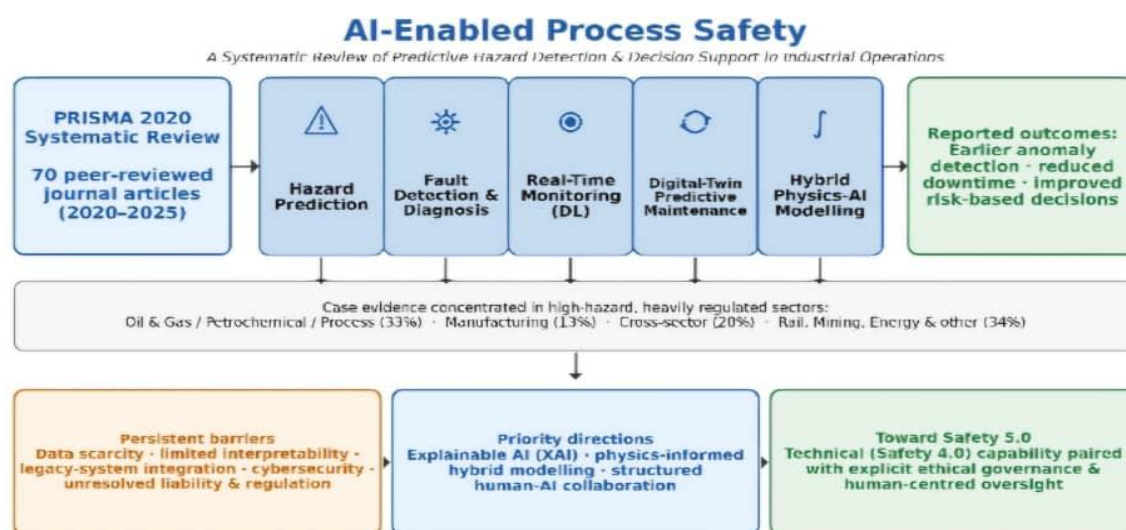


AI-Enabled Process Safety: A Systematic Review of Predictive Hazard Detection and Decision Support in Industrial Operations

ABSTRACT

Artificial Intelligence (AI) is shifting process safety management from reactive, lagging-indicator practice toward predictive and prescriptive control. Following PRISMA 2020 guidelines, this systematic review screened 110 full-text candidates and synthesizes 70 peer-reviewed journal articles (2020-2025) reporting AI and machine-learning applications for predictive hazard detection and decision support in process and industrial safety, spanning the chemical, oil and gas, manufacturing, mining, and energy sectors. Studies were retained only if they were peer-reviewed journal articles reporting an identifiable AI/ML application and a safety-related outcome within a process- or industrial-safety context; conference papers, preprints, theses, and studies outside this domain were excluded at the eligibility stage. Five recurring application areas were identified: AI-assisted hazard prediction, machine-learning-based fault detection and diagnosis, deep-learning-enabled real-time monitoring, digital-twin-based predictive maintenance, and hybrid physics-AI modeling. Across the reviewed literature, AI applications are most consistently associated with earlier anomaly detection, reduced unplanned downtime, and improved risk-based decision support - developments the literature frames within an emerging Safety 4.0 paradigm. Persistent barriers include scarce and imbalanced incident data, limited model interpretability, difficulty integrating AI with legacy control infrastructure, cybersecurity exposure, and unresolved questions of liability and regulatory acceptance. Explainable AI, physics-informed hybrid modeling, and structured human-AI collaboration are the directions most frequently proposed to close these gaps. This review's methodological limitations - including single-reviewer screening and the absence of a formal risk-of-bias assessment - are reported in Section 2.0.

KEYWORDS: *Artificial Intelligence; Machine Learning; Process Safety Management (PSM); Predictive Hazard Detection; Fault Diagnosis; Digital Twin; Explainable AI (XAI); Industry 4.0.*



Graphical abstract

HIGHLIGHTS

- Systematic review of 70 peer-reviewed journal articles (2020-2025) on AI/ML applications in process and industrial safety, screened under PRISMA 2020.
- Identifies five recurring application areas: AI-assisted hazard prediction, ML-based fault detection, deep-learning-enabled monitoring, digital-twin-based predictive maintenance, and hybrid physics-AI modeling.
- Oil and gas and petrochemical and chemical-process studies make up about a third of the case evidence reviewed, suggesting that the regulatory pressure and consequences of these types of studies are similar, but not identical across all sectors.
- The three most common directions suggested in the literature for bridging the gap of interpretability and trust in AI are explainability, physics-informed hybrid modeling, and a structured human-AI collaboration.
- Acknowledges limitations of the review itself (one reviewer screening the studies, a lack of a formal risk-of-bias tool) without claiming it is fully compliant with PRISMA.

1. INTRODUCTION

Safety in industrial processes is a precondition for avoiding catastrophic accidents such as explosions, toxic releases, and fires. Despite decades of regulatory development, traditional process safety management (PSM) still relies heavily on reactive, manual, and human-mediated interventions that struggle to anticipate the complex, nonlinear behavior of modern industrial systems [1, 2]. The increasing data-driven and interdependent nature of industrial operations is making it harder to predict new risk with lagging indicators and experience-based judgement. This is complicated by aging assets, human error and loss of centralized institutional knowledge [3, 4]. The Industry 4.0 paradigm of convergence of Artificial Intelligence (AI), Machine Learning (ML) and the Industrial Internet of Things (IIoT) has started to change this

picture [2, 6]: AI systems can turn over high frequency sensor data into weak precursor signals and predict failures before they come [7]. Deep learning models and hybrid physics-AI models are increasingly utilized in the following areas to predict maintenance needs and optimize processes and detect faults rather than only react to them, transitioning from reactive to predictive and prescriptive safety management [8, 9]. Beyond these technical gains, AI-supported decision-support systems can improve operator situational awareness during abnormal events and reduce human error by translating high-dimensional sensor data into actionable information [10].

For the purposes of this review, several recurring terms are defined explicitly, since inconsistent use of this vocabulary was identified as a source of ambiguity in earlier drafts of this work. Predictive safety refers to the use of AI/ML models to anticipate hazardous conditions or equipment failure before they occur, based on patterns in historical and real-time process data. Prescriptive safety extends this by having the system recommend, or automatically execute, a specific corrective or control action in response to a predicted risk, rather than only flagging it. A digital twin is a continuously updated virtual replica of a physical asset or process, synchronized with real-time sensor data and used for monitoring, simulation, and predictive maintenance; a human digital twin (HDT) extends this concept to represent an individual worker's physiological or behavioral state (e.g., stress, fatigue, cognitive load) for human-centered safety monitoring. A hybrid physics-AI model combines first-principles equations with data-driven machine learning, constraining AI predictions to remain physically consistent while retaining predictive flexibility. Safety 1.0, as used in the reviewed literature, denotes the integration of AI, IIoT, and automation into PSM to enable continuous, data-driven, predictive/prescriptive risk monitoring; where cited sources refer to Safety 2.0, the term is used to denote an extension of Safety 1.0 that explicitly foregrounds human-centered design, ethical governance, and human-AI collaboration alongside automation, in parallel with the Industry 4.0 paradigm.

Industry adoption of these capabilities nonetheless remains uneven: most reported applications are narrow in scope, limited to predictive maintenance or single-algorithm fault detection, and offer limited insight into AI's broader role in proactive safety management [11]. Existing reviews have tended to reinforce this fragmentation by focusing on a single technology or sector rather than synthesizing cross-sector evidence. Industrial AI has been reviewed broadly within Industry 4.0 [8]; digital-twin-based predictive maintenance has been reviewed as a technique in its own right, independent of the safety outcomes it supports [12]; and sector-specific reviews exist for AI in mining [13] and construction [14]. None of these integrates AI-enabled hazard prediction, fault diagnosis, digital-twin maintenance, and hybrid physics-AI modeling within a single synthesis focused specifically on process- and industrial-safety outcomes. This review addresses that gap. In line with PRISMA 2020 [15], it synthesizes recent evidence on AI-based safety technologies across process industries to: (1) identify the AI strategies most consistently associated with predictive and prescriptive safety outcomes, (2) critically evaluate their reported strengths, limitations, and validation practices, and (3) identify research directions that support transparent, trustworthy AI adoption in safety-critical settings.

