ISSN-L 2664-9969; E-ISSN 2706-5448

UDC 6.67.672
DOI: 10.15587/2706-5448.2025.339322

Viacheslav Berezutskyi

ASSESSING THE RISKS OF APPLYING ARTIFICIAL INTELLIGENCE TO OCCUPATIONAL SAFETY

This paper explores the opportunities, advantages, and risks of integrating artificial intelligence (AI) into occupational health and safety management systems. It is noted that the use of intelligent technologies contributes to improved workplace safety by enabling automatic monitoring of working conditions, detection and prediction of hazardous situations, and real-time analysis of workers' behavior. The potential of AI is demonstrated in identifying safety violations, monitoring the use of personal protective equipment, responding to dangerous events, and organizing preventive actions. Special attention is given to technical, legal, ethical, and organizational risks associated with AI implementation in industrial settings. The study analyzes risks related to AI-based systems in occupational safety using the example of a food processing plant with an automated packaging line. An incident involving worker injury due to the AI system's failure to detect human presence in the manipulator zone is examined. The application of the FMEA (failure modes and effects analysis) method identified key risk sources: failure to detect a person in the hazardous zone (RPN = 270), lack of integration between AI and emergency stop systems (RPN = 192), and loss of communication between modules (RPN = 140). All risks exceeded the RPN > 100 threshold, indicating high priority. The relevance of a multisensor approach, implementation of fail-safe protocols, and redesigning human – machine interaction architecture is substantiated. A comparison is made between the FMEA method and the PTSR (Probability – Time – Severity Risk), which incorporates the time factor of hazard exposure, increasing risk assessment accuracy in dynamic environments. A combined risk management approach is proposed, integrating preventive analysis (FMEA) and real-time operational evaluation (PTSR), which enhances safety control effectiveness when using adaptive AI systems.

Keywords: *artificial intelligence, occupational safety, risk management, ethics, legal responsibility, automation.*

Received: 09.06.2025
Received in revised form: 22.08.2025
Accepted: 11.09.2025
Published: 30.10.2025

How to cite
Berezutskyi, V. (2025). Assessing the risks of applying artificial intelligence to occupational safety. Technology Audit and Production Reserves, 5 (2 (85)), 26–32. https://doi.org/10.15587/2706-5448.2025.339322

1. Introduction

Artificial intelligence (AI) is rapidly reshaping contemporary manufacturing processes, unlocking new opportunities for managing complex technological systems. Its integration into such domains as industry, energy, transportation, and healthcare has been driving substantial improvements in operational efficiency, enabling decision-making automation, and reducing dependence on human factors. At the same time, the accelerated pace of digitalization and the proliferation of autonomous systems create an urgent need for a critical examination of AI's implications for occupational safety.

In highly automated environments, occupational safety acquires new dimensions. Traditional risk management frameworks, historically grounded in the statistical analysis of past incidents and in monitoring the technical condition of equipment, are proving increasingly inadequate for anticipating hazards within digital ecosystems. In such contexts, where decisions are executed in real time by self-learning models, novel categories of risk emerge: technical (algorithm failures), organizational (human-system coordination breakdowns), ethical (employee behavior monitoring), and legal (lack of regulatory accountability).

Global trends demonstrate the active deployment of AI in safety management systems, ranging from computer vision tools for monitoring the use of personal protective equipment to predictive analytics for accident prevention. Large-scale initiatives in countries such as China, Germany, the United States, and Japan are aimed at developing AI-driven autonomous safety systems. However, despite these advancements, incident rates linked to system imperfections or inconsistencies with established safety protocols are on the rise.

A review of the current literature reveals an abundance of studies addressing the advantages of AI in production process management, whereas considerably less attention is given to the occupational safety implications of AI integration.

For example, [1] examines the application of AI technologies in predictive maintenance (PdM) within industrial enterprises. While demonstrating the benefits of PdM in the context of Industry 4.0, the study overlooks risks stemming from erroneous machine learning model decisions that could impact worker safety.

In [2], the use of computer vision and image recognition algorithms in the construction industry is explored. The authors emphasize the effectiveness of such approaches in detecting safety violations, yet fail to address false alarm rates and questions of legal liability.

Study [3] investigates the deployment of automated monitoring systems for mine working conditions, describing technologies for detecting hazardous behavior, but omitting any assessment of risks posed by misclassifications, which may result in false responses or the neglect of actual dangers.

The ethical challenges inherent in automated personnel monitoring systems are discussed in [4], which stresses the importance of balancing automation with worker autonomy, though without offering concrete strategies to prevent conflicts between AI-generated and human operator decisions.

Publication [5] underscores the necessity of fail-safe mechanisms in adaptive control systems, raising concerns about unanticipated failures or erroneous outputs, but without providing a quantitative methodology for hazard level assessment.

