Human Systems Engineering and Design (IHSED 2026), Vol. 237, 2026, 777–790
https://doi.org/10.54941/ahfe1008216

Human Authority in the Age of Intelligent Systems: A Sociotechnical Framework for Decision Assurance and Psychosocial Safety

Dr Peta Chirgwin

Chameleon Mettle Group, Perth, WA, Australia

ABSTRACT

The rapid integration of intelligent systems into high-risk industries is reshaping how work is performed, decisions are made, and responsibility is distributed. While automation and artificial intelligence promise gains in efficiency, safety, and productivity, they also introduce under-recognised sociotechnical risks, including degraded human decision authority, cognitive overload, loss of practical expertise, and increasing psychosocial strain associated with uncertainty and workforce transition. This paper argues that current approaches to human factors engineering and system design are insufficient to address these conditions. While technical assurance frameworks for AI-enabled systems continue to mature, equivalent structures for ensuring the integrity, clarity, and sustainability of human decision-making remain underdeveloped. This gap presents a critical risk in high-reliability environments where human judgement is essential for managing variability, uncertainty, and system failure. Drawing on applied experience in automation-intensive operations and informed by an emerging global research agenda, this research introduces the **Human Authority Index (HAI)** as a novel sociotechnical framework developed by the author for assessing and designing decision-making in human–machine systems. The HAI positions human authority as a safety-critical variable across three dimensions: Decision Clarity, Cognitive Sustainability, and Psychosocial Safety. Together, these dimensions extend traditional human factors approaches by integrating organisational psychology, sociotechnical systems theory, and resilience engineering into a unified model. Through conceptual analysis and industry-informed scenarios, the paper demonstrates how poorly integrated automation can create hidden system fragility, while systems designed with clear human authority exhibit greater resilience and workforce engagement. This research contributes a practical framework to support safe, effective, and meaningful human participation in increasingly autonomous environments.

Keywords: Human systems integration, Sociotechnical systems, Human decision authority, Cognitive systems engineering, Automation, Psychosocial safety, Resilience engineering, AI governance

INTRODUCTION

The increase in the deployment of intelligent systems across high-risk sectors such as mining, energy, construction and manufacturing is fundamentally revolutionising the way we work. Automation tools are now able to handle

Received May 15, 2026 Revised June 28, 2026 Accepted July 13, 2026 Available online September 25, 2026

complex tasks, generating predictive insights and supporting important operational decisions at unprecedented speed and scale. In the resources sector alone, autonomous haulage systems and remote operations centres as well as AI-assisted condition monitoring have become vital for production decisions as well as risk management (Burgess-Limerick et al., 2024). While these advancements have significantly increased throughput and reduced certain categories of human error, they also fundamentally alter the way humans interact with technology and where human authority resides.

Human Factors Engineering (HFE) has historically aimed to improve the interaction between human and system in terms of human usability, workload management and error reduction (Wickens et al., 2015). As systems become increasingly autonomous, however, the role of the human operator gradually shifts from active controller to supervisory monitor and exceptions manager (Parasuraman and Riley, 1997). Bainbridge (1983) identified this transition as an irony of automation in that the very system that was designed to remove human error, at the same time, causes humans to be placed in the same types of conditions that are most likely to produce these errors, including skill atrophy, reduced situational awareness and lowered sense of responsibility. More recently, Endsley (2017) demonstrated that increasing automation does not eliminate the need for human judgement, rather, it changes the environment in which that judgement is to be exercised, usually under time pressure and with degraded informational support.

Despite such progress, the system designs continue to favour technical performance over the integrity of human decision making. Current frameworks for evaluating human-machine interaction address workload, situational awareness, and automation capabilities, but do not provide a formal definition or measure for human decision authority as a system property. In high reliability environments such as those discussed within this paper, this omission carries significant risk. When human authority is ambiguous, operators cannot exercise effective judgement when a system fails, the unexpected happens, or when events occur outside of the scope of the automated proficiency. The result is a class of latent organisational risk that standard technical processes are ill equipped to catch.

