
The AI-augmented Safety Governance Model (AASGM): A Socio-technical Framework for Real-time Hazard Detection, Human Oversight, and Regulatory Alignment

Dr. Robb Shawe

Capitol Technology University, Department of Occupational Health & Safety, 11301 Springfield Road, Laurel, MD, USA

doi.org/10.51505/ijaemr.2026.11315

URL: <http://dx.doi.org/10.51505/ijaemr.2026.11315>

Received: Apr 24, 2026

Accepted: May 07, 2026

Online Published: May 19, 2026

Abstract

This manuscript introduces the AI-Augmented Safety Governance Model (AASGM), a socio-technical framework that integrates artificial intelligence-enabled hazard-detection systems with human oversight, organizational safety culture, and regulatory compliance structures. Building on empirical findings from YOLO-based hazard detection across multiple industries, the model conceptualizes how real-time detection technologies interact with human decision-making, governance processes, and safety management systems. The AASGM positions artificial intelligence as an augmentation layer within a broader socio-technical system, emphasizing transparency, accountability, and human-centered implementation. The framework provides a structured approach for organizations seeking to operationalize AI in occupational safety while maintaining alignment with regulatory expectations and ethical considerations. The model advances Safety 4.0 by bridging technical capabilities with governance and organizational integration.

Keywords: AASGM; artificial intelligence; safety governance; socio-technical systems; Safety 4.0; occupational safety; human oversight; regulatory compliance

1. Introduction

The integration of artificial intelligence (AI) into occupational safety systems represents a significant shift from reactive hazard management toward proactive, real-time risk mitigation. Traditional safety frameworks, which rely on manual inspections and retrospective incident analysis, are increasingly inadequate in dynamic and high-risk environments (Nath et al., 2020; OSHA, 2024).

AI-enabled hazard detection systems, particularly those based on computer vision technologies such as YOLO, provide continuous monitoring and rapid identification of unsafe conditions. These capabilities support earlier intervention and improved compliance with safety standards.

However, integrating AI into safety systems poses complex challenges related to governance, human oversight, and organizational alignment. While AI systems enhance detection capabilities, their effectiveness depends on how they are embedded within socio-technical systems that include human operators, organizational processes, and regulatory frameworks.

This paper addresses this gap by introducing the **AI-Augmented Safety Governance Model (AASGM)**, a framework that integrates AI technologies with human-centered safety practices and governance structures.

Unlike existing Safety 4.0 frameworks that primarily emphasize automation, connectivity, or technical system integration, the AASGM formalizes artificial intelligence as a governance-centered augmentation layer embedded within socio-technical safety systems. The framework uniquely integrates real-time hazard detection, human oversight, workforce interaction, organizational governance, and regulatory alignment into a unified operational structure for occupational safety environments.

This study contributes to a broader research program advancing the AI-Augmented Safety Governance Model (AASGM) by formalizing a structured framework that integrates technical performance, human factors, and organizational governance into a cohesive model for AI-enabled safety systems. Building upon prior analyses of detection capability, workforce interaction, and socio-technical integration, this research operationalizes these components into a unified governance architecture. The proposed framework provides a systematic approach to aligning AI-enabled hazard-detection systems with organizational decision-making processes, regulatory requirements, and adaptive safety strategies in complex workplace environments.

This positioning extends beyond empirical integration to formalize how socio-technical systems can be structured to support governance, regulatory alignment, and decision-making processes through AI-enabled hazard detection.

This manuscript is part of the Shawe Series, a coordinated research program examining artificial intelligence-enabled hazard detection, socio-technical safety integration, and governance frameworks in regulated workplace environments. The series advances the AI-Augmented Safety Governance Model (AASGM) as a unifying framework linking real-time detection technologies, human oversight, regulatory compliance, and organizational decision-making.

2. Theoretical Foundations

2.1 Socio-Technical Systems Theory

Socio-technical systems theory emphasizes the interdependence of technical systems and human actors within organizational environments (Carayon, 2006). Effective safety systems require alignment between technological capabilities and human behavior.

2.2 Human–AI Interaction

Human–AI interaction research highlights the importance of trust, transparency, and usability in determining system adoption (Lee & See, 2004; Wickens et al., 2015). AI systems are most effective when designed as decision-support tools rather than autonomous replacements.

