



ISSN: 0976-3376

Available Online at <http://www.journalajst.com>

ASIAN JOURNAL OF
SCIENCE AND TECHNOLOGY

Asian Journal of Science and Technology
Vol. 17, Issue, 09, pp. 14506-14515, September, 2026

RESEARCH ARTICLE

HUMAN-CENTERED ARTIFICIAL INTELLIGENCE AND THE ARCHITECTURE OF WELL-BEING: AN EIGHT-DIMENSIONAL FRAMEWORK FOR TECHNOLOGICAL INNOVATION WITH HUMAN PURPOSE

Ignacio Bonasa Alzuria

University of Happiness and Well-Being, Mexico; Liderarte Factoría de Talento, Spain

ARTICLE INFO

Article History:

Received 20th June, 2026

Received in revised form

24th July, 2026

Accepted 19th August, 2026

Published online 30th September, 2026

Key words:

Human-Centered Artificial Intelligence;
Well-Being; Responsible Innovation;
Future of Work; Humanistic
Leadership; AI Governance; job design;
Digital transFormation; Organizational
Health.

*Corresponding author:

Ignacio Bonasa Alzuria

ABSTRACT

Artificial intelligence is entering organizations faster than most institutions can redesign the human experience around it. This conceptual article argues that the central question is no longer whether AI can increase efficiency, but whether technological innovation can improve performance without impoverishing autonomy, meaning, relationships, health, judgment, or dignity. It develops the Human-Centered AI Well-Being Architecture (HAWA), an original eight-dimensional framework integrating physical, cognitive, emotional, transcendent/meaning, social, professional, financial, and technological well-being. The model is developed through a structured integrative synthesis of established theory, international governance frameworks, and recent evidence on generative AI and work. HAWA positions well-being as a design requirement, governance criterion, and measurable organizational capability rather than a soft outcome added after deployment. Five operating principles—augmentation before substitution, agency before automation, intelligibility before opacity, human connection before frictionless isolation, and shared value before one-sided productivity—are translated into a cycle of govern, map, co-design, deploy, measure, and learn. A Well-Being Impact Matrix and Human Purpose Scorecard expose trade-offs that conventional productivity metrics can conceal. The distinctive contribution is a multidimensional, non-compensatory bridge between responsible-AI governance and the lived experience of workers and other stakeholders. Trustworthy AI, the paper concludes, requires an architecture in which technology expands human capability while preserving the conditions that make work and life worth living.

Citation: Ignacio Bonasa Alzuria. 2026. "Human-centered artificial intelligence and the architecture of well-being:an eight-dimensional framework for technological innovation with human purpose", *Asian Journal of Science and Technology*, 17, (09), 14506-14515.

Copyright©2026, Ignacio Bonasa Alzuria. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Artificial intelligence has moved from the laboratory to the meeting room, the classroom, the hospital, the bank, the factory floor, and the mobile phone. Generative systems can summarize, predict, classify, translate, draft, code, recommend, simulate, and converse at a speed that changes the economics of knowledge work. Yet the most important transformation is not computational. It is human. Every time an organization introduces an intelligent system, it quietly renegotiates who decides, who learns, who is visible, who is trusted, who carries responsibility, and what kind of effort is considered valuable. The dominant narrative of digital transformation has traditionally privileged measurable gains: productivity, cost reduction, cycle time, accuracy, scalability, and return on investment. These outcomes matter. Organizations cannot survive by ignoring efficiency. The difficulty appears when efficiency becomes the only language through which progress is judged. A system may save ten minutes per task while increasing anxiety, weakening professional judgment, reducing social contact, intensifying

monitoring, or making employees feel replaceable. A dashboard can report the minutes saved and remain completely silent about the human price. This article begins from a simple proposition: technological innovation is incomplete when it optimizes the process but degrades the person. Human-centered AI therefore cannot mean merely creating interfaces that people find easy to use. It must mean designing, governing, and evaluating AI in relation to human flourishing. That shift expands the unit of analysis from the algorithm to the socio-technical system and from immediate performance to the quality of the life and work that the technology helps to create. International frameworks already point in this direction. The OECD AI Principles connect trustworthy AI with inclusive growth, sustainable development and well-being, while the NIST AI Risk Management Framework treats AI as a socio-technical phenomenon whose risks and benefits emerge through interaction among technical systems, people, institutions, and contexts. These frameworks provide an essential governance foundation. The present paper complements them by asking a more intimate organizational question: what dimensions of human well-being should leaders

observe when AI enters everyday work?. The answer developed here is an eight-dimensional architecture. It considers the body, the mind, emotions, meaning and transcendence, social relationships, professional life, financial security, and the technological relationship itself. These dimensions are analytically distinct but practically intertwined. An AI tool that reduces repetitive work may improve professional satisfaction and cognitive load, but if its implementation creates fear of redundancy it may damage emotional and financial well-being. A virtual assistant may increase accessibility while reducing human contact. A predictive tool may support better decisions while encouraging over-reliance and skill atrophy. The purpose of the framework is not to romanticize work before AI or to resist automation. Many jobs already contain unnecessary bureaucracy, unsafe tasks, avoidable errors, and cognitive overload. AI can remove drudgery, democratize expertise, personalize learning, support people with disabilities, and help small teams achieve outcomes previously available only to large organizations. The argument is instead that these benefits should be designed intentionally and distributed responsibly. Accordingly, this paper develops the Human-Centered AI Well-Being Architecture (HAWA), a conceptual model for evaluating and guiding AI adoption. It contributes three elements: first, an eight-dimensional definition of well-being for AI-enabled organizations; second, five design principles that translate humanistic intent into operational choices; and third, a governance and measurement cycle that allows organizations to detect trade-offs rather than celebrating productivity in isolation.

