Advanced AI OHS Professional Certification (Part II) (Canada)

Implementing AI-Based Predictive Analytics for OHS

AI (Artificial Intelligence) – The simulation of human intelligence processes by machines, especially computer systems. Related terms: machine learning, deep learning, cognitive computing. In OHS predictive analytics, AI enables the analysis of complex safety data to uncover hidden risk patterns and support proactive interventions.

AI Ethics – Principles guiding the responsible development and deployment of AI systems, ensuring fairness, transparency, accountability, and respect for privacy. Related terms: bias mitigation, algorithmic fairness, responsible AI. For OHS, ethical AI safeguards worker data and prevents discriminatory risk assessments.

AI Governance – Framework of policies, procedures, and oversight mechanisms that direct AI strategy, risk management, and compliance within an organization. Related terms: AI policy, AI oversight committee, AI risk register. Effective governance ensures predictive models align with occupational health regulations and corporate safety objectives.

Algorithmic Bias – Systematic and unfair discrimination embedded in AI models due to biased training data or design choices. Related terms: bias detection, fairness metrics, mitigation strategies. In OHS analytics, unchecked bias could misrepresent risk for certain worker groups, leading to inequitable safety measures.

Algorithmic Transparency – The degree to which the inner workings of an AI model are understandable and explainable to stakeholders. Related terms: explainable AI, model interpretability, documentation. Transparent models enable safety managers to justify predictive alerts and regulatory filings.

Apache Spark – An open-source distributed computing system for big data processing, supporting real-time analytics and machine learning pipelines. Related terms: RDD, DataFrame, Spark MLlib. Spark can ingest large OHS sensor streams, perform feature engineering, and feed predictive models at scale.

ArcGIS – Geographic information system (GIS) software for mapping and spatial analysis. Related terms: spatial clustering, heat maps, geocoding. Integrating ArcGIS with predictive analytics visualizes injury hotspots across worksites, guiding targeted interventions.

Asset Management – Systematic approach to maintaining and optimizing physical assets throughout their lifecycle. Related terms: CMMS, predictive maintenance, reliability engineering. AI-driven predictive analytics forecast equipment failures that could cause safety incidents, supporting proactive maintenance schedules.

Association Rule Mining – Data mining technique for discovering relationships (if-then statements) among variables in large datasets. Related terms: Apriori algorithm, confidence, support. In OHS, rule mining can reveal co-occurring risk factors such as “night shift” and “manual handling” leading to higher injury rates.

Automation Bias – Tendency of users to over-trust automated decisions, potentially overlooking errors. Related terms: human-in-the-loop, decision support, verification. Safety officers must validate AI alerts to prevent reliance on false positives or negatives.

Balancing Scorecard – Strategic management tool linking performance metrics across financial, customer, internal process, and learning dimensions. Related terms: KPI, leading indicators, lagging indicators. Predictive analytics enrich the “learning” quadrant with forward-looking safety risk scores.

Bayesian Networks – Probabilistic graphical models representing variables and their conditional dependencies. Related terms: inference, belief propagation, DAG. In OHS, Bayesian networks model causal chains from environmental exposure to injury outcomes, updating risk estimates as new data arrives.

Benchmarking – Process of comparing an organization’s performance metrics against industry standards or best practices. Related terms: KPIs, gap analysis, performance dashboard. AI predictive scores can be benchmarked to identify safety performance relative to peers.

Big Data – Extremely large and complex data sets that exceed traditional processing capabilities. Related terms: volume, velocity, variety, veracity. OHS data streams from wearables, incident logs, and environmental sensors constitute big data for AI analytics.

Binary Classification – Machine-learning task that assigns observations to one of two categories (e.G., “High risk” vs “low risk”). Related terms: logistic regression, ROC curve, confusion matrix. Most OHS predictive models output a binary safety alert to trigger immediate action.

Black-Box Model – AI model whose internal logic is not readily interpretable, often due to complex architectures. Related terms: deep neural network, opacity, explainability. While powerful, black-box models pose challenges for regulatory compliance in occupational safety reporting.

Bluetooth Low Energy (BLE) Sensors – Wireless devices that transmit low-power data such as location, temperature, or motion. Related terms: IoT, proximity detection, signal strength. BLE sensor data feed predictive algorithms that detect unsafe proximity or fatigue-related movement patterns.