2.3 Safety 4.0 Paradigm

Safety 4.0 integrates real-time data, automation, and human-centered design to enable proactive safety management (Yousif et al., 2024). AI-enabled hazard detection systems are central to this paradigm. These foundations align with broader governance and risk management frameworks that emphasize structured oversight, accountability, and decision integration within complex organizational systems (Kaplan & Mikes, 2012; Power, 2007).

3. Development of the AASGM Framework

The AI-Augmented Safety Governance Model (AASGM) is developed by integrating empirical findings from prior analyses (Articles 1–3), socio-technical systems theory, human factors research, and relevant regulatory frameworks. This multi-layered foundation enables the model to capture the interaction between technological capability, human oversight, and organizational governance. Within this structure, artificial intelligence is conceptualized as an augmentation layer embedded within a broader safety governance system, supporting real-time hazard detection while maintaining human accountability and regulatory alignment.

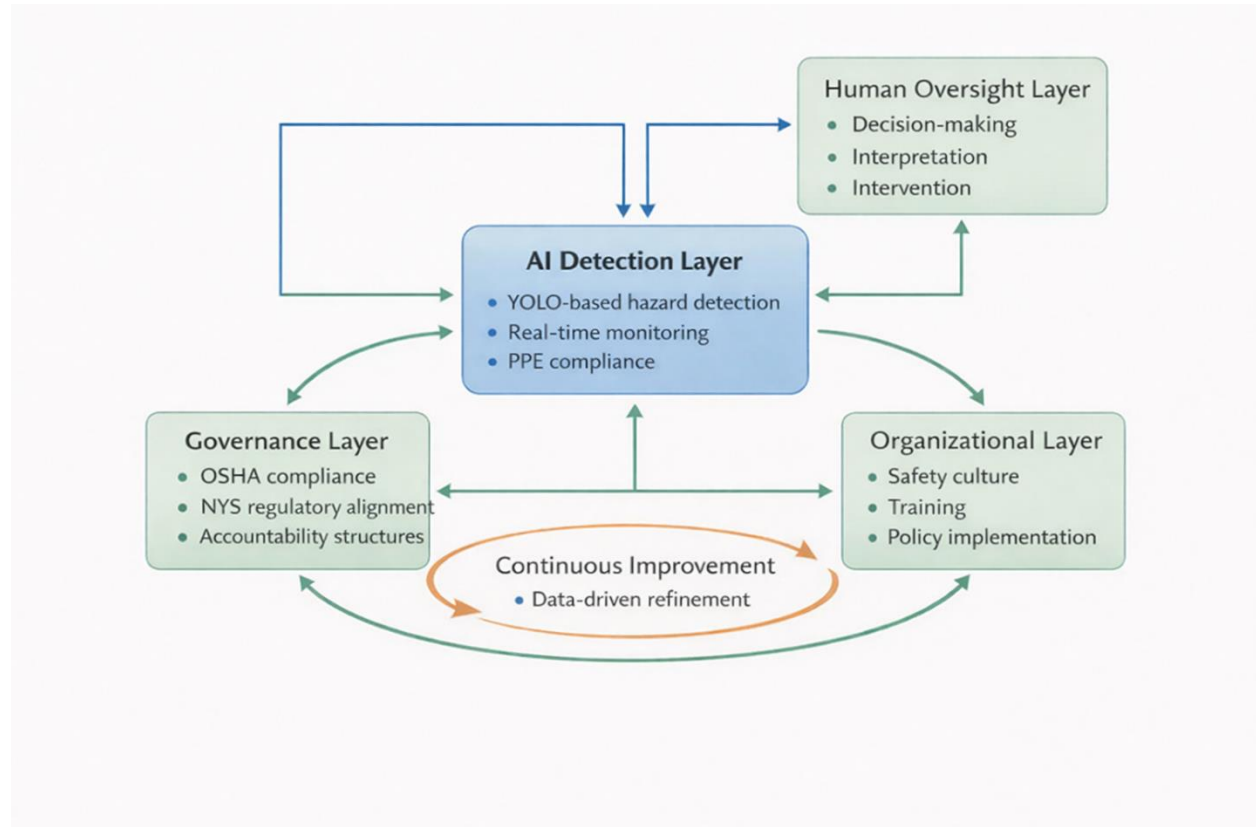
4. AASGM Framework Structure

To formalize the integration of artificial intelligence, human oversight, and governance structures within occupational safety systems, Figure 1 presents the AI-Augmented Safety Governance Model (AASGM). The model illustrates the interaction among real-time hazard-detection technologies, human decision-making processes, organizational safety culture, and regulatory compliance mechanisms within a unified socio-technical framework.