RESEARCH AIM AND QUESTIONS

The article addresses a gap between responsible-AI governance and organizational well-being. Technical frameworks are increasingly sophisticated in relation to validity, safety, security, transparency, explainability, privacy, fairness, and accountability. Well-being research, meanwhile, offers mature theories of motivation, positive functioning, job quality, mental health, and meaningful work. What remains underdeveloped is a practical bridge that enables leaders to evaluate an AI deployment simultaneously as a technical intervention and as a redesign of human experience. The primary aim is therefore to formulate an integrative framework that places multidimensional human well-being inside the design and governance of AI-enabled work. Four research questions guide the analysis: (RQ1) Which dimensions of well-being are most relevant when AI changes tasks, decisions, relationships, and careers? (RQ2) Through what mechanisms can AI enhance or diminish those dimensions? (RQ3) Which design and leadership principles can reduce avoidable harm while preserving innovation? (RQ4) How can organizations measure human outcomes without turning well-being itself into a surveillance mechanism?

Theoretical contribution and boundary of novelty: HAWA does not claim to replace established theories of motivation, occupational health, human-computer interaction, or AI governance. Its novelty lies in integrating them around a specific unit of analysis: the human consequences of organizational AI adoption. The framework contributes (1) an eight-domain map that includes technological and financial well-being alongside psychological and occupational dimensions; (2) a non-compensatory logic in which serious deterioration in dignity, safety, fairness, or agency cannot be

justified merely by aggregate productivity gains; and (3) an implementation architecture linking human outcomes to governance decisions before, during, and after deployment. HAWA is therefore proposed as a mid-range conceptual framework designed to generate testable hypotheses rather than as a universal theory of human flourishing.

METHODOLOGICAL APPROACH: AN INTEGRATIVE CONCEPTUAL REVIEW

Methodological transparency is important because integrative reviews can otherwise appear more systematic than they are. This study is not presented as a systematic review or meta-analysis, and no claim of exhaustive database coverage is made. Source selection was purposive and theory-oriented: foundational theories were retained when they offered established constructs for autonomy, competence, relatedness, meaningful work, positive functioning, job design, or human-centered system design; authoritative institutional frameworks were retained when they translated AI risk into governance practice; and recent evidence was prioritized when it addressed generative AI, productivity, work quality, augmentation, algorithmic management, worker agency, or skill development. The synthesis used four analytical stages. First, recurrent human outcomes in the literature were identified and grouped into domains of lived experience. Second, those domains were compared with self-determination theory, positive psychology, occupational health, and work-design research to reduce conceptual redundancy. Third, AI mechanisms—automation, augmentation, prediction, personalization, monitoring, conversational interaction, and generative assistance—were mapped against plausible positive and negative pathways. Fourth, the domains were translated into design principles, governance questions, indicators, and propositions for empirical testing.

Recent evidence strengthens the need for this synthesis. OECD workplace analysis reports substantial perceived performance and enjoyment benefits from AI while also identifying concerns about work intensity, data collection, inequality, privacy, and autonomy. Its 2025 synthesis of generative AI evidence emphasizes that outcomes depend on task complexity, user skill, trust, and the quality of human-AI integration. NIST's Generative AI Profile likewise frames risk management as an organizational lifecycle responsibility rather than a one-time technical test (OECD, 2024a; OECD, 2025; Autio et al., 2024). HAWA adopts a non-compensatory principle for severe harms. Positive effects in several dimensions should not automatically offset a major deterioration in safety, dignity, discrimination, livelihood, or meaningful human control. This is especially important for high-consequence applications because a single aggregate score can conceal unacceptable distributional effects. This is a conceptual and theory-building article rather than an empirical report. No survey responses, interviews, experiments, or organizational datasets are presented; consequently, no artificial empirical claims are made. The method is an integrative review designed to connect bodies of knowledge that are often discussed separately: human-centered artificial intelligence, responsible AI and governance, occupational and psychological well-being, self-determination theory, positive psychology, job design, human-computer interaction, and the future of work. An integrative approach is appropriate because the problem is inherently multidisciplinary. AI adoption is simultaneously a technical choice, a work-design intervention,

a leadership decision, an ethical act, and a change in the distribution of knowledge and power. The synthesis gives particular weight to established theoretical contributions and institutional frameworks with broad international relevance, including the OECD AI Principles, NIST AI RMF, WHO guidance on mental health at work, and scholarship on autonomy, competence, relatedness, meaningful work, and human-centered system design. The framework was developed through four analytical moves. First, recurrent human outcomes associated with digitalization and AI were identified and grouped into domains of lived experience. Second, these domains were compared with established well-being theories to avoid reducing well-being to momentary happiness. Third, AI-specific mechanisms—automation, augmentation, prediction, personalization, monitoring, conversational interaction, and generative assistance—were mapped against potential benefits and harms. Fourth, governance principles and indicators were derived to make the framework usable by organizations. The resulting model should be understood as a structured hypothesis for research and practice. It is intended to be testable, adaptable across sectors, and open to falsification. Its value does not depend on claiming that eight dimensions are the only possible taxonomy. Rather, its value lies in preventing important human consequences from disappearing simply because they are difficult to express in the language of technical performance.