CMMS (Computerized Maintenance Management System) – Software that tracks maintenance activities, work orders, and asset histories. Related terms: preventive maintenance, work order scheduling, asset lifecycle. Integrating CMMS data with AI models predicts equipment-related injury risks before breakdowns occur.

Confusion Matrix – Tabular representation of classification outcomes: True positives, false positives, true negatives, false negatives. Related terms: precision, recall, F1-score. In OHS predictive analytics, the matrix quantifies model effectiveness in correctly flagging imminent hazards.

Continuous Integration/Continuous Deployment (CI/CD) – Automated pipeline for building, testing, and releasing software updates. Related terms: DevOps, version control, automated testing. CI/CD ensures AI model updates for safety analytics are rolled out reliably without service interruption.

Correlation vs Causation – Distinction between variables that move together (correlation) and those where one directly influences the other (causation). Related terms: spurious correlation, causal inference, confounding. Predictive analytics must avoid mistaking correlated OHS factors for causal injury drivers.

Cross-Validation – Technique for assessing model performance by partitioning data into training and testing subsets multiple times. Related terms: k-fold, hold-out set, model stability. Robust cross-validation prevents overfitting of OHS risk models to historical incident data.

Data Augmentation – Process of artificially expanding training datasets using transformations or synthetic generation. Related terms: SMOTE, oversampling, noise injection. For rare OHS events like severe injuries, augmentation helps AI models learn from limited examples.

Data Governance – Policies, standards, and processes that ensure data quality, security, and compliance. Related terms: data stewardship, metadata management, data lineage. Strong governance protects worker privacy while enabling trustworthy predictive analytics.

Data Lake – Centralized repository that stores raw, unstructured, and structured data at scale. Related terms: schema-on-read, Hadoop, object storage. OHS data from video feeds, sensor logs, and health records can be ingested into a data lake for downstream AI modeling.

Data Normalization – Scaling numeric features to a common range to improve model convergence. Related terms: min-max scaling, z-score, feature scaling. Normalizing exposure intensity measurements ensures fair weighting across diverse OHS variables.

Data Pipeline – Automated workflow that extracts, transforms, and loads (ETL) data from source to destination. Related terms: Airflow, Prefect, streaming ingestion. Real-time OHS pipelines deliver sensor readings to predictive models with minimal latency.

Data Privacy Impact Assessment (DPIA) – Systematic evaluation of privacy risks associated with data processing activities. Related terms: GDPR, consent management, anonymization. Conducting a DPIA before deploying AI for worker health data safeguards compliance with Canadian privacy law.

Decision Support System (DSS) – Interactive software that assists users in making informed decisions based on data analysis. Related terms: dashboards, alerts, scenario modeling. AI-driven OHS DSS provides safety managers with risk scores, recommended controls, and mitigation timelines.

Deep Learning – Subset of machine learning using multi-layer neural networks to model complex patterns. Related terms: convolutional neural network, recurrent neural network, backpropagation. Deep learning excels at processing unstructured OHS data such as video footage of unsafe behaviors.

Deployment Architecture – Structural design of hardware, software, and network components that host AI models in production. Related terms: microservices, containers, edge computing. Choosing an architecture that balances latency and security is critical for real-time OHS hazard detection.

Dimensionality Reduction – Techniques that compress high-dimensional data into fewer variables while retaining essential information. Related terms: PCA, t-SNE, feature selection. Reducing sensor feature space improves model speed for on-site predictive analytics.

Discrete Event Simulation – Modeling approach that represents systems as a sequence of distinct events in time. Related terms: Monte Carlo, queuing theory, process flow. Simulating worksite operations helps validate AI-generated risk forecasts against realistic scenarios.

Docker – Platform for packaging applications and their dependencies into portable containers. Related terms: containerization, image registry, orchestration. Containerizing OHS predictive models ensures consistent execution across development, testing, and production environments.

Edge Computing – Processing data near its source rather than in centralized clouds. Related terms: IoT gateway, latency reduction, fog computing. Edge deployment of AI models enables instant hazard alerts from wearable sensors without network delays.

Ensemble Learning – Method that combines predictions from multiple models to improve accuracy. Related terms: bagging, boosting, stacking. An ensemble of decision trees, logistic regressions, and neural networks can robustly predict injury likelihood across diverse OHS datasets.