In [6], the integration of AI into real-time risk recognition systems is analyzed, with examples of successful hazard detection on construction sites. However, recognition accuracy limitations, stemming from training dataset quality, are noted.

The legal dimensions of liability for AI-driven autonomous decisions are examined in [7], where the absence of a coherent regulatory framework is identified as a source of legal uncertainty in incident resolution.

Study [8] draws parallels between AI-based decision, support systems in healthcare, a domain characterized by high-stakes responsibility, and similar risks in industrial contexts, noting that high accuracy alone does not mitigate the absence of shutdown and oversight mechanisms.

Finally, [9] compares traditional and AI-oriented safety systems, concluding that the high-performance demands imposed by AI SW on the underlying HW. Its inherent complexity (both in terms of hardware (HW) and SW) is challenging for safety standards compliance.

From this review, several reasons emerge as to why occupational safety in the context of AI adoption remains insufficiently studied:

– the majority of publications emphasize AI's benefits while neglecting associated risks;

– there is no systematic framework for classifying threats arising from autonomous system operations;

– regulatory and legal ambiguities hinder the formulation of enforceable requirements for AI system developers and users in the occupational safety domain;

– empirical studies specifically addressing industrial safety – rather than solely technical performance – remain scarce.

Consequently, the identification, quantitative evaluation, and prioritization of risks associated with AI integration into occupational safety systems constitute a timely and complex scientific challenge. Key obstacles include the absence of adapted risk analysis tools, the difficulty of modeling adaptive system behavior, and the deficiencies of existing regulatory mechanisms.

To address these challenges, this study proposes the integration of multiple risk assessment methodologies, specifically failure modes and effects analysis (FMEA) and potential task safety risk (PTSR) assessment, toward the development of a scientifically grounded framework for analyzing and evaluating risks in occupational safety systems.

Research aim: to develop a scientifically substantiated approach for the analysis and assessment of risks associated with the implementation of AI technologies in occupational safety systems, employing FMEA and PTSR methodologies.

2. Materials and Methods

Research Object: industrial environments in which Artificial Intelligence (AI) technologies are deployed for automated monitoring, decision-making, and occupational safety management.

Research Hypothesis. This study hypothesizes that the integration of the *failure modes and effects analysis* (FMEA) and *potential task safety risk* (PTSR) methodologies into a unified analytical framework enhances the capability to systematically identify, evaluate, and prioritize risks associated with the application of adaptive AI systems in occupational safety contexts.

Assumptions. The research is conducted under the following assumptions:

– AI systems operate autonomously within predefined operational scenarios;

– machine vision training datasets possess limited situational diversity;

– the technical environment does not support complete multisensor data verification;

– explicit regulatory requirements for AI-emergency system integration are absent.

Simplifications. The following methodological simplifications were applied:

– risk assessment is confined to individual production operations, excluding macro-level system interactions;

– the AI system behavioral model excludes real-time self-learning capabilities;

– threat prioritization is determined by fixed weighting coefficients without adaptive recalibration.

Research Materials. The materials used in the study were obtained from open-access sources, including:

– technical documentation from industrial enterprises;

– occupational safety reports;

– peer-reviewed scientific publications;

– international standards (ISO 31010:2019);

– documented case studies of AI implementation (e. g., Tesla, Siemens, and AI-assisted Chinese mining operations).

*Methodological Framework***.** Within the developed risk analysis approach, the following methods were applied:

– the FMEA method, used to assess severity (S), occurrence (O), and detectability parameters (D), with subsequent calculation of the risk priority number;

– the PTSR method, which incorporates an additional parameter of system reaction time (T), enabling the calculation of an aggregated risk index.

All calculations were carried out manually on an applied case study, which allowed for a comparison of results and the justification of the effectiveness of combining both methods within a unified approach.

3. Results and Discussion

3.1. Risk analysis of artificial intelligence applications in occupational safety

As part of the first stage of the proposed approach, the primary types of risks associated with the implementation of artificial intelligence (AI) technologies in occupational safety systems were identified and classified.

The analysis was conducted based on publicly available information, including case descriptions, technical documentation of industrial systems, analytical reports, and peer-reviewed publications on the digital transformation of industry.

The study identified four main categories of risks:

1. *Technical risks* – related to false triggers, software malfunctions, misinterpretation of input data, or loss of connectivity between system components. Such failures can lead to incorrect AI system behavior under critical conditions that require high precision and rapid response.

2. *Legal risks* – stemming from the absence of a clearly defined regulatory framework governing liability for the actions of autonomous systems. In particular, the issue of legal responsibility for decisions made by AI without human operator involvement remains unresolved.

