



Human-centred artificial intelligence and organisational well-being: An integrative framework for responsible technology management

Dr. Ignacio Bonasa Alzuria
University of Happiness and Well-Being, Mexico
Liderarte Talent Factory, Spain
ibonasa@hotmail.com

Corresponding author: Dr. Ignacio Bonasa Alzuria
DOI Link: <https://doi.org/10.64971/j.cph.eijtem.v13.i3.3.2026>

Abstract

Artificial intelligence is increasingly embedded in organisational decisions, workflows and employee interactions, yet technical performance and productivity gains do not necessarily improve the quality of work. This conceptual study develops the Human-Centred Artificial Intelligence for Organisational Well-Being (HCAI-WB) framework and explains the mechanisms through which AI configurations can support or erode sustainable human performance. A structured integrative review synthesises research on human-centred AI, work design, the job demands-resources model, meaningful work, psychological safety, procedural justice and responsible AI governance. The synthesis identifies six interdependent dimensions: purpose and human agency; work design and cognitive load; trust, fairness and data dignity; capabilities and AI literacy; social connection and psychological safety; and learning, measurement and accountability. Eight falsifiable propositions connect these dimensions to engagement, exhaustion, meaningfulness, trust, innovation and long-term performance. The article contributes a six-stage implementation architecture, non-compensatory stage gates, a multidimensional indicator dashboard, a five-level maturity model and an empirical validation roadmap. Its distinctive contribution is to bridge model-centred governance with the lived organisation of work: an AI system is not considered human-centred merely because a person remains nominally in the loop, but because that person retains intelligible authority, adequate time, capability, voice and access to review. The framework offers managers and researchers a testable architecture for aligning responsible technology management with organisational well-being.

Keywords:

Artificial intelligence; organisational well-being; human-centred AI; work design; responsible innovation; AI governance; employee experience; sustainable performance.

1. Introduction

Artificial intelligence (AI) has moved from a specialist technology to a general organisational infrastructure. It now assists writing, coding, customer service, recruitment, scheduling, forecasting, performance monitoring and strategic decision-making. Evidence from controlled and field studies shows that generative AI can increase speed and quality in selected knowledge tasks, often with larger gains for less experienced workers (Brynjolfsson et al., 2025; Noy & Zhang, 2023). At the same time, algorithmic management can intensify control, reduce discretion and obscure the grounds on which decisions are made (Kellogg et al., 2020). The same technological family can therefore function as a resource or a demand, an instrument of empowerment or surveillance, and a source of learning or insecurity.

This duality creates a management problem rather than a purely technical problem. Organisations frequently assess AI through operational indicators such as accuracy, throughput, cost, response time and return on investment. These indicators matter, but they are incomplete. A system can meet a technical target while degrading autonomy, increasing cognitive fragmentation, displacing social contact or creating uncertainty about roles and careers. Conversely, a carefully

designed system can remove repetitive work, improve safety, support learning and free attention for judgement, empathy and creativity. Recent studies suggest that employee well-being is shaped less by the mere presence of AI than by the work conditions through which AI is introduced, including task optimisation, security, efficacy, engagement and technostress (Chuang et al., 2025; Valtonen et al., 2025).

The gap addressed in this article is the limited integration between responsible AI governance and organisational well-being. Existing governance frameworks emphasise validity, safety, accountability, transparency, privacy and fairness. Work psychology emphasises autonomy, meaningfulness, social support, manageable demands and opportunities for development. Both traditions are necessary, yet they are often implemented by different professional communities using different vocabularies and metrics. As a result, ethics may remain detached from work design, while well-being initiatives may remain detached from technology architecture.

The purpose of this article is to develop an integrative framework that connects responsible AI governance with organisational well-being and translates both into decisions about work design. It addresses three questions: (1) Which organisational conditions enable AI to support rather than erode well-being? (2) Through which psychological and work-system mechanisms do those conditions influence sustainable performance? and (3) How can organisations assess implementation quality without allowing strong technical results to conceal serious human harm?

The article makes four contributions. First, it changes the unit of analysis from the AI tool to the configured work system in which tasks, interfaces, incentives, data practices, decision rights and relationships interact. Second, it specifies a multilevel mechanism chain linking design choices to job demands and resources, psychological needs, employee behaviour and organisational outcomes. Third, it introduces a non-compensation principle: critical failures involving dignity, discrimination, unsafe workload or absent appeal cannot be averaged away by efficiency gains. Fourth, it converts the synthesis into propositions, implementation gates, indicators and a validation agenda. The resulting HCAI-WB framework complements technical risk frameworks and occupational well-being models without duplicating either.

