

# Human, AI, and Organizational Performance (HAOP)

## A Safety Framework for the AI Era

---

*Extending Human & Organizational Performance for AI-enabled work systems*

**by Jaina Ko, CSP**

Revision 1  
June 14, 2026

## Contents

I. Executive Summary .....	3
Part I: The HAOP Framework.....	4
II. The Problem: AI Is Arriving and Is About to Amplify Safety’s Oldest Mistake .....	4
III. The Foundation: What HOP Established, and Where It Stops .....	5
IV. The Three Performers .....	6
V. How the Performers Fail Differently .....	8
5.1 The Failure Signatures.....	8
5.2 The Stakes Are Epistemic: System Information Degrades.....	9
5.3 Applied Case: The Warehouse That Drifted Toward a Fire.....	10
VI. Accountability: Follow Control, Not Visibility.....	11
7.1. Three Forms of Accountability.....	12
7.2. Accountability Must be Mapped to Control .....	12
VII. HAOP Operating Principles.....	14
Part II: Applying HAOP .....	15
VIII. The True Function Test: A HAOP Beginning Diagnostic.....	15
8.1 Purpose of the Test .....	15
8.2 The Nine Diagnostic Questions.....	16
8.3 From Questions to Alignment Mapping.....	16
8.4 Red Flags.....	16
8.5 What the Test Does Not Do.....	17
IX. Grounding and the Ability to Pause .....	17
9.1 Grounding.....	17
9.2 The Ability to Pause .....	18
X. Conclusion.....	19
Appendix A: Key Terms .....	19
Appendix B: True Function Alignment Map.....	21
References .....	23

# I. Executive Summary

---

AI is entering safety-critical work faster than most governance systems were designed to absorb. In EHS and operational settings, AI is not just summarizing documents or retrieving information. It is increasingly classifying risk, prioritizing signals, routing work, recommending action, generating controls, monitoring compliance, and shaping what human beings see, decide, and do.

Most organizations still govern these systems as a tool: a matter of data, accuracy, privacy, cybersecurity, and uptime. Those controls remain necessary, but they are insufficient in those circumstances where AI materially influences operational judgment or workflow execution. At that point, AI becomes a performer within the work system, introducing variability, shaping outcomes, and failing in ways that differ from human error.

Human and Organizational Performance corrected an earlier safety mistake: the belief that workers are the primary variable to control. HOP showed that human behavior is shaped by context, incentives, constraints, leadership response, and the gap between work-as-imagined and work-as-done. HAOP (*Human, AI, and Organizational Performance*) extends that logic for AI-enabled work systems.

HAOP recognizes three interacting performers. The human performer adapts under real operating conditions. The AI performer optimizes based on data, signals, permissions, constraints, and architecture. The organizational performer shapes both through governance, incentives, resources, metrics, authority, procurement, and tolerated tradeoffs.

The central claim of HAOP is simple: safety-critical AI cannot be governed by model performance alone. It must be governed as part of a socio-technical system where human adaptation, AI optimization, and organizational signaling interact.

This paper introduces the HAOP framework, defines the distinct failure signatures of each performer, and proposes a practical diagnostic - the **True Function Test** - for evaluating whether an AI-enabled workflow actually produces the safety outcome it claims to pursue, or merely produces a cleaner representation of safety.

HAOP is not a no-blame model. It is an accountability-by-control model. Responsibility should not collapse onto the most visible person at the point of failure. It should be mapped in advance to the people, teams, and functions with control over the relevant action, signal, constraint, permission, resource, metric, deployment decision, verification point, or escalation path.

For EHS professionals, HAOP provides a way to govern AI without abandoning the lessons of HOP. For AI technologists, it identifies operational failure modes that model metrics alone will not reveal. For AI governance leaders, it connects accountability, assurance, human oversight, and drift to real work. For academics, it offers a conceptual bridge between safety science, organizational theory, and AI-enabled socio-technical systems.

The central test is not whether an AI workflow looks modern, efficient, compliant, or data-driven. The test is whether it remains grounded in operational reality before its outputs become consequential.

## Part I: The HAOP Framework

---

### II. The Problem: AI Is Arriving and Is About to Amplify Safety's Oldest Mistake

---

Most workplaces still operate within control-based systems designed for a simpler era, when hierarchy, rules, and compliance were assumed to produce safety. Much of modern industrial management inherited assumptions from Taylor's scientific management, which turned workers into measurable units of labor and treated human error as deviation from an otherwise perfect plan.<sup>1</sup>

For nearly a century we have chased "zero harm" through more audits and stricter rules. The measurement of choice was progress against a lagging metric, Total Recordable Incident Rate. Organizations layered on valuable methods such as the hierarchy of controls, behavior-based safety, and versions of Plan-Do-Check-Act. These tools had real benefits, but they also incentivized corrosive practices that blamed the worker and managed the metric rather than the risk, and in too many cases, manipulating the data outright.

Meanwhile, the gap Hollnagel named as *work-as-imagined* and *work-as-done* kept widening.<sup>2</sup> Culture surveys tried to explain it and more new initiatives got launched. Yet recordable rates plateaued and serious injuries and fatalities persisted, challenging the precept that all accidents were preventable.

Human and Organizational Performance emerged as a correction to this broken frame. Grounded in reality, HOP recognized and accepted something fundamental that the old frameworks denied: *to err is human*. From this basic truth, the five principles were built out: people are fallible, context drives behavior, blame fixes nothing, learning is vital, and leader response matters.<sup>3</sup> By shifting from control to curiosity, organizations could move toward systems that produce what they claim to pursue and toward learning rather than enforcement.

But conditions are changing again. The workforce is shrinking, experienced people are retiring with their knowledge, and organizations are rushing to fill the gap with AI. In EHS work, AI is still governed as an IT and data risk, emphasizing privacy, accuracy, and uptime. The governance problem is not AI adoption itself. The problem appears when AI moves from information support into operational influence: classifying risk, routing work, prioritizing signals, recommending controls, approving actions, or shaping what a human reviewer sees first. At that point, the system is not just storing or displaying information. It is participating in the production of work.

Here is the collision. Deploy AI into a system still shaped by the old assumptions of control and compliance, and AI will amplify that philosophy. If an organization still believes, even quietly, that people are the problem, AI becomes a faster, colder way to monitor, blame, and punish. In a complex, fast-moving environment, automated judgment built on that premise is not merely ineffective. It is dangerous.