This paper addresses this gap by introducing the Human Authority Index (HAI), a sociotechnical framework for assessing and designing human decision authority within intelligent systems. The paper proceeds as follows: Section 2 describes the literature landscape and presents key points as well as identifying existing frameworks' limitations for authority allocation. Section 3 introduces the HAI framework, defines its three constituent dimensions, and describes a preliminary operationalisation approach. Section 4 presents a structured illustrative scenario demonstrating HAI application in the context of an autonomous mining operation. Section 5 presents implications for system design, AI governance, and future empirical development. Section 6 ends with some thoughts on the role of human authority in the age of intelligent systems.

THEORETICAL BACKGROUND AND RESEARCH GAP

Sociotechnical systems theory holds that technical and social components of work systems are interdependent and must be jointly optimised to achieve effective, safe and sustainable performance (Trist and Bamforth, 1951; Baxter and Sommerville, 2011). This principle has become a guiding line of approach for Human Systems Integration in industrial settings for decades. However, the rise of intelligent systems including AI and automation, introduces a new form of complexity that challenges the classic sociotechnical design rules. Decision making is no longer just a human operator's job, but is dispersed through algorithms, sensors, interfaces, workflows and organisational processes in ways that can obscure the boundaries of authority and accountability (Leveson, 2011).

The Levels of Automation (LOA) model by Parasuraman et al. (2000) describes a continuum of fully manual systems to fully automated decisions, providing a structured basis for allocating tasks between human and machine. Although the LOA framework has been influential, it is mostly focused on aviation and process control, and does not specifically discuss how authority is exchanged from one actor (human and system) to another with respect to the system, nor how operators experience and sustain those interactions over time. Billings (1997) built on this idea through the concept of human-centred automation and claimed that the human operator still has a right to govern even if a system functions autonomously. In terms of how this has been used within industry, however, Billings' principles have not been operationalised as a measurement framework.

Situation Awareness theory (SA), most fully developed by Endsley (1995; 2017), is primarily used as an operational approach to the way that operators view, understand, and perceive the condition of a system. SA deterioration has been found to be a contributing factor to a number of automation related high consequence accidents (Endsley, 2017). Although SA literature richly describes the circumstances in which human judgement becomes compromised, it does not articulate the organisational connection between the design of the automation and the distribution and assignment of decision authority. For example, an operator could have accurate situational awareness yet still lack a clear sense of their authority to act. The two constructs are related but distinct.

By employing resilience engineering (Hollnagel 2014; Dekker 2006), the emphasis of safety analysis shifts away from the failure avoidance system to adaptive capacity, focusing on factors under which operators manage variability and surprise well. Hollnagel's Safety-II approach recognises that performance variability is not intrinsically pathological and resilient systems are a function of the adaptive capacity humans demonstrate on a daily basis. Nevertheless, resilience engineering has yet to develop operational tools for assessing whether automation designs preserve or erode the conditions necessary for effective human adaptation, in particular when it comes to authority and accountability.

Research on workplace psychosocial risk has shown that uncertainty, lack of control and role ambiguity are leading causes of occupational stress, underachieved performance, and work disengagement (Dollard and

Bakker, 2010; Karasek, 1979). Dollard and Bakker (2010) constructed the Psychosocial Safety Climate (PSC) model, which translates organisational-level polices, practices and procedures that protect worker psychological health. Although PSC is an acceptable organisational measure, it does not adequately deal with the operator level experience of working within automated systems, specifically, the relationship between system design, perceived decision authority, and psychological wellbeing. Edmondson's (1999) work on psychological safety additionally establishes that supporting, empowering and enabling an employee to take responsibility and to act, are associated with enhanced productivity, learning and reporting of errors. These conditions pertain to the construction of human automation interfaces at high consequence sites.

Previous studies on trust in automation (Lee and See, 2004) have shown that the interaction between human operators and automated systems is neither passive nor static. Operators adjust their dependence on automated recommendations based on perceived systems reliability, predictability, and transparency. Mis-calibrated trust results from either over-emphasis (automation bias) or lack of use (disuse) and can contribute to degraded system outputs. Lee and See (2004) find that appropriate trust can be obtained only through clear system intent and operator competence. The HAI's Decision Clarity and Cognitive Sustainability Dimensions are conceptually compatible with these findings in the allocation of authority questioning.