Feature Engineering – Creation, transformation, and selection of variables that improve model performance. Related terms: one-hot encoding, interaction terms, lag features. Crafting a "cumulative exposure index" from daily noise level readings enhances predictive power for hearing loss risk.

Feature Importance – Metric quantifying the contribution of each input variable to a model's predictions. Related terms: SHAP values, permutation importance, gain. Highlighting top risk factors (e.G., "Lifting weight > 30 kg") guides targeted safety interventions.

Fisher Exact Test – Statistical significance test used for small sample sizes and categorical data. Related terms: contingency table, p-value, odds ratio. The test evaluates whether a rare injury event is disproportionately associated with a specific work shift.

FMEA (Failure Modes and Effects Analysis) – Systematic method for identifying potential failure points and their impacts. Related terms: risk priority number, mitigation planning, reliability. AI-enhanced FMEA incorporates predictive failure probabilities to prioritize safety controls.

Forecasting Horizon – Time span into the future for which predictions are generated. Related terms: lead time, short-term horizon, long-term planning. In OHS, a 7-day horizon may trigger immediate PPE checks, whereas a 12-month horizon informs strategic safety budgeting.

Frequentist vs Bayesian Statistics – Two philosophical approaches to inference: Frequentist relies on long-run frequency, while Bayesian incorporates prior beliefs. Related terms: posterior distribution, confidence interval, prior probability. Bayesian updating allows OHS risk models to refine predictions as new incident data become available.

General Data Protection Regulation (GDPR) – EU regulation governing personal data processing, with extraterritorial effect. Related terms: right to erasure, data minimization, consent. While a Canadian course, understanding GDPR principles helps align AI OHS solutions with global privacy expectations.

Geofencing – Virtual boundary defined by GPS coordinates that triggers actions when devices enter or exit. Related terms: location-based alerts, geo-tags, perimeter monitoring. Geofencing can alert workers when they approach high-risk zones identified by predictive models.

Gradient Boosting Machines (GBM) – Ensemble technique that builds sequential decision trees, each correcting errors of the prior. Related terms: XGBoost, LightGBM, learning rate. GBM often yields high accuracy for binary OHS risk classification tasks.

Hazard Identification (HAZID) – Process of systematically recognizing potential sources of injury or illness. Related terms: risk matrix, JSA, safety audit. AI can augment HAZID by surfacing previously unnoticed patterns from historical incident data.

Hazardous Materials Management System (HMMS) – Software that tracks inventory, storage, and handling of dangerous substances. Related terms: SDS, chemical safety, compliance reporting. Predictive analytics applied to HMMS data forecast chemical exposure incidents before they occur.

Health Surveillance – Ongoing monitoring of workers' health status to detect early signs of occupational disease. Related terms: biomonitoring, medical exams, exposure tracking. AI models analyze trends in surveillance data to predict emerging health issues such as respiratory disorders.

Hierarchical Clustering – Unsupervised learning method that groups data points into a tree-like structure based on similarity. Related terms: dendrogram, agglomerative, linkage criteria. Clustering incident reports can reveal natural groupings of injuries by department or equipment type.

Human-in-the-Loop (HITL) – Design approach that retains human oversight over automated decisions. Related terms: review workflow, escalation, validation. In OHS, HITL ensures that AI-generated hazard alerts are examined by safety professionals before action.

Hyperparameter Tuning – Optimization of model settings that are not learned during training. Related terms: grid search, random search, Bayesian optimization. Proper tuning of tree depth, learning rate, and regularization improves OHS risk model generalization.

Incident Command System (ICS) – Standardized management framework for emergency response coordination. Related terms: incident action plan, role assignment, escalation protocol. AI predictive alerts can be fed into the ICS to pre-position resources before an incident escalates.

Incident Rate (IR) – Metric representing the number of occupational injuries per a standard number of work hours (e.G., Per 200,000 hours). Related terms: TRIR, lost-time injury frequency rate, severity rate. Predictive models aim to reduce IR by identifying high-risk periods in advance.

Inference Engine – Component of an AI system that applies logical rules or probabilistic calculations to generate conclusions. Related terms: rule-based system, Bayesian inference, decision logic. In OHS, the inference engine translates sensor inputs into actionable risk scores.