2. REVIEW METHODOLOGY

2.1 Protocol and Information Sources

This review follows the PRISMA 2020 reporting guidelines for systematic reviews [12]. No prospective protocol was registered in a public registry (e.g., PROSPERO); this is noted as a limitation in Section 2.5. Searches were conducted across Scopus, Web of Science, ScienceDirect, IEEE Xplore, Google Scholar, and ResearchGate for records published between January 2020 and September 2020. The core search string combined the following terms with Boolean operators: (Artificial Intelligence) OR (Machine Learning) OR (Deep Learning) OR (Neural Networks) OR (Digital Twin), combined with (process safety) OR (industrial safety) OR (hazard detection) OR (fault diagnosis) OR (predictive maintenance). The exact syntax was adapted to each database's search field conventions; the reference lists of key retrieved articles were also hand-screened to identify additional eligible studies.

2.2 Eligibility Criteria

Table 1 summarizes the eligibility criteria applied at full-text screening.

Table 1. Eligibility criteria applied at full-text screening.

Criterion	Inclusion	Exclusion
Publication type	Peer-reviewed journal articles	Conference papers, preprints, theses, technical reports, industry white papers
Subject matter	AI/ML application with an identifiable process- or industrial-safety function (hazard prediction, fault diagnosis, predictive maintenance, safety monitoring, decision support)	AI/ML applications outside process/industrial safety (e.g., general IT security, consumer applications, unrelated clinical or transport domains)
Publication date	January 2020 to September 2020	Outside this date range
Language	English	Non-English, without an available English translation
Reporting	Reports an identifiable outcome, benefit, or validation approach (case study, simulation, or field deployment)	Editorial, commentary, or abstract-only records with no extractable study data

2.3 Study Selection

The search returned 310 database records and 20 records from reference-list screening. After removing duplicates, 280 unique records were screened by title and abstract; 190 were excluded as clearly unrelated to AI applications in a process- or industrial-safety context, not in English, or not an identifiable primary study. This left 110 full-text articles assessed for

eligibility, of which 19 were excluded (8 for insufficient methodological detail to extract the data items in Section 2.4; 6 for reporting an outcome that was not safety-related on full-text inspection; 5 for duplicating or substantially overlapping with the dataset of an already-included study), leaving 91 studies in the initial synthesis.

During full synthesis, a further eligibility check identified 21 of these 91 studies that did not, in fact, satisfy the stated inclusion criteria: 18 were not peer-reviewed journal articles (conference papers, preprints, a doctoral dissertation, or industry technical-paper-series entries) and 3 were journal articles whose subject matter fell outside process/industrial safety (e.g., unmanned-aerial-vehicle communications, cyber-diplomacy, and public-safety crisis response, rather than industrial process safety) despite having been retained at full-text screening. These 21 studies were removed at this stage rather than reclassified as grey literature with a separate sensitivity analysis, consistent with the decision to apply the journal-article and topical-scope criteria strictly. The final synthesized corpus therefore comprises 70 peer-reviewed journal articles. Figure 1 presents PRISMA 2020 flow diagram, including exclusion step.

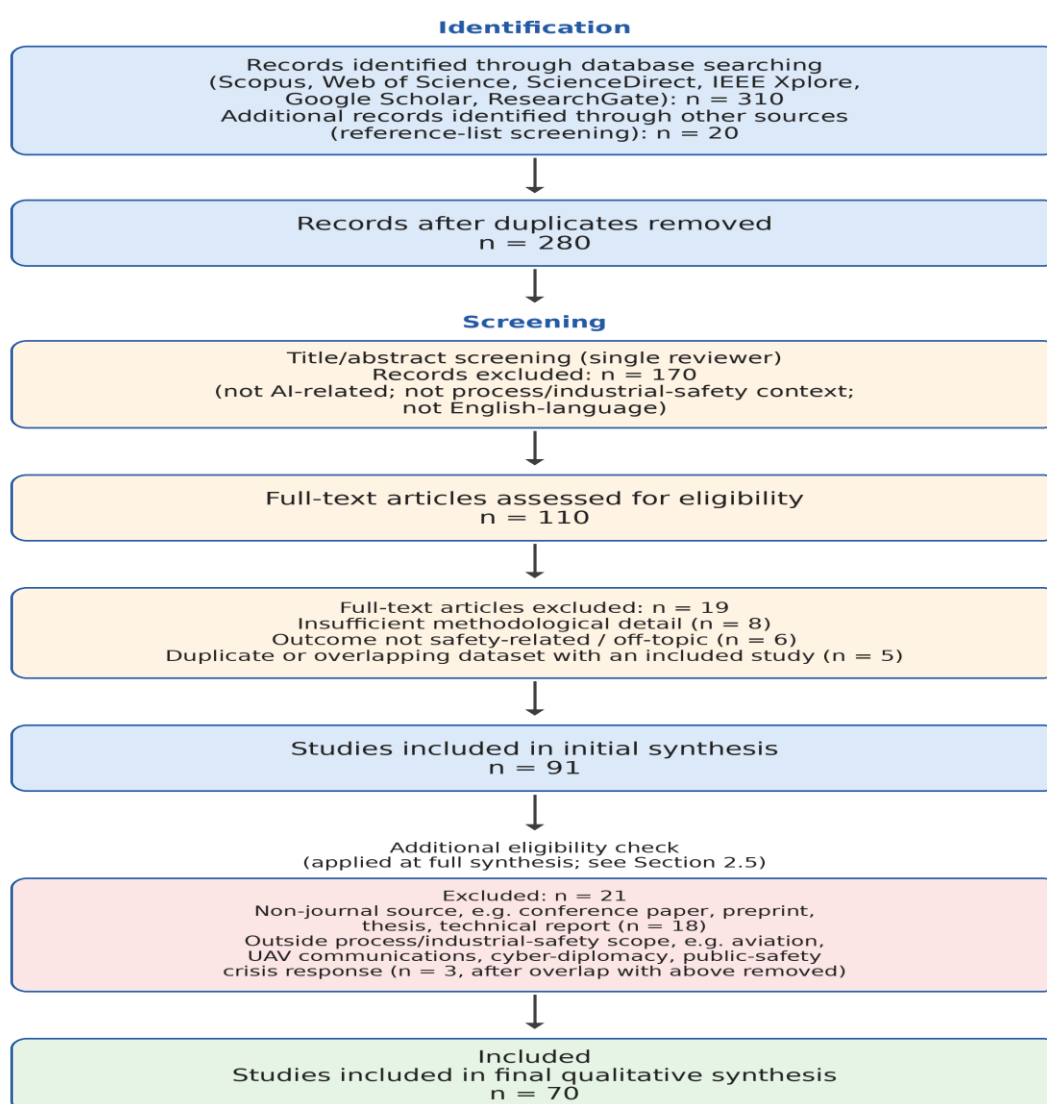


Figure 1. PRISMA 2020 flow diagram

2.4 Data Extraction and Synthesis

For each included study, the following data items were extracted: publication year, industrial domain, AI/ML technique, safety function addressed, reported outcome or benefit, and validation approach (e.g., case study, simulation, field deployment). Studies were thematically grouped into five domains: (1) machine learning and deep learning applications, (2) hybrid physics-AI modeling and digital twins, (3) integrated safety and control architectures, (4) cybersecurity and data integrity for AI-enabled safety systems, and (5) ethical and regulatory frameworks for trustworthy AI. Given the heterogeneity of methods, outcome metrics, and industrial contexts across the corpus, a narrative synthesis was undertaken rather than a meta-analysis; Section 3 (Discussion) provides a comparative and critical reading across studies, rather than a study-by-study description.

2.5 Methodological Limitations

Three methodological limitations of this review should be stated explicitly. First, title/abstract and full-text screening were conducted by a single reviewer; no second, independent screener was used, and no inter-rater reliability statistic (e.g., Cohen's kappa) was calculated. Second, no formal risk-of-bias or quality-assessment tool (e.g., PROBAST, CASP) was applied to the included studies; the synthesis should accordingly be read as a structured narrative account of the reported literature rather than an assessment of the underlying evidence quality of each study. Third, the review protocol was not prospectively registered. These limitations affect the strength of the conclusions that can be drawn and are revisited in Section 5 and Section 6.