Collectively, existing frameworks provide rich but incomplete accounts of human performance in AI and automated systems. One clear and significant gap exists in that no current framework or model defines human decision authority as a quantifiable and designable system feature that brings together cognitive, organisational and psychosocial dimensions in a structured assessment model. The Human Authority Index (HAI) is proposed to fill this gap.

THE HUMAN AUTHORITY INDEX (HAI)

To address the gap identified, this study introduces the Human Authority Index (HAI), a structured sociotechnical framework designed for assessing, measuring and designing human authority within intelligent systems (see Figure 1). The HAI conceptualises authority not as a binary condition, but as an emergent system property shaped by the interaction of design, organisational and human factors. It evaluates human authority across three interdependent dimensions: Decision Clarity (DC), Cognitive Sustainability (CS), and Psychosocial Safety (PS). The framework is intended to be applicable at the system design stage, operational review and post-incident analysis phases.

The HAI can be represented as a composite function:

$$HAI = f(DC, CS, PS)$$

where Decision Clarity (DC), Cognitive Sustainability (CS), and Psychosocial Safety (PS) each contribute to an overall measure of human authority in a given system. In practice, the function f can be represented as a weighted composite

score with dimension weights optimised to each of the risks, regulatory environment, and operational characteristics of the system reviewed. For example, in an emergency response context, Decision Clarity may carry a greater weight. In a sustained monitoring role, Cognitive Sustainability may be primary. The framework does not prescribe a single universal weighting but provides a structured basis for context-sensitive evaluation.

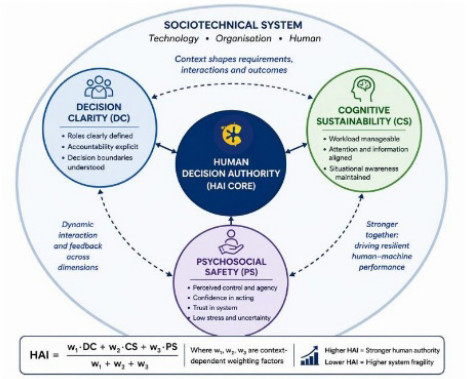

Figure 1: Human Authority Index Framework (HAI) (Chirgwin, 2026). Human decision authority is conceptualised as an emergent property of three interacting dimensions – Decision Clarity, Cognitive Sustainability, and Psychosocial Safety – within a sociotechnical system comprising technological, organisational, and human elements.

Dimension 1: Decision Clarity (DC)

Decision Clarity is the extent to which roles, responsibilities and decision boundaries between human operators and automated systems are clear, explicitly defined, and understood in a sociotechnical system. It includes three closely related components: (1) role clarity-the degree to which each actor in the human-machine system understands their decision making context and its boundaries; (2) system transparency-the degree to which automated systems logic, state and intent are interpretable by the human operator; and (3) override and escalation clarity- whether conditions when human intervene, override or escalate are unambiguous and actionable.

Low Decision Clarity is characterised by ambiguous authority boundaries, opaque automated reasoning ('black-box' decision support), poorly defined escalation triggers, and unclear accountability structures. It is associated with automation bias, delayed intervention, and diffusion of responsibility (Parasuraman and Riley, 1997; Leveson, 2011). High Decision Clarity supports informed operator judgement by ensuring that humans know when to act, what they are authorised to do, and how the automated system arrived at its current state. Indicators for DC assessment may include: documented role definitions, interface transparency ratings, escalation protocol clarity audits, and incident reports attributable to authority ambiguity.