Information Security Management System (ISMS) – Set of policies and procedures for managing information security risks. Related terms: ISO 27001, confidentiality, access control. Protecting OHS data from breaches is essential when deploying AI models that contain personal health information.

IoT (Internet of Things) – Network of interconnected devices that collect and exchange data. Related terms: sensor nodes, MQTT, edge analytics. Wearable IoT devices capture physiological and environmental metrics that feed predictive safety algorithms.

JIT (Just-In-Time) Maintenance – Maintenance strategy that performs repairs exactly when needed, minimizing downtime. Related terms: predictive maintenance, condition-based monitoring, reliability. AI forecasts equipment failure risk, enabling JIT interventions that also reduce injury exposure.

JSA (Job Safety Analysis) – Systematic examination of job tasks to identify hazards and prescribe controls. Related terms: SWMS, risk assessment, task breakdown. AI can auto-populate JSA templates by extracting task-specific risk factors from historical data.

K-Fold Cross-Validation – Specific form of cross-validation that partitions data into *k* equal subsets. Related terms: validation set, overfitting prevention, model robustness. Using 5-fold cross-validation balances computational cost with reliable performance estimates for OHS models.

KPI (Key Performance Indicator) – Quantifiable measure used to evaluate success against objectives. Related terms: metric, target, dashboard. AI-derived safety KPIs, such as “predicted high-risk incidents per month,” provide early insight into program effectiveness.

Label Imbalance – Situation where one class (e.G., Injuries) is far less frequent than the other (non-injuries). Related terms: class weighting, SMOTE, under-sampling. Addressing imbalance prevents models from defaulting to the majority class and missing rare but critical safety events.

Latent Variable – Unobserved factor inferred from observed data, often representing underlying risk. Related terms: hidden state, factor analysis, EM algorithm. In OHS, latent variables may capture “worker fatigue” derived from sleep patterns, heart rate variability, and shift length.

LightGBM – Gradient boosting framework that uses leaf-wise tree growth for speed and efficiency. Related terms: GBDT, histogram-based splitting, parallel learning. LightGBM’s low latency makes it suitable for real-time OHS hazard prediction on streaming data.

Logistic Regression – Statistical model that estimates the probability of a binary outcome using a logistic function. Related terms: odds ratio, maximum likelihood, regularization. Logistic regression remains a baseline for OHS risk scoring due to interpretability and simplicity.

Machine Learning Operations (MLOps) – Set of practices that combine machine learning development with DevOps principles for reliable production deployment. Related terms: model registry, continuous training, monitoring. MLOps pipelines automate retraining of OHS predictive models as new incident data are recorded.

Meta-Learning – “Learning to learn” approach where algorithms improve their ability to adapt to new tasks quickly. Related terms: few-shot learning, model-agnostic meta-learning (MAML), transfer learning. Meta-learning can accelerate OHS model adaptation when entering a new industry sector with limited data.

Missing Data Imputation – Techniques for estimating absent values in datasets. Related terms: mean substitution, multiple imputation, k-NN imputer. Proper imputation ensures AI models do not bias risk predictions due to gaps in sensor logs or health records.

Model Drift – Degradation of model performance over time as underlying data distributions change. Related terms: concept drift, performance monitoring, retraining triggers. Detecting drift in OHS risk models is vital when workplace processes, equipment, or regulations evolve.

Model Explainability – Ability to articulate how an AI model reaches its conclusions. Related terms: SHAP, LIME, counterfactual analysis. Explainable OHS models help safety officers understand why a particular worker is flagged as high-risk.

Monte Carlo Simulation – Computational technique that uses random sampling to estimate probability distributions of outcomes. Related terms: stochastic modeling, risk quantification, scenario analysis. Monte Carlo can assess the uncertainty of predicted injury rates under varying safety investment levels.

Multi-Label Classification – Predictive task where each observation may belong to multiple categories simultaneously. Related terms: binary relevance, classifier chain, label powerset. OHS models may assign both “chemical exposure” and “ergonomic strain” labels to a single incident.

Natural Language Processing (NLP) – Suite of techniques for analyzing and deriving meaning from human language. Related terms: sentiment analysis, entity extraction, topic modeling. NLP extracts risk indicators from free-text incident reports, safety meeting minutes, and employee feedback.

