

An Artificial Intelligence Based Predictive Safety Management Framework for High Risk Industries in Saudi Arabia

ABDUL RAUF

Abstract: *High-risk industries in Saudi Arabia are expanding while major construction, mining, energy, logistics and industrial programmes increase the scale and complexity of occupational risk. Conventional safety management remains essential, yet lagging indicators, periodic inspections and rule-based controls often identify deterioration only after exposure has accumulated. This structured review develops an artificial intelligence-based predictive safety management framework suited to Saudi operating conditions. It integrates leading indicators from permits, inspections, sensors, equipment systems, workforce records, environmental monitoring and incident narratives; applies transparent risk models; and routes predictions through human review and proportionate controls. The synthesis shows that artificial intelligence can improve prioritisation, early-warning capability and learning across dispersed sites, but only when data quality, privacy, workforce participation, model drift, cybersecurity and accountability are governed explicitly. The proposed framework therefore treats artificial intelligence as a decision-support layer within an ISO 45001-aligned management system, not as an autonomous replacement for competent safety judgement. Its principal contribution is a risk-based operating model linking data governance, prediction, verification, intervention and continuous assurance to Saudi Vision 2030 objectives.*

Keywords: *artificial intelligence; predictive safety; occupational health and safety; high-risk industries; Saudi Arabia; risk governance; leading indicators; Industry 4.0.*

I. INTRODUCTION

Saudi Arabia is undertaking simultaneous expansion in infrastructure, energy, mining, manufacturing, logistics, tourism and large-scale urban development. The resulting project portfolios involve contractors, mobile workforces, complex interfaces, harsh climatic conditions and safety-critical assets. These conditions create strong demand for safety systems that identify weak signals before they develop into injuries, process events or operational disruption. The national

emphasis on digital transformation and human-capital protection provides a strategic setting for predictive safety, while the moral and legal obligation to prevent harm remains unchanged [1-4].

Traditional safety programmes depend on hazard identification, risk assessment, hierarchy-of-control decisions, inspections, audits, permits, training and incident investigation. These controls are indispensable, but their information is frequently fragmented among spreadsheets, enterprise systems, contractor portals and paper records. Lagging measures such as recordable injury rates describe outcomes after the event and can be statistically unstable when severe events are rare. Leading indicators provide earlier insight, yet organisations often count activities rather than test whether those activities reduce exposure. Artificial intelligence can connect heterogeneous records, detect non-linear patterns and estimate changing risk, but its output is useful only when managers understand what the prediction means and act within a controlled process [5-8].

This paper reviews relevant evidence and proposes a Saudi-specific framework. It uses the term artificial intelligence broadly for machine learning, natural-language processing, computer vision and analytical methods that support classification, anomaly detection, forecasting and prioritisation. The framework does not assume that every site requires sophisticated deep learning. In many settings, interpretable statistical models, carefully designed thresholds and disciplined data governance may create more reliable value. Ultimately,

II. AIM OBJECTIVES AND RESEARCH QUESTIONS

The aim is to develop an artificial intelligence-based predictive safety management framework for high-risk industries in Saudi Arabia that strengthens prevention while preserving human accountability.

The objectives are to identify predictive safety data and leading indicators; compare appropriate analytical methods; define a risk-based architecture for multi-site operations; examine ethical, legal and organisational controls; and propose implementation and assurance measures aligned with ISO 45001 and Saudi digital-transformation priorities.

The review addresses four questions. Which combinations of operational, behavioural, environmental and asset data provide credible early warning? How should predictive models be validated when severe incidents are rare and costs are asymmetric? How can predictions trigger timely controls without encouraging surveillance, automation bias or punitive use? What governance arrangements are required for scalable deployment across Saudi high-risk industries?

III. REVIEW METHODOLOGY

A structured integrative review was used because the subject combines occupational safety, engineering, data science, human factors and governance. Searches were conceptually organised across Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, Google Scholar and official repositories. Search strings combined terms including artificial intelligence, machine learning, predictive safety, occupational risk, construction, mining, oil and gas, computer vision, wearables, leading indicators, explainability, Saudi Arabia and safety management.