Dimension 2: Cognitive Sustainability (CS)

Cognitive Sustainability reflects the alignment between the demands placed on human operators by a system and their sustained cognitive capacity to meet those demands effectively over time. The construct is informed by cognitive load theory (Sweller, 1988), attention resources theory (Wickens, 2002) and the concept of skill-based, rule-based and knowledge-based performance modes (Rasmussen, 1983). It captures not only instantaneous workload, but the cumulative and longitudinal dimension of cognitive demand; whether operators can sustain adequate performance across a full operational period, over a career, and through periods of heightened system complexity.

Critically, Cognitive Sustainability includes the conditions that sustain decision-relevant expertise over time. Bainbridge (1983) identified that automation can erode the very skills it is designed to supplement, creating a latent fragility in the human component of the system. Cognitive Sustainability therefore encompasses skill maintenance, meaningful engagement with system processes, and the adequacy of training environments to support continued competence. Indicators for CS assessment may include: workload measures (eg., NASA-TLX; Hart and Staveland, 1988), error rates attributable to attentional failures, skill decay indicators, and training currency records.

Dimension 3: Psychosocial Safety (PS)

Psychosocial Safety, as used in the HAI, refers to the subjective and organisationally shaped experience of the human operator working within an automated system. It is distinct from, though related to, the organisational-level Psychosocial Safety Climate construct (Dollard and Bakker, 2010), which assesses the degree to which management policies and practices protect psychological health. In the HAI, PS is operationalised at the human-system interface level and encompasses three sub-constructs. (1) perceived control – the operator's sense that they retain meaningful influence over system outcomes, (2) role identity and purpose- the degree to which operators experience their work as meaningful and their expertise as valued within an automated system, and (3) psychological safety in the team and organisational context- the extent to which operators feel safe to raise concerns, deviate from procedure when necessary, and report errors without fear of blame (Edmondson, 1999).

Low PS environments are associated with reduced operator engagement, increased error concealment, heightened occupational stress, and attrition of experienced personnel, each of which constitutes a safety risk in high-reliability settings (Karasek, 1979; Zohar, 2010). High PS environments support candid reporting, adaptive decision making and sustained workforce capability. Indicators for PS assessment may include: validated psychosocial survey instruments, workforce turnover, and absenteeism data, safety reporting rates, and qualitative assessments of supervisor-operator communication norms.

While this paper presents the HAI as a conceptual framework, a preliminary operational path is outlined. The DC, CS, PS dimensions will be measured using a structured instrument combining observable indicators (i.e. documentation audit, workload measurement) and self-reported operator data. All dimensions are rated on a standardised scale, and a composite HAI score is computed using context specific weighting. High composite scores suggest high human authority while low scores indicate some risks that warrant design or organisational action. Empirical validation of this instrument is identified as a priority for future research. The operationalised instrument, comprising four dimensions and fourteen consequence-weighted scoring sub-dimensions, is documented in Appendix A.

APPLICATION IN HIGH-RISK INDUSTRIES

To illustrate the application of the HAI framework, this section provides an illustrative, structured scenario taken from the context of an autonomous surface mining operation. The scenario is illustrative rather than empirical, created through publicly documented patterns of human behaviour in the resources sector (Burgess-Limerick et al., 2024; Chirgwin, 2021). It aims to demonstrate how HAI domains can diagnose authority-based risks and inform design recommendations. This scenario describes two opposing organisational configurations; a low HAI environment and a high HAI environment to highlight the practical consequences of varying levels of human authority.

Scenario Context: Remote Operations Centre, Autonomous Haulage System

A remote operations centre (ROC) oversees a fleet of autonomous haulage trucks at a large open pit mine. In addition to other tasks, operators monitor truck telemetry, production throughput, and safety zone incursion alerts across multiple screens simultaneously. Fleet management systems govern truck routing, speed regulation, production throughput and collision avoidance. Operator intervention is required when (a) the system generates a fault alarm; (b) trucks operate outside their automated envelope or a human operated machine interrupts the autonomous operation, or (c) a hazard is detected. Operators work 12 hour shifts, in front of the systems.