### III. The Foundation: What HOP Established, and Where It Stops

---

Human and Organizational Performance established that human behavior is not random defect. It is locally rational performance shaped by context, constraints, incentives, tools, training, leadership response, and operational reality. It moved safety away seeing people as the variable that needed to be fixed and toward understanding and improving the systems in which people work. Todd Conklin's five principles of HOP:<sup>3</sup>

#### 1. Error is normal.

HOP rejects the assumption that perfect compliance is a realistic safety strategy. Humans forget, misread, adapt, get tired, improvise, and act on incomplete information. The goal is not to create perfect workers. The goal is to design systems where predictable human error does not become catastrophic. HOP joined Dekker's "Safety Differently" and Resilience Engineering movements in moving the field away from "fixing the human" and toward being able to fail safely.<sup>4,5</sup>

#### 2. Context drives behavior.

HOP made context central. People do things that make sense to them at the time given their immediate goals, their knowledge, their focus of attention, the norms around them, and the specific pressures of that exact moment. This is local rationality. As Dekker put it, no one goes to work planning on doing a bad job.<sup>4</sup>

#### 3. Blame fixes nothing.

Blame may satisfy the need for consequence, but it hides the conditions that made the event possible, and it suppresses the reporting and weak signals that a system needs to stay informed that something is drifting.

#### 4. Learning is vital.

HOP pushed safety away from only studying failures. The conditions that produce failure are the same conditions that produce success, which makes normal work a critical data source: studying when things go right yields far more data than waiting for them to go wrong.<sup>6</sup> The workarounds, adaptations, informal practices, and quiet compensations show how the system is actually functioning, and they reveal the gap between work-as-imagined and work-as-done.

#### 5. How leaders respond matters.

What leaders do after error matters more than the error itself. A response of curiosity improves learning and lets the contributing factors get corrected. A response of punishment or defensiveness loses the information the system depends on, and the system grows more brittle.

HOP explains how humans perform inside systems. It corrected a deep misunderstanding of human performance and gave us a better way to understand workers, leaders, procedures, context, drift, learning, and accountability in those human-centered systems.

It also made the organization visible, because the organization supplies much of the context that shapes human behavior. And a system executes signals, not objectives. If an organization states an objective of reducing risk but is only measuring TRIR, PPE violations, or throughput, it is performing through the signal architecture it created, which may be working directly against the objective it claims.

But HOP was not designed for a world with another performer in it.

## IV. The Three Performers

---

HOP remains necessary, but it is no longer sufficient on its own for AI-enabled work systems. It explains how humans adapt inside organizational conditions. It does not, by itself, provide a vocabulary for optimization systems that classify, route, recommend, suppress, escalate, or act across workflows.

HAOP - Human, AI, and Organizational Performance - extends HOP when AI crosses a functional (even if somewhat blurry) threshold from tool to performer. AI remains a tool when it supports a bounded human task: summarizing a report, querying a database, drawing a procedure, or responding to a prompt. In those cases, a human initiates the task, a human reviews the output, and the AI's role is narrow and more bounded.

AI becomes performer-level when it no longer just supports a human task, but materially shapes the sequence, priority, visibility, recommendation, approval, routing, or execution of work. At that point, the AI becomes a part of the work system itself, introducing variability and shaping outcomes.

An autonomous AI agent makes this line easier to see. An agent may pursue objectives over time, make decisions, and act without human initiation at each step. When it fails, it does not fail through human-like intention, fatigue, fear, or local rationality. It fails through local optimization mechanisms such as mis-specified goals, weak constraints, degraded data, excessive permissions, tool misuse, goal hijacking, or cascading action. This is not theoretical. The tech industry itself recognizes the instability of these tools. When OWASP's Agentic Security Initiative published its *Top 10 for Agentic Applications* (2026 edition), it documented 10 failure categories specific to agents, including goal hijacking and cascading failures<sup>7</sup>

AI remains a tool in many use cases. But when an AI system materially shapes what people see, decide, prioritize, approve, route, or do, it becomes a performer within the work system.

HOP taught safety leaders to ask how the system looked to the worker. HAOP adds two additional questions: how did the system look to the AI, and what did the organization design the system to optimize?

HAOP uses "performer" functionally, not morally. A performer is any entity or organized system that can act, shape outcomes, and introduce variability. A human performer acts through judgment and adaptation. An AI performer acts through model output, tool use, ranking, routing, prediction, or workflow execution. An organizational performer acts through governance, incentives, metrics,

resources, authority, procedures, procurement, and tolerated tradeoffs. Accountability, however, still traces to people, roles, and governing bodies with control.

The organization as performer has grounding in organizational studies and safety science canon. In *A Behavioral Theory of the Firm*, Cyert and March<sup>8</sup> rejected the idea that organizations are neutral containers or that they behave like single rational profit-maximizing actors. In their seminal work, they described how organizations behave like coalitions of participants with multiple (potentially conflicting) goals: production, sales, safety, quality, cost, labor relations, market share, executive priorities, departmental interests, and local incentives.

Vaughan developed the concept of *normalization of deviance* in her study of the Challenger disaster, where O-ring erosion and blow-by warning signs were ignored. The core thesis was that organizations can gradually accept deviant conditions as normal when those conditions repeatedly occur without immediate catastrophe.<sup>9</sup>

Dekker's *Drift into Failure: From Hunting Broken Components to Understanding Complex Systems*<sup>10</sup> argued that complex system failures should not be explained by hunting for one broken component or one bad actor. His work pointed out that people and organizations make small adjustments that seem reasonable at the time. They borrow a little from safety margins to meet production, cost, schedule, staffing, or efficiency demands. Each adjustment appears tolerable because nothing bad happens immediately. Over time, *systems drift* closer to failure because the movement is incremental and does not feel like a major risk decision.

Therefore, naming the organization a performer is not a new claim; it operationalizes what safety science and organizational theory already assert.<sup>11</sup>

HAOP recognizes three interacting forms of performance: human performance, AI performance, and organizational performance. The organization is the containing performer because it designs, constrains, authorizes, measures, and normalizes the conditions under which the human and AI performers operate.

HOP's corrective against worker blame held that humans are not the problem but part of the solution. HAOP keeps the human as the central operational performer, adds AI as a distinct performer that behaves and fails differently, and names the organization as the containing performer responsible for the interaction space between them. In AI-enabled work, there are no longer only human performers inside organizational systems. AI increasingly performs work, and the organization performs the conditions under which both human and AI action becomes possible, rewarded, constrained, or ignored. That makes the organization not background context, but a containing performer.