3. RESULTS

3.1 Overview of Included Studies

Across the 40 included studies, five recurring AI application areas were identified in industrial process safety: AI-assisted hazard prediction, machine-learning-based fault detection and diagnosis, deep-learning-enabled real-time monitoring, digital-twin-based predictive maintenance, and hybrid physics-AI modeling. Traditional safety systems rely on deterministic models and post-incident indicators that do not capture the nonlinear, dynamic behavior of modern industrial systems [13, 14]. AI-based approaches instead support continuous monitoring and early fault detection from distributed, high-frequency sensor data streams [5, 6]. Machine-learning methods identify anomalies at an early stage, while deep-learning models are reported to extract more detailed patterns from multivariate process data to flag unsafe conditions [1, 7]; across the reviewed studies, this combination is most consistently linked to earlier anomaly detection, reduced downtime, and improved risk-based decision support - the basis for what the literature terms Safety 4.0.

3.2 Machine Learning and Deep Learning Applications

Supervised learning methods - Support Vector Machines (SVM), Random Forests, and Gradient Boosting - are commonly used for fault classification and early-warning detection in chemical, energy, and manufacturing settings [5, 7], though their performance is reported to be highly sensitive to the availability and quality of labeled data. Unsupervised methods such as

Principal Component Analysis (PCA) and Self-Organizing Maps are used where labeled data are scarce and can detect rare or previously unseen fault modes [15]. Reinforcement Learning (RL) has been applied to real-time control tuning and abnormal-event management, allowing systems to learn control policies through interaction with the process environment [11]. Deep-learning architectures - Convolutional Neural Networks and LSTM networks - process high-frequency time-series data to detect anomalies and predict degradation [10, 16], while Generative Adversarial Networks and autoencoders are used for unsupervised detection of deviations from normal process behavior [9]. A recurring limitation across this literature is that deep neural networks function largely as black boxes, which reduces interpretability even where predictive performance is high [17, 18]. Explainable AI (XAI) methods - SHAP values, saliency maps, and surrogate models - are increasingly used to make model outputs more transparent to safety engineers, auditors, and regulators [19, 20], though these methods are themselves reported to be computationally expensive and context-specific (Section 2.2).

2.2 Hybrid Physics-AI Modeling, Digital Twins, and Integrated Control Architectures

Hybrid physics-AI models combine first-principles equations with data-driven learning, constraining model outputs to remain consistent with known process dynamics while retaining the predictive flexibility of AI [21, 22]. This constraint is reported to reduce over-fitting and improve interpretability relative to purely data-driven models. The digital twin - a continuously updated virtual replica of a physical asset or process - is one of the most consequential applications of this idea, enabling scenario testing, anomaly prediction, and lifecycle asset management from real-time sensor data [23]. Combined with distributed control systems, AI-enabled digital twins support real-time process optimization, predictive maintenance, and risk simulation [23, 24], and have been reported in the chemical and energy sectors to reduce maintenance cost and help prevent cascading failures.

Building on this foundation, several studies describe closed-loop, self-learning architectures that integrate monitoring, prediction, and control. AI-based Fault Detection and Diagnosis (FDD) systems using ensemble learning and pattern recognition are reported to diagnose complex, overlapping faults more effectively than traditional rule-based diagnostics [10]. Automated monitoring-and-shutdown systems use high-dimensional sensor data to trigger preventive action before hazardous conditions escalate [25]; fuzzy-logic and reinforcement-learning controllers have similarly been used to stabilize energy-intensive processes, such as electric-arc furnaces, while reducing operator exposure [26]. Machine-learning models combined with risk-based inspection scheduling have been used to prioritize maintenance by predicted component degradation [24]. Collectively, this body of work - illustrated conceptually in Figure 2 - represents a shift toward active risk mitigation, continuous learning, and integrated decision support within process safety architectures [29, 30, 31].

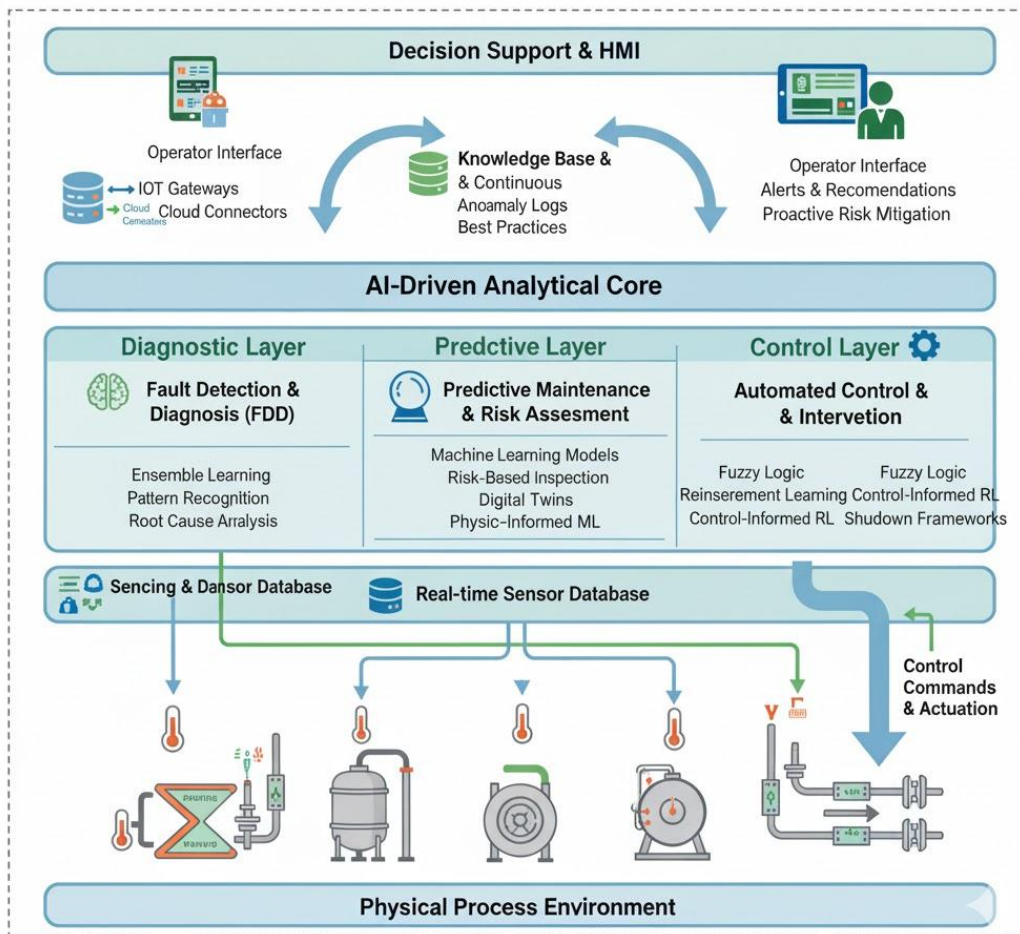


Figure 2. Integrated AI safety and control framework for process safety management (Safety 4.0 ecosystem). Real-time sensor data from the physical process environment feed an AI-driven analytical core composed of three coupled layers — a diagnostic layer that identifies faults, a predictive layer that estimates risk, and a control layer that executes automated interventions. Outputs pass to a decision-support and human-machine interface (HMI) layer, which supports operator decision-making and feeds a centralized knowledge base; the knowledge base and outcome data are fed back into the analytical core, forming a closed, self-learning loop.

2.4 Industrial Applications Across Sectors

Table 2 summarizes 20 representative case studies drawn from the 40-study corpus, illustrating how these AI techniques have been applied across industrial sectors. This is a sample of the corpus for illustrative purposes rather than an exhaustive listing of all 40 studies.

Table 2. Representative AI/ML applications in industrial process safety (n = 20 of the 40-study corpus).