Low HAI Configuration

In the low-HAI configuration, the ROC was designed primarily around system efficiency metrics. *Decision Clarity is low*: operator roles and decision boundaries are generic within the safety management plan and not reflected in the interface design or daily operational protocols. The fleet management system's decision logic is not actively displayed to operators. Trucks stop or re-route without explanation and reasoning may require investigation. Escalation triggers are not codified in the interface and depend on individual operator interpretation of alarm severity. Operators report uncertainty about when they are 'supposed to' step in as opposed to waiting for the system

to self-resolve. *Cognitive Sustainability is low:* operators manage over 20 trucks and other mobile equipment, with simultaneous truck feeds during peak production and alerts frequently exceeding comfortable attentional capacity. Operators report many alarms are false positives leading to increase desensitisation. Opportunities for active control are rare, resulting in skill erosion among operators who rarely visit site, or rarely perform manual override procedures. **Psychosocial Safety is low**: operators describe feeling like babysitters to the system or report being handed tasks not aligned to role due to perception of decision authority within the role. Operators report a sense of limited professional agency. Near-miss events are infrequently reported due to perceived blame culture and uncertainty about whether intervention decisions will be scrutinised. Composite HAI assessment indicates significant authority risk across all three dimensions.

High-HAI Configuration

In the high-HAI configuration, the same operational context is supported by a redesigned interface and revised organisational protocols informed by human factors review. Decision Clarity is high: operator authority boundaries are explicitly mapped to interface states. When a truck triggers an anomalous condition, the system displays the reason for automated action, the current system recommendation, and the specific operator authority options available (observe, intervene, escalate). Escalation pathways are embedded in the interface with single action access. Cognitive Sustainability is achieved through alert rationalisation (i.e. the number of simultaneous alarm categories has been reduced from 11 to 4 priority tiers, with lower priority events consolidated in summary dashboards rather than individual alerts). Operators rotate through active monitoring and higher-level decision-making task-based work within each shift to maintain engagement and skills currency. Periodic simulation exercises maintain manual override proficiency. Psychosocial Safety is supported by a fair and just culture framework. For example, operators are explicitly informed that intervention decisions made in good faith will not attract blame. Near miss reporting rates have increased by 34% since the redesign, consistent with improved psychological safety (Edmondson, 1999). Operators describe clearer sense of professional identity, respect, and decision ownership. Composite HAI assessment indicates robust human authority across all three dimensions.

This scenario illustrates that the conditions characterising low and high HAI environments are not determined solely by automation technology but by integrated design of technical systems, organisational protocols, and culture. The HAI provides a structured vocabulary and assessment process for identifying which dimensions are at risk and where targeted intervention is most warranted.

DISCUSSION

The HAI extends existing human factors and sociotechnical frameworks by introducing authority as a measurable and designable system property. Unlike the Levels of Automation framework (Parasuraman et al., 2000),

which classifies task allocation along a manual-automated continuum, the HAI evaluates the quality and sustainability of human authority within that allocation. It builds on Billings' (1997) human-centred automation principles through a structured, multi-dimensional assessment mechanism. It complements Situation Awareness theory (Endsley, 1995) by addressing the structural conditions shaping operator judgement and operationalises at the human-system interface level the psychosocial insights of Dollard and Bakker (2010) and Edmondson (1999), previously applied primarily at the organisational climate level. The framework also aligns with emerging AI governance requirements in both the EU AI Act (2024) and ISO/IEC 42001, mandating demonstrable human oversight in high-risk systems, and the HAI provides a practical structure to support such compliance.

Several limitations must be acknowledged. The HAI remains a conceptual model. Dimensions were derived through theoretical synthesis and practitioner experience rather than empirical validation, and the illustrative scenario cannot be treated as confirmatory evidence. Generalisability beyond the heavy industry context - to healthcare, aviation, or financial services, for example - requires dedicated investigation. Authority is also a dynamic construct, and a longitudinal assessment warrants future development. Priorities for subsequent research include psychometric validation of the HAI instrument (Appendix A), empirical testing of dimension scores against safety outcomes, cross-industry application, and exploration of approaches suitable for continuous operational monitoring.