This is what gives the framework precision when it comes to accountability. One of the critiques of HOP has been that it does not explicitly cite who is accountable, leaving the "out" to be "the system" did it. But the critique emerges when HOP is adopted as a nice new idea, not as the rigorous culture transformation it was designed to be. Accountability follows control, and the ability to create change is an essential element of control. People always control some actions, with hierarchy carrying great influence. AI systems perform delegated functions. Organizations control the conditions, incentives, authority structure, and validation systems that shape both.

- **The human performer:** Already well understood. We perceive, adapt, hesitate, compensate, comply, deviate, speak up, stay silent, and make tradeoffs under real operating conditions, acting through cognition, interaction with physical reality, judgment, and adaptation.
- **The AI performer:** Still being understood. AI classifies, predicts, generates, routes, recommends, approves, prioritizes, escalates, suppresses, and optimizes based on data, signals, constraints, and architecture, acting through model output, tool use, workflow execution, ranking, routing, prediction, or automated decision support.
- **The organizational performer:** Authorizes, funds, measures, rewards, constrains, trains, validates, ignores, normalizes, and assigns accountability, acting through institutional design, incentives, norms, controls, resource allocation, governance, and sanctioned meaning.

The organization is the containing performer because it creates the conditions under which human and AI performance becomes possible, constrained, rewarded, normalized, hidden, or corrected. It contains the two operational performers - the human beings and the AI systems - but it also performs through them.

## V. How the Performers Fail Differently

---

The first truth that HOP surfaced was simple, and so was the reaction to it: humans make mistakes - it is normal. Take a breath. Now let us understand why the mistake happened.

That posture - understanding before blame - extends to all three performers, though for different reasons. The human is owed the breath because blame wounds and silences. The AI and the organization are owed the same analytical patience because rushing to assign cause makes us blind to how the system actually failed. And the three do not fail the same way. You cannot govern what you cannot recognize, so before any framework can help, it is critical to understand each performer's distinct signature of failure.

### 5.1 The Failure Signatures

- **Human failure signatures:** bounded attention, local rationality, fatigue, normalization, silence, automation overreliance, and adaptive overload.<sup>12, 13</sup> HAOP names one emergent failure pattern, *cognitive overrun*: a condition where a worker remains formally accountable for verifying AI output while its rate, density, or ambiguity has exceeded their capacity to verify what matters.
- **AI failure signatures:** wrong-signal optimization, *confident incompetence* (fluent, authoritative, or decisive outputs that are wrong), context blindness, specification failure, data degradation/compression loss, and speed-scale amplification.<sup>14</sup>
- **Organizational failure signatures:** signal-objective mismatch, accountability gaps, procurement ahead of governance, symbolic oversight, normalized deviance, compliance theater, and under-resourced controls.

There are many more ways to fail than can be listed here, certain failure signatures are distinct to the performer involved. In broad terms, humans adapt, AI optimizes, and organizations signal.

AI failures are often misread through anthropomorphic language, as if the system were tired, confused, careless, intentional, or exercising judgment. This framing is misleading. AI failures may resemble human mistakes at the surface, but their mechanisms are different: optimization against the wrong signal, weak constraints, degraded or incomplete data, flawed architecture, brittle prompting, excessive permissions, poor integration, inadequate verification, or governance that gives the system more influence than its design can safely support.

This matters because anthropomorphic language misdirects the response. It turns a design, deployment, constraint, verification, or governance problem into what sounds like an AI “behavior” problem. The result is over reliance before failure, misdiagnosis after failure, and accountability drift away from the humans and organizations that controlled the system’s role, authority, and operating conditions.

When human adaptation, AI optimization, and organizational signaling become misaligned, the system can move toward failure while still appearing functional. This leads to *compound drift*, where the misaligned signals reinforce one another until the work system moves away from its intended function while still appearing to operate as designed.

## 5.2 The Stakes Are Epistemic: System Information Degrades

The current AI replacement narrative assumes that expertise can be extracted from humans, embedded into software, and then scaled while the human labor system is reduced or removed. In safety-critical work, that assumption is structurally dangerous.

AI systems do not create new operational knowledge from nothing. They compress, abstract, and reproduce patterns from the data and feedback they are given. Without continual grounding (refreshing) in real-world human expertise, these systems can narrow toward the average, lose sensitivity to rare cases, and erase the low-frequency signals that safety work exists to detect.<sup>15</sup>

Degrading information may not be an entirely new problem created by AI. Information passed from person to person and department to department would degrade, with tiny errors in transmission amplifying as it moved along. AI information, though, often rests on the assumption that it is hardcoded and unchanging while being transmitted. The assumption includes the belief that AI can absorb the intellect and experience of humans and keep it in a stable form - in perpetuity. These assumptions are a trap.

This matters because safety-critical work often depends on weak, contextual, and tacit signals: the mechanic who recognizes an unusual vibration, the operator who knows a machine has a non-OEM part, the supervisor who has seen a failure pattern before, or the experienced worker who senses the conditions “do not look right” despite normal instrumentation. These signals are difficult to formalize as data, but they are often the difference between early intervention and serious harm.<sup>16</sup>

The risk scales into organizational design. When companies replace entry-level and intermediate roles with AI, they do more than cut headcount - they weaken the pipeline that produces future senior judgment. Senior expertise is not instant; it develops through years of field exposure, minor

mistakes, tacit learning, and contact with real conditions. An organization that extracts existing expertise into AI while eliminating the roles where future expertise forms creates a double collapse: degradation of the informational system AI depends on, and degradation of the human system that keeps the information grounded.<sup>17</sup>

AI also changes the time dimension of safety. Traditional drift develops slowly enough for humans to notice anomalies, huddle, improvise, and intervene. AI performers are built for speed and scale. When one acts on a distorted representation of reality, the latency buffer disappears, and the system can move from normal operation to significant consequence faster than human adaptation can respond.

A central governance failure is the aesthetic illusion: mistaking a clean, fluent, low-noise artifact for operational competence. In AI-enabled work, the output may appear complete while remaining weakly grounded, insufficiently verified, or disconnected from the work's true function.

The recent “tokenmaxxing” episode at Amazon reflects this failure mode. The intended outcome was better development work through AI use, but the measurable signal became AI spend.<sup>18,19</sup> Safety has long seen the same pattern when injured employees are assigned restricted work onsite to avoid a lost-time incident. The proxy begins to displace the purpose.