Priority was given to peer-reviewed and institutional material published from 2020 to 2025, while earlier seminal studies and standards were retained where they defined methods or management requirements. Eligible sources addressed prediction of incidents or unsafe conditions, extraction of safety knowledge, sensing and computer vision, trustworthy artificial intelligence, occupational safety governance or the

Saudi transformation context. Sources making unsupported performance claims, lacking methodological transparency or addressing generic prediction without a safety connection were excluded. The evidence was synthesised thematically rather than statistically because datasets, industries, outcomes and evaluation metrics were heterogeneous. Each source was considered for relevance, methodological clarity, operational realism and transferability. No fabricated screening counts are reported. The resulting synthesis focuses on five linked domains: information foundations, prediction, decision control, intervention and assurance. This method supports framework development but cannot establish a pooled causal effect.

IV. RISK CONTEXT IN SAUDI HIGH RISK INDUSTRIES

The Saudi risk landscape is diverse. Construction combines changing work fronts, subcontractor interfaces, lifting, work at height, temporary works and vehicle movement. Mining adds geotechnical, mobile-equipment, blasting and remote-area risks. Oil, gas and petrochemical operations combine process hazards, maintenance, isolation, ignition control and contractor work. Logistics and industrial manufacturing introduce traffic, machinery, ergonomics and production-pressure interactions. Heat exposure and dust can amplify risk across sectors. A predictive framework must therefore use a common governance core while allowing sector-specific hazard models.

Major programmes also create scale effects. A weak signal repeated across hundreds of sites may be invisible to local teams but detectable at portfolio level. Conversely, a central model may misinterpret local work practices, language, equipment or reporting culture. Saudi deployment should therefore combine enterprise learning with site validation. Arabic and English terminology, contractor data maturity, regional climate and differing technology environments must be represented. A model trained only on one contractor or mature facility should not be presumed portable.

Safety performance is shaped by organisational conditions as much as immediate acts. Schedule

pressure, supervision span, competence gaps, maintenance backlog, fatigue, permit quality and repeated corrective-action delays can interact. Predictive analysis should therefore avoid reducing risk to worker behaviour. The more defensible target is the configuration of work: hazards, controls, resources, environment and organisational signals. This approach is compatible with prevention principles and reduces incentives to use artificial intelligence as a disciplinary surveillance mechanism.

V. DATA FOUNDATION AND LEADING INDICATORS

A predictive system begins with a defined risk question, not with available data. Relevant sources include hazard registers, permits to work, inspection findings, audit nonconformities, corrective-action ageing, training and competence records, shift patterns, maintenance work orders, alarms, process deviations, telematics, environmental measurements, near-miss reports and incident narratives. Each field requires an owner, definition, timestamp, provenance and quality rule. Without this discipline, a sophisticated model merely automates inconsistent reporting.

Leading indicators should represent control health. Examples include overdue safety-critical maintenance, repeated permit deviations, unresolved high-potential findings, supervisor-to-worker ratios, abnormal equipment states, heat-stress exposure,

fatigue proxies and changes in near-miss reporting. Counts alone can mislead: a rise in reported near misses may indicate deteriorating conditions or improved reporting culture. Models need context, denominators and temporal direction. They should distinguish exposure volume from control failure and should never treat silence as proof of safety.

Natural-language processing can structure free-text observations and investigation reports, identify recurring hazards and connect semantically similar events [9,10]. Computer vision can identify limited observable conditions such as missing personal protective equipment or entry into restricted zones [11,12]. Wearables and connected sensors can add physiological, location or environmental information, but proportionality is critical. The organisation should collect the minimum data needed for a defined prevention purpose, restrict access and specify retention. Worker consultation improves both legitimacy and interpretation.

Data quality must be monitored like a safety-critical control. Completeness, timeliness, label consistency, class balance and sensor reliability require dashboards and escalation. Missingness is often informative: a site that stops submitting inspections may have a system problem. Yet missing data can also reflect connectivity or workload. The framework therefore records missingness explicitly and requires operational confirmation before intervention.

