

The Last Metre of Predictive Safety: A Competency-Centred Framework for Responsible AI in Occupational Health and Safety Management

Ashish Mulraj Ved¹, Ravindra Kumar²

¹ PhD Research Scholar, Faculty of Commerce and Management, Asian International University, Manipur, India

² Assistant Professor, Asian International University, Manipur, India

Abstract- Artificial intelligence (AI) is becoming easier to install in workplaces than it is to use well. A camera can identify a missing hard hat, and a predictive system can flag a risky shift, yet neither result explains who should respond, how quickly action must follow, or whether the control actually worked. This article examines the often-overlooked interval between an AI-generated signal and a competent preventive decision. Drawing on a structured, non-exhaustive critical synthesis of occupational health and safety (OHS), human factors, training and AI-governance literature, it sets out a six-dimension Predictive Safety Management Framework (PSMF). The framework links data governance, AI selection and validation, predictive monitoring, human-AI collaboration, training and competency, and ethical governance with continuous performance evaluation. Its contribution is to make competency a continuing operational requirement rather than a single training event at system launch. A role-based competency matrix and six decision gates show what evidence an organisation should seek before moving from a promising demonstration to supervised use. The paper also proposes measures that separate alert quality from response quality and from verified control outcomes. This is conceptual framework development, not a field evaluation: no employees were surveyed, no training intervention was tested, and no injury reduction is claimed. The central argument is practical: AI assists prevention only when workers retain the knowledge, authority, time and trust needed to act on its information.

Keywords— Artificial intelligence; occupational health and safety; Predictive Safety Management Framework; training and competency; human-AI collaboration; safety culture; ISO 45001; responsible AI

I. INTRODUCTION: THE ALERT IS NOT THE OUTCOME

Imagine a site supervisor at the start of a busy shift. An automated dashboard highlights a possible equipment-pedestrian conflict. The warning reaches the right screen, but the supervisor is handling a permit-to-work problem, the operator believes the sensor is unreliable, and nobody is certain who should stop the job. By the time the alert is discussed, the vehicle has moved on. Nothing in this hypothetical scene proves that the AI performed badly. It shows why technical performance alone is too narrow a test of a safety system.

OHS professionals have long worked with the difference between identifying a hazard and controlling it. AI makes this difference more visible. Computer vision, natural language processing and

predictive analytics can review large quantities of operational data; they can also create false positives, miss genuine hazards or produce outputs that are difficult to interpret (Fang et al., 2020; Shah & Mishra, 2024). A prediction is an invitation to act, not evidence that action has taken place.

A further difficulty is the gap between the speed of technology development and evidence of benefit to workers. Trask and Linderroth (2023) found limited direct evidence of improved conditions or fewer injuries from digital OHS interventions in construction. A later systematic review identified very little high-quality direct evidence of AI-based OHS tools reducing worker illness or injury (Jetha et al., 2025). Neither finding argues against AI. Both require authors and managers to be careful about what they claim.

This article asks a management question: what has to be in place for an AI-generated safety signal to become timely, competent and accountable preventive action? It develops a practical framework in which technology and training are designed together, with the hierarchy of controls and worker participation remaining central.

II. APPROACH AND SCOPE OF THE EVIDENCE

The article develops a conceptual framework from a structured, non-exhaustive critical synthesis presented in the underlying 2026 doctoral thesis on AI in OHS management. That study examined peer-reviewed research and authoritative sources relevant to hazard detection, predictive risk analytics, human factors, training, organisational readiness, management systems and responsible AI. Sources were selected for their relevance to the management problem and the complementary perspectives they offered, rather than through a complete database-by-database systematic search.

The synthesis is intentionally not described as a PRISMA systematic review or meta-analysis. No claim is made that every eligible study was retrieved, that independent reviewers screened a specified number of records, or that pooled intervention effects were calculated. The procedure is analytical: identify recurring barriers between detection and action, compare them with established OHS responsibilities, and organise the resulting requirements into a framework that can subsequently be tested in practice.

Published literature supplies background and critique; ISO 45001:2018 supplies OHS management-system principles; ISO/IEC 42001:2023 and the NIST AI Risk Management Framework supply governance concepts. None is treated as proof that the proposed PSMF has already improved safety outcomes (International Organization for Standardization, 2018; International Organization for Standardization & International Electrotechnical Commission, 2023; National Institute of Standards and Technology, 2023).