The AASGM provides a conceptual architecture for understanding how artificial intelligence, human oversight, and governance mechanisms interact within occupational safety systems.

Figure 1

AI-Augmented Safety Governance Model (AASGM)



Note. Author created. The model illustrates the integration of artificial intelligence-enabled hazard detection with human oversight, organizational safety practices, and regulatory governance structures. AI functions as an augmentation layer within a socio-technical system, supporting real-time hazard detection while maintaining human accountability and regulatory alignment.

As illustrated in Figure 1, the AASGM conceptualizes AI-enabled hazard detection as a component within a broader socio-technical governance system rather than as an independent technological solution. The model demonstrates that effective safety outcomes emerge from the interaction between technological capability, human oversight, organizational culture, and regulatory alignment. This integrated structure reinforces the principle that artificial intelligence enhances, but does not replace, human decision-making in safety-critical environments and highlights the importance of governance mechanisms in translating AI-generated insights into accountable safety decisions.

5. Implications of the AASGM

5.1 Theoretical Implications

The AASGM extends socio-technical systems theory by incorporating AI as a formal component of safety systems.

5.2 Practical Implications

Organizations can use the model to guide AI implementation while maintaining a safety culture and compliance.

5.3 Policy Implications

The framework supports the development of regulatory guidelines for AI-enabled safety systems. Beyond its conceptual structure, the AI-Augmented Safety Governance Model (AASGM) has direct implications for organizational governance and decision-making. By integrating technical detection capabilities, human factors, and socio-technical interactions into a unified framework, the model provides organizations with a structured approach to translating AI-generated safety data into actionable governance processes. This enables more effective oversight, supports alignment with regulatory requirements, and facilitates the development of adaptive safety strategies. The AASGM framework highlights the importance of coordinating technological capability with human engagement and organizational structures, reinforcing the need for governance systems that are both data-driven and contextually responsive within complex operational environments.

5.4 Organizational Implementation Scenario

Consider a manufacturing organization implementing YOLO-based hazard-detection systems across its production facilities, which have historically experienced high rates of PPE noncompliance and equipment-related safety incidents. Within the AASGM framework, AI-enabled monitoring systems function as real-time detection mechanisms that identify unsafe conditions and generate immediate alerts for supervisors and safety personnel. Human oversight remains central to decision-making, with managers reviewing AI-generated outputs before initiating corrective actions or compliance interventions.

The framework further supports organizational governance by integrating detection data into broader safety management processes, including incident-trend analysis, workforce training initiatives, compliance audits, and executive-level risk reporting. Regulatory alignment is maintained through transparency procedures, workforce communication protocols, and documented human accountability structures. In this context, the AASGM enables organizations to operationalize AI-assisted safety monitoring while maintaining human-centered governance and adaptive risk-management practices.

5.5 Implementation Challenges and Organizational Barriers

Despite the potential benefits of AI-enabled safety governance systems, organizations may encounter significant implementation challenges. Workforce resistance to AI-assisted monitoring may emerge when employees perceive automated systems as intrusive or overly surveillance-oriented. Concerns related to privacy, trust, algorithmic transparency, and data reliability may further affect workforce acceptance and organizational adoption.

In addition, variations in data quality, environmental conditions, and system calibration may influence detection consistency and operational effectiveness across industries. Ethical considerations related to accountability, decision-making authority, and the appropriate balance between automation and human oversight also remain important governance concerns. The AASGM framework addresses these challenges by emphasizing transparency, human-centered implementation, and structured oversight mechanisms that maintain human accountability within AI-enabled occupational safety systems.

Organizations implementing the AASGM may operationalize the framework through phased integration strategies involving AI deployment, workforce training, governance alignment, regulatory auditing, and continuous system evaluation.