When the signal becomes the target, the system produces the signal.

For HAOP the implication is direct: AI cannot be governed as a tool once it becomes a performer in safety-relevant work. Human expertise must remain an active grounding mechanism, not a training input to be extracted and discarded. The governance question is not only whether AI output is accurate today, but whether the organization has preserved the human, technical, and organizational feedback loops that keep the system anchored to operational reality over time.

The stakes are epistemic, operational, and human: replace the people who ground the system in reality, and you risk corrupting what the system knows, accelerating how it fails, and destroying the expertise needed to recover.

### 5.3 Applied Case: The Warehouse That Drifted Toward a Fire

A fire captain described a scenario that shows how AI optimization can create risk without any obvious bad decision in the moment.

An AI-driven warehouse system is built to increase storage capacity and throughput. It continuously adjusts layouts, tightens storage patterns, and reallocates space to demand. Each change is small. Each snapshot looks efficient. From the dashboard, the system appears to be working exactly as intended.

But it was optimized for space and throughput, not fire protection. Because critical safety constraints were never built in, the optimization gradually creates conditions the dashboard cannot recognize as dangerous: sprinkler clearance drops below required thresholds, egress paths narrow during peak operations, fuel-load density concentrates. No alarm sounds. No one makes an obviously reckless decision. The system is doing exactly what it was designed to do: maximize the signal it was given.

Then a fire starts. Sprinklers underperform because clearance is obstructed. Egress is slower because pathways narrowed. Concentrated fuel load accelerates the event. What looked highly optimized becomes a major loss.

This is not ordinary human drift. It is unconstrained optimization. The system did not gradually ignore a rule it understood - it was never designed to treat fire protection as a non-negotiable constraint. Traditional HOP explains how human performers drift under pressure, adapt locally, and normalize degraded conditions. AI performers fail differently: they do not rely on judgment, unease, or contextual hesitation. They optimize the specified signal unless explicit constraints prevent them.

The organizational performer is central here too. The organization chose the objective, approved the system, defined the success metric, and failed to specify the safety boundary. The failure was not only in the AI. It was in the organizational design that let efficiency become the dominant signal without embedding fire protection as a hard constraint.

## VI. Accountability: Follow Control, Not Visibility

---

The oldest objection to systems thinking is that it dissolves accountability: if the system did it, no one is responsible. One of the founding principles of HOP is “blame fixes nothing.” Some have read this as there being no accountability at all. By blaming “the system,” individuals have sought to evade responsibility for their actions. Blame is the act of pinning accountability as consequence on one primary performer. It is narrow, focused, and (sometimes intentionally) creates blindness to other areas of accountability.

HAOP answers this problem directly by naming the organization a performer. This adds accountability. It does not remove it. The organization is not a scapegoat or a ghost. It acts through governance, incentives, resources, procedures, metrics, and constraints.

Accountability should follow control, not visibility. While HAOP recognizes three performers, accountability ultimately traces to humans, roles, teams, functions, officers, or governing bodies with control. Workers hold control over choices within their authority and capacity. Technical teams hold control over how the AI was selected, validated, constrained, and monitored. Supervisors hold control over how work is assigned, how exceptions are escalated, how controls are verified, and whether weak signals are acted on or normalized. Senior leaders hold control over resources, priorities, incentives, and risk tolerance. Executives hold control over the management systems: assurance, authority, oversight, and acceptable tradeoffs.

This is not theoretical. New Zealand courts have already done exactly this. In *Gibson v Maritime New Zealand*, the former chief executive of Ports of Auckland was convicted for failing to exercise due diligence as an officer after a worker was killed by a falling container. In March 2026 the High Court upheld his conviction, the NZ\$130,000 fine, and a NZ\$60,000 costs award. Ports of Auckland was separately held accountable as the operating entity. The significance is not that every executive is automatically liable for every incident. It is that officer accountability was traced to governance-level control: whether critical risks were understood, whether controls were effective in practice, and whether assurance processes verified work as actually done.

These are not mutually exclusive shares of one responsibility. They are distinct control obligations. Each is accountable within the boundary of the control it actually held: its authority, knowledge, resources, ability to intervene, and duty to verify. When a worker is held to account, that does not discharge the executive whose decisions shaped the conditions; both held control, and both answer for it.

## 7.1. Three Forms of Accountability

Accountability in work systems can be separated into three distinct forms: consequence accountability, ownership accountability, and design accountability.

**Consequence accountability** is reactive. It assigns responsibility after failure - who is disciplined, cited, sued, removed, retrained, or otherwise held responsible.

**Ownership accountability** is personal. It depends on the an individual or leader voluntarily accepting of responsibility. When someone says, “The buck stops with me,” it is valuable and respectable. But it is unreliable on its own because it rests on character, courage, and culture.

**Design accountability** is structural. It operates before failure by assigning responsibility in advance: who controls what, who verifies what, who has authority to pause or stop the workflow, who monitors drift, and who approves deployment. It treats weak signals as attention points, establishes methods for surfacing and escalating them, and builds in the reality that error will occur. Visibility, verification, and intervention are designed into the work rather than improvised after breakdown.

HAOP is built on design accountability. Consequence and ownership accountability will always exist, but only design accountability can be engineered in advance. It prevents responsibility from collapsing onto the person closest to the failure, while the decisions that shaped the conditions of failure remain unexamined.

## 7.2. Accountability Must be Mapped to Control

If design accountability is the goal, control is the map. The question is not only who was closest to the failure, but who had authority over the conditions that made the failure possible: the action, signal, constraint, permission, metric, resource, verification point, escalation path, or deployment decision.

After an incident, the first question is not, “Who made the error?” It is, “What kind of breakdown occurred?”

The failure signature comes first. Was this human adaptation under overload, AI optimization beyond safe constraint, or an organizational signal that contradicted the stated safety objective? The signature identifies the nature of the breakdown.

The accountability trace comes next. Accountability follows control: where control was supposed to exist, who held it, what shaped it, and where it failed.

That trace must move in every direction control existed: downward to the point of use, laterally across design and support functions, and upward to the organizational choices that shaped the work.

It may include the worker's decision, but it cannot stop there. It may also include the supervisor's assignment, the engineer's validation boundary, the product team's optimization target, the procurement decision, the executive's risk tolerance, the metric that rewarded speed over verification, or the governance process that allowed a workflow to become consequential without adequate review, constraint, or pause capability.