III. WHAT GETS LOST BETWEEN PREDICTION AND PREVENTION?

Three breaks in the chain deserve attention. The first is a data break. A predictive system may learn from incident records containing inconsistent definitions, missing near misses or old work practices. When the source changes, the meaning of its outputs can change as well. Good data governance therefore includes recording where information came from, who owns it and how incomplete or unrepresentative records will be handled.

Technology acceptance is shaped by perceived usefulness, effort and organisational support, not only by the quality of the software (Davis, 1989; Venkatesh et al., 2003). The second is a decision break. Alerts compete with alarms, permit approvals, production questions and human attention. A technically correct notification might reach a person who lacks authority to stop work. Another might be delivered without a clear explanation of urgency. When workers receive repeated low-value alerts, they may begin to ignore the whole system. Human factors research on automation helps explain why role clarity and carefully chosen boundaries for automated decisions matter (Parasuraman et al., 2000; Niehaus et al., 2022).

The third is a competence break. Safety training cannot end with a presentation on how to read the dashboard. A worker needs to recognise the task-specific hazard, decide whether the warning fits the situation, follow the safe system of work and report a possible model error. A supervisor needs to triage competing alerts, control the activity and document closure. An OHS professional needs to understand false negatives, false positives, model drift and assurance records, while a manager needs to resource the response system. Adult learning principles support practice based on real decisions and feedback, rather than knowledge recall alone (Knowles, 1984; Anderson & Krathwohl, 2001).

These breaks are related. Poor data can reduce trust; low trust can weaken reporting; weaker reporting can further damage data quality. For this reason, AI in OHS should be treated as a socio-technical

change within the existing OHS management system, not as an isolated IT purchase.

IV. THE SIX DIMENSIONS OF THE PREDICTIVE SAFETY MANAGEMENT FRAMEWORK

The ILO (2025) also emphasises that digitalisation at work creates both safety opportunities and new risks requiring participation and oversight. The PSMF organises the work into six connected dimensions (Table 1). The first two ask whether data and the selected model are suitable for a clearly defined safety problem.

The third asks whether the results can be incorporated into existing inspection, maintenance, permit-to-work and risk-assessment processes. The fourth and fifth ask who acts and how their competence is demonstrated. The sixth establishes the obligation to check for unexpected harm, poor performance, unfairness and loss of confidence over time.

These dimensions should not be read as six departments working independently. A frontline worker can reveal a data problem; a data analyst may identify a competency problem; a supervisor may discover that a perfect classification provides no opportunity to intervene before exposure. Each discovery should be able to change the design. Worker consultation is therefore an input to governance rather than a message sent after deployment.

The framework respects the hierarchy of controls. If a hazardous interaction can be eliminated by segregating vehicle and pedestrian routes, an AI warning must not be offered as a cheaper substitute.

Monitoring may still support assurance, but the case for using it must be explained in relation to stronger controls and the actual exposure being managed.

Table 1. Proposed six-dimension Predictive Safety Management Framework (adapted from the doctoral thesis).

Dimension	What an organisation must be able to demonstrate
1. Data infrastructure and governance	Defined safety purpose; records with known origin; data quality, access, privacy and retention controls.
2. AI selection, development and validation	Representative testing; explicit error consequences; documented limitations and change control.
3. Predictive risk monitoring and decision support	Useful alerts embedded in OHS workflows, with an owner, urgency and escalation route.
4. Human-AI collaboration and oversight	Clear authority, worker feedback, human override, preserved situational awareness.
5. Training and competency	Role-specific scenario practice, demonstrated judgement, reassessment and refresher learning.
6. Ethical governance, performance evaluation and continual improvement	Bias, drift, privacy, worker wellbeing, control closure, audit and withdrawal decisions monitored.

V. TRAINING AS AN OPERATING CONTROL, NOT AN EVENT

Training is sometimes listed among the actions required to introduce new software. In a safety-critical setting, this is insufficient. The required capability changes with the role and may deteriorate as the system, workforce or task changes. The PSMF therefore treats training and competency as a living control with entry, observation, reassessment and refresher requirements.

For frontline workers, a valid assessment might present a realistic alert and ask what to check, when to stop, how to obtain support and what to do if the device fails. For supervisors, a scenario might include three simultaneous warnings and limited response resources. The assessor should observe whether the learner can prioritise safely, not simply identify which coloured symbol indicates the highest score. OHS

practitioners should be able to explain the limits of accuracy, recall and precision and the consequences of missed hazards. Senior leaders should be able to ask whether a deployment has enough people and authority to close the alerts it creates.

Table 2 proposes a role-based assessment structure. These are proposed training outputs and not results from a completed course or trial. Their value would need to be tested with real learners, suitable tasks and independent competency criteria. A multiple-choice score alone would not show that a worker can make a good decision under pressure.