FROM HUMAN-CENTERED AI TO HUMAN FLOURISHING: Human-centered AI has emerged as a response to the limitations of purely technology-centered development. Shneiderman (2020) argues for systems that combine high levels of automation with high levels of human control rather than assuming that greater automation necessarily requires less human agency. Human-centered design similarly emphasizes the needs, capabilities, contexts, and limitations of users. In organizational settings, however, the relevant 'user' is not a single person interacting with a screen. It is a network of employees, managers, customers, professionals, citizens, and communities whose incentives and vulnerabilities differ.

This distinction matters because a system can be usable and still be harmful. An employee-monitoring tool may have an elegant interface but create a climate of suspicion. A recommendation engine may be accurate but narrow human judgment. A generative assistant may be convenient while quietly shifting expertise from people to platforms. Conversely, a system can initially require effort yet strengthen competence, inclusion, and autonomy over time. Human-centered evaluation must therefore ask not only 'Can people use this system?' but 'What does repeated use of this system do to people?'. Self-determination theory offers a powerful lens. Ryan and Deci identify autonomy, competence, and relatedness as basic psychological needs associated with high-quality motivation and well-being. AI can support autonomy by removing administrative burden and expanding choice, or undermine it through algorithmic control. It can support competence through feedback and tutoring, or weaken it through deskilling and passive dependence. It can enable collaboration across distance, or replace meaningful human interaction with synthetic convenience. Positive psychology adds a broader perspective. Seligman's PERMA model emphasizes positive emotion, engagement, relationships, meaning, and accomplishment. These elements remind technology leaders that well-being is not synonymous with

comfort. Challenging work can be deeply satisfying when people experience agency, growth, connection, purpose, and achievement. The design objective should not be to remove every difficulty, but to remove pointless friction while preserving the forms of effort through which people learn and become capable. Occupational health research introduces another necessary correction: well-being is affected not only by individual mindset but by the structure of work. Excessive demands, low control, insecurity, unfairness, role ambiguity, and poor social environments can generate strain. WHO guidance on mental health at work emphasizes organizational interventions, not merely individual resilience. Applied to AI, this means that organizations should not deploy destabilizing systems and then respond with mindfulness workshops. The architecture of work itself must be examined.

THE HUMAN-CENTERED AI WELL-BEING ARCHITECTURE (HAWA): HAWA proposes eight interdependent dimensions. The framework is deliberately broader than occupational satisfaction because AI increasingly crosses the boundary between work and life. Notifications follow workers home; recommender systems shape attention; automated decisions influence income and opportunity; conversational systems affect learning and social behavior. A responsible organization needs a map wide enough to see these interactions.

Figure 1. Human-Centered AI Well-Being Architecture (HAWA)

PHYSICAL Energy • safety • recovery	COGNITIVE Attention • learning • judgment	EMOTIONAL Safety • confidence • stress
SOCIAL Belonging • trust • empathy	HUMAN PURPOSE + AI Agency • dignity • meaning Responsible augmentation	PROFESSIONAL Mastery • autonomy • growth
FINANCIAL Security • fairness • opportunity	TECHNOLOGICAL Privacy • intelligibility • boundaries	TRANSCENDENT / MEANING Purpose • contribution • coherence

Note. The eight dimensions are interdependent. Severe deterioration in safety, dignity, fairness, or meaningful human control is not treated as automatically compensable by productivity gains.

Physical well-being: Physical well-being concerns energy, sleep, ergonomics, movement, safety, fatigue, and the bodily consequences of digitally mediated work. AI can reduce exposure to dangerous tasks, optimize schedules, detect safety risks, and automate repetitive physical or clerical activities. It can also increase sedentary behavior, screen time, work intensity, and the expectation of continuous availability. The key design question is whether AI returns time and energy to the person or merely fills newly created capacity with more work. Productivity gains that increase pace without increasing recovery can convert efficiency into exhaustion. Organizations should therefore monitor workload, break patterns, after-hours activity, ergonomic demands, and perceived fatigue alongside throughput.

Cognitive well-being: Cognitive well-being refers to attention, mental load, clarity, learning, curiosity, memory, judgment, and the experience of being mentally capable. Generative AI can reduce search costs, synthesize information, scaffold complex tasks, and accelerate learning. Yet cognitive offloading can become cognitive atrophy when people stop

practicing the skills required to verify, challenge, or replace the system. A human-centered deployment distinguishes between supportive offloading and corrosive dependency. Calculators did not eliminate mathematics, but they changed which mental operations remained essential. Generative AI creates a similar transition for writing, analysis, coding, and decision support. The organization must identify the capabilities that employees should continue to exercise even when the machine can produce an acceptable first answer.

Emotional well-being: Emotional well-being includes psychological safety, stress, confidence, hope, fear, frustration, and the capacity to regulate emotion under uncertainty. AI adoption can create excitement and relief, but also status anxiety, identity threat, and fear of redundancy. These reactions are not irrational obstacles to change; they are information about how change is being experienced. Leaders should communicate what the technology will change, what it will not change, how roles may evolve, what support will be provided, and how decisions about people will be made. Ambiguity is often more damaging than difficult news. A humane implementation reduces avoidable uncertainty and creates channels where employees can question the system without being labeled resistant.