2. Literature review

The literature relevant to human-centred AI at work is broad and fragmented. This section reviews five streams that are particularly important for theory building: automation and augmentation, work design, employee well-being, human-AI interaction and responsible governance. Their integration reveals that positive outcomes depend on sociotechnical choices rather than technological capability alone.

2.1 Automation, augmentation and the redistribution of expertise

Management discussions often frame AI as a choice between replacing labour and assisting labour. Raisch and Krakowski (2021) argue that automation and augmentation are interdependent rather than mutually exclusive. Automation may remove routine subtasks while augmentation reshapes the remaining human role. This redistribution matters because task removal can either enrich or impoverish work. If AI absorbs administrative burden and leaves employees with greater discretion, learning and relational responsibility, the new role may become more meaningful. If it removes the coherent core of an occupation and leaves workers to monitor exceptions, correct errors or carry accountability without authority, the role may become more fragmented.

Empirical evidence also shows heterogeneous effects. In customer support, generative AI improved average productivity and appeared to diffuse expert practices to less experienced agents (Brynjolfsson et al., 2025). In professional writing, access to generative AI reduced completion time and improved rated quality (Noy & Zhang, 2023). These findings are important, but they do not imply that all uses create sustainable value. The transfer of tacit knowledge may support development when employees understand and interrogate suggestions; it may create dependency when outputs are accepted without reflection. Productivity benefits must therefore be interpreted alongside capability formation, role clarity and the distribution of judgement. These results also support a symbiotic view in which value depends on the allocation of judgement between people and systems rather than on automation alone (Jarrahi, 2018).

2.2 Work design as the missing transmission mechanism

Work design explains how technology becomes a lived experience. Parker and Grote (2022) show that digital technologies affect task, knowledge, social and contextual characteristics of work. The job demands-resources model similarly proposes that job resources support motivation and engagement, whereas excessive demands increase strain and exhaustion (Bakker & Demerouti, 2007). AI can operate on both sides of this model. It may provide cognitive support, faster access to information and safer processes, while also generating interruption, surveillance, ambiguity, pace pressure and learning demands.

Recent evidence supports this dual-pathway view. Chuang et al. (2025) found that AI-related resources and demands have different effects on productivity, satisfaction, exhaustion and work-family conflict. Valtonen et al. (2025) found that AI adoption influenced well-being indirectly through task optimisation and safety rather than through adoption itself. A recent systematic review similarly shows that worker attitudes and experiences vary with use context and organisational conditions (Savolainen et al., 2026). These findings suggest that the unit of analysis should not be the technology in isolation. The more useful unit is the configured work system: tasks, decision rights, interfaces, feedback, incentives, data practices, learning support and social relationships.

2.3 Meaningful work, agency and psychological safety

Well-being at work is more than the absence of distress. It includes vitality, belonging, growth, fairness, purpose and the capacity to influence one's environment. Job-quality research also indicates that work characteristics are central to general well-being (Green et al., 2024). Self-determination theory identifies autonomy, competence and relatedness as basic psychological needs that support internalised motivation (Deci & Ryan, 2000). Meaningful work research similarly highlights significance, self-realisation, contribution and coherence (Rosso et al., 2010). AI adoption can affect each of these needs. It can increase competence by providing feedback, reduce autonomy through prescriptive recommendations, strengthen contribution by improving service quality, or weaken coherence when employees no longer understand how their work connects to outcomes.

Psychological safety is equally important because AI systems are fallible and context-dependent. Employees must be able to question outputs, report failure modes and challenge high-stakes decisions without fear of punishment (Edmondson, 1999). Where the system is treated as unquestionable, errors may be hidden and human expertise may atrophy. Where dissent is invited and documented, human judgement becomes part of the control environment. Employee voice is therefore not a cultural accessory; it is a mechanism of technological reliability.

2.4 Human-AI interaction and calibrated trust

Human-AI interaction research emphasises expectation setting, intelligibility, correction, feedback and recovery from errors. Amershi et al. (2019) propose design guidelines that include making clear what the system can do, showing contextually relevant information, supporting efficient correction and learning from user behaviour. These principles are especially relevant in organisations because trust is shaped not only by interface quality but also by power. An employee may comply with an opaque recommendation because it is embedded in a performance system, even when personal confidence is low. This emphasis is consistent with human-centred AI principles that combine high levels of human control with reliable automation (Shneiderman, 2020).