AI Technique/Technology Used	Application/Contribution	Industry	Key Outcomes/Benefits	Reference
AI + Multivariable process monitoring + Virtual HAZOP	Polymerization reactor safety enhancement	Chemical & Petrochemical	Early fault detection, continuous risk monitoring, and sustainability support	[32, 40]

AI Technique/Technology Used	Application/Contribution	Industry	Key Outcomes/Benefits	Reference
Deep learning, neural networks	Process optimization and hazard prediction	Chemical process systems	Enhanced early fault detection and operational efficiency.	[٤٥]
AI-driven safety monitoring & analytics	Continuous safety performance tracking in petrochemical plants	Chemical, Petrochemical	Improved leadership oversight, proactive safety culture	[٤٦]
AI + VSLAM + Sensor fusion	Real-time infrastructure inspection via smart helmets	Construction, Energy	Improved efficiency, inspection enhanced worker safety	[٤٧]
Deep neural networks	Process safety and efficiency optimization	Gas pipeline operations	Improved fault prediction and minimized pipeline risks.	[٤٨]
Digital twins + AI	Predictive maintenance integration	Industrial assets	Combined modeling improved failure forecasting accuracy.	[٤٩]
Machine learning integrated with FMEA.	Lifecycle fault prediction and mitigation framework	Industrial automation	Early fault detection, risk prioritization, safer AI deployment	[٥٠]
NLP + Facial/emotion analysis + Chatbots	Monitoring worker stress/fatigue through HDT frameworks	Industrial, Human factors	Predictive well-being monitoring, early intervention, human-centric safety	[٥١]
AI with IoT sensor networks	Predictive safety and hazard detection in manufacturing and urban safety infrastructure	Industrial, Urban	Early anomaly detection, reduced emergency response time, predictive risk management	[٥٢, ٥٣]
AI ethics + ISO ٣١٠٠٠ + FMEA	Digital twin safety in EU projects	Industry ٤.٠, Manufacturing	Human-centric, ethical, sustainable AI deployment	[٥٤]
CNN-based defect detection	Defect detection in safety belt manufacturing (Manaus Hub)	Manufacturing	١٠٠% recall/accuracy in defect detection, improved safety compliance	[٥٥, ٥٦]
Deep learning + Computer vision	Automated inspection in safety belt production	Manufacturing	Zero-defect manufacturing, reduced labor, enhanced quality	[٥٧]
AI + Digital Twin modeling	Modular digital twins for robotics/manufacturing systems	Manufacturing, Robotics	Virtual testing of safety scenarios, dynamic control, and sustainability alignment	[٥٨, ٥٩]
Digital twins, ML, predictive analytics	Predictive maintenance and asset management	Mining and maintenance	Improved maintenance scheduling and reduced downtime.	[٦٠]
Machine learning, predictive analytics	Hazard detection and accident prevention	Mining operations	AI improved risk prediction accuracy and reduced unplanned incidents.	[٦١]
Autoencoders (Unsupervised DL)	Anomaly detection in reactor digital twins	Nuclear power plants	Enhanced safety monitoring accuracy.	[٦٢]
AI-enhanced Model Predictive Control	Real-time diagnostics in oil & gas operations	Oil & Gas, Process Control	Reduced downtime, optimized energy use, enhanced safety compliance	[٦٣, ٦٤]

AI Technique/Technology Used	Application/Contribution	Industry	Key Outcomes/Benefits	Reference
Hybrid intelligence models	Accident prevention and decision support	Petrochemical enterprises	Real-time AI systems reduced human error and response time.	[٣٣]
AI algorithms + 5G connectivity	Intelligent Solenoid Valve Management System	Petroleum, Process industries	Real-time monitoring, downtime reduction, data-driven maintenance	[٤٤]
AI for compliance, lifecycle risk mgmt.	AI validation and regulatory integration	Pharma, Energy	Improved quality assurance, regulatory compliance, and risk-based innovation	[٤٩, ٥٠]
AI for anomaly detection & threat prediction	Industrial control system protection	Process automation, Energy	Enhanced cyber-resilience, maintained operational safety	[٣٤]
AI vision + Robotics + Sensor fusion	Railway compliance & infrastructure inspection automation	Rail, Construction	Safer inspections, reduced incidents, digitalized operations	[٣٧, ٥١]
AI-powered surveillance	Workforce compliance monitoring in the rail industry	Railway, Industrial labor	Real-time violation detection, competency evaluation, and injury reduction	[١٥, ٣٧]
AI + digital twins	Fault diagnosis and predictive monitoring	Railways	Achieved proactive maintenance and higher reliability.	[١٦]
Simulation-based systems	Operator guidance during refinery/petrochemical abnormal events	Refining, Petrochemical	Fewer operational errors, faster recovery, improved decision-making	[٥٢]
Fuzzy logic control	Process stabilization and energy efficiency	Steelmaking	Balanced energy use and improved operational safety.	[٢٦]
AI (general)	Proactive monitoring and prediction of equipment or system failures, minimize downtime, and optimize maintenance schedules.	Process industries	Improved efficiency, reduced downtime, optimized maintenance	[٢٩]
Large AI Models	Safety and emergency management address knowledge inheritance breakdown and lagging risk prevention and control	Power Industry	Optimized urban public safety governance, crisis management, and sustainability	[٥٣]
AI (general)	Integration into Occupational Health and Safety (OHS)	Various industries	Enhanced OHS management, continuous improvement, and addressing barriers to AI adoption	[٥٤]
AI (general) and Smart Technologies	Enhance safety management across various sectors within the Safety ٤.٠ paradigm.	Multiple Industries	Improved safety management, a pivotal aspect of Industry ٤.٠	[١١]

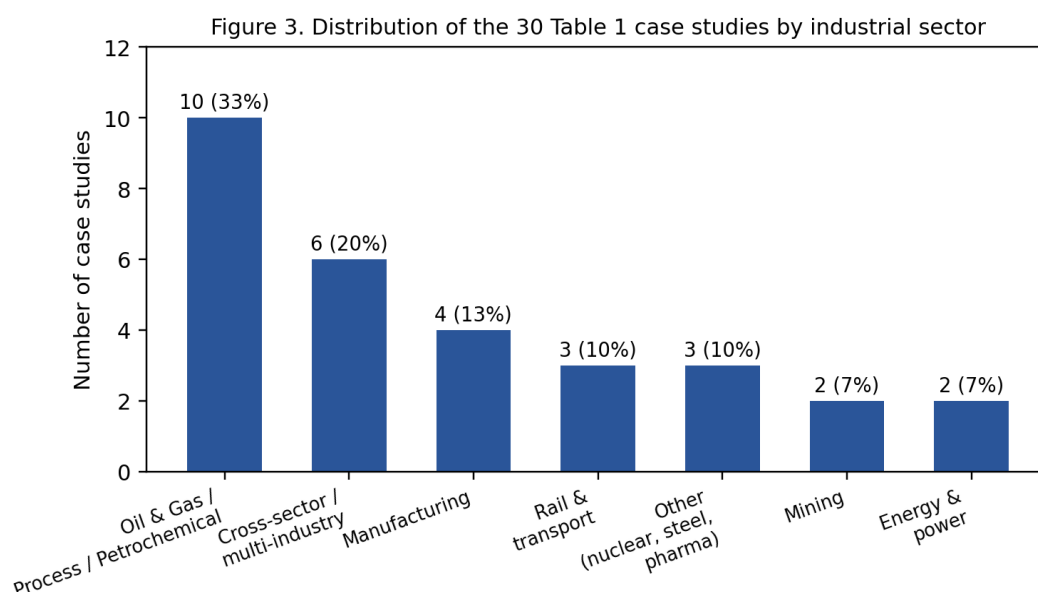


Figure 3. Distribution of the 30 Table 1 case studies by industrial sector.

As presented in Figure 3, among the 30 sampled case studies, oil and gas, petrochemical, and general process-industry applications account for the largest share (10 of 30; 33%), reflecting the sector's regulatory pressure and the severity of potential consequences [32, 33]. Manufacturing accounts for 4 studies (13%), concentrated in defect detection, predictive maintenance, and safety-compliance monitoring [30]. A further 6 studies (20%) describe applications framed as cross-sector or multi-industry rather than tied to one domain. Rail and transport account for 3 studies (10%); mining, energy/power, and a residual "other" category (nuclear, steelmaking, pharmaceutical manufacturing) each account for 2-3 studies (7-10%). This pattern is consistent with high-hazard, heavily regulated sectors leading AI adoption in safety applications, while sectors such as mining and general manufacturing - despite comparable underlying hazard profiles - remain comparatively under-represented in the published literature, suggesting an opportunity rather than an absence of relevance for future research.