Transcendent and meaning well-being: This dimension concerns purpose, values, coherence, contribution, and the sense that one's activity connects to something larger than immediate output. The term transcendent is used in a non-denominational sense: it refers to the human capacity to locate work within a wider horizon of meaning, service, legacy, or moral commitment. AI can liberate people from low-value tasks and create more space for judgment, creativity, care, and contribution. It can also produce the opposite effect if workers become supervisors of opaque machines, correcting outputs without understanding why their role matters. The relevant question is not only 'What task remains for the human?' but 'Is the remaining human task worth doing?'

Social well-being: Social well-being includes belonging, trust, reciprocity, empathy, collaboration, recognition, and access to human support. AI-mediated communication can help multilingual teams, make knowledge more accessible, and connect distributed organizations. At the same time, every automated interaction creates a substitution decision: should this moment be faster, or should it remain human? Not all friction is waste. Informal conversations, mentoring, shared problem-solving, and even certain disagreements create social capital. If AI removes too many of these moments, an organization may become efficient but relationally thin. Human-centered design protects 'high-human-value interactions'—moments where empathy, trust, identity, or moral responsibility matter more than speed.

Professional well-being: Professional well-being concerns mastery, role quality, employability, achievement, autonomy, fair evaluation, career development, and pride in work. AI can expand individual capability and enable workers to perform tasks previously outside their reach. Evidence from field and experimental studies increasingly suggests that generative AI can improve performance in selected knowledge tasks, often with heterogeneous effects across skill levels. The distribution of gains matters. If the organization captures all productivity benefits while employees experience increased monitoring, reduced discretion, and insecurity, adoption may be

economically rational in the short term but socially fragile. Professional well-being requires a credible development pathway: people need to know how they can become more valuable with AI rather than simply compete against it.

Financial well-being: Financial well-being includes income security, predictability, perceived fairness, access to opportunity, and confidence in one's economic future. AI affects this dimension through job redesign, wage structures, hiring, scheduling, performance evaluation, and the distribution of productivity gains. Algorithmic decisions can improve consistency, but they can also scale bias or make consequential judgments difficult to contest. Organizations should treat financial impact as part of responsible AI, especially when systems influence hiring, promotion, compensation, workload allocation, or redundancy. A human appeal path is essential. Shared-value mechanisms—training investment, gain-sharing, reduced hours without proportional loss of pay where feasible, or clearer internal mobility—can convert technological productivity into human security.

Technological well-being: Technological well-being is the quality of the person's relationship with digital systems themselves. It includes digital agency, privacy, intelligibility, boundaries, trust calibration, accessibility, and the ability to disengage. This dimension is necessary because technology is no longer merely an external tool; it increasingly structures attention, choices, communication, and access to institutions. Healthy technological relationships require calibrated trust. Users should neither reject AI reflexively nor accept it uncritically. They need enough information to understand when a system is reliable, where uncertainty remains, what data are used, and when human review is appropriate. Technological well-being also requires the right to periods of non-connectivity and protection against unnecessary surveillance.

THE WELL-BEING IMPACT MATRIX

The eight dimensions become operational when they are used prospectively, before deployment, and diagnostically, after deployment. Table 1 summarizes representative opportunities, risks, and indicators. The purpose is not to create a universal metric but to force balanced inquiry.

FIVE DESIGN PRINCIPLES FOR AI WITH HUMAN PURPOSE

Augmentation before substitution: Automation should begin with the question of which burdens can be removed so that human capability becomes more valuable. Substitution may sometimes be justified, particularly for dangerous, degrading, or purely repetitive tasks. But substitution should not be the default intellectual frame. An augmentation-first approach asks how AI can help people perceive more, learn faster, make better judgments, create more, or serve others more effectively.

Agency before automation: People need meaningful control over consequential work. Agency does not require manual approval of every machine action; that would make automation pointless. It requires clear boundaries: which decisions are delegated, which are advised, which require human authorization, and which can be challenged. Human override should be real rather than ceremonial.

Dimension	Potential AI benefit	Potential AI risk	Illustrative indicator
Physical	Less hazardous/repetitive work	Work intensification; screen fatigue	Fatigue; after-hours use
Cognitive	Lower search load; learning support	Deskilling; automation bias	Perceived mental load; verification quality
Emotional	Reduced frustration; support	Anxiety; identity threat	Stress; psychological safety
Meaning	More time for high-value contribution	Loss of purpose; hollowed roles	Meaningful-work score
Social	Accessible collaboration	Isolation; synthetic substitution	Belonging; human-contact quality
Professional	Augmented capability; mobility	Reduced autonomy; unfair evaluation	Skill growth; discretion
Financial	Productivity and opportunity	Insecurity; opaque allocation	Income security; fairness
Technological	Accessibility; useful assistance	Surveillance; dependency; opacity	Trust calibration; right to disconnect

Intelligibility before opacity: Not every model can be made fully transparent to every user, but every deployment can communicate purpose, limits, confidence, provenance, escalation routes, and responsibility more clearly. Intelligibility is relational: the explanation must be sufficient for the person who bears the consequence. A data scientist, a manager, a customer, and a citizen may need different forms of explanation.