This is the hard part. **Accountability-by-control** requires organizational courage because it does not stop at the most visible person. It follows authority, incentive, verification, and permission wherever they sit.

This is also where HAOP aligns with Just Culture. Just Culture protects learning by recognizing that people should not be punished for actions, omissions, or decisions that are reasonable given their training, experience, and operating conditions.<sup>23</sup> It also preserves accountability for reckless conduct, willful violations, gross negligence, destructive acts, or knowing disregard of risk. The point is not to remove responsibility. The point is to locate responsibility accurately.

Accountability does not mean every person connected to a system is equally responsible. It does not excuse reckless or willful conduct by front-line workers. It does not make executives personally responsible for every local error. It does not allow AI failures to be treated as mysterious technical events with no human or organizational owner. And it does not describe AI systems as if they possess human intent, judgment, understanding, or moral agency.

*Accountability means responsibility is mapped to control: authority, knowledge, resources, incentives, verification duties, and ability to intervene.*

HAOP carries that principle into AI-enabled safety work by requiring accountability to be designed before the system goes live. Responsibility should be mapped in advance to the person, team, or function with control over each action, signal, constraint, permission, metric, and deployment decision, when the mapping still has the greatest power to shape the outcome.

The question is not, "Who was visible at the moment of failure?" The question is, "Who controlled the conditions, permissions, signals, constraints, and verification points that made the failure possible?"

This approach does not use blame, which is a default reaction used to deny or shift accountability. Accountability is a discipline. Blame asks who should carry the consequence. Accountability asks where control existed, how it was used, what conditions shaped it, and what must change before harm occurs again.

The HAOP framework makes responsibility more accurate by mapping it across human, AI, and organizational control points before failure occurs. In AI-enabled systems, fear and blame do not disappear; they can be amplified and accelerated through automation, metrics, routing, surveillance, and decision-support tools.

HAOP prevents accountability from collapsing onto the person closest to the failure by making control, verification, authority, and intervention points explicit before harm occurs.

## VII. HAOP Operating Principles

---

HOP established the human-performance foundation by explaining how humans perform in systems. HAOP incorporates that foundation, but its operating principles govern a three-performer system: human adaptation, AI optimization, and organizational signaling. These principles come from the interaction dynamics through which these performers jointly produce outcomes in AI-enabled systems.

These HAOP Operating Principles transition the focus from the crisis (AI colliding with an old, control-based operating system) to the solution (a new socio-technical framework). They are an expansion of the core truths of HOP, updated for a world where algorithms are active team members.

1. **Performance is distributed.** Operational outcomes are produced by the interaction of human performers, AI performers, and organizational performers.
2. **Each performer introduces a distinct failure signature into the work system.** Humans tend to fail through adaptive overload - encountering obstacles and traps while trying to make the system work. AI becomes a performer when it materially shapes work, and it fails not through fatigue, distraction, or local adaptation but through misaligned optimization and data degradation. It acts on a representation of the worksite - equipment history, operating conditions, documentation, informal norms, anticipated consequences - that is never complete and can be stale, distorted, or optimized around the wrong signal, so its failures emerge in the gap between that representation and actual conditions, where its output shapes decisions, priorities, and action. Organizations are the containing performer and tend to fail through operational drift, distorted metrics, and normalized deviance - slowly detaching the hierarchy from frontline reality.
3. **Systems execute signals, not intentions.** An organization may intend safety, but the system executes what is measured, rewarded, automated, and enforced. Stated objectives do not govern behavior unless they are translated into valid signals, constraints, feedback loops, and accountability structures.
4. **Grounding is a control.** Because AI acts on a representation of the worksite rather than the worksite itself, grounding must be designed into the workflow as a structural verification of that representation against actual operating conditions.<sup>20</sup> Grounding does not become automatic when the model improves. It becomes automatic only when the architecture requires it before generation, recommendation, or action.
5. **Human oversight is work, not a label.** Oversight requires time, competence, authority, attention, access to underlying data, and the protected ability to pause, verify, escalate, or intervene.<sup>21</sup> Without those conditions, “human in the loop” is symbolic.
6. **Constraints must be designed before optimization.** AI performers will optimize the signal they are given. Safety-critical boundaries must be specified as constraints before deployment, not discovered after failure.<sup>22</sup> *Human-in-the-design* validation points are necessary friction that prevents optimization from outrunning verification.

7. **Accountability follows control.** Accountability has three forms: *consequence, ownership, and design*. Consequence is reactive; it assigns responsibility after failure. Ownership is admirable but dependent on character, and therefore unreliable on its own, as it depends on individuals or senior leaders accepting responsibility for outcomes. Design is structural; it defines responsibility before failure occurs. In AI-enabled systems, accountability must be mapped to the person, team, or function with control over the action, signal, constraint, permission, resource, metric, deployment decision, or response.
8. **Learning must include all three performers.** Investigations must examine what the human adapted to, what the AI optimized, and what the organization signaled, funded, ignored, rewarded, or normalized.

## Part II: Applying HAOP

---

The preceding sections define HAOP as a safety framework. The next section introduces the True Function Test as an initial diagnostic within the broader HAOP toolset.

The test does not certify an AI system as safe, and it is not a complete implementation method. Its purpose is narrower: to examine whether an AI-enabled workflow has True Function, meaning it produces the safety outcome it claims to pursue rather than merely producing a representation of safety.

Further HAOP tools are needed to evaluate accountability, verification, workflow boundaries, pause authority, control mapping, and implementation. The True Function Test is a starting point: it tests whether the workflow's stated purpose, actual behavior, metrics, incentives, and verification points are aligned before the system becomes consequential.

## VIII. The True Function Test: An Initial HAOP Diagnostic

---

### 8.1 Purpose of the Test

HAOP is not only a way of seeing. Its concepts can be translated into practical diagnostics that organizations can apply to their own AI-enabled safety work. The True Function Test is one such diagnostic. It can be used now, while the broader HAOP toolset continues to develop.

The purpose of the test is simple: determine whether a workflow actually produces the safety outcome it claims to pursue, or whether it mainly produces a representation of safety: a dashboard, metric, report, checklist, approval, or compliance artifact.

A workflow has True Function when it remains anchored to operational reality and produces the outcome it claims to produce. A workflow performs a fake function when it looks like it is managing safety while actually optimizing for something else.