3. *Ethical risks* – involving violations of employee privacy, continuous monitoring, and potential discrimination based on automated assessment of behavior or physiological indicators.

4. *Cybersecurity risks* – associated with potential internal or external attacks on AI systems, data manipulation, or exploitation of vulnerabilities in network infrastructure, which may result in destabilization or compromise of the safety system.

All identified risks were systematized, and the results are summarized in Table 1.

Table 1

Key risks associated with the application of AI in occupational safety

Risk type	Description	Possible mitigation measures
Technical	Algorithm failure, false triggers	Testing, backup systems, manual confirmation
Legal	Uncertainty of liability for AI decisions	Development of legal norms, action log recording
Ethical	Violation of employee privacy	Data anonymization, ethical protocols
Cybersecurity	Possibility of hacking or AI manipulation	Encryption, access restriction, system audit

The results of the analysis indicate that the integration of AI technologies into occupational safety systems involves a complex set of interrelated risks encompassing technical, legal, ethical, and informational security dimensions. Timely identification and classification of these risks is a necessary prerequisite for subsequent quantitative assessment, prioritization, and the development of effective risk management measures based on a systems approach combining the FMEA and PTSR methods.

3.2. FMEA analysis of potential failures in artificial intelligence systems used for occupational safety

In the course of implementing the next stage of the proposed approach, a *failure modes and effects analysis* (FMEA) was conducted for the main components of artificial intelligence systems applied in the field of industrial safety.

The evaluation was performed according to three criteria:

– *S* (Severity) – the seriousness of the failure consequences;

– *O* (Occurrence) – the likelihood of occurrence;

– *D* (Detection) – the ability to detect a potential failure.

Based on these evaluations, the *risk priority number* (*RPN*) was calculated as the product of the three indicators

$$RPN = S \cdot O \cdot D. \tag{1}$$

The results are presented in Table 2.

The evaluation was based on scenarios of typical violations recorded in industrial environments. Explanation of criteria:

– *S* (Severity): severity of consequences, with 1 indicating negligible impact and 10 indicating fatal outcome;

– *O* (Occurrence): likelihood of occurrence, with 1 being rare and 10 being frequent;

– *D* (Detection): probability of detecting the defect before manifestation, with 1 meaning always detected and 10 meaning never detected.

Table 2

Example of FMEA analysis for AI systems in occupational safety

Failure type	Consequences	*S*	*O*	*D*	*RPN*
Incorrect recognition	Failure to detect hazard, worker injury	9	6	5	270
Loss of connection to sensors	Lack of data, erroneous situation assessment	8	5	4	160
Model update failure	Use of outdated or incorrect algorithm	7	4	6	168
Biased training	Incorrect threat prioritization, discrimination	6	5	6	180
Cyber threat or attack	System compromise or external control	10	3	7	210

The *RPN* is calculated as the product of *S*, *O*, and *D*, allowing for risk prioritization. A limitation of this approach is that different combinations of factors can result in identical *RPN* values, complicating the differentiation of critical threats.

The FMEA method is widely applied in functional safety practices (e. g., ISO 14971, IEC 61508) and in the development of intelligent technical systems. Its use enables structuring information on potential hazards, assessing failure criticality, and justifying the need for preventive measures in critical industrial environments employing autonomous or semi-autonomous AI systems.

3.3. Analysis of occupational safety incident and risk assessment using FMEA at a food production facility

This study examined a production incident at a food manufacturing plant equipped with an automated packaging line c ontrolled by an artificial intelligence system. The AI system was responsible for visual product identification, monitoring packaging quality, and coordinating the operation of manipulators. During scheduled maintenance, a worker entered the manipulator's operating zone without stopping the line. The system failed to detect the person's presence, resulting in the worker's hand injury.

The FMEA method was applied to analyze the incident causes and prioritize associated risks. Numerical values for severity (*S*), occurrence (*O*), and detection (*D*) parameters were derived from typical values used in international practice and adapted to the specific case conditions. All assessments were expert judgments intended for preliminary risk ranking.

The analysis results are presented in Table 3.

The obtained *RPN* values (270, 192, and 140) significantly exceed the approximate threshold of 100, indicating high risk priority. The highest risk is associated with failure to detect the worker, primarily due to limitations of computer vision systems that struggle to reliably recognize objects in complex poses, specific protective clothing, or partial occlusions. It is recommended to expand training datasets and employ multisensor systems combining visual, thermal, and depth data.

The second priority risk concerns the lack of communication between the AI system and traditional safety mechanisms. The absence of hardware integration between AI and emergency stop devices creates critical conditions where the system cannot adequately respond to a person entering a hazardous zone. Technical integration of the AI software core with physical control systems is advised.