Table 1. Predictive information domains and preventive use

Domain	Representative signals	Preventive decision
Work planning	Permits, simultaneous operations, lift complexity	Verify conflicts and controls
Assets	Maintenance backlog, alarms, safeguard status	Prioritise inspection and restoration
Workforce systems	Competence, supervision, shift patterns	Adjust resources and authorisation
Environment	Heat, dust, wind, visibility	Modify schedule and protective controls

VI. PREDICTIVE METHODS AND MODEL SELECTION

Model choice should follow the decision. Logistic regression and decision trees provide interpretable

baselines for event likelihood and risk categories. Random forests and gradient-boosting methods can capture interactions in mixed tabular data. Sequence models can analyse temporal patterns from sensors or

work orders, while anomaly detection can identify unusual combinations when labelled incidents are scarce. Natural-language models support classification and thematic extraction. Computer-vision models support bounded visual tasks. Hybrid systems combine engineering rules with learned estimates and are often preferable in safety-critical settings [13-17].

Rare severe events create a central challenge. Overall accuracy can appear excellent when a model predicts the majority safe class. Evaluation should therefore include precision, recall, specificity, area under the precision-recall curve, calibration and cost-weighted error. False negatives can miss emerging catastrophe, while excessive false positives create alarm fatigue and interrupt legitimate work. Thresholds should vary with consequence, reversibility and response burden. A process-safety warning may justify a lower threshold than a low-consequence observation, provided review capacity exists.

Validation must separate training from future or independent sites to avoid leakage. Temporal validation tests performance after operational change; site-based validation tests transfer across locations; subgroup analysis tests whether errors concentrate among contractors, occupations or shifts. Calibration matters because managers use probabilities to allocate scarce attention. A model that ranks cases well but overstates absolute risk may distort intervention. Confidence, data freshness and applicability boundaries should accompany each score.

Explainability must be designed for the user. Senior managers need portfolio drivers and uncertainty; safety specialists need contributing factors and comparable cases; supervisors need a short statement of the control concern and required verification. Feature importance is not causal proof. Explanations should be treated as prompts for investigation and never as automatic blame attribution. Where a model uses proxies for nationality, health or employment status, governance review should test necessity and discriminatory impact [18-20].

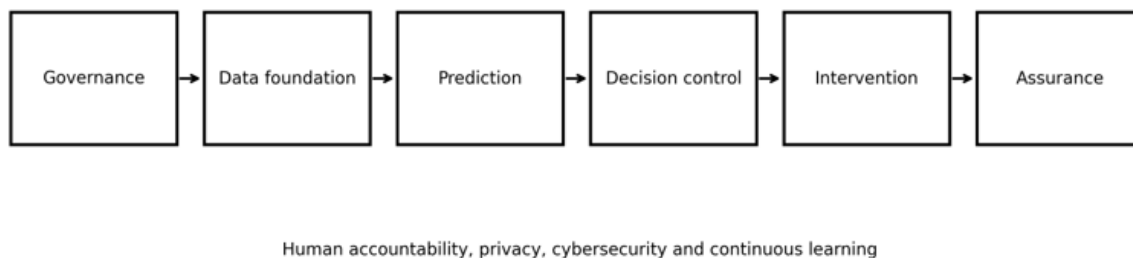


Figure 1. Six-layer predictive safety management framework

VII. PROPOSED PREDICTIVE SAFETY MANAGEMENT FRAMEWORK

The proposed framework has six layers. The first is governance: accountable executives define prevention purposes, prohibited uses, risk appetite and oversight. The second is the information foundation: controlled data pipelines align operational, environmental, asset and workforce records. The third is analytical processing: validated models generate risk estimates, anomalies and explanations. The fourth is decision orchestration: rules combine consequence,

probability, confidence and control status. The fifth is intervention: competent persons verify the signal and implement hierarchy-of-control actions. The sixth is assurance: performance, drift, bias, cybersecurity and learning are reviewed.

Predictions enter a tiered workflow. Low-risk signals enrich local planning. Moderate signals require supervisor verification within a defined time. High-risk signals prompt safety-specialist review and may restrict or pause work when existing authority already permits that action. Critical signals involving credible

imminent danger follow emergency arrangements. Artificial intelligence does not create new disciplinary authority and cannot independently declare a person unsafe. Every intervention records the evidence reviewed, action taken and outcome, creating feedback without assuming the original prediction was correct.