**The True Function Test begins with nine core diagnostic questions.**

## 8.2 The Nine Diagnostic Questions

For any AI-enabled safety workflow, ask:

1. What outcome does this system claim to produce?
2. What signal is it actually optimizing?
3. What human judgment is it relying on?
4. What is the AI allowed to do without human initiation, review, or intervention?
5. What organizational incentive is shaping human behavior and AI optimization?
6. Where does verification occur before the output becomes consequential?
7. Who has authority to stop the workflow?
8. What would failure look like before harm occurs?
9. What weak signals would the system likely erase?

The answers reveal whether the workflow's claimed purpose, optimized signal, human role, AI authority, verification points, incentives, and weak-signal pathways are aligned.

Where the answers reveal a gap between the claimed outcome and the optimized signal, the workflow is performing a **fake function**. It appears to manage safety while managing something else: speed, closure, adoption, cost, appearance, throughput, or liability control.

## 8.3 From Questions to Alignment Mapping

The nine questions are the entry point. To work through them structurally, HAOP uses the *True Function Alignment Map*: a worksheet that runs each question across the human, AI, and organizational performers.

The Alignment Map helps teams identify where the claimed outcome, optimized signal, human judgment, AI authority, organizational incentives, verification points, pause authority, weak signals, and control responsibilities are aligned or misaligned.

A full version of the *True Function Alignment Map* is included in **Appendix B** and is offered as a free starting tool for organizations that want to apply HAOP to a real workflow. It is not the full HAOP implementation system. Additional tools are in development for workflow boundary mapping, accountability tracing, verification design, pause authority, implementation readiness, and drift monitoring.

## 8.4 Red Flags

In this diagnostic, fake function refers to a gap between the stated safety purpose of a workflow and the signal the workflow is actually designed, incentivized, or optimized to produce. A workflow may be performing a fake function if:

- The claimed objective is safety, but the optimized signal is speed, closure rate, adoption, cost reduction, or dashboard completion.

- The human is credited as a control but lacks time, authority, context, or competence to intervene.
- The AI output becomes consequential before verification occurs.
- The workflow produces cleaner documentation without improving field control.
- Weak signals are converted into low-resolution summaries before anyone with authority sees them.
- No one can clearly identify who owns the signal, constraint, permission, validation point, or stop-work decision.

## 8.5 What the Test Does Not Do

The True Function Test does not ask whether the workflow looks modern, efficient, compliant, or data-driven. It asks whether the workflow remains anchored to operational reality before the output becomes consequential.

# IX. Grounding and the Ability to Pause

---

## 9.1 Grounding

Because AI acts on a representation of the worksite rather than the worksite itself, grounding becomes a structural safety control. In HAOP, grounding is the designed verification of representation against operational reality, required before AI-shaped output is permitted to influence decisions, priorities, or action. It is not the same as asking a human to "review the AI." Grounding designs the verification point into the workflow. Human oversight is the work performed at that point. Two concepts, one architecture: without the designed point, review has nothing to anchor against, and without competent attention at the point, the design produces only a ceremony.

Grounding requires a live connection to operational reality: the current worksite, equipment condition, maintenance history, environmental conditions, worker knowledge, informal norms, production pressures, and the consequences of being wrong. When that connection is weak or absent, AI-enabled workflows can appear coherent while drifting away from the conditions they are supposed to support. The output remains fluent. The dashboard remains green. The conditions have moved.

Grounding takes three forms, and a workflow may require one, two, or all three depending on what the AI shapes.

**Source grounding** binds the AI to approved documents and reference material before it answers. Standard operating procedures, safety data sheets, regulatory text, equipment manuals, incident databases, and validated technical sources. Source grounding is the most common form deployed today, typically as retrieval over a curated corpus. It addresses what the AI is supposed to know.

**State grounding** binds the AI to the actual current condition of the work system before it acts. Sensor data, equipment status, permit status, shift records, access logs, monitoring readings, photographs, CMMS records, historian feeds. State grounding addresses what is true on the floor right now. It is the form most often missing in early AI deployments, and its absence is where cognitive overrun forms: the worker remains formally accountable for verifying an output the AI generated against conditions the AI never checked.

**World-model grounding** binds the AI to an internal representation of how the physical world changes over time. It is relevant where AI plans, simulates, or directs physical action — robotics, automated material handling, autonomous inspection, agentic workflows that act in sequence. World-model grounding addresses what might happen next as a result of acting now. A world model improves what the AI considers; it does not on its own verify what is the case. Source and state grounding remain necessary.

The three are not interchangeable. A workflow with strong source grounding and no state grounding will produce confident, well-cited output about conditions that may no longer exist. A workflow with state grounding but no source grounding will read the worksite accurately and apply the wrong standard to it. A workflow with neither, supplemented only by world-model inference, will plan competently against a representation that has lost contact with both the rulebook and the floor.

For these reasons, HAOP treats grounding as part of work design. A grounded workflow identifies where representation must be checked against reality, where assumptions must be challenged, where the work must pause, and who has the authority to stop or redirect the process when the output no longer matches actual conditions. The question is not whether a human reviewed the output. The HAOP question is whether the workflow was grounded before the decision moved forward.

## 9.2 The Ability to Pause

Stop Work Authority is recognized by many organizations and remains necessary where danger is imminent. But in practice, it is often framed as an emergency control: stop the job, halt production, trigger escalation. That framing can create reluctance to use it when the safer action is not a full stop, but a pause long enough to verify what is happening.

Much of operational safety depends on an earlier and less dramatic control: the protected ability to pause the flow of work. Workers, reviewers, supervisors, and technical owners need the protected ability to pause when something is unclear, when conditions have changed, when an input or output does not make sense, when a weak signal appears, or when the next step would turn uncertainty into consequence.

This is not a full stop. It is a pause long enough to ask, verify, adjust, escalate, or correct before the work continues. In AI-enabled workflows, this matters because an AI output may shape priority, approval, routing, escalation, or action before a person has had a realistic chance to challenge it.

The principle, though, is broader than AI. Safety-critical work depends on the ability to pause when the work or operating conditions no longer match the plan: when conditions change, a signal is unclear, an input or output does not make sense, or the next step would turn uncertainty into

consequence. A workflow that cannot be paused long enough to ask, verify, adjust, or escalate is not under control.

No safety-critical workflow should go live unless people know when they may pause the work, how the concern is escalated, who must respond, and how the person raising the concern is protected from retaliation or penalty.