CONCLUSION

As intelligent systems become more capable and autonomous, human authority within those systems becomes simultaneously less visible and more consequential. The HAI responds to this condition by providing a structured, theoretically grounded framework for assessing and designing human decision authority across three dimensions of Decision Clarity, Cognitive Sustainability, and Psychosocial Safety. In doing so, it offers practitioners, designers, and organisations a practical vocabulary for identifying where authority is at risk and where intervention is warranted: with direct relevance to both operational safety management and emerging AI governance requirements.

The challenge of automation is not ultimately technical. It is the challenge of designing systems in which human beings remain capable, accountable, and empowered to exercise judgement, even as the systems around them grow more capable and autonomous. The Human Authority Index is offered as a contribution towards meeting that challenge.

ACKNOWLEDGMENT

The author acknowledges practitioners and operational leaders in the Australian resources sector whose applied experience informed the development of this framework.

REFERENCES

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779.

Baxter, G. and Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1), 4–17.

Billings, C.E. (1997). *Aviation automation: The search for a human-centred approach*. Lawrence Erlbaum Associates.

Burgess-Limerick, R. et al. (2024). *Human Aspects of Automation and New Technology in Mining: Integrating People and Technology Through Human-Centred Design*. University of Queensland/ACARP.

Chirgwin, P. (2021). Skills development and training of future workers in mining automation control rooms. *Computers in Human Behavior Reports*, 4, 100115.

Dekker, S. (2006). *The field guide to understanding human error*. Ashgate.

Dollard, M.F. and Bakker, A.B. (2010). Psychosocial safety climate as a precursor to conducive work environments, psychological health problems, and employee engagement. *Journal of Occupational and Organizational Psychology*, 83(3), 579–599.

Edmondson, A.C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.

Endsley, M.R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, 59(1), 5–27.

European Parliament (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union.

Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). North-Holland.

Hollnagel, E. (2014). *Safety-I and Safety-II: The past and future of safety management*. Ashgate.

ISO/IEC (2023). *ISO/IEC 42001:2023 Information technology – Artificial intelligence – Management system*. International Organization for Standardization.

Karasek, R.A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24(2), 285–308.

Lee, J.D. and See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.

Leveson, N.G. (2011). *Engineering a safer world: Systems thinking applied to safety*. MIT Press.

New South Wales Parliament (2026). Work Health and Safety Amendment (Digital Work Systems) Act 2026 (NSW).

Parasuraman, R. and Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.

Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3), 286–297.

Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3), 257–266.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.

Trist, E.L. and Bamforth, K.W. (1951). Some social and psychological consequences of the Longwall method of coal-getting. *Human Relations*, 4(1), 3–38.
Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159–177.
Wickens, C.D., Hollands, J.G., Banbury, S. and Parasuraman, R. (2015). *Engineering Psychology and Human Performance*. Routledge.
Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis and Prevention*, 42(5), 1517–1522.

APPENDIX A: OPERATIONALISATION OF THE HAI – FROM THEORETICAL DIMENSIONS TO SCORING INSTRUMENT

This appendix documents the operationalisation of the Human Authority Index (HAI) from the conceptual framework presented in the body of this paper to its current instrument form. Two developments are recorded. First, the three theoretical dimensions – Decision Clarity (DC), Cognitive Sustainability (CS), and Psychosocial Safety (PS) – have been operationalised as scored sub-dimensions in a deployed assessment instrument. Second, a fourth dimension, Work Design (WD), has been added to complete the framework, and the PS dimension has been expanded to incorporate obligations arising from the New South Wales Work Health and Safety Amendment (Digital Work Systems) Act 2026. These developments are additive: they do not alter or supersede the three-dimension theoretical construct presented in the body of the paper, which remains the lens for understanding why human authority matters. Work Design addresses whether the system was built to support it.