The objective is calibrated trust: neither blind acceptance nor reflexive rejection. Calibrated trust requires adequate explanations, visible uncertainty, contestability, role-specific training and clarity about accountability. It also requires data dignity—the recognition that employee data represent people whose privacy, reputation and opportunities may be affected. Technical explainability is insufficient if workers cannot appeal, obtain human review or understand how data are used.

2.5 Responsible AI governance and the well-being gap

The NIST AI Risk Management Framework organises responsible practice around governance, mapping, measurement and management, while UNESCO and the OECD emphasise human rights, dignity, human oversight, fairness and well-being. These frameworks provide an essential foundation, but organisational implementation often concentrates on legal, security and model risks. The quality of work may receive less attention because it falls across the boundaries of technology, human resources, operations and occupational health.

This separation creates two blind spots. First, a technically compliant system may still create harmful work design. Second, a well-being programme may focus on individual coping while ignoring the digital architecture that creates demands. A human-centred approach must connect system characteristics to employee experience and organisational outcomes.

2.6 The unresolved integration problem

Across these literatures, the principal unresolved issue is not a shortage of ethical principles or well-being variables, but the absence of a shared causal architecture. Responsible AI frameworks typically begin with systems and risks; work psychology begins with jobs, relationships and psychological experience. Neither perspective alone explains why the same technical capability can create learning and autonomy in one setting, but surveillance, fragmentation and insecurity in another.

The synthesis therefore treats workplace AI as a multilevel sociotechnical configuration. Design and governance choices influence proximal work conditions, including discretion, pace, cognitive load, transparency, data exposure, skill use and social contact. These conditions activate psychological

mechanisms such as autonomy, competence, relatedness, meaningfulness, justice and safety, which then influence behavioural and organisational outcomes. Sectoral risk, task interdependence, employment security, leadership behaviour and worker representation are treated as contextual moderators rather than background details.

3. Materials and method: integrative theory-building review

This study uses a structured integrative review and conceptual theory-building design. Integrative reviews are appropriate when a problem spans mature and emerging literatures and the objective is to combine theoretical perspectives, expose tensions and generate a new framework rather than estimate a pooled effect (Snyder, 2019; Torraco, 2005). The review is deliberately problem-driven and does not present itself as an exhaustive systematic review.

Source assembly combined foundational theories with recent peer-reviewed studies and authoritative governance standards. Eligible sources had to contribute at least one of four elements: evidence about AI-enabled work outcomes; a mechanism explaining job quality or employee well-being; a construct relevant to human-AI interaction; or an actionable governance requirement. Sources were excluded when they treated adoption as an undifferentiated binary variable, offered unsupported technological claims, or could not clarify a mechanism, boundary condition or design implication.

Analysis proceeded through five iterative stages. First, reported benefits and harms were coded as proximal work-system changes rather than final outcomes. Second, those changes were mapped to job demands and resources and to autonomy, competence, relatedness, meaningfulness, procedural justice and psychological safety. Third, overlapping concepts were consolidated into higher-order dimensions while preserving discriminant boundaries. Fourth, negative configurations were examined to identify conditions under which apparently beneficial practices could fail. Fifth, the dimensions were translated into propositions, implementation practices, observable indicators and falsification criteria.

Four quality checks guided theory construction: conceptual parsimony, traceability to established literature, non-redundancy among dimensions and practical observability. The framework also had to explain divergent outcomes from similar technologies and specify evidence that could disconfirm its claims. Because no formal meta-analysis or exhaustive database search was conducted, the contribution should be judged as a transparent, falsifiable synthesis whose constructs and causal relations require empirical validation.

4. Results: the HCAI-WB framework

The HCAI-WB framework proposes that sustainable human performance emerges when six dimensions are designed as a system. None is sufficient by itself. A transparent interface cannot compensate for coercive use; training cannot compensate for excessive surveillance; and a well-being survey cannot compensate for absent appeal mechanisms. The dimensions are connected by governance, participation and continuous dialogue.

Its causal logic follows four linked layers: AI and governance characteristics shape the organisation of work; work-system conditions alter resources, demands and perceived justice; these mechanisms influence agency, capability, trust, connection and meaning; and these experiences affect engagement, exhaustion, learning, innovation, retention and service quality. Feedback loops are expected because employee use, resistance, adaptation and incident reporting subsequently change the system. The framework is therefore dynamic and multilevel rather than a static checklist.