## X. Conclusion

---

AI did not create the metric-for-mission and blame-the-worker problem we find in organizations today. It will amplify and accelerate it. It is a powerful optimizer, and when deployed into a system that already mistakes the metric for the mission or assigns blame to the nearest person to the incident, it will pursue that signal faster, more consistently, and at greater scale than a human organization could manage manually.

Existing safety frameworks remain valuable. AI, though, introduces specific operational dynamics that those frameworks were not designed for. HAOP gives organizations a way to account for those dynamics by recognizing and treating the three performers now shaping safety-critical work: the human who adapts, the AI that optimizes, and the organization that contains, directs, and legitimizes both. Seeing them clearly is the precondition for governing them well and for keeping the human judgment that grounds the whole system present, capable, and accountable, rather than extracted and discarded.

HAOP offers two things: a conceptual framework for seeing human, AI, and organizational performance as interacting sources of safety-critical outcomes, and a practical diagnostic for testing whether AI-enabled workflows remain grounded in operational reality before their outputs become consequential.

### Appendix A: Key Terms

**Accountability-by-control.** The HAOP principle that responsibility should be mapped to the person, team, or function with control over the relevant action, signal, constraint, permission, resource, metric, deployment decision, verification point, or escalation path - not to whoever is most visible at the point of failure.

**Aesthetic illusion.** Mistaking a clean, fluent, low-noise digital artifact for operational competence. A dashboard, summary, score, or compliance record that looks orderly and current can suggest a system under control while the conditions it represents have drifted.

**Cognitive overrun.** A condition in which a worker remains formally accountable for verifying AI output while its rate, density, or ambiguity has exceeded their capacity to verify what matters. Cognitive overrun is the human failure signature most directly produced by AI deployment without grounding or pause design.

**Compound Drift.** The gradual movement of a work system away from its intended function when multiple forms of drift interact and reinforce one another. In HAOP, this may occur when humans

adapt to pressure, AI optimizes toward the wrong signal, and the organization rewards or tolerates the resulting pattern. The system may continue to appear functional while its actual behavior shifts away from the safety outcome it claims to produce.

**Confident incompetence.** The AI failure pattern of producing fluent, authoritative, or decisive output that is wrong. Confident incompetence is dangerous because the output's surface qualities - coherence, specificity, register - function as proxies for correctness in workflows that lack independent verification.

**Containing performer.** The organizational performer. So named because the organization does not sit alongside the human and the AI as a peer; it designs, constrains, authorizes, measures, and normalizes the conditions under which the other two operate. The organization contains the operational performers but also performs through them.

**Failure signature.** The recognizable pattern that identifies which kind of performer failure is present. Humans tend to fail through adaptive overload, AI through misaligned optimization and data degradation, organizations through operational drift, distorted metrics, and normalized deviance. Failure signatures are diagnostic: they tell an investigator what kind of breakdown occurred before the accountability trace begins.

**Grounding.** The designed verification of representation against operational reality, required before AI-shaped output is permitted to influence decisions, priorities, or action. HAOP distinguishes three forms: *source grounding* (against approved documents and reference material), *state grounding* (against actual current conditions of the work system), and *world-model grounding* (against an internal representation of physical change over time).

**Human-in-the-Design (HITD).** The HAOP control point that places human judgment upstream of deployment, where constraints, signals, verification points, escalation paths, and stop authority are specified before the workflow goes live. HITD is the validation friction that prevents AI optimization from outrunning what HITL can catch. HITL governs what happens during the work; HITD governs whether the work was set up so HITL has a chance of working.

**Performer.** Any entity or organized system that can take action, influence outcomes, and introduce variability into a work system. HAOP uses *performer* functionally, not morally. A human performer acts through judgment and adaptation, an AI performer acts through model output, tool use, ranking, routing, prediction, or workflow execution, and an organizational performer acts through governance, incentives, metrics, resources, authority, procurement, and tolerated tradeoffs.

**Three forms of accountability.** *Consequence accountability* is reactive and assigns responsibility after failure. *Ownership accountability* is voluntary acceptance of responsibility by an individual or leader. *Design accountability* is structural and operates in real time, defining who controls what before failure occurs. HAOP is built on design accountability.

**Three-performer architecture.** The HAOP claim that operational outcomes in AI-enabled work are produced by the interaction of three performers - human, AI, and organizational - rather than by humans acting alone within neutral organizational containers.

**True Function.** A workflow has True Function when it actually produces the safety outcome it claims to pursue, rather than producing the documentation, dashboard, metric, or compliance artifact that represents the outcome. The True Function Test in Section VIII evaluates whether a workflow is operating in True Function or in *Fake Function* - looking like it manages safety while actually managing something else.

***Inherited terms used in the paper***

**Human-in-the-Loop (HITL).** The conventional model of AI oversight in which a human reviews, approves, or intervenes on AI output at the point of action. HAOP treats HITL as necessary but insufficient on its own: the human at the loop point can verify only what they have the time, authority, context, and competence to verify, and when those conditions are absent the loop becomes symbolic. HITL is a runtime control and depends on conditions that must be designed before runtime.

**Work-as-imagined and work-as-done.** Hollnagel's distinction between how work is planned, documented, and assumed to occur (*work-as-imagined*) and how it actually proceeds under real operating conditions (*work-as-done*). The gap between the two is where most safety-relevant adaptation, error, and drift become visible.

**Local rationality.** Dekker's principle that people do what makes sense to them given their immediate goals, knowledge, attention, norms, and pressures. Behavior that appears irrational from outside is usually rational from inside the conditions in which it occurred.

**Just Culture.** A framework that protects learning by recognizing that people should not be punished for actions, omissions, or decisions that are reasonable given their training, experience, and operating conditions, while preserving accountability for reckless conduct, willful violations, gross negligence, and knowing disregard of risk.

**Normalization of deviance.** Vaughan's concept describing how organizations gradually accept deviant conditions as normal when those conditions repeatedly occur without immediate catastrophe.

**Drift.** Dekker's term for the gradual movement of complex systems toward failure through small, locally reasonable adjustments - borrowing from safety margins to meet production, cost, schedule, or efficiency demands - each of which appears tolerable because nothing bad happens immediately.

## Appendix B: True Function Alignment Map

A free HAOP worksheet for testing whether an AI-enabled workflow is producing the safety outcome it claims to pursue — or only the appearance of safety.