The unit of analysis should match the hazard. Person-level prediction is rarely necessary and creates privacy and fairness concerns. Task, work-front, asset, location or control-level predictions often provide actionable prevention with less intrusion. For example, a lifting-risk model can combine lift complexity, equipment condition, wind, crew competence and simultaneous operations. A heat-risk model can combine temperature, workload,

acclimatisation controls, work-rest scheduling and hydration arrangements. A permit-risk model can identify unusual isolation patterns or conflicting activities.

Integration with ISO 45001 embeds the system within planning, support, operation, performance evaluation and improvement [3]. Predictive outputs inform risk reviews and operational planning; they do not replace legal compliance, hazard assessment, worker consultation or emergency preparedness. Management of change applies when data, algorithms, thresholds or connected equipment change. Internal audit samples both technical performance and whether users respond appropriately.

Table 2. Risk tiers and required response

Tier	Interpretation	Required response
Low	Planning signal; no immediate control failure	Include in routine review
Moderate	Possible deterioration or unusual pattern	Supervisor verifies within defined time
High	Credible control weakness with material consequence	Safety specialist reviews and restores control
Critical	Credible imminent danger or severe barrier failure	Apply emergency or stop-work authority

VIII. HUMAN FACTORS ETHICS PRIVACY AND CYBERSECURITY

Trust is an operational requirement. Workers and contractors should know what information is collected, why it is used, how long it is retained and how a prediction can be challenged. Participation should occur before deployment and during review. If reporting data are used punitively, under-reporting will increase and model quality will deteriorate. A just-culture policy should separate good-faith reporting and human error from reckless or intentional violations.

Automation bias can cause users to accept incorrect alerts, while algorithm aversion can cause them to ignore useful warnings after visible errors. Training must therefore cover limitations, confidence and verification. Interfaces should show the reason for

escalation and the required check, not a mysterious score. Workload matters: a model that produces more alerts than teams can investigate is unsafe by design. Alert budgets, prioritisation and closure quality should be monitored.

Saudi privacy requirements and organisational policies govern personal data. The framework uses purpose limitation, data minimisation, role-based access, encryption, audit logs and controlled retention. Health and biometric data require heightened protection. Cross-border processing and vendor access require legal and security review. De-identification reduces risk but may not prevent re-identification when location and shift data are granular [21-23].

Cybersecurity is inseparable from predictive safety. Compromised sensors, manipulated labels or unavailable dashboards can produce unsafe decisions.

Threat modelling should cover data poisoning, evasion, credential compromise, ransomware, insecure interfaces and vendor dependencies. Safety-critical operations need fail-safe procedures for model or network outage. Model versions, code, datasets and thresholds must be change-controlled, tested and recoverable.

IX. IMPLEMENTATION ROADMAP AND ASSURANCE

Implementation should begin with one bounded, high-value use case where data and response processes exist. A baseline should document current performance, reporting behaviour and intervention cost. A multidisciplinary team then defines the decision, outcome, acceptable errors and escalation. A shadow phase runs predictions without operational consequence, allowing comparison with expert review. Only after validation should controlled live use begin.

Scaling requires a product operating model rather than a one-time project. Safety, operations, data, technology, human resources, legal, privacy and cybersecurity functions share defined responsibilities. Site champions support interpretation, while a central

model-risk group approves material changes. Vendors should disclose training provenance, performance limits, security controls and update practices. Contracts should preserve access to audit evidence and prevent unapproved secondary use of workforce data. Assurance covers technical and safety outcomes. Technical measures include discrimination, calibration, latency, missingness, drift and subgroup error. Operational measures include alert acceptance, verification time, action completion, recurrence and user workload. Safety measures examine control effectiveness, high-potential exposure and event severity, not merely injury counts. A successful pilot should demonstrate that decisions improve; high model accuracy alone is insufficient.

Independent review is appropriate before high-impact deployment and after material incidents or changes. Reviewers should be able to reproduce performance from controlled data, examine false negatives and false positives, and trace alerts to decisions. The organisation should retain an incident process for harmful or systematically misleading model behaviour. Periodic retirement review prevents obsolete models from remaining active because they are embedded in workflow.