Human connection before frictionless isolation: Organizations should identify interactions where human presence creates value that cannot be reduced to information transfer. Coaching, difficult feedback, grief, conflict resolution, ethical deliberation, recognition, and care are examples. AI may support preparation or accessibility, but convenience should not automatically replace relationship.

Shared value before one-sided productivity: AI-generated value should be evaluated in terms of who benefits and who carries the risk. Sustainable adoption is more likely when employees see a credible connection between higher productivity and better work: learning opportunities, safer tasks, improved schedules, career mobility, recognition, or economic participation. The principle is not a fixed formula for distribution; it is a requirement to make distribution visible.

HUMANISTIC LEADERSHIP IN THE AGE OF AI: Technology does not enter a neutral organization. It enters a culture. The same AI system can produce different human outcomes depending on leadership behavior, trust, participation, and the pre-existing quality of work. In a high-trust environment, an assistant may be experienced as empowerment. In a climate of surveillance, the same data collection may be experienced as control. Humanistic leadership treats people as ends and not merely as productive resources. This does not weaken accountability. On the contrary, it raises the standard of accountability by requiring leaders to consider both results and the conditions under which results are achieved. In AI transformation, leaders become architects of boundaries: they decide what should be automated, what must remain human, which risks are acceptable, whose voice counts, and how benefits are shared. Three leadership practices are especially important. The first is participatory sense-making. Employees should be involved early enough to influence the design, not merely trained after the decision is irreversible. Frontline workers often know where data are misleading, where exceptions occur, and where a technically efficient process could harm customers or colleagues. Participation is therefore both ethical and epistemic. The second practice is narrative clarity. AI creates a vacuum that people fill with stories: 'the company wants fewer of us,' 'my expertise no longer matters,' or 'the machine will make the decisions.' Leaders must replace rumor with a credible account of purpose, boundaries, and transition. Credibility requires specificity and consistency with actual decisions.

The third practice is protected dissent. Responsible AI needs people who can say that a system is wrong, unfair, unsafe, or dehumanizing without career penalty. Psychological safety becomes part of technical safety because many failures are first detected by humans who notice anomalies. A culture that silences dissent destroys one of its most valuable risk sensors.

THE HAWA IMPLEMENTATION CYCLE

Govern: define purpose, rights, and red lines

Before selecting a tool, the organization should define the legitimate purpose of the use case, the affected stakeholders, decision rights, prohibited uses, data boundaries, accountability, and escalation mechanisms. This aligns with the governance logic of the NIST AI RMF while adding explicit well-being objectives.

Map: understand tasks, people, and consequences

Map the existing workflow before automating it. Identify pain points, meaningful tasks, informal coordination, vulnerable groups, skill dependencies, and potential second-order effects. The objective is to avoid automating a bad process or removing activities that quietly sustain learning and connection.

Co-design: involve the people who live the work

Co-design sessions should include users, managers, technical specialists, risk and legal functions, and where appropriate customers or worker representatives. Participants should examine each of the eight dimensions and record expected benefits, concerns, unknowns, and mitigation ideas.

Deploy: start bounded and observable

Pilots should be narrow enough to learn from. High-consequence decisions require stronger human oversight, documentation, and contestability. The deployment plan should specify what evidence would justify expansion, modification, pause, or withdrawal.

Measure: combine performance and human outcomes

Measurement should include operational indicators and human indicators. A productivity increase is not sufficient if error correction, stress, turnover intention, or customer dissatisfaction rises. Likewise, subjective satisfaction alone cannot substitute for reliability or safety. The unit of success is balanced performance.

Learn: institutionalize feedback and adaptation

AI systems, user behavior, and organizational contexts change. Governance should therefore be continuous. Post-deployment reviews should examine unexpected uses, workarounds, emerging dependencies, inequities, and changes in skill. The organization should be willing to redesign both the technology and the surrounding work.

THE HUMAN PURPOSE SCORECARD: To complement conventional business cases, this paper proposes a Human Purpose Scorecard. For each AI use case, the organization rates expected and observed effects across the eight HAWA dimensions using a simple scale from -2 (material deterioration) to +2 (material improvement), with 0 indicating no meaningful change or insufficient evidence. Scores should never be collapsed mechanically into a single number for high-stakes decisions; a severe negative in one domain cannot always be compensated by small positives elsewhere.

The scorecard is best used as a conversation architecture. It forces cross-functional teams to articulate assumptions and evidence. A customer-service copilot, for example, may score positively on cognitive load and professional capability, neutrally on physical well-being, and negatively on emotional well-being if workers believe performance data will be used punitively. The appropriate response is not to average the numbers and proceed. It is to redesign the governance of performance data. Measurement must also respect privacy. Organizations should prefer aggregated, proportionate, voluntary, and purpose-limited methods. Well-being programs become self-defeating when employees believe that emotional or behavioral data may be used to evaluate them. The measurement system must embody the same human-centered principles as the AI system it evaluates.

PROPOSITIONS FOR FUTURE EMPIRICAL RESEARCH

P1: AI deployments characterized by higher perceived human agency will be associated with stronger professional and emotional well-being than functionally similar deployments characterized by low perceived agency.

P2: The relationship between AI-enabled productivity and employee well-being will be moderated by perceived fairness in the distribution of benefits and risks.

P3: AI augmentation will support cognitive well-being when it reduces unnecessary load while preserving active verification, learning, and skill practice; excessive unverified offloading will be associated with lower skill retention and poorer trust calibration.