Table 3

FMEA risk analysis for AI use incident in production

Potential failure	Cause	Possible consequences	*S*	*O*	*D*	*RPN*	Preventive measures
Failure to detect worker presence	Insufficient training dataset, unusual poses	Worker injury	9	6	5	270	Expand dataset, add LiDAR and thermal imaging sensors
Lack of AI integration with emergency systems	Inadequate system design	Inability to stop equipment	8	4	6	192	Install physical emergency stops, integrate AI with safety circuits
Loss of communication between AI modules	Network failure or communication error	Unpredictable system behavior	7	5	4	140	Develop safe shutdown protocols, implement fail-safe mechanisms

The third risk involves loss of communication between control system modules. In cases of network or software environment failures, the AI may act inappropriately if safe shutdown protocols are not implemented. Introducing fail-safe algorithms to switch equipment to a safe state upon communication loss is recommended.

These results demonstrate the practical applicability of FMEA as a tool for risk assessment in the operation of adaptive AI systems within industrial settings. This approach allows not only identification of critical hazard zones but also formulation of recommendations to prevent injuries and enhance overall technological safety.

Comparison of risk assessment methods: FMEA and PTSR.

To broaden the analytical capabilities for occupational safety risk assessment in AI applications, the PTSR (probability – time – severity risk) method proposed in [10] was utilized. The difference from the method proposed in [10] is that instead of the number of personnel, the factor of the time of exposure of the worker to the hazard is introduced, which increases the accuracy of risk assessment in a dynamic environment. This method calculates an integral risk indicator considering three factors:

- P – probability of hazard occurrence;
- T – duration of hazard exposure;
- S – severity of consequences.

Risk is defined as the product of these three factors

$$R = f(P, T, S) = P \cdot T \cdot S. \qquad (2)$$

Based on the numeric value of the integral risk, situations are classified as acceptable, moderately hazardous, or unacceptable.

The main advantage of the PTSR method is inclusion of the time parameter (T), enabling more precise evaluation of the temporal characteristics of risk. This is particularly relevant for adaptive AI systems that may interact with risks repeatedly, maintaining or increasing hazardous impact over time.

Thus, combined use of FMEA and PTSR methods is advisable, allowing coverage of a broad spectrum of factors and enhancing risk assessment accuracy (Table 4).

The FMEA method is appropriate for identifying technical failures and weaknesses during AI system design. In contrast, the PTSR approach, which incorporates exposure duration, allows more flexible risk evaluation during actual operation when AI behavior is dynamic and interaction scenarios with humans and environments are unpredictable.

Table 4

Comparative characteristics of FMEA and PTSR methods

Criterion	FMEA (failure modes and effects analysis)	PTSR (probability – time – severity risk)
Purpose	Identification of potential failures and effects	Integral hazard assessment considering time
Type of assessment	Quantitative (*RPN*)	Quantitative (*PTS*)
AI application focus	Technical failures, algorithmic errors	Dynamic situations, prolonged risks
Attention focus	Component risk structure	Context and dynamics of hazard manifestation
Advantages	Ease of implementation, standard support	Higher accuracy for repeated exposure risks
Limitations	Static approach, limited time consideration	Does not cover causal relationships
Recommended usage	Design phase	Operation and real-time monitoring

3.4. Analysis of risk assessment methods applied to artificial intelligence systems

In the field of safety related to the application of artificial intelligence (AI), hybrid approaches that combine traditional risk analysis methods with novel techniques adapted to digital and autonomous systems are most commonly employed. Below are the most frequently used methods in industry, IT, and safety standards:

Commonly applied safety analysis methods for AI:

1. *FMEA (failure modes and effects analysis)* – easily adaptable to new technologies, allowing for the rapid identification of critical failure points within a system. Application in AI includes the development of safe autonomous systems, machine vision, drones, robotics, etc.

2. *HAZOP (hazard and operability study)* – enables systematic investigation of potential unexpected system behaviors. It is applied in high-risk industries (energy, chemical industry) where AI is integrated into critical processes.

3. *ISO/PAS 21448 (SOTIF – safety of the intended functionality)* – a modern standard for AI safety analysis in transportation (automobiles, autonomous systems). Its importance lies in addressing the fact that even a "correctly functioning" AI can be hazardous if operating conditions are outside the scope of training.

4. *STRIDE (threat modeling for AI/ML)* – ensures cybersecurity and protects models against manipulation (e. g., data poisoning, adversarial attacks). This is particularly relevant in IoT, facial recognition systems, and financial models.

5. *ALARP (As low as reasonably practicable)* – a practical engineering approach to determine when a risk level is sufficiently low. It is widely applied in risk acceptability assessments when implementing new technologies, particularly in aviation and healthcare.