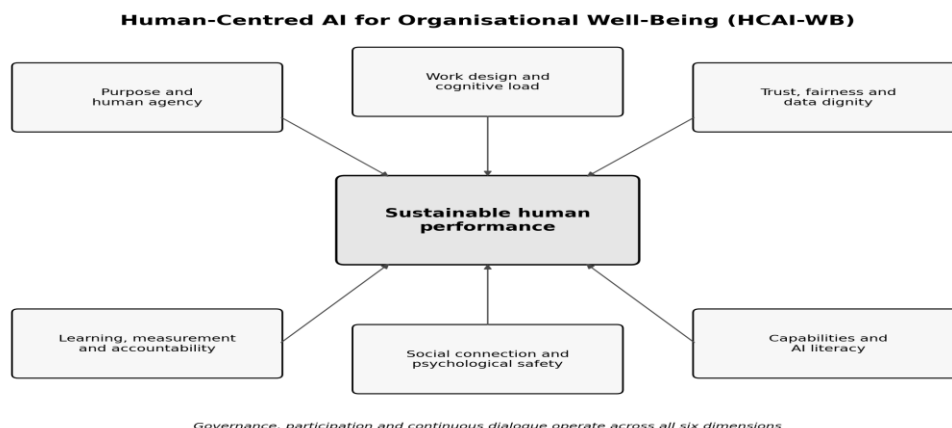


Figure 1. The Human-Centred AI for Organisational Well-Being framework. Source: author's conceptual development.

Table 1. AI mechanisms and organisational well-being pathways

AI mechanism	Potential resource pathway	Potential demand pathway	Design question
Task automation	Reduces repetitive work and frees time for judgement	Removes coherent tasks and increases exception monitoring	What work disappears, what remains, and does the remaining role become richer?
Generative assistance	Provides ideas, drafts and rapid access to knowledge	Creates dependency, verification burden and output homogenisation	How are outputs checked, contextualised and challenged?
Algorithmic recommendations	Supports consistency and decision speed	Obscures reasoning and shifts accountability without authority	Who can override, appeal or request human review?
Performance analytics	Makes patterns visible and supports coaching	Enables surveillance, gaming and loss of trust	Which data are necessary, who sees them and for what purpose?
AI-enabled scheduling	Improves matching and operational coordination	Reduces predictability and worker control over time	Are preferences, constraints and recovery needs represented?
Conversational agents	Expands access and immediate support	Displaces human contact and may simulate empathy without responsibility	Which interactions require a real human relationship?

4.1 Purpose and human agency

The first dimension asks why AI is being introduced and which human capabilities it should enlarge. A human-centred purpose is expressed in outcome terms—safer work, better decisions, more accessible service, reduced administrative burden or stronger learning—rather than in adoption metrics alone. Human agency requires meaningful choice, the ability to override or escalate, and clarity about where human judgement is mandatory. Agency does not mean that every recommendation is optional; it means that decision rights are explicit, proportionate to risk and supported by accountable review.

Proposition 1. The positive relationship between AI use and employee engagement will be stronger when employees retain meaningful decision latitude and understand the human purpose of the system.

4.2 Work design and cognitive load

The second dimension concerns the architecture of tasks. Managers should examine what the system automates, what it adds, what it accelerates and what new monitoring or correction work it creates. Net workload may rise even when individual tasks become faster because employees face more cases, more channels or continuous verification. Cognitive load should therefore be assessed across the full workflow, including interruptions, context switching, exception handling and emotional labour.

Proposition 2. AI-enabled task optimisation will improve well-being when efficiency gains are reinvested in job quality, recovery, learning or relational work rather than converted entirely into higher work intensity.

Proposition 3. The relationship between AI use and exhaustion will be mediated by cognitive fragmentation and moderated by employees' ability to pace, prioritise and recover from digitally accelerated work.

4.3 Trust, fairness and data dignity

The third dimension integrates technical trustworthiness with procedural justice. Employees need understandable information about data sources, system limitations, monitoring practices and the consequences of outputs. Fairness includes bias testing, but also consistent procedures, opportunities to be heard and access to human review. Data dignity requires minimisation, purpose limitation, security and restraint in inferring emotions, personality or future behaviour from weak signals.

Proposition 4. Perceived procedural justice and contestability will mediate the effect of AI-supported management decisions on organisational trust.

Proposition 5. Intrusive or poorly explained employee-data practices will reduce the well-being benefits of AI even when the system improves operational performance.