Table 2. Suggested role-based AI/OHS competency evidence (proposed, not field-tested).

Role	Competency to demonstrate	Suitable evidence
Frontline worker	Recognises alert meaning; checks context; follows safe system; knows fallback	Observed task scenario and feedback
Supervisor	Prioritises alerts; can stop work; assigns and verifies controls	Timed scenario and reviewed closure log
OHS professional	Interprets error types, hazards, drift and worker concerns	Case review with defended risk decisions
Technical specialist	Tests model limits; diagnoses changes; manages access	Validation record and controlled change review
Senior manager	Resources response; retains accountability; reviews outcomes	Recorded management decision and assurance review

VI. FROM A PILOT TO ROUTINE USE: SIX EVIDENCE GATES

The strongest way to make the framework useful is to turn its principles into decisions that can be audited. A proposed six-gate sequence (Table 3) asks the organisation to document a safety purpose, prove that the data can support it, test model behaviour, demonstrate human competence, show

that the response workflow is practical and review performance after implementation.

These gates are not a certification scheme. Their purpose is to make it difficult to move from an attractive demonstration to unsafe reliance without answering basic questions. An organisation that has validated a camera system but not assigned an alert owner should not describe the system as fully ready. One that cannot explain how workers can challenge errors should reconsider its governance. A deployment that produces a flood of unresolved warnings should be redesigned rather than celebrated for increased detection.

The approach fits ISO 45001 management review, operational control and continual improvement while incorporating AI-specific change control, accountability and risk assessment from ISO/IEC 42001 and NIST AI RMF. The records should remain integrated with existing OHS responsibilities rather than establishing a second, parallel safety bureaucracy.

Table 3. Six assurance gates for controlled implementation (proposal).

Gate	Evidence to review	Decision
Purpose	Safety problem and stronger control alternatives	Accept or revise use case
Data	Lineage, legal access, missingness and representation	Allow restricted development
Validation	Realistic tests and consequence-sensitive error analysis	Allow controlled trial
People	Role-based competency and override authority	Allow supervised use
Operation	Response workload, capacity, escalation and fallback	Allow bounded deployment
Review	Drift, worker feedback and independently checked controls	Continue, improve, restrict or withdraw

VII. MEASURING WHETHER THE SYSTEM HELPS

An organisation can report 96% detection accuracy and still have a weak safety process. Overall accuracy may conceal poor recognition of uncommon but high-consequence hazards. Therefore performance monitoring needs to distinguish three layers of evidence.

At the technical layer, record data completeness, recall or sensitivity, precision, false-negative patterns, model drift and performance under realistic working conditions. At the workflow layer, measure whether alerts reach the intended person, the time between alert and acknowledgement, the proportion receiving a competent response, unresolved workload, override reasons and completion of role-based competency reassessment. At the OHS layer, examine independently verified control closure, exposure changes, repeat near misses and, where the study design permits, injury and ill-health outcomes.

The distinction matters because a faster alert may be valuable without yet demonstrating reduced injury, while a reduced incident count may be due to unrelated changes or under-reporting. A useful evaluation sets the baseline before deployment, records concurrent changes, includes worker feedback and considers unintended effects such as fatigue, intrusive surveillance and pressure to work faster (European Agency for Safety and Health at Work, 2022; El-Helaly, 2024).

Workers should be able to report that an alert was wrong, late or disruptive without fear that questioning the system will be treated as misconduct.

Such feedback is a source of control intelligence. It can support calibration, better working arrangements and stronger safety culture, provided management visibly responds to it.

VIII. DISCUSSION: WHAT THE FRAMEWORK ADDS—AND WHAT IT CANNOT YET PROVE

The framework contributes a way to connect questions that are often managed separately. AI assurance specialists may focus on classification error, OHS teams on compliance and control measures, trainers on course delivery, and senior management on the cost of deployment. The PSMF insists that these questions meet at the point where an unsafe condition could be prevented. Its main conceptual contribution is not a new algorithm; it is an accountable route from information to action.

The framework also changes the meaning of successful training. Completion records, attendance and knowledge quizzes remain useful administrative evidence, but they do not establish operational capability by themselves. A more demanding test asks whether a person notices uncertainty, selects a proportionate control and can explain their decision after a simulated or real alert. This makes training a component of OHS assurance rather than an optional addition to technical rollout.