6. *STAMP (systems-theoretic accident model and processes)* – a novel systemic method for analyzing incidents and risks, including control and information flows. Applied in autonomous vehicles, drones, and highly automated production systems.

Additional methods to consider include:

7. *Bayesian risk analysis* – suitable for models involving uncertainty and probabilistic reasoning.

8. *Ethical AI checklists/AI explainability tools* – critical for fostering trust in AI decisions, especially in healthcare and judicial systems.

Given the complexity of ensuring safety in industrial settings, it is recommended to consider a combination of several methods. For example, if selecting 2–3 risk assessment methods for AI implementation in an enterprise, the best combination (depending on the context, industrial or digital) would be: FMEA, HAZOP, and either STRIDE or SOTIF.

Key risk groups are typically distinguished in research, encompassing both technical and social dimensions [11]. Below is a summary of the main risks most frequently investigated:

1. *Technical risks:* Incorrect or unpredictable AI behavior; models may make wrong decisions in complex or novel situations not covered in training. This risk is especially critical in robotics, autonomous systems, and machine vision. The quality of training data significantly affects technical risk: if training data or sensor inputs contain errors or outdated information, hazardous working conditions may arise. Explainability issues: inability to justify decisions complicates human oversight. The risk of failures during model or software updates is also notable, as unchecked updates can introduce new vulnerabilities or faulty algorithms.

2. *Cyber risks (AI security):* Attacks on models (adversarial attacks) can deceive AI by manipulating inputs, leading to incorrect system responses (e. g., failing to recognize a person near a crane). Leakage of personal or industrial data may result in automated systems handling sensitive information requiring protection [12].

3. *Social and organizational risks:* Overautomation ("automation bias"), people tend to trust AI decisions without verification, which is

dangerous in production or healthcare environments. Psychosocial risks to employees include loss of process control, stress, risk of dismissal, or role changes due to AI implementation. Ethical risks, such as discrimination arising from biased algorithms, may create unsafe or unfair working conditions for certain employee groups [12].

4. *AI-based decision-making risks:* Automatic violation detection may penalize or warn employees without context, infringing labor rights; conflicts between algorithmic and human decisions under uncertainty (e. g., in construction or mining). These risks are systematized in reports and studies [13].

The summarized analysis of primary AI risks in occupational safety with examples and research sources is presented in Table 5.

An approximate distribution in percentages can be made for each risk group based on the frequency of their mention in scientific publications and reports (especially EU-OSHA, NIST, ArXiv, ScienceDirect). This will give an idea of which risks researchers pay the most attention to in the context of occupational safety.

An approximate distribution of the risks of AI application in occupational safety (by frequency of research) is shown in Fig. 1.

The distribution is approximate because it is based on a content analysis of scientific publications on occupational safety and AI. The frequencies reflect how often specific risks appear in scientific sources, not the actual probability of their occurrence. All values are presented in 5% increments for clarity and ease of perception.

According to Fig. 1, labor law conflicts constitute the smallest share, only 5.0% of total studies, indicating relatively low attention to this issue in the context of AI implementation in the workplace.

Next, groups of psychosocial, ethical, legal risks related to AI's impact on employees, overautomation risks, and cybersecurity threats each account for approximately 15%. A similar 15% share is assigned to data quality issues, algorithmic bias, and lack of explainability, reflecting the complexity of evaluating AI systems' objectivity and transparency. The highest research focus is on technical failures and errors, recognized as among the most critical risk categories due to their direct impact on the safety, reliability, and stability of AI systems in industrial environments.

3.5. Discussion of findings on the risks associated with the application of artificial intelligence in occupational safety

The proposed risk assessment framework departs fundamentally from conventional methodologies in several respects. First, it employs a hybrid model integrating the principles of ISO 31010:2019 with the failure mode and effects analysis (FMEA) methodology and advanced artificial intelligence (AI) tools. Second, rather than relying solely on expert judgment, the approach incorporates real-time algorithms capable of processing large-scale data streams from sensors, video surveillance systems, and digital event logs. Third, AI-based systems facilitate not only reactive control but also predictive analytics, exceeding the capabilities of traditional safety mechanisms that are limited to post-incident interventions.

Previous works [1, 4, 6] have documented early attempts to integrate AI into occupational safety management. However, the present study demonstrates that coupling intelligent monitoring with scenario-based failure analysis and active human operator engagement offers superior effectiveness in preventing critical incidents.