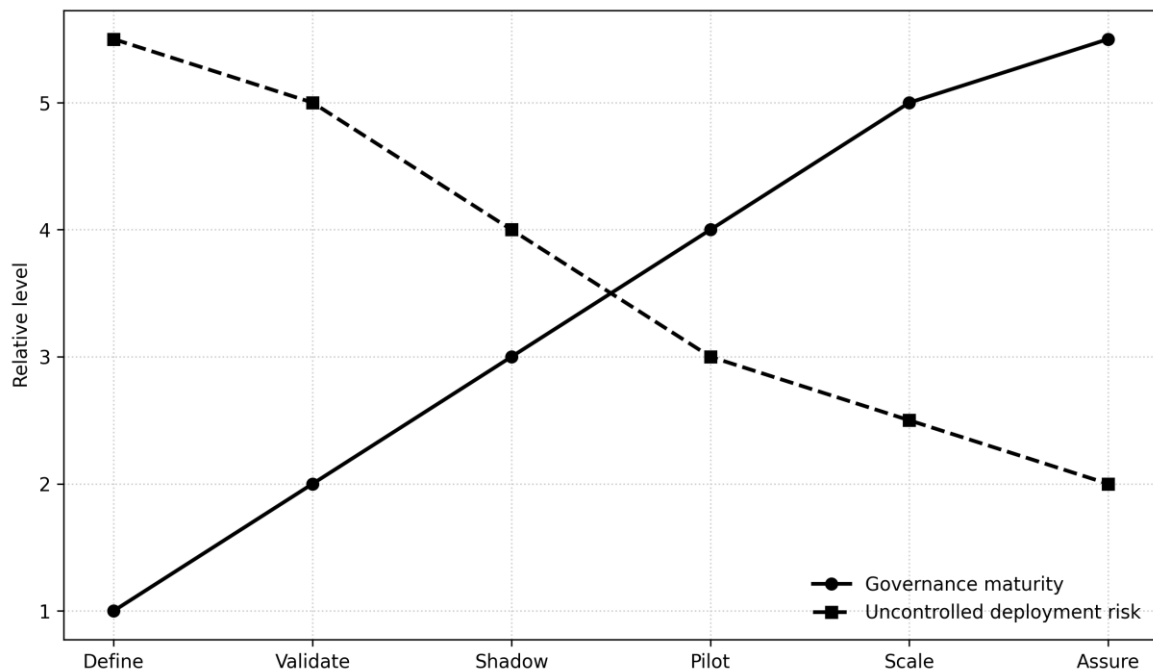


Figure 2. Controlled implementation and risk reduction pathway

X. DISCUSSION

The evidence supports a cautious but positive conclusion. Artificial intelligence can convert fragmented safety information into earlier, prioritised signals, especially where operations are distributed and hazards interact. However, the technology does not discover risk independently of definitions, labels and organisational practice. Predictions inherit the strengths and weaknesses of reporting culture. High-quality safety management is therefore a precondition for reliable predictive safety, not an outcome that software can supply by itself.

The framework differs from technology-centred approaches by placing intervention and assurance on equal footing with modelling. This is important in Saudi megaproject and multi-contractor environments, where an accurate central prediction may still fail if responsibility is unclear or if local teams lack authority and resources. The framework also favours task and control-level prediction over person-level scoring. That choice strengthens actionability while reducing surveillance and fairness risks.

Saudi Arabia has an opportunity to build interoperable safety learning across sectors, but standardisation should focus first on definitions and governance rather than a single national algorithm. Common taxonomies for hazards, controls, event potential and corrective actions could enable benchmarking without exposing unnecessary personal or commercially sensitive data. Sector regulators and professional bodies can encourage validation protocols, competence requirements and reporting of algorithm-related failures.

The main practical tension concerns sensitivity. Lower thresholds may detect weaker signals but create false alarms and operational burden. Higher thresholds reduce burden but may miss precursors. The defensible solution is consequence-based thresholding, calibrated uncertainty, human verification and continuous monitoring. No single metric resolves this trade-off. Governance must make it explicit.

Safety case design. A predictive use case should be

documented as a safety case that states the hazard, decision, intended user, response time and evidence required. The case should explain why prediction adds value beyond existing controls, how uncertainty is displayed, and which conditions make the output invalid. Linking each feature to a plausible control pathway prevents teams from collecting data merely because it is available. The safety case also identifies foreseeable misuse, including punitive worker scoring, unsupported causal interpretation and automatic shutdown without verification. Approval should depend on a proportionate demonstration that the combined technical and organisational system reduces risk.