P4: Participatory implementation will be positively associated with trust, adoption quality, and psychological safety, partly because frontline knowledge improves contextual fit and perceived procedural fairness.

P5: The replacement of high-human-value interactions with automated interactions will predict lower social well-being even when users report short-term convenience gains.

P6: Clear communication about role evolution, reskilling, and decision rights will reduce emotional strain during AI adoption compared with ambiguous transformation narratives.

P7: Organizations that monitor multidimensional well-being alongside productivity will detect negative second-order effects earlier and will show more sustainable AI adoption over time.

P8: Technological well-being—particularly privacy, intelligibility, contestability, and the right to disengage—will predict calibrated trust more strongly than generalized attitudes toward AI.

ILLUSTRATIVE APPLICATIONS

Customer service copilot: A generative copilot can summarize customer history, suggest responses, and retrieve policy information. The immediate business case is faster resolution. HAWA broadens the case. Cognitive benefits may include reduced search load; professional benefits may include faster onboarding; emotional benefits may arise from support during difficult interactions. Risks include script dependence, surveillance through granular productivity metrics, loss of authentic customer connection, and skill erosion. A human-centered design would preserve discretion, make suggestions editable, separate coaching data from punitive monitoring where possible, and create regular skill practice without the copilot.

AI-supported education: An AI tutor can provide explanations, examples, translation, and personalized practice. Potential benefits include accessibility, cognitive scaffolding, confidence, and lower barriers to asking questions. Risks include inaccurate content, reduced struggle that is necessary for learning, overdependence, weakened teacher-student relationships, and privacy concerns. HAWA suggests that the relevant outcome is not merely answer accuracy but growth in competence, agency, curiosity, and connection.

Clinical administration: In healthcare, AI can reduce documentation burden, summarize records, and support scheduling. Removing clerical work can return attention to patients, which creates professional, emotional, and social value. Yet poorly designed systems may add alerts, increase cognitive fragmentation, or create uncertainty about accountability. The framework therefore evaluates whether the technology gives clinicians more presence with patients rather than simply enabling a higher volume of encounters.

Human resources and talent decisions: AI can assist in matching skills, identifying development opportunities, or organizing large applicant pools. Because these systems influence professional and financial well-being, fairness, transparency, contestability, and human review are central. The organization should distinguish between low-consequence administrative support and decisions that shape livelihood. Efficiency in screening cannot justify opaque exclusion.

Executive decision support: AI can synthesize scenarios, challenge assumptions, and broaden access to information. The danger is authority laundering: a leader may present a model output as objective in order to avoid responsibility. Humanistic leadership rejects this move. AI may inform judgment, but accountability remains human, especially when decisions involve values, trade-offs, or consequences that cannot be fully represented in data.

THE PARADOXES OF AI-ENABLED WELL-BEING

Several paradoxes explain why well-intentioned AI programs can produce unexpected outcomes. The first is the efficiency paradox: saving time at the task level does not guarantee more time at the human level. Organizations often convert saved minutes into higher volume. Unless leaders make deliberate choices about workload, recovery, learning, or reduced friction, employees may experience acceleration rather than relief. The second is the personalization paradox. AI can tailor learning, recommendations, and support to individual needs, yet excessive personalization can isolate people from shared experiences and reduce exposure to difference. Human flourishing requires both personal fit and participation in communities. The third is the assistance paradox. The better the assistant becomes, the easier it is to stop practicing the skills required to evaluate it. This creates a long-term resilience problem. Organizations need 'cognitive reserve': people who remain capable of reasoning when the system is unavailable, wrong, compromised, or confronted with a novel situation. The fourth is the measurement paradox. AI makes more behavior measurable, but what can be measured is not identical to what matters. Metrics can improve visibility while narrowing attention. When people know they are continuously scored, they may optimize for the indicator rather than the purpose. Good governance therefore combines data with judgment and protects spaces that should not become performance telemetry.

The fifth is the humanization paradox. Conversational AI can sound empathic, patient, and supportive. This can be genuinely useful, particularly when it increases access. Yet simulated empathy is not the same as mutual human relationship. The more convincing the interface becomes, the more important it is to preserve clarity about what the system is and where human care is indispensable.

ETHICS, FAIRNESS, AND DIGNITY

Well-being cannot be separated from justice. An organization could theoretically improve average satisfaction while imposing disproportionate harms on a smaller group. Responsible AI therefore requires distributional analysis. Who receives the convenience? Who performs the hidden correction work? Whose data train the system? Who is more likely to be misclassified? Who has the resources to appeal? Dignity adds a dimension that is difficult to reduce to utility. People care not only about outcomes but about how they are treated. Being evaluated by an opaque system, monitored continuously, or denied meaningful explanation can be experienced as disrespect even when the statistical process is efficient. Human-centered AI should preserve the person's status as a participant capable of reason, voice, and contestation.

Fairness is also procedural. Participation, explanation, consistent rules, and accessible appeals can strengthen legitimacy. This is particularly important in employment, finance, education, healthcare, and public services. The higher the consequence, the stronger the expectation that a person can understand the basis of a decision and reach a responsible human authority.

ORGANIZATIONAL CAPABILITIES REQUIRED: Implementing HAWA requires more than an ethics statement. Organizations need capability in at least five areas. First, AI

literacy: employees and leaders must understand basic capabilities, limitations, uncertainty, data issues, and appropriate verification. Second, work-design competence: teams need the ability to redesign roles rather than simply insert a tool into an old process.