Neural Architecture Search (NAS) – Automated process for discovering optimal neural network structures. Related terms: hyperparameter optimization, reinforcement learning, search space. NAS can tailor deep-learning models to the specific sensor modalities used in OHS environments.

Normalization (Database) – Process of organizing data to reduce redundancy and improve integrity. Related terms: first normal form, relational schema, foreign key. Normalized OHS databases facilitate accurate joins between incident logs, personnel records, and equipment maintenance tables.

Ontology (Domain Knowledge) – Formal representation of concepts and relationships within a specific field. Related terms: semantic model, RDF, knowledge graph. Building an OHS ontology enables AI to reason about hierarchical hazard categories and regulatory requirements.

Outlier Detection – Identification of data points that deviate markedly from the norm. Related terms: Z-score, isolation forest, DBSCAN. Detecting outliers in sensor streams can signal abnormal equipment vibrations indicative of imminent failure.

Overfitting – Situation where a model captures noise instead of underlying patterns, performing poorly on unseen data. Related terms: regularization, validation set, early stopping. Guarding against overfitting is essential when training OHS models on limited injury records.

Parameter Server – Distributed system that stores and updates model parameters during large-scale training. Related terms: distributed learning, synchronization, gradient aggregation. Parameter servers enable training of deep OHS models on massive sensor datasets across multiple sites.

Partial Dependence Plot (PDP) – Visualization that shows the marginal effect of a feature on the predicted outcome. Related terms: ICE plot, feature effect, model interpretation. PDPs help safety analysts understand how increasing noise exposure influences predicted hearing loss risk.

Personal Protective Equipment (PPE) Compliance Monitoring – Use of AI to verify that workers are wearing required safety gear. Related terms: computer vision, RFID tagging, compliance score. Real-time alerts reduce incidents caused by PPE non-use.

Predictive Maintenance – Strategy that anticipates equipment failures before they occur using data analytics. Related terms: condition monitoring, remaining useful life, failure prediction. Preventing equipment breakdowns directly lowers the likelihood of injury in OHS contexts.

Predictive Modeling – Process of creating statistical or machine learning models that forecast future events. Related terms: regression, classification, time-series analysis. In OHS, predictive modeling estimates the probability of incidents based on historical and real-time inputs.

Probabilistic Graphical Model – Framework that represents random variables and their conditional dependencies via graphs. Related terms: Markov network, Bayesian network, factor graph. Such models capture complex interrelations among exposure, behavior, and injury outcomes in safety analytics.

Process Mining – Extraction of process models from event logs to discover actual workflow behavior. Related terms: conformance checking, bottleneck analysis, process discovery. Applying process mining to OHS work orders reveals deviations that may increase injury risk.

Python – High-level programming language widely used for data science and AI development. Related terms: Pandas, Scikit-learn, TensorFlow. Python scripts automate data cleaning, feature extraction, and model training for OHS predictive analytics.

Quantum Machine Learning – Emerging field that leverages quantum computing to accelerate certain ML algorithms. Related terms: qubits, quantum annealing, hybrid quantum-classical. While still experimental, quantum approaches could eventually handle the combinatorial optimization problems in large-scale OHS risk assessment.

Random Forest – Ensemble learning method that constructs multiple decision trees on bootstrapped samples and aggregates their predictions. Related terms: bagging, feature importance, out-of-bag error. Random forests provide robust OHS risk classification with built-in measures of variable relevance.

Recall (Sensitivity) – Metric measuring the proportion of actual positives correctly identified. Related terms: true positive rate, false negative rate, detection rate. High recall in OHS models ensures most genuine hazards are flagged, even at the cost of some false alarms.

Regulatory Compliance – Adherence to laws, standards, and guidelines governing occupational health and safety. Related terms: OSHA, CSA, audit trail. AI systems must generate documentation that demonstrates compliance with reporting and mitigation obligations.

Reinforcement Learning – Learning paradigm where an agent interacts with an environment, receiving rewards for desirable actions. Related terms: policy, Q-learning, Markov decision process. In OHS, reinforcement learning can optimize the scheduling of safety inspections to maximize hazard reduction.

Remote Sensing – Acquisition of information about a surface without direct contact, often via satellites or drones. Related terms: LIDAR, multispectral imaging, aerial surveys. Remote sensing data enrich predictive models with terrain and environmental risk factors such as slope steepness.