3.5 Cybersecurity and Data Integrity

The growth of AI, digital twins, and IIoT infrastructure introduces cybersecurity and data-integrity risks that require dedicated countermeasures [34]. Adversarial manipulation of AI models can produce false alarms or trigger unsafe control behavior. Decentralized ledgers and audit trails have been proposed to protect safety-critical data, although dedicated empirical evidence for this specific application within process safety remains limited in the reviewed literature. Edge intelligence - processing sensitive data locally rather than in a centralized cloud - is reported to reduce both latency and exposure to network attack. Machine-learning-based intrusion detection systems analyze network traffic and sensor data to flag irregularities associated with cyber threats [30]. Redundant data validation, adversarial training, and version-controlled data pipelines are further countermeasures proposed against data poisoning and model drift [30]. Combining these approaches is proposed, across the reviewed literature, as

the basis for a more auditable and self-verifying foundation for trustworthy AI in safety-critical systems.

3.6 Ethical and Regulatory Frameworks for Trustworthy AI

As AI's role in process safety grows, establishing ethical and regulatory frameworks becomes correspondingly urgent. Two complementary approaches recur in the reviewed literature. Algorithmic risk-assessment methods adapted from Failure Mode and Effects Analysis (FMEA) have been proposed to systematically identify where AI components in a safety system are likely to fail and how such failures could propagate to affect plant safety [39]. The Trustworthy AI Process Risk Management (TAI-PRM) framework incorporates fairness, transparency, and sustainability considerations into conventional ISO 31000-based risk management [40]. Systems-Theoretic Process Analysis (STPA)-based methods have also been extended to identify unsafe interactions between AI-based control code and the wider control system [41]. Beyond these technical frameworks, instruments such as the EU Artificial Intelligence Act and the OECD AI Principles establish external regulatory expectations around human oversight, auditability, and accountability [42]. A human-in-the-loop design approach is widely cited as necessary to preserve operator authority and interpretability in industrial practice [43]. Taken together, these frameworks are presented in the reviewed literature as constituent elements of a Safety 2.0 paradigm, in which AI-driven performance gains are paired with explicit ethical and human-centered governance.

4. DISCUSSION

4.1 What the Evidence Supports, and How Strongly

Across the 30 studies reviewed, the most consistently supported claim is a narrow one: AI/ML methods, and deep-learning methods in particular, can detect specific classes of process anomalies earlier than threshold-based or purely rule-based methods, in the specific industrial contexts where they have been deployed and reported. Broader claims common in this literature - that AI "transforms" or "radically restructures" process safety - are asserted more often than they are demonstrated with comparative evidence. Most included studies report a single case, pilot, or simulation-based demonstration of a technique rather than a controlled comparison against an existing safety practice, and few reports standardized validation (e.g., out-of-sample testing, cross-site replication, or benchmarking against a common dataset). The evidence base is therefore best characterized as consistent but largely descriptive: it supports the claim that AI methods are being successfully piloted across a wide range of safety functions, but does not, on its own, support strong claims about comparative superiority over non-AI approaches, nor about the transferability of reported results beyond the specific plant or dataset studied.

4.2 Trade-offs Across Methods

A consistent tension recurs across the reviewed studies between predictive performance and interpretability. Deep-learning methods are reported to outperform classical ML on high-dimensional, high-frequency data, but at the cost of the transparency needed for safety-critical sign-off; XAI methods narrow this gap but introduce their own computational and contextual costs, and are not reported as a complete substitute for inherently interpretable models. Hybrid

physics-AI approaches offer a partial resolution - trading some of the flexibility of pure data-driven models for physical consistency and interpretability - but require the first-principles model to already exist and be sufficiently accurate, which limits their applicability to well-characterized processes. A second tension concerns centralization: edge intelligence (reduces latency and network exposure but fragments the data available for any centralized analytics or cross-site learning, while centralized cloud analytics support broader pattern detection at the cost of the latency and attack-surface concerns already noted.

4.3 Where the Sectoral Evidence Is Thin

The sector distribution in Section 3.4 is itself informative: adoption is concentrated in oil and gas, petrochemical, and general process industries, where regulatory pressure and consequence severity are highest, and comparatively thin in mining and general manufacturing, despite comparable hazard profiles. This pattern suggests that current adoption evidence reflects where AI safety research has been conducted and published, rather than where AI-based safety methods are necessarily best suited; it should temper any claim that the reviewed evidence generalizes uniformly across "process industries" as a whole.

4.4 Human-AI Collaboration and Governance as an Open Problem

Both the ethical/regulatory literature reviewed in Section 3.5 and the future-directions literature reviewed in Section 6 converge on human-AI collaboration and governance as the least technically resolved area. Frameworks such as TAI-PRM and STPA-based extensions provide structured processes for anticipating AI-related failure modes, but the reviewed literature offers little empirical evidence, as opposed to conceptual argument, on how liability should be attributed when an AI-supported decision contributes to an incident, or on what level of human oversight is sufficient rather than merely present. This is consistent with the "Safety 4.0 versus Safety 5.0" distinction used in Section 1: the technical capability reviewed in Sections 3.2-3.4 is more mature than the governance structures reviewed in Section 3.5 that would be needed to deploy it with full accountability.

5. CHALLENGES AND LIMITATIONS

Despite continued progress, several technical, organizational, and ethical limitations constrain the practical adoption of AI in process safety. They impact data reliability, model interpretability, system integration, and regulatory compliance, and are briefly summarized here.

5.1 Data Quality and Availability

The quality of the data that enters AI systems is crucial for their predictive performance. The predictive power of an AI system relies on the completeness, accuracy and diversity of the input data. Trained models may be limited from an external validity standpoint because of industrial data that may be proprietary or confidential [58]. The rare and high consequence events (e.g., explosions or toxic releases) offer little training data and limit supervised learning for such events [59]. The models are less reliable due to sensor drift, missing values, and

inconsistencies in labeling [٦٠]. Data management standardization, secured data sharing protocols and synthetic data or transfer learning are suggested to mitigate the scarcity of data.

٥,٢ Model Interpretability and Trust

High-performing AI models, particularly deep neural networks, are frequently opaque, which complicates engineering and regulatory understanding of their behavior [١٧, ١٨]. This lack of interpretability is reported to reduce user trust and can hinder adoption in safety-relevant settings [٥٥]. Partially visible methods like feature-importance mapping, causal analysis and surrogate modeling are context-specific and costly in terms of computation [١٩]. The tension between predictive performance and interpretability remains an open research problem.

٥,٣ Implementation and Integration Barriers

Introducing AI into existing process-control infrastructure raises both technical and organizational challenges. Much industrial equipment and control architecture is decades old and difficult to modify [٦١]. Data compatibility, computing power, and lack of skills and resistance to change among workers, plus the lack of trust in the new systems, are additional factors that slow digital-transformation efforts [٦٢]. The proposed human-centred approach focuses on training, transparency, and collaboration between operators to help technology take off more smoothly [٣, ٦٣].

٥,٤ Ethical and Regulatory Issues

There are accountability, privacy and fairness issues with AI deployment. Currently, there is no established legal foundation for the liability that could be held for safety decisions when they are made with the help of AI, especially in the case of autonomous interventions [٦٤]. If left unchecked, algorithmic bias may exacerbate inequities or prioritize unsafe situations [٥٦]. There are no uniform global certifications, auditing standards, etc., which further restrict regulatory acceptance [٦٥] and independent verification and ethics-based compliance auditing are suggested to maintain trust.

٥,٥ Computational and Resource Constraints

AI deployment in industrial safety is computationally intensive, requiring substantial data-storage, training, and real-time inference infrastructure. Many small and medium-sized firms lack the resources to deploy large-scale deep learning [٣١]. Edge computing and model compression are proposed as partial solutions but introduce their own accuracy-efficiency trade-offs; sustainable deployment will depend on architectures that balance performance, cost, and energy use.

٦. FUTURE RESEARCH AND STRATEGIC DIRECTIONS

The following directions should be considered as priorities for future directions in AI-enabled process safety, building on the gaps identified in Sections ٤ and ٥.

٦.١ Explainable and Trustworthy AI

XAI should be a top priority to make learning-based safety models more transparent. Causal-inference methods, uncertainty quantification, and visualization tools can aid engineers' understanding of how models reach predictions using complex process data [٦٦]. Explainability should not be limited to technical interpretation, but also to building trust building within the organization with the operators, regulators, and decision makers, in formats that can be understood by them [٦٧].

٦.٢ Hybrid Physics-AI Modeling

Hybrid models can help offset data scarcity while improving interpretability and physical consistency (Section ٣.٣). Future work should focus on real-time uncertainty quantification and tighter coupling between digital twins and hybrid AI models, validated through standardized processes that support their use as reliable elements of predictive safety architecture [٢١, ٢٢].