Third, organizations need socio-technical risk assessment. Technical testing remains essential, but it must be complemented by analysis of incentives, behavior, power, workload, and culture. Fourth, leaders need change capability rooted in participation and psychological safety. Fifth, organizations need measurement discipline that integrates productivity, quality, risk, learning, and well-being without creating invasive surveillance. These capabilities should not be concentrated in a single AI office. Responsible adoption is a distributed competence. Technical teams understand systems; frontline employees understand context; HR understands roles and development; legal and risk functions understand obligations; leaders allocate resources and define purpose. The architecture works when these perspectives meet before harm occurs.

DISCUSSION

The central contribution of HAWA is to reposition well-being from a peripheral benefit to a design variable. This matters because organizations often discover human consequences only after deployment, when habits, contracts, workflows, and expectations have already formed. By placing eight dimensions into the initial business case, HAWA makes hidden trade-offs discussable before they become embedded. The framework also challenges a false choice between innovation and humanity. Recent OECD evidence supports a conditional interpretation of AI at work: many workers report improvements in performance and enjoyment, while concerns persist around work intensity, data use, inequality, privacy, and autonomy (OECD, 2024a). The OECD's 2025 synthesis of generative AI evidence further stresses that outcomes depend on task complexity, worker skills, trust calibration, and the quality of human-AI integration rather than on access to the technology alone (OECD, 2025). Human-centered constraints can improve technical and organizational performance because they reveal contextual information, strengthen adoption, preserve critical expertise, and reduce resistance generated by poor implementation. At the same time, the framework resists simplistic optimism. AI will create genuine displacement in some tasks and roles. Human-centered leadership cannot promise that every existing job will remain unchanged. Its obligation is different: to avoid unnecessary harm, communicate truthfully, invest in adaptation, preserve dignity, and distribute transition costs fairly. A mature well-being framework must be able to hold both possibility and loss.

HAWA also extends the concept of technological well-being. Digital well-being is often discussed as screen time or healthy device use. In AI-enabled organizations, the concept must include epistemic and institutional questions: Do people know when AI is involved? Can they contest it? Can they understand its limits? Can they preserve independent judgment? Can they disconnect? Are they becoming more capable or merely more dependent? Finally, the framework suggests that the future of work should be evaluated by the quality of human capability it produces. A technologically advanced organization in which people are anxious, isolated, deskilled, and constantly

monitored cannot be considered fully successful. Progress should enlarge the space for people to think, create, relate, contribute, and live with greater security and purpose.

ALIGNMENT WITH CONTEMPORARY AI GOVERNANCE: HAWA is designed to complement established AI governance rather than compete with it. NIST's AI RMF organizes risk management around Govern, Map, Measure, and Manage, while its 2024 Generative AI Profile extends this logic to risks distinctive to generative systems (Autio et al., 2024). HAWA adds a human-outcomes lens: mapping includes lived work and stakeholder consequences; measurement includes multidimensional well-being; and management includes redesign of tasks, relationships, skills, and boundaries. This alignment is deliberate. Human well-being should not become a parallel ethics exercise detached from operational risk management; it should be one of the outcomes through which trustworthy deployment is judged.

LIMITATIONS

This paper is conceptual and therefore does not establish causal relationships. The eight dimensions require empirical validation across cultures, sectors, job types, and levels of AI maturity. Some dimensions overlap, and their relative importance may vary. Financial security may dominate during restructuring, while cognitive and professional dimensions may be more salient in stable expert work. Measurement also creates methodological challenges. Self-report data can capture lived experience but may be influenced by expectations and organizational climate. Behavioral indicators can add evidence but risk becoming intrusive. Future research should therefore use mixed methods, longitudinal designs, and participatory measurement strategies. The framework is intentionally technology-agnostic. This increases transferability but means that specific applications—biometric monitoring, autonomous systems, generative AI, decision engines, robotics—will require additional technical and legal controls. HAWA should complement, not replace, domain-specific safety, security, privacy, and regulatory frameworks.

FUTURE RESEARCH AGENDA

Future research should develop and validate multidimensional scales for AI-related well-being, test the propositions proposed here, and examine whether different AI configurations create distinctive well-being profiles. Longitudinal studies are particularly important because novelty effects may produce early enthusiasm or anxiety that changes with experience. Comparative research should investigate cultural variation. Concepts such as autonomy, privacy, hierarchy, collective belonging, and meaningful work are interpreted differently across societies. A globally useful framework must allow local expression without abandoning core protections of dignity and agency. Researchers should also examine distribution. Average effects can conceal important differences by age, gender, disability, occupation, contract status, digital skill, or economic vulnerability. The central empirical question should not only be whether AI 'works,' but for whom, under what conditions, and at what human cost. A further research priority is the relationship between AI assistance and long-term expertise. Short-term productivity gains are easier to measure than changes in judgment, creativity, memory, or professional identity. Organizations need evidence about when AI functions as a tutor that develops capability and when it becomes a

crutch that erodes it. Finally, future studies should evaluate interventions: participatory design, AI literacy, protected dissent, gain-sharing, human-contact safeguards, right-to-disconnect policies, and periodic no-AI practice. The field will advance when it can identify not only risks but practical organizational choices that reliably improve both performance and flourishing.