Resampling Techniques – Methods for creating multiple training/test splits to assess model stability. Related terms: bootstrapping, jackknife, Monte Carlo cross-validation. Resampling helps quantify uncertainty in OHS risk predictions.

RNN (Recurrent Neural Network) – Neural architecture designed to process sequential data by maintaining internal state. Related terms: LSTM, GRU, time-series forecasting. RNNs model temporal patterns in worker fatigue signals derived from wearable heart-rate monitors.

Root Cause Analysis (RCA) – Systematic investigation to determine underlying reasons for an incident. Related terms: 5 Whys, fishbone diagram, corrective action. AI can prioritize probable root causes by correlating incident attributes with historical RCA outcomes.

Safety Climate Survey – Questionnaire that gauges employees' perceptions of safety culture. Related terms: psychological safety, engagement index, Likert scale. AI aggregates survey responses to predict areas where low safety climate may precede incidents.

Safety Data Sheet (SDS) – Document providing detailed information about hazardous chemicals. Related terms: GHS, hazard classification, exposure limits. NLP extracts hazard codes from SDS to enrich predictive models of chemical exposure incidents.

Scaling (Machine Learning) – Adjusting model complexity or computational resources to handle larger datasets. Related terms: horizontal scaling, GPU acceleration, distributed training. Scaling ensures OHS predictive pipelines remain responsive as sensor networks expand.

Scenario Analysis – Exploration of "what-if" situations to assess potential outcomes under varying assumptions. Related terms: sensitivity analysis, stress testing, decision trees. Scenario analysis evaluates how changes in staffing levels affect predicted injury rates.

Seamless Integration – Design principle ensuring that AI components connect smoothly with existing OHS information systems. Related terms: API, middleware, data interoperability. Seamless integration reduces data silos and accelerates adoption of predictive safety tools.

Sensor Fusion – Combination of data from multiple sensor modalities to produce more accurate estimates. Related terms: Kalman filter, data assimilation, multimodal analytics. Fusing accelerometer, temperature, and ambient noise readings improves detection of unsafe conditions.

Sequential Pattern Mining – Identification of frequent subsequences within ordered event data. Related terms: prefixspan, SPADE, temporal rules. Mining sequential patterns from incident logs uncovers typical event chains that precede accidents.

Shapley Additive Explanations (SHAP) – Game-theoretic method for assigning each feature an importance value for a particular prediction. Related terms: local interpretability, contribution plot, model agnostic.

SHAP values enable OHS managers to see why a specific worker received a high-risk score.

Signal-to-Noise Ratio (SNR) – Measure comparing the level of desired signal to the level of background noise. Related terms: filtering, data quality, thresholding. High SNR in sensor streams ensures reliable inputs for AI hazard detection.

Singular Value Decomposition (SVD) – Matrix factorization technique used for dimensionality reduction and noise filtering. Related terms: latent semantic analysis, PCA, low-rank approximation. SVD can compress large OHS exposure matrices while preserving essential variance.

Skewed Distribution – Probability distribution where most observations cluster to one side, creating imbalance. Related terms: log-normal, power-law, transformation. Many OHS metrics, such as injury severity, exhibit skewness, requiring appropriate modeling techniques.

Smart PPE – Protective equipment embedded with sensors and connectivity for real-time monitoring. Related terms: IoT-enabled helmets, vibration-damping gloves, temperature-sensing vests. Smart PPE streams data to AI models that predict fatigue or exposure exceedances.

Social Network Analysis (SNA) – Study of relationships and interactions among individuals within a group. Related terms: centrality, community detection, graph metrics. SNA reveals informal safety communication pathways that influence hazard awareness.

Software Development Kit (SDK) – Collection of tools, libraries, and documentation for building applications. Related terms: API, sample code, integration guide. An OHS AI SDK accelerates custom development of risk dashboards and mobile alert apps.

Spatial Autocorrelation – Statistical property where nearby locations exhibit similar values. Related terms: Moran’s I, Geary’s C, hotspot analysis. Accounting for spatial autocorrelation prevents overestimation of risk clusters in OHS predictive maps.

SQL (Structured Query Language) – Standard language for managing relational databases. Related terms: SELECT, JOIN, stored procedure. SQL queries extract historical incident records for training OHS machine-learning models.