There are important limits. The underlying literature synthesis was selective and did not estimate pooled effect sizes. The PSMF has not been externally validated by expert consensus, a workplace pilot or a controlled before-and-after comparison. Its proposed indicators do not establish that AI changes accident rates. Sector, workforce, language, resources and legal context may alter which dimensions are most demanding. These uncertainties should be reported rather than concealed by a tidy diagram.

Agenda for field testing

Field evaluation should begin with a narrow, clearly defined hazard and a baseline of existing controls. Workers and safety representatives should participate in purpose-setting and in a privacy review. A shadow-mode period can record AI outputs without making the work dependent on them. That would allow analysts to review false

positives, false negatives, workload and performance differences across tasks or conditions.

If the initial evidence is acceptable, a supervised pilot should record response times, choices made, override decisions and independently verified completion of controls. A competency assessment should test the relevant roles before and after training, using comparable scenarios and transparent criteria. The evaluation should also record unintended outcomes, including alert fatigue, reduced job autonomy, privacy complaints and missed work created by unnecessary investigations. Only after these intermediate measures are credible should researchers attempt stronger outcome evaluation. Comparative studies across sites or periods could investigate exposure and injury outcomes while accounting for concurrent changes. Unfavourable results would be informative: a model that adds work without improving timely control is a useful reason to revise or withdraw the intervention.

IX. CONCLUSION

The important question is not whether AI can see something earlier than a person. It is whether the workplace can make good use of that earlier knowledge. A predictive safety system needs reliable information, sensible alert thresholds, competent people, clear decision rights, sufficient response capacity and an honest way to check whether hazards were controlled.

The PSMF brings these requirements into one management conversation. It presents a practical, testable approach to responsible AI in OHS, with training and competency at its centre. It remains a proposal until tested in actual workplaces. For now, its most useful instruction is simple: judge a safety alert by the quality of the prevention it helps people deliver, not by the sophistication of the screen on which it appears.

Declarations and data availability

Study design: conceptual and literature-based; no participants recruited and no organisational data collected. Ethics approval: not applicable to this conceptual manuscript. Data availability: no

empirical dataset was generated for this article. Competing interests and funding: to be completed and approved by both authors before submission.

REFERENCES

1. Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
2. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
3. El-Helaly, M. (2024). Artificial intelligence and occupational health and safety, benefits and drawbacks. *La Medicina del Lavoro*, 115(2), e2024014. <https://doi.org/10.23749/mdl.v115i2.15835>
4. European Agency for Safety and Health at Work. (2022). Artificial intelligence for worker management: An overview. EU-OSHA.
5. Fang, W., Ding, L., Love, P. E. D., Luo, H., Li, H., Peña-Mora, F., Zhong, B., & Zhou, C. (2020). Computer vision applications in construction safety assurance. *Automation in Construction*, 110, 103013. <https://doi.org/10.1016/j.autcon.2019.103013>
6. International Labour Organization. (2025). Revolutionizing health and safety: The role of AI and digitalization at work. ILO. <https://doi.org/10.54394/KNZE0733>
7. International Organization for Standardization. (2018). ISO 45001:2018 occupational health and safety management systems—Requirements with guidance for use. ISO.
8. International Organization for Standardization & International Electrotechnical Commission. (2023). ISO/IEC 42001:2023 information technology—Artificial intelligence—Management system. ISO.
9. Jetha, A., Bakhtari, H., Irvin, E., et al. (2025). Do occupational health and safety tools that utilize artificial intelligence have a measurable impact on worker injury or illness? Findings from a systematic review. *Systematic Reviews*, 14, 146. <https://doi.org/10.1186/s13643-025-02869-1>

10. Knowles, M. S. (1984). *The adult learner: A neglected species* (3rd ed.). Gulf Publishing.
11. Niehaus, S., Hartwig, M., Rosen, P. H., & Wischniewski, S. (2022). An occupational safety and health perspective on human in control and AI. *Frontiers in Artificial Intelligence*, 5, 868382. <https://doi.org/10.3389/frai.2022.868382>
12. National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1)*. <https://doi.org/10.6028/NIST.AI.100-1>
13. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A*, 30(3), 286–297.
14. Shah, I. A., & Mishra, S. (2024). Artificial intelligence in advancing occupational health and safety: An encapsulation of developments. *Journal of Occupational Health*, 66(1), uiad017. <https://doi.org/10.1093/joccuh/uiad017>
15. Trask, C. M., & Lindereth, H. C. J. (2023). Digital technologies in construction: A systematic mapping review of evidence for improved occupational health and safety. *Journal of Building Engineering*, 80, 108082. <https://doi.org/10.1016/j.jobe.2023.108082>
16. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>