CONCLUSION

TECHNOLOGY SHOULD LEAVE PEOPLE MORE HUMAN, NOT LESS

The decisive question of the AI era is not whether machines will become more capable. They will. The decisive question is what humans and institutions choose to do with that capability. Technology can compress time, expand access to knowledge, reduce dangerous work, personalize support, and amplify creativity. It can also centralize control, intensify work, weaken skills, isolate people, and make consequential decisions harder to understand. The Human-Centered AI Well-Being Architecture offers a way to keep the human consequences visible. By examining physical, cognitive, emotional, transcendent, social, professional, financial, and technological well-being, organizations can move beyond the narrow equation of innovation with efficiency. The framework does not ask leaders to choose between performance and people. It asks them to design performance through people, with people, and for people.

Responsible AI therefore becomes a leadership discipline as much as a technical one. Governance defines boundaries; design shapes agency; culture determines whether dissent is heard; measurement determines what becomes visible; and distribution determines whether progress feels legitimate. These choices create the moral architecture around the algorithm. The ambition should be higher than harmless technology. The goal should be technology that increases human capability while protecting dignity, connection, meaning, and freedom. The most successful AI-enabled organizations will not be those that automate the greatest number of tasks. They will be those that understand which tasks should disappear, which capabilities should grow, which relationships must remain human, and how technological power can be converted into shared well-being. In that sense, the future of AI is not only a question of artificial intelligence. It is a test of human intelligence: our capacity to decide what deserves optimization, what must be protected, and what kind of future is worth building.

PRACTICAL CHECKLIST FOR ORGANIZATIONS

- **Purpose:** Can we state the human and organizational problem this AI use case is meant to solve without mentioning the technology?
- **Stakeholders:** Have we identified everyone who may benefit, carry risk, or lose discretion because of the system?
- **Agency:** Is it clear which decisions are automated, advised, human-authorized, or contestable?
- **Skills:** Which human capabilities will this system strengthen, and which may deteriorate through disuse?
- **Workload:** Will saved time become recovery, learning, better service, or simply more volume?

- **Relationships:** Are we automating an interaction whose human value exceeds its transactional cost?
- **Fairness:** Have we examined differential impacts and created an accessible appeal path for consequential decisions?
- **Privacy:** Are data collection and monitoring proportionate to the legitimate purpose?
- **Meaning:** Does the redesigned role preserve a credible sense of contribution, mastery, and responsibility?
- **Security:** Do affected people understand how AI may influence their income, role, evaluation, or career?
- **Transparency:** Can users understand the system's purpose, limits, uncertainty, and escalation route?
- **Trust:** Are we designing for calibrated trust rather than maximum adoption?
- **Participation:** Have frontline users had real influence before deployment?
- **Dissent:** Can employees report problems without fear of punishment?
- **Measurement:** Are we tracking quality, risk, learning, and the eight well-being dimensions alongside productivity?
- **Distribution:** Who captures the gains, and how are transition costs shared?
- **Human contact:** Have we identified interactions that should remain meaningfully human?
- **Boundaries:** Are there reasonable rights to disengage and periods without algorithmic monitoring?
- **Review:** Is there a defined point at which we will reconsider, pause, redesign, or withdraw the system?
- **Learning:** What evidence will tell us that the technology is making people more capable rather than merely more dependent?

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

Funding: No external funding was received for this conceptual study.

Data availability: No new empirical dataset was generated or analyzed because this article is a conceptual integrative review.

Ethics statement: Ethics approval was not required because the study did not involve human participants, identifiable personal data, or animal subjects.

AUTHOR NOTE

Ignacio Bonasa Alzuria works at the intersection of humanistic leadership, organizational transformation, learning through the arts, happiness, and well-being. His current research agenda explores how institutions can combine technological innovation with human purpose, multidimensional well-being, and responsible leadership. The conceptual framework presented in this manuscript is intended as a foundation for subsequent empirical validation and cross-cultural research.

REFERENCES

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, 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>
- 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>
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. NBER Working Paper No. 31161. <https://doi.org/10.3386/w31161>
- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraye, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Technology & Operations Management Unit Working Paper.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- International Labour Organization. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. ILO Working Paper 96. <https://doi.org/10.54394/FHEM8239>
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. <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). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.600-1>
- OECD. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). Organisation for Economic Co-operation and Development.
- OECD. (2024). The impact of artificial intelligence on productivity, distribution and growth. OECD Artificial Intelligence Papers.
- 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>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Free Press.
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- World Health Organization. (2022). WHO guidelines on mental health at work. World Health Organization.
- World Health Organization & International Labour Organization. (2022). Mental health at work: Policy brief.

- World Health Organization and International Labour Organization.
- Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.600-1>
- OECD. (2023). OECD Employment Outlook 2023: Artificial intelligence and the labour market. OECD Publishing. <https://doi.org/10.1787/08785bba-en>
- OECD. (2024a). Using AI in the workplace: Opportunities, risks and policy responses. OECD Artificial Intelligence Papers, No. 11. OECD Publishing. <https://doi.org/10.1787/73d417f9-en>
- OECD. (2024b). Job Creation and Local Economic Development 2024: The geography of generative AI. OECD Publishing.
- OECD. (2025). The effects of generative AI on productivity, innovation and entrepreneurship. OECD Artificial Intelligence Papers. OECD Publishing.
- 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>
- Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.