4.4 Capabilities and AI literacy

The fourth dimension treats capability development as an outcome of adoption, not merely a prerequisite. AI literacy includes functional use, critical evaluation, domain verification, ethical awareness and knowledge of when not to use the tool (Long & Magerko, 2020). Training should be role-based and experiential. Employees need practice identifying hallucinations, bias, privacy risks and unsuitable delegation. Managers need additional capability in work redesign, accountability and interpreting performance data without confusing correlation with causation. Recent evidence also suggests that meaningful work and creative self-efficacy may help explain how employee-AI collaboration influences engagement (Sun et al., 2026).

Proposition 6. AI literacy will strengthen the positive effect of human-AI collaboration on creative self-efficacy and meaningful work while reducing anxiety and uncritical reliance.

4.5 Social connection and psychological safety

The fifth dimension recognises that work is social. AI may reduce routine coordination, but it can also remove conversations through which tacit knowledge, support and identity are formed. Teams should identify which interactions may be automated and which should remain intentionally human. Psychological safety enables employees to disclose uncertainty, question a model, report adverse effects and admit their own errors in using AI.

Proposition 7. The positive relationship between AI-enabled productivity and team innovation will be stronger in climates of psychological safety and weaker where employees fear that questioning the system signals incompetence or resistance.

4.6 Learning, measurement and accountability

The final dimension converts good intentions into an operating system. Organisations should define owners, review cycles, incident channels and thresholds for pausing or redesigning use. Measurement must include leading indicators such as training completion, override patterns and reported friction, as well as outcomes such as engagement, error rates, absenteeism, retention and service quality. Metrics should be segmented by role and demographic group where lawful and ethically justified, because average results can hide unequal burdens.

Proposition 8. Organisations that jointly monitor technical performance, work quality and distributional effects will identify harmful trade-offs earlier and achieve more sustainable AI value than organisations that monitor efficiency alone.

4.7 Boundary conditions and failure configurations

The propositions are expected to vary with task risk, employee discretion, interdependence and employment context. In tightly regulated or safety-critical work, autonomy may appropriately be constrained, but legitimacy still depends on intelligible rules, proportionality, competence and access to escalation. In highly standardised work, automation may improve well-being only when saved effort is not immediately converted into a higher pace. In precarious employment contexts, even well-designed assistance may be experienced as threatening when workers reasonably expect that productivity data will support workforce reduction.

Three failure configurations deserve particular attention. First, nominal oversight occurs when a human approves decisions without time, information or authority to challenge them. Second, benevolent surveillance occurs when intrusive monitoring is justified as coaching or well-being support without meaningful consent and purpose limitation. Third, capability extraction occurs when AI learns from expert workers while the organisation reduces opportunities for those workers and their successors to maintain or develop expertise. These configurations are theoretically important because they can satisfy superficial governance indicators while violating the human-centred mechanism of the framework.

5. A six-stage implementation architecture

The HCAI-WB framework can be implemented through six stages. The sequence is iterative rather than linear because deployment generates new evidence. Each stage should involve technology, operations, human resources, risk functions and employees who perform the affected work.

1. Discover the work system. Map tasks, pain points, decision rights, data flows, affected groups and existing well-being risks before selecting a tool.
2. Design for augmentation and dignity. Specify the human purpose, non-delegable judgements, acceptable data uses, workload assumptions, accessibility needs and recovery mechanisms.
3. Deliberate with affected stakeholders. Use structured dialogue, scenario testing and pre-mortems to surface failure modes, unequal effects and hidden work.

4. Deploy proportionately. Begin with bounded use cases, clear escalation routes, reversible settings and evidence thresholds that reflect the level of risk.
5. Develop capability. Provide role-specific learning, protected experimentation, peer support and opportunities to practise verification and ethical judgement.
6. Demonstrate value responsibly. Evaluate operational performance together with employee experience, job quality, capability growth, fairness and unintended consequences.

A stage-gate decision should be used before scaling. The gate asks whether the system is technically reliable, legally and ethically acceptable, usable in context, compatible with healthy work and supported by adequate capability. Failure on a critical dimension should trigger redesign rather than compensation by gains elsewhere.

The stage gate should be non-compensatory. A use case should not scale when it produces unresolved discriminatory outcomes, prevents meaningful human review, creates an unsafe or unsustainable workload, relies on disproportionate monitoring, or leaves affected people without a credible appeal route. These are redesign conditions rather than variables to be traded against cost savings. The principle is compatible with AI management systems such as ISO/IEC 42001, psychosocial risk management under ISO 45003 and risk-based regulatory duties, while extending them into the concrete experience of work.