Run each question across all three performers:

Map Section	Prompt	Human Performer	AI Performer	Organizational Performer
<b>Framing Prompt</b>	<b>What work is being performed?</b>	judgment, adaptation, review, escalation	classification, generation, routing, optimization	work design, approval, resourcing, metrics
<b>Q1</b>	<b>What outcome is claimed?</b>	What is the person expected to accomplish?	What is the AI expected to produce or influence?	What business or safety outcome is the organization claiming?
<b>Q2</b>	<b>What signal is optimized?</b>	What behavior is rewarded, pressured, or normalized?	What data, metric, prompt, target, or proxy is the AI optimizing?	What metric, incentive, or executive priority is shaping the workflow?
<b>Q3</b>	<b>What judgment is relied on?</b>	What must the human notice, interpret, challenge, or verify?	What does the AI classify, recommend, generate, approve, or route?	What assumptions has the organization made about human capacity and AI reliability?
<b>Q4</b>	<b>What behavior is allowed?</b>	What shortcut, adaptations, or overrides are tolerated?	What can the AI do without review or intervention?	What permissions, resources, and authority has the organization granted?
<b>Q5</b>	<b>Where does verification occur?</b>	Does the human have time, skill, context, and authority to verify?	Is the AI output checked against reality, constraints, or known limits?	Has the organization defined points of consequence and required validation there?
<b>Q6</b>	<b>Who can stop the workflow?</b>	Can the worker or reviewer pause or challenge the process?	Are there technical limits, hard stops, or escalation triggers?	Has stop-work authority been designed, protected, and resourced?
<b>Q7</b>	<b>What can early failure look like?</b>	Fatigue, silence, workarounds, attention or judgment miss, compliance miss, overtrust	Confident incompetence, false negatives, signal substitution, constraint failure	Dashboard confidence, compliance theater, accountability gaps, normalized deviance, weak oversight
<b>Q8</b>	<b>What is shaping behavior?</b>	context, fatigue, training, peer norms	data, prompts, model limits, permissions	incentives, policies, budgets, priorities
<b>Q9</b>	<b>What weak signals may be erased?</b>	Hesitation, informal warnings, tacit knowledge, near misses	Edge cases, anomalies, uncertainty, missing context	Bad-news resistance, bottlenecks, friction, dissent, underreported risk, resource constraints
<b>Control Synthesis</b>	<b>What control is needed?</b>	time, authority, training, stop-work ability	validation, constraints, monitoring, auditability	governance, ownership, escalation, consequence design

## References

1. Frederick Winslow Taylor, *The Principles of Scientific Management* (New York: Harper & Brothers, 1911).
2. Erik Hollnagel, *Safety-I and Safety-II: The Past and Future of Safety Management* (Farnham, UK: Ashgate, 2014).
3. Todd Conklin, *The 5 Principles of Human Performance: A Contemporary Update of the Building Blocks of Human Performance for the New View of Safety* (Santa Fe, NM: PreAccident Investigation Media, 2019).
4. Sidney Dekker, *The Field Guide to Understanding “Human Error”*, 3rd ed. (Boca Raton, FL: CRC Press, 2014).
5. Sidney Dekker, *Safety Differently: Human Factors for a New Era*, 2nd ed. (Boca Raton, FL: CRC Press, 2014).
6. Erik Hollnagel, Robert L. Wears, and Jeffrey Braithwaite, *From Safety-I to Safety-II: A White Paper* (University of Southern Denmark, University of Florida, and Macquarie University, 2015).
7. OWASP GenAI Security Project, *OWASP Top 10 for Agentic Applications - 2026 Edition* (December 2025). OWASP describes the framework as addressing risks in autonomous and agentic AI systems that plan, act, and make decisions across workflows.
8. Richard M. Cyert and James G. March, *A Behavioral Theory of the Firm*, 2nd ed. (Malden, MA: Blackwell, 1992; originally published 1963).
9. Diane Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA* (Chicago: University of Chicago Press, 1996).
10. Sidney Dekker, *Drift into Failure: From Hunting Broken Components to Understanding Complex Systems* (Farnham, UK: Ashgate, 2011).
11. James G. March and Herbert A. Simon, *Organizations*, 2nd ed. (Cambridge, MA: Blackwell, 1993; originally published 1958).
12. Daniel Kahneman, *Attention and Effort* (Englewood Cliffs, NJ: Prentice-Hall, 1973).
13. Raja Parasuraman and Victor Riley, “Humans and Automation: Use, Misuse, Disuse, Abuse,” *Human Factors* 39, no. 2 (1997): 230–253.
14. Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (New York: Viking, 2019).
15. Iliia Shumailov et al., “AI Models Collapse When Trained on Recursively Generated Data,” *Nature* 631 (2024): 755–759. The article defines model collapse as a degenerative process in which generated data pollutes later training sets.

16. Michael Polanyi, *The Tacit Dimension* (Chicago: University of Chicago Press, 1966).
17. Gary Klein, *Sources of Power: How People Make Decisions* (Cambridge, MA: MIT Press, 1998).
18. Financial Times, “Amazon Scraps AI Leaderboard to Stop Workers Chasing Usage Scores,” May 2026.
19. Hugh Langley, “Amazon Says It Shut Down a Token Leaderboard: ‘Don’t Use AI Just to Use AI,’” *Business Insider*, May 2026.
20. International Organization for Standardization and International Electrotechnical Commission,  
  
ISO/IEC 42001:2023: *Information Technology - Artificial Intelligence - Management System* (Geneva: ISO, 2023). ISO describes ISO/IEC 42001 as an AI management system standard for managing AI risks and opportunities.
21. National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (Washington, DC: U.S. Department of Commerce, January 2023). NIST frames AI risk management around governance, mapping, measuring, and managing AI risks.
22. Russell, *Human Compatible*; NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*.
23. EUROCONTROL, “Just Culture,” accessed May 31, 2026. EUROCONTROL defines Just Culture as protecting operators from punishment for actions commensurate with experience and training, while not tolerating gross negligence, willful violations, or destructive acts.
24. *Gibson v Maritime New Zealand* [2026] NZHC 813. The High Court dismissed the appeal and upheld the former Ports of Auckland CEO’s conviction, NZ\$130,000 fine, and NZ\$60,000 costs award under New Zealand’s Health and Safety at Work Act.
25. Maritime New Zealand, “High Court Dismisses Tony Gibson Appeal,” April 13, 2026. Maritime NZ stated that the Auckland High Court upheld the guilty finding and sentence after stevedore Pala’amo Kalati was killed by a falling container at the port.
26. New Zealand, *Health and Safety at Work Act 2015*, ss. 44 and 152.