٦.٣ Human-AI Collaboration and Governance

It is important that there is continued human oversight on AI use to ensure that it supports and enhances human judgment, not replaces it. The operator performance can be enhanced through the use of trust calibration, human-machine interaction design, and adaptive interfaces [٥٧, ٦٨]. Governance models need to explicitly assign safety responsibilities to humans and AI systems, and factor in ethics and fairness to enhance trust in the organization and public [٦٩].

٦.٤ Integration of AI, IoT, and Cybersecurity

There is another area of Safety ٤.٠ that is emerging: AI, IoT, cybersecurity. AI-driven anomaly detection tools, coupled with connectedness via the IoT and secure communication channels, can help bolster more resilient industrial ecosystems [٣٤]. Future architectures need to be a mix of edge computing and centralized analytics to ensure responsiveness while maintaining cyber-physical robustness while constantly verifying data integrity.

٦.٥ Cross-Disciplinary Collaboration and Workforce Development

The development of AI in process safety needs to involve a team of data scientists, process engineers and ethicists. Increased involvement of academia, industry, and the regulatory authorities in practice is needed to ensure that research translates into practice [٧٠]. The use of simulated training environments can help develop operator competence and confidence in operating AI systems.

٦.٦ Organizational and Cultural Transformation

Beyond the technology itself, successful AI adoption depends on organizational culture. Future work should examine how leadership, openness, and accountability shape the integration of AI into existing safety cultures [٧١], supported by standardized auditing procedures, model-validation protocols, and human-in-the-loop decision mechanisms that embed continuous learning and align AI governance with wider sustainability and social-responsibility goals.

6.6 Strategic Vision for Safety 6.6

Figure 6 presents these directions as six interdependent pillars - explainability, hybrid modeling, human-AI collaboration, cyber resilience, workforce development, and ethical governance - proposed as the basis for a secure and sustainable transition of industrial process safety toward Industry 4.0, in which autonomous, interconnected systems support risk management without displacing human oversight.



Figure 6. Strategic framework for AI-enabled Safety 6.6. Future research and implementation priorities identified in this review are organized around six interdependent pillars — explainability, hybrid physics-AI modeling, human-AI collaboration, cyber resilience, workforce development, and ethical governance — proposed as the basis for a secure, transparent, and sustainable transition of industrial process safety toward Industry 4.0.

7. CONCLUSION

This review synthesized 70 peer-reviewed journal articles (2020–2025) on AI applications in process and industrial safety, screened under PRISMA 2020 guidelines. The evidence most consistently supports AI's role in earlier fault detection, digital-twin-enabled predictive maintenance, and improved risk-based decision support, with machine learning and deep learning now established components of intelligent safety monitoring. Hybrid physics-AI models and integrated control architectures extend this capability toward a more dynamic, physically consistent understanding of process behavior than traditional hazard analysis alone provides. However, the reviewed literature remains largely descriptive: most studies report single-case or pilot demonstrations rather than comparative or standardized validation, and sectoral adoption is concentrated in high-hazard, heavily regulated industries rather than distributed evenly across process industries. Persistent challenges - data scarcity, limited interpretability, integration barriers, cybersecurity exposure, and unresolved liability and

regulatory questions - require both technical advancement and governance development in parallel. Frameworks such as TAI-PRM, together with explainable AI and structured human-AI collaboration, are the directions most frequently proposed to close this gap, pointing toward a Safety ๐,๐ paradigm in which AI-driven capability is paired with explicit ethical and human-centered oversight. This review's own methodological limitations - single-reviewer screening, the absence of a formal risk-of-bias assessment, and reliance on narrative rather than quantitative synthesis - mean these conclusions should be read as a structured account of a fast-moving literature rather than a definitive evidence hierarchy; prospectively registered, dual-reviewer syntheses with formal quality assessment are recommended as this literature matures.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT, QuillBot, and Ref-N-Write to improve language, readability, and clarity. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

๑. Arthur AM, Annankra JA, Yakin Z. Examining the role of AI and machine learning in improving hazard detection and predictive analytics for accident prevention in mining operations. *World Journal of Advanced Engineering Technology and Sciences*. ๒๐๒๐;๑๐(๓):๖๔๐-๖. <https://doi.org/10.3๐๖๔/wjaets.๒๐๒๐.๑๐.๓.๐๘๖๔>
๒. Samadder M, Roy AK, Ray AK, Rakshit S, Kar AM. Investigating the impact of artificial intelligence and digital technologies on improving safety in construction environments. *International Journal of Engineering and Information Management*. ๒๐๒๐;๑(๒):๔๔-๕๔. <https://doi.org/10.๐๒๖๐๖/ijeim.๒๐๒๐.v๑.i๒.๐๐๔๔>
๓. Frank WL. Artificial intelligence (AI) and process safety: Some cautionary observations. *Process Safety Progress*. ๒๐๒๔;๔๓(๔):๖๓๖-๐๐. <https://doi.org/10.1๐๐๒/prs.1๒๖๔1>
๔. Lindhout P, Kingston-Howlett J, Reniers G. Learning from language problem-related accident information in the process industry: A literature study. *Process Safety and Environmental Protection*. ๒๐๑๙;๑๓๐:๑๔๐-๐๒. <https://doi.org/10.1๐๑๖/j.psep.๒๐๑๙.๐๖.๐๑๖>
๕. Chowdhury S, Ghosh A, Acharya S, Saha S, Pal PK, Roy S, et al. Transforming chemical process engineering: The role of AI and machine learning in revolutionizing process systems. *The Canadian Journal of Chemical Engineering*. ๒๐๒๐. <https://doi.org/10.1๐๐๒/cjce.๒๐๒๐.๐๐๐๐>
๖. Soori M, Arezoo B, Dastres R. Internet of things for smart factories in Industry ๔.๐: A review. *Internet of Things and Cyber-Physical Systems*. ๒๐๒๓;๓:๑๙๒-๒๐๔. <https://doi.org/10.1๐๑๖/j.iotcps.๒๐๒๓.๐๔.๐๐๖>
๗. Wang B, Wang J. Energy futures prices forecasting by novel DPFWR neural network and DS-CID evaluation. *Neurocomputing*. ๒๐๑๙;๓๓๘:๑-๑๐. <https://doi.org/10.1๐๑๖/j.neucom.๒๐๑๙.๐๑.๐๑๖>
๘. Peres RS, Jia X, Lee J, Sun K, Colombo AW, Barata J. Industrial artificial intelligence in Industry ๔.๐— Systematic review, challenges and outlook. *IEEE Access*. ๒๐๒๐;๘:๒๒๐๑๒๑-๓๙. <https://doi.org/10.1๑๐๑/ACCESS.๒๐๒๐.๓.๔๒๘๖๔>
๙. Benotsmane R, Kovács G, Dudás L. Economic and social impacts of smart factories in Industry ๔.๐: Simulation and artificial intelligence of collaborating robots. *Social Sciences*. ๒๐๑๙;๘(๐):๑๔๓. <https://doi.org/10.๓๓๙๐/socsci8.๐๐.๑๔๓>