The Four-Dimension Instrument

The operationalised instrument comprises four theoretical dimensions, each weighted equally at 25 points within a 100-point composite score. Decision Clarity, Cognitive Sustainability, and Psychosocial Safety retain the definitions given in the body of the paper. Work Design (WD) is defined as the extent to which the work system – tasks, roles, interfaces, and competency architecture – has been deliberately designed to preserve and enable human authority, rather than treating human involvement as residual to system capability. The first three dimensions ask whether human authority is present; Work Design asks whether the system was built to allow it. Without WD, the instrument would score authority states without accounting for whether those states were ever achievable by design. A low WD score indicates that targeted work redesign, rather than procedural or training intervention alone, is the appropriate response.

Consequence-Based Weighting

Within each dimension, scoring sub-dimensions are weighted by the safety consequence of failure rather than distributed equally. Override Capability (DC), Operational Skill Currency (CS), and Task and Authority Allocation (WD) carry the highest weights within their dimensions because their failure modes are acute and least recoverable. Weights are context-adjustable at both levels, consistent with the context-sensitive evaluation principle described in the body of the paper.

Psychosocial Safety and the NSW WHS Amendment (Digital Work Systems) Act 2026

The NSW Work Health and Safety Amendment (Digital Work Systems) Act 2026 (s. 21A) creates explicit duties for persons conducting a business or undertaking to assess whether algorithmic work allocation creates excessive workloads, unreasonable performance metrics, excessive surveillance, or

discriminatory decision-making. The Surveillance and Monitoring Integrity sub-dimension (PS-4) operationalises these obligations directly: a low PS-4 score indicates a potential exposure under the Act, providing a bridge between the HAI assessment and emerging Australian regulatory compliance obligations.

Complete Scoring Reference

Table A1 lists all fourteen scoring sub-dimensions across the four theoretical dimensions, with the governance question each addresses and its consequence-based point allocation (total: 100 points).

Table A1: HAI scoring sub-dimensions, governance questions, and consequence-based point allocations (Chirgwin, 2026).

#	Scoring Sub-Dimension	Governance Question	Dim.	Pts
1	Override capability	Can a human meaningfully override the system output, and is that pathway documented and tested?	DC	8
2	Decision ownership	Is there a named human who owns this AI decision outcome – and who knows it?	DC	7
3	Escalation architecture	Are escalation triggers defined, tested, and owned before an incident occurs?	DC	6
4	Explainability access	Does the accountable executive understand enough of the system logic to exercise judgement?	DC	4
5	Operational skill currency	If this system fails or produces unexpected output, do operators retain the practised capability to intervene effectively – now, not theoretically?	CS	10
6	Cognitive demand calibration	Has the cognitive load this system places on operators been assessed, and is it within limits that sustain reliable decision-making under pressure?	CS	8
7	Role integrity	Does the human role require genuine expertise and judgement, or has automation reduced it to monitoring without meaningful authority?	CS	7
8	Agency and influence	Do operators experience their decisions as genuinely consequential – or has system design rendered human input performative?	PS	7
9	Organisational safety climate	Do operators feel safe to raise concerns, override procedures, and report errors without fear of blame – and does the organisation actively protect that safety?	PS	7
10	Eudaimonic work quality	Does this work allow operators to exercise meaningful capability, develop expertise, and experience their contribution as purposeful?	PS	6
11	Surveillance and monitoring integrity	Are the metrics, monitoring practices, and performance systems applied to workers through digital systems reasonable, transparent, and free from discriminatory effect?	PS	5

(Continued)

Table A1: Continued.

#	Scoring Sub-Dimension	Governance Question	Dim.	Pts
12	Task and authority allocation	Are decisions requiring human judgement explicitly assigned to humans, not defaulted to the system by convenience or cost?	WD	10
13	Competency architecture	Are human roles meaningful, and are people equipped and trained to exercise the authority they nominally hold?	WD	9
14	Interface and information design	Does the system surface the right information, at the right time, in a form that supports rather than overrides human judgement?	WD	6

A Note on the Integrity of the Framework

The addition of Work Design as a fourth theoretical dimension, the expansion of Psychosocial Safety to four scored sub-dimensions, and the introduction of consequence-based weighting are each additive, documented, and grounded in either the safety literature or emerging regulatory obligations. Future iterations of the HAI instrument will continue to be documented as versioned addenda under a maintained version history.