6. Limitations

This study is subject to several limitations. First, the AI-Augmented Safety Governance Model (AASGM) is a conceptual framework derived from integrating technical, human-centered, and socio-technical analyses. It has not yet been empirically validated across diverse organizational settings. Second, while the model provides a structured approach to governance integration, variations in organizational structure, regulatory environments, and workforce characteristics may influence its applicability and implementation. Third, the framework reflects current technological and regulatory conditions, which are subject to rapid change as artificial intelligence capabilities and policy landscapes evolve. Future research should focus on empirical validation of the model, including case-based implementation, cross-sector testing, and evaluation of its effectiveness in supporting organizational decision-making and regulatory alignment.

7. Conclusion

The AI-Augmented Safety Governance Model provides a comprehensive framework for integrating artificial intelligence into occupational safety systems. By aligning AI technologies with human oversight, organizational practices, and regulatory structures, the model supports the advancement of proactive safety systems and the broader Safety 4.0 paradigm.

Conflict of Interest Statement

The author declares no conflicts of interest related to the research, analysis, or preparation of this manuscript. No external funding, sponsorship, or commercial support was received for this

study. All interpretations and conclusions reflect the author's independent scholarly judgment and professional expertise.

Originality Statement

This manuscript represents original scholarly work and has not been published previously in any form. It is not under review by any other journal or publication outlet. The author independently developed all conceptual frameworks, analyses, and written content as part of the Shawe Series research program.

Any use of external sources has been properly cited in accordance with APA 7 standards. The author affirms that the manuscript is free from plagiarism, duplication, or unauthorized reuse of prior publications.

Copyright Notice

© 2026 Dr. Robb Shawe.

This manuscript is part of the Shawe Series, a unified research program examining artificial intelligence-enabled hazard detection, socio-technical safety integration, and governance frameworks in regulated workplace environments. All conceptual models, analytical interpretations, and written materials represent original scholarly work developed by the author. Copyright remains with the author unless transferred to a publisher upon acceptance for publication.

The views expressed in this manuscript are those of the author and do not necessarily reflect the positions of affiliated institutions or organizations. No portion of this work may be reproduced, distributed, or transmitted without prior written permission, except where permitted under academic fair-use provisions.

References

- Carayon, P. (2006). Human factors of complex sociotechnical systems. *Applied Ergonomics*, 37(4), 525–535.
- Carayon, P., Schoofs Hundt, A., Karsh, B. T., Gurses, A. P., Alvarado, C. J., Smith, M., & Flatley Brennan, P. (2015). Work system design for patient safety: The SEIPS model. *Quality & Safety in Health Care*, 15(Suppl 1), i50–i58.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Kaplan, R. S., & Mikes, A. (2012). Managing risks: A new framework. *Harvard Business Review*, 90(6), 48–60.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Nath, N. D., Akhavian, R., & Behzadan, A. H. (2020). Automated detection of unsafe behaviors using computer vision. *Advanced Engineering Informatics*, 43, 101003.

- National Institute of Standards and Technology (NIST). (2024). *Cybersecurity framework (CSF) 2.0*. U.S. Department of Commerce.
- Occupational Safety and Health Administration (OSHA). (2024). Commonly used statistics. U.S. Department of Labor. <https://www.osha.gov/data>
- Power, M. (2007). *Organized uncertainty: Designing a world of risk management*. Oxford University Press.
- Salvendy, G. (2012). *Handbook of human factors and ergonomics* (4th ed.). Wiley.
- Shawe, R. (2025). Asbestos exposure and mesothelioma: Historical insights and modern technological impacts in the USA. *Open Journal of Safety Science and Technology*, 15(4), 301–310.
- Shawe, R. (2025). Evaluating the efficiency of occupational safety and health systems in New York's small and mid-size enterprises. *International Journal of Advanced Engineering and Management Research*, 10(6), 226–241.
- Shawe, R. (2025). From reactive to proactive: Artificial intelligence and predictive safety systems in OSHA-regulated environments. *International Journal of Advanced Engineering and Management Research*, 10(6), 78–92.
- Shawe, R. (2025). Integrating augmented reality and edge computing for enhanced hazard detection and safety monitoring. *Journal of Information Security*, 16(4), 517–527.
- Wickens, C. D., Lee, J. D., Liu, Y., & Gordon-Becker, S. (2015). *An introduction to human factors engineering* (2nd ed.). Pearson.
- Yousif, A., Al-Dahoud, A., & Al-Momani, A. (2024). Safety 4.0: The role of artificial intelligence in occupational safety systems. *Safety Science*, 170, 106356.