Table 5

Risks of AI application in occupational safety

Risk category	Description	Examples	Source/research
Technical	Malfunction or AI limitations	Incorrect person recognition, sensor failures	EU-OSHA (2023) [14], ScienceDirect (2024) [15], ISO 21448 (SOTIF) [16]
Unreliable data	Low quality or limited representativeness	Biased training data, data from limited cases	ArXiv: 2401.09459 [17], OSHA [14], NIST [18]
Lack of explainability	Model cannot justify decisions	Autonomous drone reroutes without operator logic	Explainable AI (XAI) [19], EU-OSHA [14]
Overautomation	Blind trust in AI	Operator ignores alerts assuming "AI never errs"	MEDIUM (2025) [20], EU-OSHA [14]
AI cybersecurity	External attacks or algorithm manipulation	Adversarial attacks substituting helmet/person images	ISO/IEC 24028 [21], ScienceDirect [15]
Discrimination/Bias	Algorithm creates unequal conditions	Underestimation of risk for women or novices	OECD AI Principles [22], EU AI Act [23]
Reduced well-being	Psychological stress, loss of autonomy	Workers fear constant AI surveillance	BauA (2022) [24], arXiv [17]
Labor law conflicts	Automated penalties, dismissals without human input	Algorithm detects "violations" without explanation or appeal	EU-OSHA [14], ILO (International Labour Organization) [25]

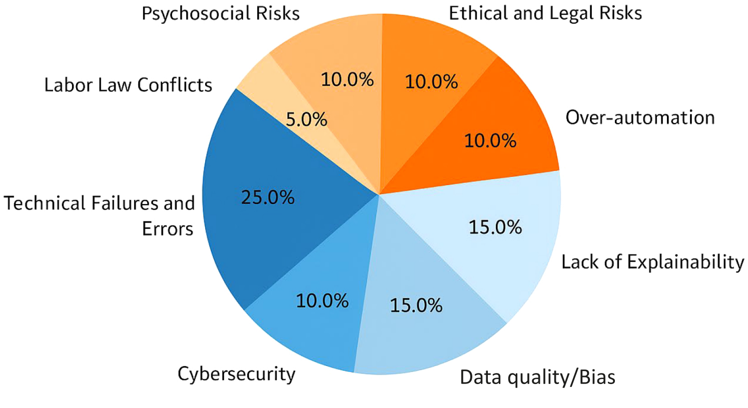

Fig. 1. The distribution of major AI application risks in occupational safety by frequency of study

The empirical results (Fig. 1, Tables 1 and 3) identify the most prominent threats as technical failures (25%), inter-module communication breakdowns, errors in human presence detection, and insufficient integration with emergency systems. FMEA results (Table 3) reveal that the highest-priority risk is the failure to detect a worker within the operational range of the AI system (risk priority number, $RPN = 270$), which significantly exceeds the acceptable threshold ($RPN > 100$). This finding underscores that even minor algorithmic inaccuracies may lead to severe consequences. The evidence aligns with documented workplace injury cases [26–30], where the deployment of intelligent systems could have reduced or prevented harm.

To enhance the analytical scope, the probability – time – severity – risk (PTSR) approach was also applied, which accounts not only for the probability and severity of hazards but also for their duration. For instance, the aforementioned failure to detect a worker received a higher PTSR rating due to prolonged human presence in a hazardous area, thereby increasing the likelihood of injury. Consequently, FMEA proves particularly effective during the design stage, whereas PTSR is more suitable for operational monitoring, especially in dynamic industrial environments. The combined application of these methods allows for a more robust risk management framework: FMEA for structured preventive analysis and PTSR for real-time adaptive monitoring and response.

In comparison with existing practices, the proposed method successfully integrates the analytical depth of FMEA with the adaptability and temporal sensitivity of PTSR. This enables the consideration of both design constraints and real-time operational conditions. The importance of such an approach grows in automated settings involving robotic manipulators and machine vision systems, where environmental changes occur rapidly and the human factor remains a significant source of risk.

Key limitations of the study include:

- *Scope of applicability*: the model is tailored primarily to automated sectors (manufacturing, logistics, mining).
- *Technical dependency*: the performance of AI models relies on the quality of input data, the maintenance of sensor systems, and software stability.
- *Reproducibility*: replication requires standardization of hardware configurations and training protocols.
- *Context sensitivity*: models may yield unreliable outputs under unanticipated scenarios not represented during training (e. g., natural disasters, cyberattacks, atypical human behavior).

Further limitations:

- *Limited experimental validation*: not all operational scenarios were tested under real-world conditions; some findings are based on simulations.
- *Lack of universality*: systems require customization for individual enterprises, entailing additional resources.
- *Ethical ambiguity*: there is an absence of established norms defining AI accountability in safety-critical contexts.

Potential strategies to address these constraints involve pilot-scale implementations, the formulation of ethical guidelines, and the development of open-access datasets for training and validating AI-driven safety models.