6. Measurement dashboard and maturity model

A human-centred dashboard should avoid a single composite score that allows serious harm in one domain to be hidden by strong performance in another. A small balanced set of indicators is preferable. Quantitative data should be interpreted with qualitative evidence from interviews, focus groups, incident narratives and observation. The purpose is organisational learning, not employee surveillance.

Measurement should combine leading indicators (training quality, unresolved concerns, override friction, workload forecasts), process indicators (actual override use, verification time, appeal resolution, task redistribution) and lagging outcomes (exhaustion, retention, error, service quality and unequal impact). Baseline data should be collected before deployment and repeated at intervals that capture adaptation rather than only launch enthusiasm. Wherever possible, comparisons should use matched teams, phased roll-outs or interrupted time-series designs. Qualitative evidence is necessary to identify hidden work and explain why a numerical trend occurred.

Maturity should be assessed through observable practices. The five levels below distinguish experimentation from institutionalisation. Progress is not defined by the number of AI tools deployed, but by the quality of governance, work design and learning.

Table 2. HCAI-WB dimensions, practices and indicative measures

Dimension	Core practices	Illustrative indicators
Purpose and human agency	Human-value thesis; explicit decision rights; meaningful override; non-delegable judgements	Role clarity; perceived autonomy; override access; proportion of high-risk cases receiving human review
Work design and cognitive load	End-to-end workflow mapping; workload simulation; recovery time; job enrichment	Work intensity; interruption load; time saved and reinvested; exhaustion; error-correction time
Trust, fairness and data dignity	Data minimisation; explanation; bias testing; appeal; proportional monitoring	Trust; perceived fairness; appeal resolution; privacy incidents; outcome disparities
Capabilities and AI literacy	Role-based training; verification practice; peer learning; career pathways	Critical AI literacy; self-efficacy; skill growth; dependency indicators; participation in learning
Social connection and psychological safety	Protected human contact; team reflection; speak-up channels; no-blame learning	Psychological safety; quality of collaboration; reported concerns; loneliness; learning from incidents

Dimension	Core practices	Illustrative indicators
Learning, measurement and accountability	Named owners; stage gates; multidimensional dashboard; periodic review	Review completion; action closure; well-being trends; technical drift; documented redesign decisions

Table 3. HCAI-WB maturity model

Level	Characteristics	Primary risk	Next step
1. Tool-led	Isolated pilots, vendor metrics, unclear ownership and limited employee input	Hidden work and uncontrolled experimentation	Map use cases, owners, data and affected roles
2. Compliance-led	Basic legal, security and privacy checks; well-being considered after deployment	Technically compliant but poor work design	Add job-quality and participation criteria to approval
3. Participatory	Employees involved in design; training and appeal routes established	Inconsistent practice across teams	Standardise governance and shared indicators
4. Integrated	Technical, ethical and well-being metrics reviewed together; roles redesigned deliberately	Metric overload or ritualistic review	Focus on learning, distributional effects and adaptive controls
5. Regenerative	AI systematically expands capability, meaningful work and collective learning while preserving constraints	Complacency as technology and contexts change	Maintain external challenge, horizon scanning and periodic revalidation

Table 4. Empirical operationalisation and falsification logic

Construct	Indicative operationalisation	Test and falsification signal
Human agency	Decision latitude, override accessibility, role clarity, perceived control	Compare equivalent use cases with different decision rights; no improvement despite genuine authority weakens the proposed mechanism.
Cognitive load and work design	Work intensity, interruption frequency, verification time, recovery opportunity, task completeness	Workflow and time-series evidence should show whether saved time becomes enrichment or acceleration; persistent benefit under severe fragmentation challenges Proposition 3.
Trust, fairness and data dignity	Procedural justice, explanation adequacy, privacy concern, appeal quality, outcome disparities	Mediation models should test justice and contestability; trust that remains high despite opaque, non-contestable decisions would qualify Proposition 4.
Capability and AI literacy	Critical evaluation, calibrated reliance, domain verification, creative self-efficacy, skill growth	Longitudinal assessment should distinguish learning from dependency; rising output with declining unaided competence indicates capability substitution.
Connection and psychological safety	Team safety, speaking-up behaviour, collaboration quality, social isolation, incident learning	Multilevel models should test whether team climate changes the productivity-innovation relationship; absent moderation weakens

Construct	Indicative operationalisation	Test and falsification signal
		Proposition 7.
Sustainable value	Engagement, exhaustion, retention, error, service quality and distributional outcomes	Joint improvement across technical and human indicators supports the framework; efficiency with persistent critical harm does not qualify as sustainable value.