10. Singh D. Harnessing artificial intelligence to safeguard food quality and safety. *Journal of Food Protection*. 2020;88(1):100-621. <https://doi.org/10.1177/j.fjp.2020.100.621>
11. Park J, Kang D. Artificial intelligence and smart technologies in safety management: A comprehensive analysis across multiple industries. *Applied Sciences*. 2024;14(2):11934. <https://doi.org/10.3390/app14211934>
12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. <https://doi.org/10.1136/bmj.n71>
13. Ayomoh M, Ongwae B. A systematic review of asset integrity and process safety management sustainability for onshore petrochemical installations. *Sustainability*. 2020;12(1):286. <https://doi.org/10.3390/su12010286>
14. Larbi JA, Tang L, Larbi RA, Abankwa DA, Danquah RD. Developing an integrated digital delivery framework and workflow guideline for construction safety management in a project delivery system. *Safety Science*. 2024;170:106486. <https://doi.org/10.1016/j.ssci.2024.106486>
15. Krainiuk O, Buts Y, Barbashyn V, Yatsiuk M. Use of artificial intelligence for work safety management. *Municipal Economy of Cities*. 2023;1(18):207-13. <https://doi.org/10.33044/mec-18-207-13>
16. Li X, Zhu M, Zhang B, Wang X, Liu ZA, Han L. A review of artificial intelligence applications in high-speed railway systems. *High-Speed Railway*. 2024;2(1):11-6. <https://doi.org/10.1016/j.hspr.2024.01.002>
17. Ihejirika CJ. Explainable artificial intelligence in deep learning models for transparent decision-making in high-stakes applications. *International Journal of Research Publication and Reviews*. 2020;7(2):1944-50. <https://doi.org/10.50248/gengpi.7.220.900>
18. Nguyen QT, Truong T, Nguyen TP, Tran T, Mašek J, Pag M. Explainable artificial intelligence in additive manufacturing: A systematic review on method convergence and assessment of standardization gaps. *MM Science Journal*. 2020;20(October). https://doi.org/10.17973/mmsj.2020_10_2020.90
19. Kostopoulos G, Davrazos G, Kotsiantis S. Explainable artificial intelligence-based decision support systems: A recent review. *Electronics*. 2024;13(14):2842. <https://doi.org/10.3390/electronics13142842>
20. Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, et al. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*. 2023;56(9):1-33. <https://doi.org/10.1145/3561048>
21. Gill SS, Xu M, Ottaviani C, Patros P, Bahsoon R, Shaghaghi A, et al. AI for next generation computing: Emerging trends and future directions. *Internet of Things*. 2022;19:100514. <https://doi.org/10.1016/j.iot.2022.100514>
22. Tsay C. Sobolev trained neural network surrogate models for optimization. *Computers & Chemical Engineering*. 2021;153:107419. <https://doi.org/10.1016/j.compchemeng.2021.107419>
23. Elele TY, Nwulu EO, Erhueh OV, Akano OA, Aderamo AT. Exploring advanced techniques in process automation and control: A generic framework for oil and gas industry applications. *Engineering Science & Technology Journal*. 2024;5(1). <https://doi.org/10.51094/estj.v5i1.1704>
24. Dinter RV, Tekinerdogan B, Catal C. Predictive maintenance using digital twins: A systematic literature review. *Information and Software Technology*. 2022;151:107008. <https://doi.org/10.1016/j.infsof.2022.107008>
25. Rojas L, Peña Á, García J. AI-driven predictive maintenance in mining: A systematic literature review on fault detection, digital twins, and intelligent asset management. *Applied Sciences*. 2020;10(6):3337. <https://doi.org/10.3390/app10063337>
26. Paranchuk Y, Jancarczyk D, Fałat P. Study and analysis of dynamics and energy efficiency of arc steelmaking furnace electrical mode with a fuzzy control algorithm. *Energies*. 2023;16(8):3401. <https://doi.org/10.3390/en16083401>
27. Werbińska-Wojciechowska S, Giel R, Winiarska K. Digital twin approach for operation and maintenance of transportation system—Systematic review. *Sensors*. 2024;24(18):6069. <https://doi.org/10.3390/s24186069>

٢٨. Faria RDR, Capron BDO, Secchi AR, Souza MBD. Where reinforcement learning meets process control: Review and guidelines. *Processes*. ٢٠٢٢;١٠(١١):٢٣١١. <https://doi.org/10.3390/pr10112311>
٢٩. Arunthavanathan R, Sajid Z, Amin MT, Tian Y, Khan F, Pistikopoulos E. Process safety ϵ : Artificial intelligence or intelligence augmentation for safer process operation? *AIChE Journal*. ٢٠٢٤;٧٠(٧):e١٨٤٧٥. <https://doi.org/10.1002/aic.18475>
٣٠. Wahab NHA, Hasikin K, Lai KW, Xia K, Bei L, Huang K, et al. Systematic review of predictive maintenance and digital twin technologies: Challenges, opportunities, and best practices. *PeerJ Computer Science*. ٢٠٢٤;١٠:e١٩٤٣. <https://doi.org/10.7717/peerj-cs.1943>
٣١. Dong Z, Xia Z, Zhu Y, Duan J. Overview of predictive maintenance based on digital twin technology. *Heliyon*. ٢٠٢٣;٩(٤):e١٤٥٣٤. <https://doi.org/10.1016/j.heliyon.2023.e14534>
٣٢. Moiseev V, Manoilo E, Zhuha O, Davydov D, Moskal M. Industry ϵ technologies in the design of reliable systems in chemical engineering. *Bulletin of the National Technical University "KhPI." Series: Chemistry, Chemical Technology and Ecology*. ٢٠٢٥;١(٢). <https://doi.org/10.26907/2079-0821.2025.01.02>
٣٣. Wang B, Wang Y, Xu F, Shi Z. Intelligence-led accident prevention and its application in petrochemical enterprises. *Process Safety and Environmental Protection*. ٢٠٢٤;١٨٤:٦٩٠-٧٠٢. <https://doi.org/10.1016/j.psep.2024.02.022>
٣٤. Dontha JD. AI and machine learning applications in industrial automation and cybersecurity. *Journal of Artificial Intelligence in Control and Computing*. ٢٠٢٣;٢(١):٢٥٨. [https://doi.org/10.47373/jaicc/2023\(2\)e208e208](https://doi.org/10.47373/jaicc/2023(2)e208e208)
٣٥. Souza KSD, Santos ESD, Alencar D, Nascimento M, Santos ADJ. Application of automation and computer vision in reducing failures in the production process of safety belts. *IOSR Journal of Business and Management*. ٢٠٢٤;٢٦(١٢):٥٥-٦٢. <https://doi.org/10.9790/2474-261213062>
٣٦. Cancemi SA, Frano RL, Santus C, Inoue T. Unsupervised anomaly detection in pressurized water reactor digital twins using autoencoder neural networks. *Nuclear Engineering and Design*. ٢٠٢٣;٤١٣:١١٢٥٠٢. <https://doi.org/10.1016/j.nucengdes.2023.112502>
٣٧. Pazukha AA. Artificial intelligence for safe maintenance, operation and repair technologies for power supply devices of JSC Russian Railways. *Occupational Safety in Industry*. ٢٠٢١;٦:٤٦-٥١. <https://doi.org/10.24002/2474-2612.21-6-46-51>
٣٨. Ahmed MF, Islam MR, Khan MRAA, Islam MN. Smart IoT and AI for sustainable safety management: Enhancing efficiency in industrial and urban environments. *Journal of Artificial Intelligence and Green Systems*. ٢٠٢٤;٦(١):٣٥٦. <https://doi.org/10.30887/jaigs.v6i1.356>
٣٩. Davila-Gonzalez S, Martín S. Human digital twin in Industry ٥.٠: A holistic approach to worker safety and well-being through advanced AI and emotional analytics. *Sensors*. ٢٠٢٤;٢٤(٢):٦٥٥. <https://doi.org/10.3390/s24020655>
٤٠. Mokhtarnane R, Safavi AA, Urbas L, Salimi F, Zerafat MM, Harasi N. Toward HAZOP ϵ approach for managing the complexities of the hazard and operability of an industrial polymerization reactor. *IFAC-PapersOnLine*. ٢٠٢٠;٥٣(٢):١١٦٦٥-٧٠. <https://doi.org/10.1016/j.ifacol.2020.12.802>
٤١. Dasgupta A, Islam MM, Nahid OF, Rahmatullah R. Engineering management perspectives on safety culture in chemical and petrochemical plants: A systematic review. *Asian Journal of Industrial Engineering and Environmental Technology*. ٢٠٢٤;١(١):١٢١. <https://doi.org/10.69093/ajiet.v1i1.121>
٤٢. Merchán-Cruz EA, Moveh S, Pasha O, Tocolovskis R, Grakovski A, Krainyukov A, et al. Smart safety helmets with integrated vision systems for industrial infrastructure inspection: A comprehensive review of VSLAM-enabled technologies. *Sensors*. ٢٠٢٥;٢٥(١٥):٤٨٣٤. <https://doi.org/10.3390/s25010483>
٤٣. Magdin M. Literature review: Current trends and advances in the use of artificial intelligence for ensuring the safety and efficiency of gas pipeline operations. *Results in Engineering*. ٢٠٢٥;٢٨:١٠٧٣٠٩. <https://doi.org/10.1016/j.rineng.2025.107309>
٤٤. Nar ZM, Nuhu AA, Zeeshan Q, Korhan O, Asmael M, Safaei B. Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry ϵ . *Sustainability*. ٢٠٢٠;١٢(١٩):٨٢١١. <https://doi.org/10.3390/su12198211>