Future research directions include:

- Designing adaptive hybrid models that combine AI with human expert oversight.
- Advancing intelligent video analytics systems to detect occupational safety violations in real time.
- Developing digital twins of production processes equipped with predictive and risk analysis capabilities.

Anticipated challenges:

- The mathematical complexity of building adaptive models in dynamic environments.
- The requirement for large-scale datasets for model training and validation.
- Resistance from personnel toward the adoption of novel technologies.
- Heightened cybersecurity risks in deeply digitized industrial infrastructures.

4. Conclusions

It has been established that the most characteristic threats associated with the integration of artificial intelligence (AI) technologies into occupational safety systems include: erroneous information recognition, degradation of training data quality, inability to interpret AI-driven decisions, and limitations in accounting for environmental context.

The analysis revealed that the scientific community predominantly focuses on the technical aspects of risks, while social, ethical, and legal implications remain underexplored. This imbalance can be attributed to the difficulty of formalizing social risks within the framework of technical analysis, as well as the absence of integrated interdisciplinary risk management models.

The adaptation of the failure mode and effects analysis (FMEA) method to the assessment of risks associated with autonomous AI behavior proved to be effective. This approach enabled the formalization of novel failure types arising from input data quality and algorithmic properties, as well as the quantitative evaluation of threats using the risk priority number (RPN). Such an approach allows for the identification of critical risks already at the design stage of intelligent safety systems.

Quantitative RPN values obtained in the study include: 270 – uncontrolled algorithmic intervention in critical operations; 192 – erroneous product classification; 140 – accuracy degradation due to environmental changes. All values exceed the critical threshold of $RPN = 100$, confirming the need for additional human oversight mechanisms.

The integration of FMEA with the probability – time – severity – risk (PTSR) method, which accounts for the temporal dynamics of hazards, significantly improved the accuracy and relevance of risk assessments in real-time systems.

Comparative analysis of various methods demonstrated that hybrid strategies, combining FMEA with other approaches (PTSR, HAZOP, STRIDE, etc.) tailored to the specific application domain, are the most effective. Such integration ensures a comprehensive consideration of technical, cybersecurity, and temporal aspects of risks. Unlike traditional assessment methods, hybrid models better align with the dynamic nature of AI in industrial environments, enabling greater analytical flexibility, more precise identification of failures, and more effective prioritization of safety measures.

Conflict of interest

The author declares no conflict of interest regarding this research, including financial, personal, authorship-related, or any other factors that could have influenced the research and its results as presented in this article.

Financing

This research received no financial support.

Data availability

No associated datasets are available with this manuscript.

Use of artificial intelligence

The author used artificial intelligence technologies within acceptable limits to provide original, verified data, as described in the research methodology section.