7. Discussion

The HCAI-WB framework reframes AI adoption as the design of a human-technology institution. Its main theoretical contribution is to connect responsible AI characteristics with the psychological and social mechanisms through which employees experience technology. Fairness, explainability and accountability are not only compliance principles; they shape trust, voice and meaning. Autonomy, competence and relatedness are not only well-being variables; they influence error detection, adoption quality and learning. This reciprocal relationship helps explain why technically similar implementations can produce different outcomes.

Relative to adjacent approaches, the framework adds explanatory and operational specificity. Technical risk frameworks identify what must be governed but do not fully model how implementation reshapes jobs. The job demands-resources model explains strain and motivation but does not specify AI governance requirements. Human-centred design improves usability and participation but may stop at the interface. HCAI-WB links these levels and requires evidence that governance controls are usable under real workload, that participation changes decisions, and that technical gains coexist with preserved agency and capability.

The framework also extends the automation-augmentation debate. Augmentation is often treated as a favourable category, but assistance can still be intrusive, deskilling or cognitively exhausting. The relevant question is not whether a human remains in the loop, but what authority, understanding, time and capability that human possesses. A nominal human checkpoint offers little protection when workloads make verification impossible or when challenging a recommendation carries social penalties.

A second contribution is the principle of non-compensation across critical domains. Organisations frequently justify negative employee effects by pointing to customer or financial benefits. The framework does not deny trade-offs, but it rejects the assumption that every harm can be averaged away. Violations of dignity, discriminatory outcomes, unsafe workload or absence of appeal should be treated as constraints requiring correction. This principle aligns sustainable performance with responsible innovation.

A third contribution is temporal. AI effects change as models, tasks and employee behaviour evolve. Initial enthusiasm may mask long-term dependency, while early anxiety may decline with competence and participation. Governance must therefore continue after launch. Review cycles, incident learning and capability renewal are part of the product life cycle.

8. Managerial implications

For senior leaders, the framework implies that AI strategy should include a human-value thesis alongside a business case. Boards and executive committees should ask which dimensions of job quality may improve or deteriorate and should assign accountability across functions. For technology leaders, requirements should include contestability, uncertainty communication, auditability and user correction. For human resource leaders, workforce planning should focus not only on displaced tasks but also on redesigned roles, learning pathways and identity transitions.

Line managers have a particularly important role because they translate system outputs into daily expectations. They should avoid treating recommendations as unquestionable, monitor hidden work and create routines for discussing errors. Employee representatives and occupational health professionals should be involved early, especially where monitoring, scheduling, safety or employment decisions are affected. Procurement teams should require vendors to provide evidence about limitations, data governance, accessibility and support for human review.

Finally, organisations should protect time for reflection. The speed of AI can create pressure to respond instantly, but high-quality judgement often requires pause, dialogue and contextual interpretation. Responsible technology management is not anti-efficiency; it recognises that durable efficiency depends on learning, trust and healthy human systems.

9. Limitations and future research

This article is conceptual and requires empirical testing. The six dimensions may have different relative importance across sectors, cultures, occupations and risk levels. Longitudinal studies are

needed to distinguish novelty effects from stable changes in work quality. Multilevel research should examine individual, team and organisational mechanisms simultaneously, including how leadership behaviour and labour institutions shape outcomes.

Future studies could develop and validate measurement scales for each dimension, test the eight propositions, compare participatory and top-down implementations, and examine whether maturity predicts retention, innovation, error rates and customer outcomes. Experimental and quasi-experimental designs could evaluate specific work-redesign interventions. Research should also examine distributional effects, including whether benefits accrue primarily to already advantaged employees and whether invisible verification or emotional labour is shifted to particular groups.

A staged validation programme is proposed. Stage 1 should use expert review, employee cognitive interviews and cross-sector cases to refine construct content and identify missing dimensions. Stage 2 should develop multi-item measures and test dimensionality, reliability, discriminant validity and measurement invariance across occupations and cultures. Stage 3 should examine propositions through multilevel longitudinal and quasi-experimental designs, including phased AI roll-outs with pre-deployment baselines. Stage 4 should test incremental validity: whether HCAI-WB maturity explains sustainable performance beyond technical risk ratings, general job-quality measures and conventional change-management assessments.

The framework should also be adapted for small organisations that lack specialist governance teams. Simpler tools, shared sector resources and proportionate controls may be necessary. Finally, the environmental costs of AI and their relationship to organisational purpose should be incorporated into later versions of the model.