Contractor integration. Contractors often use different taxonomies, systems and reporting thresholds. A common minimum dataset should therefore define exposure hours, work type, hazard category, control status, event potential and action closure. Local detail can remain in contractor systems while mapped fields support portfolio learning. Data-quality feedback should be supportive rather than punitive because sanctions encourage concealment. Mobilisation requirements should cover consent, access, competence, device management and incident escalation. Joint governance forums can review model errors and emerging patterns without shifting the principal organisation responsibility for coordinated work.

Heat and environmental risk. Environmental prediction should combine measured temperature, humidity, radiant heat and air movement with workload, clothing, acclimatisation, hydration and work-rest arrangements. A weather forecast alone cannot represent personal exposure. The model should produce a control recommendation tied to approved heat-stress procedures, and supervisors should confirm local conditions. Sensor failure, shaded placement or connectivity loss must be visible. Aggregated learning can identify work packages and time windows where controls repeatedly become marginal, supporting rescheduling and engineering solutions instead of relying solely on individual tolerance.

Process safety application. Process safety requires attention to low-frequency, high-consequence events.

Useful predictors include safety-critical maintenance deferral, alarm floods, bypassed safeguards, operating-envelope excursions, overdue proof tests and repeated permit conflicts. Because historical disasters are scarce, engineering rules and barrier models should constrain learning algorithms. The output should indicate which barrier is weakening and what verification is required. A high score cannot substitute for hazard and operability studies, layers-of-protection analysis or competent technical authority. The primary benefit is prioritising attention across large asset portfolios.

Learning feedback. Every alert should create structured feedback: confirmed condition, false alarm, contributing factors, action and residual risk. Analysts must distinguish a false prediction from a successful early intervention that prevented the predicted outcome. Labels should be reviewed by more than one role for high-impact cases. Feedback latency must be measured because stale labels delay adaptation. When users repeatedly override a model, governance should investigate whether the model, interface, threshold, training or local process is deficient. Continuous learning should occur through controlled releases rather than unreviewed automatic retraining.

Performance governance. A balanced scorecard should combine predictive quality, operational response and prevention outcomes. Precision and recall describe classification, calibration describes probability reliability, and drift measures changing input patterns. Response indicators include acknowledgement time, verification quality and action completion. Prevention indicators examine exposure reduction and barrier restoration. Metrics should be stratified by site, contractor, task and shift when lawful and statistically meaningful. Small samples require cautious interpretation. Governance should prohibit teams from optimising a single metric in ways that increase hidden risk or reporting suppression.

Change management. Industrial operations change through new equipment, contractors, software, schedules and procedures. Each change may invalidate historical relationships. The management-of-change process should ask whether model inputs, thresholds, interfaces or response responsibilities are affected. Major model updates need regression testing and

renewed user acceptance. Emergency patches require documented retrospective review. Version control should allow reconstruction of the prediction visible at the time of a decision. Clear rollback arrangements are necessary when performance deteriorates or a cybersecurity event compromises confidence.

Competence and culture. Competence extends beyond data scientists. Safety professionals need enough statistical literacy to understand calibration and error; analysts need hazard knowledge; supervisors need verification skills; and executives need to interpret uncertainty. Scenario-based exercises can expose automation bias and test escalation. Worker representatives should participate in design and review. Communication must explain that predictions concern conditions and controls, not personal worth. Leadership behaviour remains decisive: if production routinely overrides verified warnings, better algorithms will not create a preventive culture.

Portfolio learning. Enterprise models can reveal repeated precursor combinations across sites, but centralisation can erase context. A federated learning structure may allow common features and evaluation while sites retain sensitive records. Portfolio dashboards should permit drill-down from aggregate patterns to source evidence. Benchmarking must adjust for exposure mix and reporting maturity. Sites with high observation rates should not be labelled worse without analysis. Cross-site learning is strongest when comparable definitions, peer review and shared corrective-action libraries accompany the model.

Emergency and resilience. Predictive systems must support, not complicate, emergency response. Critical alerts need clear ownership, redundant communication and predefined actions. During outages, teams revert to established manual controls rather than waiting for system recovery. Exercises should include incorrect alerts, missing data and cyber compromise. Business continuity plans should identify dependencies on cloud services, network connectivity and vendor support. Post-event review should examine whether predictions influenced decisions appropriately, whether warnings were comprehensible and whether the system introduced new failure modes.