References

1. Falsk, R. (2023). *AI for Predictive Maintenance in Industrial Systems*. Handbuch Ansehen. https://doi.org/10.13140/RG.2.2.27313.35688
2. Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., Wang, X. (2020). Computer Vision Techniques in Construction: A Critical Review. *Archives of Computational Methods in Engineering, 28 (5),* 3383–3397. https://doi.org/10.1007/s11831-020-09504-3
3. Chen, K., Wang, C., Chen, L., Niu, X., Zhang, Y., Wan, J. (2020). Smart safety early warning system of coal mine production based on WSNs. *Safety Science, 124*, 104609. https://doi.org/10.1016/j.ssci.2020.104609
4. Cebulla, A., Szpak, Z., Howell, C., Knight, G., Hussain, S. (2022). Applying ethics to AI in the workplace: the design of a scorecard for Australian workplace health and safety. *AI & SOCIETY, 38 (2),* 919–935. https://doi.org/10.1007/s00146-022-01460-9
5. Fernández Peñalver, M. (2024). The Foundations of AI Safety: Ensuring Technical Robustness. *Nemko*. Available at: https://www.nemko.com/blog/ai-safety-and-robustness
6. Alateeq, M. M., Rajeena, F. P. P., Ali, M. A. S. (2023). Construction Site Hazards Identification Using Deep Learning and Computer Vision. *Sustainability, 15 (3),* 2358. https://doi.org/10.3390/su15032358
7. Nagda, P. (2025). Legal Liability and Accountability in AI Decision-Making: Challenges and Solutions. *International Journal of Innovative Research in Technology, 11 (11).* Available at: https://ijirt.org/publishedpaper/IJIRT174899_PAPER.pdf
8. Čartolovni, A., Tomičić, A., Lazić Mosler, E. (2022). Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. *International Journal of Medical Informatics, 161,* 104738. https://doi.org/10.1016/j.ijmedinf.2022.104738
9. Fernández, J. (2024). *Integrating AI into Functional Safety Management. Safe and Explainable*. Critical Embedded Systems based on AI. Available at: https://safexplain.eu/integrating-ai-into-functional-safety-management/
10. Berezutskyi, P. S., Horbenko, S. V. (2017). Otsenka rycka ot KhPY. *Okhrana truda, 11,* 14–16. Available at: https://www.researchgate.net/publication/394486032_Formiruem_risk-orientirovannoe_myslenie
11. Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., Sheth, A. P. (2018). Machine learning for internet of things data analysis: a survey. *Digital Communications and Networks, 4 (3),* 161–175. https://doi.org/10.1016/j.dcan.2017.10.002
12. AI in worker management: involving people to prevent risks (2025). *ENSHPO*. Available at: https://www.enshpo.eu/ai-in-worker-management-involving-people-to-prevent-risks/
13. Lialiuk, O., Osypenko, R. (2023). Features of the implementation of artificial intelligence in construction. *Modern Technology, Materials and Design in Construction, 35 (2),* 172–176. https://doi.org/10.31649/2311-1429-2023-2-172-176
14. Implementing safer AI worker management through policy and prevention (2024). *European Agency for Safety and Health at Work (EU-OSHA)*. Available at: https://osha.europa.eu/en/oshnews/implementing-safer-ai-worker-management-through-policy-and-prevention
15. Chauhan, S., Vashishtha, G., Zimroz, R. (2024). Analysing Recent Breakthroughs in Fault Diagnosis through Sensor: A Comprehensive Overview. *Computer Modeling in Engineering & Sciences, 141 (3),* 1983–2020. https://doi.org/10.32604/cmes.2024.055633
16. ISO 21448:2022. Road vehicles – Safety of the intended functionality (2022). *ISO*. Available at: https://www.iso.org/standard/77490.html
17. Ryan, P., Porter, Z., Al-Qaddoumi, J., McDermid, J., Habli, I. (2023). *What's my role? Modelling responsibility for AI-based safety-critical systems*. arXiv:2401.09459. https://doi.org/10.48550/arXiv.2401.09459
18. AI Risk Management Framework (AI RMF 1.0) (2023). *NIST*. Available at: https://www.nist.gov/itl/ai-risk-management-framework
19. Gunning, D. (2017). XAI: Explainable Artificial Intelligence. *DARPA*. Available at: https://www.darpa.mil/program/explainable-artificial-intelligence
20. Epelboim, M. (2025). *Cursor Rules: Why Your AI Agent Is Ignoring You (and How to Fix It)*. Available at: https://sdrmike.medium.com/cursor-rules-why-your-ai-agent-is-ignoring-you-and-how-to-fix-it-5b4d2ac0b1b0
21. ISO/IEC TR 24028:2020. Information technology – Artificial intelligence Overview of trustworthiness in artificial intelligence (2020). *ISO*. Available at: https://www.iso.org/standard/77608.html
22. *Recommendation of the Council on Artificial Intelligence* (2019). Paris: OECD Publishing. Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
23. Kusche, I. (2024). Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk. *Journal of Risk Research,* 1–14. https://doi.org/10.1080/13669877.2024.2350720
24. An Occupational Safety and Health Perspective on Human in Control and AI (2022). *BauA*. Available at: https://www.baua.de/EN/Service/Publications/Essays/article3454
25. Huibregtse, A. (2025). AI provides innovative ways to improve compliance with labour laws. *ILO*. Available at: https://www.ilo.org/resource/article/ai-provides-innovative-ways-improve-compliance-labour-laws
26. Volkswagen robot kills worker in Germany (2015). *The Guardian*. Available at: https://www.theguardian.com/world/2015/jul/02/robot-kills-worker-at-volkswagen-plant-in-germany
27. Yang, X., Li, Y., Chen, Y., Li, Y., Dai, L., Feng, R., Duh, Y.-S. (2020). Case study on the catastrophic explosion of a chemical plant for production of m-phenylenediamine. *Journal of Loss Prevention in the Process Industries, 67,* 104232. https://doi.org/10.1016/j.jlp.2020.104232
28. Tim, B., Zoë, D., Gerald, P. (2019). Mineworker fatigue: A review of what we know and future decisions. *Minerals Engineering, 70 (3),* 33. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC5983045/
29. Reports of Fatalities and Catastrophes – Archive. *OSHA*. Available at: https://www.osha.gov/fatalities/reports/archive
30. Travmatyzm. Statystyka. Prychyny. *Derzhavna sluzhba Ukrainy z pytan pratsi*. Available at: https://dsp.gov.ua/category/diyalnist/travmatyzm-statystyka-prychyny/

Viacheslav Berezutskyi, *Doctor of Technical Science, Professor, Department of Occupational and Environmental Safety, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine, ORCID: https://orcid.org/0000-0002-7318-1039, e-mail: viaberezuc@gmail.com*