Standard Operating Procedure (SOP) – Documented step-by-step instructions to perform routine tasks safely. Related terms: work instruction, compliance checklist, procedural audit. AI can recommend SOP revisions when predictive analytics identify emerging hazards.

Statistical Process Control (SPC) – Methodology using control charts to monitor process stability. Related terms: Shewhart chart, process capability, assignable cause. SPC charts integrated with AI predictions highlight when safety processes deviate from acceptable limits.

Stratified Sampling – Technique that divides a population into homogeneous subgroups and samples each proportionally. Related terms: cluster sampling, weighting, representation. Stratified sampling ensures OHS datasets capture diverse worker categories (e.G., Trades, office staff).

Supervised Learning – Machine-learning paradigm where models are trained on labeled input-output pairs. Related terms: classification, regression, training set. Most OHS predictive tasks rely on supervised learning using historical incident labels.

Support Vector Machine (SVM) – Classification algorithm that finds the hyperplane maximizing margin between classes. Related terms: kernel trick, soft margin, support vectors. SVMs can separate high-risk vs low-risk scenarios when feature space is complex.

Survival Analysis – Statistical approach for modeling time-to-event data, accounting for censored observations. Related terms: Cox proportional hazards, Kaplan-Meier, hazard function. Survival analysis predicts the time until a worker experiences a musculoskeletal injury based on exposure history.

Swarm Intelligence – Collective behavior algorithms inspired by natural groups such as ants or bees. Related terms: particle swarm optimization, ant colony optimization, decentralized search. Swarm techniques can optimize sensor placement across a large construction site for maximal hazard coverage.

Temporal Fusion Transformer (TFT) – Deep-learning architecture for multi-horizon time-series forecasting. Related terms: attention mechanism, variable selection network, quantile regression. TFT handles mixed static and dynamic OHS inputs, delivering accurate short- and long-term risk forecasts.

Time-Series Decomposition – Process of separating a series into trend, seasonal, and residual components. Related terms: STL, additive model, multiplicative model. Decomposition clarifies whether rising injury counts are driven by seasonal work peaks or underlying systemic issues.

Transfer Learning – Reusing a pre-trained model on a new, related task to reduce data requirements. Related terms: fine-tuning, domain adaptation, feature reuse. Transfer learning enables OHS teams to leverage large-scale image models for detecting unsafe postures from limited site-specific footage.

Tree-Based Model – Predictive model that uses hierarchical decision rules, such as decision trees or ensembles. Related terms: splitting criterion, leaf node, pruning. Tree-based models are popular in OHS because they produce intuitive rule sets for safety managers.

Unstructured Data – Information that does not follow a predefined data model, such as text, images, or audio. Related terms: free-text, multimedia, raw logs. Unstructured OHS data (e.G., Safety photos) require NLP or computer-vision techniques for analysis.

Under-Sampling – Technique that reduces the majority class size to balance class distribution. Related terms: random under-sampling, Tomek links, cluster centroids. Under-sampling helps prevent OHS models from being biased toward the abundant “no-incident” records.

Uncertainty Quantification – Process of measuring the confidence or error bounds of model predictions. Related terms: prediction interval, Bayesian posterior, Monte Carlo dropout. Quantified uncertainty informs safety leaders about the reliability of AI-generated risk scores.

Validation Set – Subset of data used to tune model hyperparameters without influencing final performance assessment. Related terms: hold-out, cross-validation, early stopping. A dedicated validation set guards against over-optimistic OHS model evaluations.

Variable Importance – Ranking of input features based on their impact on predictive performance. Related terms: Gini importance, permutation score, SHAP summary. Knowing which OHS variables drive risk empowers targeted mitigation strategies.

Version Control – System for tracking changes to code, data, and model artifacts. Related terms: Git, branching, commit history. Version control ensures reproducibility of OHS predictive analytics experiments.

Virtual Reality (VR) Safety Training – Immersive simulation used to teach hazard awareness. Related terms: simulation, immersive learning, scenario rehearsal. While not a hands-on activity, AI can personalize VR scenarios based on predicted individual risk profiles.

Vision-Based Hazard Detection – Use of computer-vision algorithms to identify unsafe conditions from video streams. Related terms: object detection, YOLO, semantic segmentation. Vision models flag missing guardrails or workers entering restricted zones in real time.

Weighted Loss Function – Modification of the loss calculation to give higher penalty to certain classes.