٤٥. Adikwu FE, Esiri AE, Aderamo AT, Akano OA, Erhuch OV. Leveraging digital technologies for health, safety, and environmental (HSE) management in industrial operations. *International Journal of Applied Research in Social Sciences*. ٢٠٢٤;٦(١):١٧٢٢. <https://doi.org/10.51594/ijarss.v6i1.1722>
٤٦. Vyhmeister E, Casta GG. TAI-PRM: Trustworthy AI project risk management framework towards Industry ٥.٠. *AI & Ethics*. ٢٠٢٤;٤(١). <https://doi.org/10.1007/s43368-023-00417-y>
٤٧. Douthwaite JA, Lesage B, Gleirscher M, Călinescu R, Aitken JM, Alexander R, et al. A modular digital twinning framework for safety assurance of collaborative robotics. *Frontiers in Robotics and AI*. ٢٠٢١;٨:٧٥٨٠٩٩. <https://doi.org/10.3389/frobt.2021.758099>
٤٨. Pandey S, Singh AK, Parhi S. Toward sustainable process safety management ٤.٠ versus process safety management. *Process Safety Progress*. ٢٠٢٤. <https://doi.org/10.1002/prs.12664>
٤٩. Ullagaddi P. Leveraging digital transformation for enhanced risk mitigation and compliance in pharma manufacturing. *Journal of Advances in Medical and Pharmaceutical Sciences*. ٢٠٢٤;٦(٦):٦٩٧. <https://doi.org/10.9734/jamps/2024/v2i6j697>
٥٠. Niazi SK. Regulatory perspectives for AI/ML implementation in pharmaceutical GMP environments. *Pharmaceuticals*. ٢٠٢٥;١٨(٦):٩٠١. <https://doi.org/10.3390/ph18.6.901>
٥١. Abioye S, Oyedele LO, Akanbi L, Ajayi A, Delgado JMD, Bilal M, et al. Artificial intelligence in the construction industry: A review of present status, opportunities, and future challenges. *Journal of Building Engineering*. ٢٠٢١;٤٤:١٠٣٢٩٩. <https://doi.org/10.1016/j.jobbe.2021.103299>
٥٢. Santoso A, Arviansyah A. Strategic operations under uncertainty: Mitigating shutdown risks and operator error through AI-driven decision support. *Locus*. ٢٠٢٥;٤(١٠). <https://doi.org/10.58344/locus.v4i10.204>
٥٣. Guang W, Yuan Y, Huang S, Zhang F, Zhao J, Hu F. Application of large AI models in safety and emergency management of the power industry in China. *Processes*. ٢٠٢٥;١٣(٨):٢٥٦٩. <https://doi.org/10.3390/pr13.82069>
٥٤. Sira M. Navigating the intersection of management and AI: Advancing occupational health and safety. *System Safety: Human - Technical Facility - Environment*. ٢٠٢٤;٦(١):٩-١٥. <https://doi.org/10.2478/czoto-2024-002>
٥٥. Kumar N. Enhancing transparency and trust in cybersecurity: Developing explainable AI models for threat detection. *International Journal of Scientific Research in Engineering and Management*. ٢٠٢٥;٩(٦):١-٩. <https://doi.org/10.50541/ijssrem.24.6>
٥٦. Raj C, Divya P, Susmita A, Sushmitha P, Ch R, Chandini K. Automation in pharmacovigilance: Artificial intelligence and machine learning for patient safety. *Journal of Integrated Advanced Pharmaceutical Sciences*. ٢٠٢٢;٧(٣):١٣-٢٠. <https://doi.org/10.37022/jiaps.v7i3.374>
٥٧. Hu X, Liu A, Li X, Dai Y, Nakao M. Explainable AI for customer segmentation in product development. *CIRP Annals*. ٢٠٢٢;٧٢(١):٨٩-٩٢. <https://doi.org/10.1016/j.cirp.2023.03.004>
٥٨. Pawlicki M, Pawlicka A, Kozik R, Choraś M. The survey on the dual nature of XAI challenges in intrusion detection and their potential for AI innovation. *Artificial Intelligence Review*. ٢٠٢٤;٥٧(١٢). <https://doi.org/10.1007/s10462-024-10972-3>
٥٩. Jiang J, Kahai SS, Yang M. Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies*. ٢٠٢٢;١٦٥:١٠٢٨٣٩. <https://doi.org/10.1016/j.ijhcs.2022.102839>
٦٠. Bozorgpanah A, Torra V. Explainable machine learning models with privacy. *Progress in Artificial Intelligence*. ٢٠٢٤;١٣(١):٣١-٥٠. <https://doi.org/10.1007/s133748-024-00310-2>
٦١. Nandipati M, Fatoki O, Desai S. Bridging nanomanufacturing and artificial intelligence—A comprehensive review. *Materials*. ٢٠٢٤;١٧(٧):١٦٢١. <https://doi.org/10.3390/ma17.071621>
٦٢. Marasinghe R, Yigitcanlar T, Mayere S, Washington TL, Limb M. Towards responsible urban geospatial AI: Insights from the white and grey literatures. *Journal of Geovisualization and Spatial Analysis*. ٢٠٢٤;٨(٢). <https://doi.org/10.1007/s41761-024-00184-2>

٦٣. Sharma S, Mishra S, Jaiswal MK, Singh V. AI-powered cybersecurity: The future of threat detection. International Journal of Scientific Research in Engineering and Management. ٢٠٢٠;٩(٤):١-٩. <https://doi.org/10.000٤١/IJSREM٤٠٩٤٣>
٦٤. Dagi TF, Barker FG, Glass J. Machine learning and artificial intelligence in neurosurgery: Status, prospects, and challenges. Neurosurgery. ٢٠٢١;٨٩(١):١٣٣-٤٢. <https://doi.org/10.1093/neuros/nyab١٧٠>
٦٥. Challen R, Denny JC, Pitt M, Gompels L, Edwards TL, Tsaneva-Atanasova K. Artificial intelligence, bias, and clinical safety. BMJ Quality & Safety. ٢٠١٩;٢٨(٣):٢٣١-٧. <https://doi.org/10.1136/bmjqs-2018-008٣٧٠>
٦٦. Salih A, Galazzo IB, Gkontra P, Rauseo E, Lee AM, Lekadir K, et al. A review of evaluation approaches for explainable AI with applications in cardiology. Artificial Intelligence Review. ٢٠٢٤;٥٧(٩). [https://doi.org/10.1007/s10446-02٤-١٠٨٥٢-w](https://doi.org/10.1007/s104٤٦-02٤-١٠٨٥٢-w)
٦٧. Thalpage N. Explainable AI for cybersecurity applications: A review article on techniques, deployments, and usability challenges. Journal of Data Science. ٢٠٢٥;٧(١):٢-١٠. [https://doi.org/10.3384٧/٢٦٨٦-٨٢٩٦,٧,١ ١](https://doi.org/10.33٨٤٧/٢٦٨٦-٨٢٩٦,٧,١ ١)
٦٨. Izumo T, Weng Y. Coarse ethics: How to ethically assess explainable artificial intelligence. AI and Ethics. ٢٠٢١;٢(٣):٤٤٩-٦١. <https://doi.org/10.1007/s٤٣٦٨١-02١-00٩١-y>
٦٩. Kaur D, Uslu S, Rittichier KJ, Durresi A. Trustworthy artificial intelligence: A review. ACM Computing Surveys. ٢٠٢٢;٥٥(٢):١-٣٨. [https://doi.org/10.114٥/3491209](https://doi.org/10.11٤٥/٣٤٩١٢٠٩)
٧٠. Sava S. Artificial intelligence in construction project management: Trends, challenges and future directions. Journal of Design for Resilience in Architecture and Planning. ٢٠٢٥;٦(٢):٣٦٥-٧٩. [https://doi.org/10.47818/drarch.202٥.v٦i216٥](https://doi.org/10.٤٧٨١٨/drarch.202٥.v٦i2١٦٥)
٧١. Sauer CR, Burggräf P. Hybrid intelligence: Systematic approach and framework to determine the level of human-AI collaboration for production management use cases. Production Engineering. ٢٠٢٤;١٨(٤):٣٤٥-٥٩. [https://doi.org/10.1007/s11740-02٤-1326-7](https://doi.org/10.1007/s11٧٤٠-02٤-١٣٢٦-٧)