10. Conclusion

AI can accelerate work, but acceleration is not the same as progress. Organisational value becomes sustainable when technology improves the quality of decisions and the quality of human experience at the same time. The HCAI-WB framework provides a structured way to pursue that alignment through six dimensions, eight propositions, a six-stage implementation architecture and a non-compensatory measurement approach.

The central managerial test is simple but demanding: after AI is introduced, do people have greater capacity to exercise responsible judgement, learn, connect and contribute—or merely greater pressure to produce? Organisations that ask this question early are better positioned to create trustworthy systems, resilient work and performance that does not consume the people on whom it depends.

Declarations

Funding: The author received no specific funding for this work.

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

Ethics statement: This conceptual article did not involve human participants, personal data, or animal subjects; therefore, ethics approval and informed consent were not required.

Data availability: No new dataset was generated. All sources used are identified in the reference list.

Author contribution: The author conceived the framework, conducted the integrative synthesis, developed the propositions and implementation tools, and wrote the manuscript.

Originality statement: This manuscript presents an original conceptual synthesis, framework and set of research propositions. It has not been published or submitted simultaneously to another journal.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Fournay, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-13. <https://doi.org/10.1145/3290605.3300233>
- Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology*, 22(3), 309-328. <https://doi.org/10.1108/02683940710733115>
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140(2), 889-942. <https://doi.org/10.1093/qje/qjae044>
- Chuang, Y.-T., Chiang, H.-L., & Lin, A.-P. (2025). Insights from the job demands-resources model: AI's dual impact on employees' work and life well-being. *International Journal of Information Management*, 83, 102887. <https://doi.org/10.1016/j.ijinfomgt.2025.102887>
- Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350-383. <https://doi.org/10.2307/2666999>
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Green, F., Felstead, A., Gallie, D., & Henseke, G. (2024). Work and life: The relative importance of job quality for general well-being. *Socio-Economic Review*, 22(2), 835-861. <https://doi.org/10.1093/ser/mwad067>

- International Labour Organization. (2025). Generative AI and jobs: A refined global index of occupational exposure. ILO Working Paper 140. <https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure>
- International Organization for Standardization. (2021). ISO 45003:2021 Occupational health and safety management—Psychological health and safety at work—Guidelines for managing psychosocial risks. ISO.
- International Organization for Standardization. (2023). ISO/IEC 42001:2023 Information technology—Artificial intelligence—Management system. ISO.
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586. <https://doi.org/10.1016/j.bushor.2018.03.007>
- 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>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-16. <https://doi.org/10.1145/3313831.3376727>
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). <https://doi.org/10.6028/NIST.AI.100-1>
- National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). <https://doi.org/10.6028/NIST.AI.600-1>
- 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>
- OECD. (2024). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Parker, S. K., & Grote, G. (2022). Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. *Applied Psychology*, 71(4), 1171-1204. <https://doi.org/10.1111/apps.12241>
- 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>
- Rosso, B. D., Dekas, K. H., & Wrzesniewski, A. (2010). On the meaning of work: A theoretical integration and review. *Research in Organizational Behavior*, 30, 91-127. <https://doi.org/10.1016/j.riob.2010.09.001>
- Savolainen, I., Ylinen, L., Grönroos, R., & Oksanen, A. (2026). AI transformation in working life: A systematic review of usage and attitudes towards AI among workers. *Digital Business*, 6(1), 100162. <https://doi.org/10.1016/j.digbus.2025.100162>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495-504. <https://doi.org/10.1080/10447318.2020.1741118>
- 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>
- Sun, L., Hu, R., & Su, H. (2026). Unlocking human potential in the AI age: How employee-AI collaboration transforms work engagement through dual psychological pathways. *Frontiers in Psychology*, 16, 1705671. <https://doi.org/10.3389/fpsyg.2025.1705671>
- 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>
- UNESCO. (2021). Recommendation on the ethics of artificial intelligence. <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>
- Valtonen, A., Saunila, M., Ukko, J., Treves, L., & Ritala, P. (2025). AI and employee wellbeing in the workplace: An empirical study. *Journal of Business Research*, 199, 115584. <https://doi.org/10.1016/j.jbusres.2025.115584>
- World Health Organization. (2022). WHO guidelines on mental health at work. World Health Organization. <https://www.who.int/publications/i/item/9789240053052>

How do I cite this article?

Dr. Ignacio Bonasa Alzuria, et.al, Human-centred artificial intelligence and organisational well-being: An integrative framework for responsible technology management, *Excel International Journal of Technology, Engineering and Management*, 2026; Volume -13, Issue-3_Page_17-27.

This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)