Incident narrative intelligence. Incident and near-miss

narratives contain contextual information that coded fields often omit, including sequence, competing goals, communication breakdown and control adaptation. Natural-language processing can group similar cases and highlight emerging terminology, but domain dictionaries and bilingual validation are necessary. Negation, abbreviations and copied text can distort classification. Analysts should review representative extracts before accepting a theme. The purpose is to direct investigation toward recurring system conditions, not to replace formal investigation or infer individual intent from writing style.

Vehicle and mobile equipment risk. Mining, construction and logistics operations can combine telematics, proximity events, speed, route geometry, maintenance condition and shift information to identify locations and periods of elevated interaction risk. Scores should lead to traffic separation, route redesign, maintenance or supervision rather than continual driver ranking. Data should be normalised for distance, operating hours and task. Short-term deviations require local confirmation because detours, emergency work or sensor faults can resemble unsafe behaviour. Camera and location data need strict access and retention controls.

Permit to work intelligence. Permit systems offer structured evidence about energy isolation, confined spaces, hot work and simultaneous operations. Predictive checks can identify unusual combinations, duplicate equipment identifiers, missing dependencies and recurring late approvals. The model should support the issuer and performing authority by presenting the conflicting fields and applicable rule. It must not approve a permit or certify an isolation. Sampling should test whether digital consistency reflects field conditions, since complete electronic records cannot prove that barriers were correctly installed.

Maintenance prioritisation. Maintenance prediction should distinguish production reliability from safety criticality. An asset may continue operating while a protective function deteriorates. The framework therefore links equipment hierarchy, failure mode, barrier role, inspection findings and deferral justification. Models can rank work orders for engineering review, but authorised technical personnel

retain decisions about continued operation. Repeated deferrals, temporary repairs and parts shortages are organisational signals. Assurance should compare recommendations with actual barrier restoration and examine whether scarce resources are diverted from less visible critical equipment.

Multilingual deployment. Saudi workforces commonly use several languages, making terminology and communication important model inputs and design constraints. Translated taxonomies should be validated by practitioners rather than generated mechanically. Interfaces should use plain language and recognised hazard symbols, while explanations should be available in relevant languages. Speech or text analytics must account for code switching and local abbreviations. Language must not become a proxy for competence or nationality. Usability testing with contractors and frontline workers is essential before live deployment.

XI. RESEARCH GAPS AND FUTURE RESEARCH

Research should test models prospectively in Saudi operating environments rather than relying only on retrospective datasets. Priority areas include heat exposure, contractor interfaces, multilingual safety text, mobile-equipment interaction, maintenance backlog, simultaneous operations and process-safety leading indicators. Studies should report data provenance, prevalence, missingness, calibration, external validation and operational consequences.

Causal evaluation is needed to determine whether model-guided intervention reduces exposure rather than merely predicts it. Cluster trials, stepped-wedge implementation and interrupted time-series designs may be feasible. Research should also examine how supervisors interpret explanations, when workers trust monitoring and how reporting culture changes after deployment. Economic evaluation should include investigation workload, downtime, avoided loss and system-maintenance cost.

Privacy-preserving learning, federated approaches and synthetic data may support cross-company learning, but their safety limitations require testing. Future work should develop assurance cases linking model

evidence to specific hazards and controls. Arabic-language models need domain evaluation, not generic language benchmarks. Finally, research should investigate safe retirement, incident reporting for algorithms and regulator access to meaningful audit evidence.

XII. PRACTICAL MANAGERIAL AND POLICY IMPLICATIONS

Executives should sponsor predictive safety as an improvement programme with clear accountability, not as an information-technology experiment. Safety leaders should define use cases and response logic. Operations leaders must ensure alerts can be investigated and controls implemented. Data and technology teams are responsible for reliability and monitoring, while legal, privacy and cybersecurity specialists establish boundaries.

Managers should begin with control-centred questions such as where critical maintenance is deteriorating or which work fronts show interacting precursors. Procurement should reject opaque claims of universal accuracy. Contracts should require documented limitations, test access, change notification, security evidence and data-return arrangements. Competence programmes should teach users to verify, override and document decisions.

