.org/10.1016/j.techfore.2016.08.019

Frontiers in Artificial Intelligence. (2024). Exploring how AI adoption in the workplace affects employees A bibliometric and systematic review. Frontiers in Artificial Intelligence, 7, Article 1473872. https //doi.org/10.3389/frai.2024.1473872

Galan, N. (2023a). Knowledge loss induced by organizational member turnover A review of empirical literature, synthesis, and future research directions (Part I). The Learning Organization, 30(2), 117–136. https //doi.org/10.1108/TLO 09 2022 0107

Galan, N. (2023b). Knowledge loss induced by organizational member turnover A review of empirical literature, synthesis and future research directions (Part II). The Learning Organization, 30(2), 137–156. https //doi.org/10.1108/TLO 09 2022 0108

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative Health Research, 15(9), 1277–1288. https //doi.org/10.1177/1049732305276687

Kim, B. J., Kim, M. J., & Lee, J. (2025). The dark side of artificial intelligence adoption Linking artificial intelligence adoption to employee depression via psychological safety and ethical leadership. Humanities and Social Sciences Communications, 12. https //doi.org/10.1057/s41599 025 05040 2

Murphy, W. M. (2012). Reverse mentoring at work Fostering cross generational learning and developing millennial leaders. Human Resource Management, 51(4), 549–573. https //doi.org/10.1002/hrm.21489

Patton, M. Q. (2002). Qualitative research and evaluation methods (3rd ed.). Sage Publications.

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

Snyder, H. (2019). Literature review as a research methodology An overview and guidelines. Journal of Business Research, 104, 333–339. https //doi.org/10.1016/j.jbusres.2019.07.039

Torraco, R. J. (2005). Writing integrative literature reviews Guidelines and examples. Human Resource Development Review, 4(3), 356–367. https //doi.org/10.1177/1534484305278283

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence Humans and AI are joining forces. Harvard Business Review, 96(4), 114–123.

World Economic Forum. (2025). The future of jobs report 2025. https //www.weforum.org/publications/the future of jobs report 2025/

Xu, S., et al. (2024). How do employees form initial trust in artificial intelligence Hard to explain but leaders help. Asia Pacific Journal of Human Resources. https //doi.org/10.1111/1744 7941.12402