Policy makers can support common definitions, trustworthy-AI guidance, privacy-compliant data collaboration and sector-specific evaluation benchmarks. Regulation should preserve employer responsibility: use of an algorithm cannot transfer accountability away from the organisation controlling work. Incentives should reward validated prevention and transparent learning rather than technology adoption alone.

XIII. LIMITATIONS

This review integrates heterogeneous research and institutional guidance and does not provide a meta-analysis. The rapid development of artificial intelligence means tools and regulation continue to evolve. Public evidence from Saudi deployments remains limited, and findings from construction or manufacturing may not transfer directly to process

industries. The framework is conceptual and requires prospective testing. It does not replace legal advice, engineering assessment or workforce consultation.

XIV. CONCLUSION

Artificial intelligence can strengthen safety management in Saudi high-risk industries by identifying changing combinations of exposure, control weakness and operational pressure before harm occurs. Its value depends on disciplined information, appropriate models, calibrated thresholds, competent verification and timely intervention. The proposed six-layer framework integrates governance, data, analytics, decision orchestration, intervention and assurance within an ISO 45001-aligned system. It protects human accountability, favours task and control-level prediction, and treats privacy, fairness and cybersecurity as safety requirements. Organisations should begin with bounded use cases, validate prospectively and scale only when decisions and controls demonstrably improve. Used in this manner, predictive safety can support Saudi Vision 2030 while reinforcing the principle that technology assists prevention but does not replace responsibility.

REFERENCES

- [1] International Labour Organization. (2022). A safe and healthy working environment is a fundamental principle and right at work. ILO.
- [2] International Labour Organization. (2025). Revolutionizing health and safety: The role of AI and digitalization at work. ILO.
- [3] International Organization for Standardization. (2018). ISO 45001:2018 Occupational health and safety management systems - Requirements with guidance for use.
- [4] Kingdom of Saudi Arabia. (2016). Saudi Vision 2030.
- [5] World Health Organization & International Labour Organization. (2021). WHO/ILO joint estimates of the work-related burden of disease and injury, 2000-2016. WHO.
- [6] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1.

- [7] International Organization for Standardization. (2023). ISO/IEC 23894:2023 Information technology - Artificial intelligence - Guidance on risk management.
- [8] OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.
- [9] Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Application of machine learning to construction injury prediction. *Automation in Construction*, 69, 102-114.
- [10] Baker, H., Hallowell, M. R., & Tixier, A. J. P. (2020). AI-based prediction of independent construction safety outcomes from universal attributes. *Automation in Construction*, 118, 103146.
- [11] Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, 85, 1-9.
- [12] Nath, N. D., Behzadan, A. H., & Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment. *Automation in Construction*, 112, 103085.
- [13] Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M. D., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review. *Journal of Building Engineering*, 44, 103299.
- [14] Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, 122, 103517.
- [15] Darko, A., Chan, A. P. C., Adabre, M. A., Edwards, D. J., Hosseini, M. R., & Ameyaw, E. E. (2020). Artificial intelligence in the AEC industry: Scientometric analysis and visualization. *Journal of Cleaner Production*, 271, 122825.
- [16] Martinez, P., Al-Hussein, M., & Ahmad, R. (2020). A scientometric analysis and critical review of computer vision applications for construction. *Automation in Construction*, 107, 102947.
- [17] Newaz, M. T., Davis, P., Jefferies, M., & Pillay, M. (2022). Developing a safety climate factor model in construction. *Engineering, Construction and Architectural Management*, 29(1), 401-420.
- [18] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. *Proceedings of KDD*, 1135-1144.
- [19] Arrieta, A. B., Diaz-Rodriguez, N., Del Ser, J., Benetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence: Concepts, taxonomies, opportunities and challenges. *Information Fusion*, 58, 82-115.
- [20] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1-35.
- [21] Kingdom of Saudi Arabia. (2021). Personal Data Protection Law, Royal Decree No. M/19.
- [22] Saudi Data and Artificial Intelligence Authority. (2020). National Strategy for Data and AI.
- [23] European Union Agency for Cybersecurity. (2023). Cybersecurity of AI and standardisation.
- [24] International Organization for Standardization. (2022). ISO/IEC 27001:2022 Information security management systems - Requirements.
- [25] European Parliament and Council. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence.