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

Applying AI in Hazard Identification and Risk Assessment

AI (Artificial Intelligence) – A branch of computer science that creates systems capable of performing tasks that normally require human intelligence, such as learning, reasoning, and perception. Machine learning, deep learning, neural networks are sub-fields. In hazard identification, AI can analyze large volumes of safety data to spot patterns that indicate potential risks. Example: An AI model processes incident reports to highlight recurring unsafe acts. Challenges include data quality, bias in training data, and ensuring the AI's decisions remain transparent to OHS professionals.

Algorithm – A step-by-step computational procedure used to solve a problem or perform a calculation. In risk assessment, algorithms sort, filter, and prioritize hazards based on severity, likelihood, and exposure. Related terms: heuristic, optimization, decision tree. Example: A scoring algorithm assigns numeric risk values to each identified hazard, allowing quick ranking. Challenges involve selecting appropriate parameters and avoiding over-fitting to historical data that may not reflect emerging risks.

Anomaly Detection – The process of identifying data points or patterns that deviate significantly from the norm. Used in OHS to flag unusual sensor readings, equipment vibrations, or sudden spikes in injury reports. Related terms: outlier analysis, statistical process control, unsupervised learning. Example: An AI system monitors temperature data from a furnace and raises an alert when readings exceed expected variance, indicating a potential fire hazard. Challenges include defining "normal" baselines in dynamic work environments and reducing false positives that can lead to alert fatigue.

Artificial Neural Network (ANN) – A computing model inspired by the human brain's network of neurons, capable of learning complex relationships from data. In hazard identification, ANNs can predict the likelihood of incidents based on multi-dimensional inputs such as worker demographics, equipment age, and environmental conditions. Related terms: deep learning, back-propagation, hidden layers. Example: An ANN predicts the probability of a slip-trip-fall on a construction site by analyzing weather, floor material, and foot traffic density. Challenges involve needing large labeled datasets and interpreting the "black-box" nature of the model for regulatory compliance.

Bayesian Inference – A statistical method that updates the probability of a hypothesis as more evidence becomes available. Applied to risk assessment, it refines hazard probability estimates when new incident data are reported. Related terms: prior probability, posterior distribution, Markov Chain Monte Carlo. Example: Initial belief that a chemical exposure risk is low is revised upward after several near-misses are logged, using Bayesian updating. Challenges include selecting appropriate prior distributions and computational intensity for complex models.

Big Data – Extremely large and diverse data sets that require advanced processing techniques to extract value. In OHS, big data may include sensor streams, wearable device logs, incident databases, and environmental monitoring records. Related terms: data lake, Hadoop, real-time analytics. Example: Aggregating 10 million sensor readings from a mining operation enables AI to map high-risk zones with unprecedented granularity. Challenges involve data storage costs, ensuring data privacy, and integrating heterogeneous sources.

Classification Model – An AI model that assigns input data to predefined categories. For hazard identification, classification models differentiate between “safe” and “unsafe” conditions, or categorize incident types (e.G., Ergonomic, chemical, mechanical). Related terms: support vector machine, logistic regression, confusion matrix. Example: A model classifies video frames from a production line as “proper PPE worn” or “PPE missing.” Challenges include class imbalance when unsafe events are rare, requiring techniques such as oversampling or cost-sensitive learning.

Confusion Matrix – A table used to evaluate the performance of a classification model by comparing predicted and actual outcomes. It reports true positives, false positives, true negatives, and false negatives. Related terms: precision, recall, F1-score. Example: In a PPE detection system, a confusion matrix shows 92 % true positives, 5 % false positives, 2 % false negatives, and 1 % true negatives, guiding model refinement. Challenges include interpreting metrics when the cost of a false negative (missed hazard) is much higher than a false positive.

Contextual Bandit – A reinforcement learning approach where an algorithm selects actions based on context, receiving immediate feedback to improve future choices. In OHS, contextual bandits can recommend targeted safety interventions for specific work zones based on current conditions. Related terms: exploration vs. Exploitation, reward function, online learning. Example: The system suggests additional signage in a high-traffic corridor after observing increased near-misses, then monitors the reduction in incidents to adjust recommendations. Challenges involve defining appropriate reward signals and preventing unsafe exploratory actions.

Cross-Validation – A technique for assessing how a predictive model will generalize to an independent data set by partitioning data into training and testing subsets multiple times. In hazard risk modeling, cross-validation helps avoid over-optimistic performance estimates. Related terms: k-fold, hold-out set, model validation. Example: A 5-fold cross-validation evaluates a risk scoring algorithm across different subsets of incident data, ensuring stability before deployment. Challenges include computational load for large data sets and maintaining temporal integrity when data have time-dependent patterns.

Data Augmentation – The process of artificially expanding a training dataset by applying transformations such as rotation, scaling, or noise injection. Useful when hazardous event data are scarce. Related terms: synthetic data, oversampling, generative adversarial network. Example: Synthetic images of faulty machine guards are generated to train a visual inspection AI, improving detection of missing guards. Challenges involve ensuring synthetic examples remain realistic and do not introduce bias.

Data Governance – The set of policies, procedures, and standards that ensure data is managed responsibly, securely, and with quality. In OHS AI projects, governance covers data ownership, privacy, retention, and audit trails. Related terms: data stewardship, compliance, metadata. Example: A governance framework mandates that all employee health data used for fatigue risk modeling is anonymized and stored for a maximum of three years. Challenges include aligning governance with multiple regulatory regimes (e.G., PIPEDA, OSHA) and maintaining stakeholder buy-in.

Decision Tree – A flowchart-like model that splits data based on feature thresholds to reach a decision or prediction. In risk assessment, decision trees can illustrate how combinations of factors (e.G., Exposure level, training, equipment age) lead to different risk categories. Related terms: entropy, Gini impurity, pruning. Example: A tree shows that when exposure > 50 ppm *and* ventilation is inadequate, the risk level escalates to “high.” Challenges include over-fitting to training data and loss of interpretability when trees become too deep.

Deep Learning – A subset of machine learning that uses multilayer neural networks to learn hierarchical representations from raw data. Enables processing of unstructured inputs such as images, audio, and video for hazard detection. Related terms: convolutional neural network, recurrent neural network, GPU acceleration. Example: A deep learning model analyses live video from a warehouse to detect unsafe lifting techniques, issuing instant alerts. Challenges involve high computational demands, need for large labeled datasets, and difficulty in explaining model decisions to safety managers.

Dimensionality Reduction – Techniques that reduce the number of variables while preserving essential information, facilitating visualization and improving model performance. Common methods include Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE). Related terms: feature extraction, latent space, variance preservation. Example: PCA compresses sensor data from 200 variables to 10 principal components, allowing a risk clustering algorithm to run efficiently. Challenges include potential loss of critical safety signals and interpreting transformed features.

Edge Computing – Processing data near the source (e.G., On a sensor or local device) rather than sending it to a central server. Reduces latency and bandwidth needs for real-time hazard monitoring. Related terms: fog computing, IoT, latency. Example: A wearable device runs a lightweight AI model locally to detect abnormal heart rate patterns, triggering an immediate evacuation alert without cloud reliance. Challenges involve limited processing power on edge devices and ensuring model updates are securely delivered.

Ensemble Methods – Techniques that combine multiple machine learning models to improve predictive performance and robustness. Examples include bagging, boosting, and stacking. Related terms: random forest, AdaBoost, model diversity. Example: An ensemble of decision trees, logistic regression, and a neural network predicts the likelihood of a workplace fire, achieving higher accuracy than any single model. Challenges include increased computational complexity and difficulty in interpreting the aggregated decision logic.

Ethical AI – The practice of designing, developing, and deploying AI systems that respect human rights, fairness, transparency, and accountability. In OHS, ethical AI ensures that risk models do not discriminate against specific worker groups and that privacy is protected. Related terms: bias mitigation, explainability, responsible AI. Example: An AI-driven fatigue monitoring system is audited to confirm it does not disproportionately flag certain demographics, and clear opt-out mechanisms are provided. Challenges include assessing hidden biases in historical safety data and balancing safety benefits against individual privacy.

Explainable AI (XAI) – Methods and tools that make the inner workings of AI models understandable to humans. Critical for OHS decision-makers who need to justify safety actions. Related terms: SHAP values, LIME, model interpretability. Example: A SHAP analysis reveals that exposure duration and lack of PPE are the top contributors to a high-risk score in a chemical handling scenario, enabling targeted interventions. Challenges include generating explanations that are both technically accurate and accessible to non-technical stakeholders.

Feature Engineering – The process of creating, selecting, and transforming variables (features) to improve model performance. In hazard identification, features may include time-of-day, equipment utilization rates, or proximity metrics. Related terms: one-hot encoding, scaling, interaction terms. Example: Combining temperature and humidity into a "heat stress index" feature improves the predictive power of a heat-related injury model. Challenges involve domain expertise to identify meaningful transformations and avoiding leakage of future information into training data.

Feature Selection – Techniques for identifying the most relevant variables for a predictive model, reducing noise and over-fitting. Methods include recursive feature elimination, mutual information, and regularization. Related terms: dimensionality reduction, L1 regularization, correlation analysis. Example: A risk model for electrical injuries retains only voltage, insulation age, and maintenance frequency after feature selection, simplifying deployment. Challenges include ensuring selected features remain stable over time and do not omit emerging risk factors.

Fuzzy Logic – A reasoning approach that handles uncertainty by allowing partial truth values between 0 and 1, rather than binary true/false. Useful for modeling qualitative safety judgments. Related terms: membership functions, inference engine, linguistic variables. Example: A fuzzy system rates "ventilation adequacy" on a scale from "poor" to "excellent," producing a risk factor that blends smoothly with other inputs. Challenges involve defining appropriate membership functions and calibrating the system against real-world outcomes.

Generative Adversarial Network (GAN) – A deep learning architecture consisting of a generator that creates synthetic data and a discriminator that evaluates authenticity. In OHS, GANs can produce realistic incident scenarios for training purposes. Related terms: synthetic data, adversarial training, latent space. Example: A GAN generates varied images of blocked emergency exits, enriching a visual detection model without exposing workers to real hazards. Challenges include mode collapse, ensuring generated data does not replicate sensitive information, and validating realism.

Geospatial Analysis – The examination of data with geographic or spatial components to identify patterns and relationships. Applied to hazard mapping, it reveals high-risk zones based on location-specific factors. Related terms: GIS, heat map, spatial autocorrelation. Example: GIS layers of incident reports, equipment locations, and ventilation maps produce a heat map that highlights a "hot spot" for chemical spills in a plant. Challenges involve data resolution, integrating disparate spatial datasets, and accounting for temporal changes.

Gradient Boosting – An ensemble technique that builds models sequentially, each correcting errors of its predecessor, often achieving high predictive accuracy. Implementations include XGBoost and LightGBM.

Related terms: learning rate, weak learner, loss function. Example: Gradient boosting predicts the probability of a crane overturn based on load weight, wind speed, and operator experience, outperforming traditional logistic regression. Challenges include tuning hyper-parameters, preventing over-fitting, and managing computational resources for large data sets.

Human-in-the-Loop (HITL) – An approach where AI systems incorporate human judgment during training, inference, or decision making. Essential for safety-critical OHS applications where final authority rests with qualified professionals. Related terms: active learning, oversight, decision support. Example: An AI model flags a potential gas leak; a safety officer reviews sensor data before initiating shutdown, providing feedback that refines the model. Challenges include designing efficient interfaces for rapid human feedback and avoiding over-reliance on automation.

Hybrid Modeling – Combining physics-based (mechanistic) models with data-driven AI models to leverage both domain knowledge and empirical patterns. In risk assessment, hybrid models can predict equipment failure by integrating wear-out equations with sensor-derived patterns. Related terms: gray-box modeling, model fusion, residual learning. Example: A hybrid model uses a fatigue-life equation for steel beams plus a neural network that captures temperature-induced degradation, improving accuracy. Challenges involve aligning scales of the two model types and ensuring consistent uncertainty quantification.

Imbalanced Data – Datasets where some classes (e.G., Incidents) are far less frequent than others (e.G., Normal operations). Common in OHS where accidents are rare. Related terms: SMOTE, class weighting, minority class. Example: An accident prediction model uses oversampling of near-miss events to balance the dataset, improving detection of rare high-severity outcomes. Challenges include avoiding over-fitting to synthetic minority samples and preserving realistic variance.

Incident Reporting System (IRS) – A digital platform for logging workplace incidents, near-misses, and unsafe conditions. Provides the primary data source for AI-driven hazard identification. Related terms: root cause analysis, data capture, workflow automation. Example: An IRS integrates with an AI engine that automatically categorizes new reports and suggests corrective actions based on historical patterns. Challenges involve ensuring consistent data entry, handling free-text fields, and protecting employee confidentiality.

Inference Engine – The component of an AI system that applies logical rules or learned patterns to draw conclusions from input data. In rule-based safety systems, the inference engine determines whether a condition meets a hazard threshold. Related terms: knowledge base, forward chaining, rule set. Example: The engine evaluates sensor inputs against a rule “if pressure > 200 psi and temperature > 150 °C, then flag high-risk.” Challenges include scaling rule sets for complex environments and maintaining rule relevance as processes evolve.

IoT (Internet of Things) – Network of interconnected devices equipped with sensors and communication capabilities, enabling continuous data collection from the workplace. IoT feeds real-time streams into AI models for dynamic risk assessment. Related terms: sensor network, MQTT, edge device. Example: Wearable IoT tags monitor workers’ posture, transmitting data to an AI model that detects repetitive strain risk. Challenges involve device security, data interoperability, and battery life management.

Knowledge Graph – A structured representation of entities (e.G., Equipment, hazards, procedures) and their relationships, enabling semantic queries and reasoning. Supports AI in linking disparate safety data sources. Related terms: ontology, RDF, SPARQL. Example: A knowledge graph connects a chemical's hazard classification to required storage protocols, allowing an AI assistant to answer "What PPE is needed for Substance X?" Challenges include curating accurate relationships and integrating legacy data.

Labeling (Data Annotation) – The process of assigning meaningful tags or categories to raw data, such as marking hazardous objects in images. Essential for supervised learning. Related terms: ground truth, annotation tool, inter-annotator agreement. Example: Safety analysts label video frames showing blocked fire exits, creating a training set for a detection model. Challenges include the time-intensive nature of labeling, ensuring consistency, and protecting privacy when annotating worker images.

Latent Variable Model – A statistical model that infers hidden (latent) factors influencing observed data. In OHS, latent variables may represent underlying safety culture or fatigue levels not directly measured. Related terms: factor analysis, probabilistic graphical model, hidden Markov model. Example: A latent variable model extracts a "safety climate" factor from survey responses, which then predicts incident rates. Challenges involve validating that latent constructs truly reflect real-world phenomena.

Learning Rate – A hyper-parameter that determines the step size at each iteration while optimizing an AI model's weights. Influences convergence speed and stability. Related terms: gradient descent, optimizer, epoch. Example: Setting a learning rate of 0.01 For a neural network training on vibration data yields steady loss reduction, whereas 0.5 Causes divergence. Challenges include selecting a rate that balances speed with accuracy and adapting it during training (e.G., Learning-rate schedules).

Lightweight Model – An AI model designed to run efficiently on devices with limited computational resources, often using model compression techniques such as pruning or quantization. Useful for edge deployment in hazardous environments. Related terms: model compression, TensorRT, mobile inference. Example: A pruned convolutional network runs on a handheld scanner to detect missing safety guards without cloud latency. Challenges include maintaining accuracy after compression and ensuring robustness to noisy inputs.

Logistic Regression – A statistical classification method that models the probability of a binary outcome using a logistic function. Frequently used for baseline risk prediction. Related terms: odds ratio, maximum likelihood, regularization. Example: Logistic regression estimates the likelihood of a fall based on age, floor slip coefficient, and footwear type, providing interpretable coefficients for each factor. Challenges include limited ability to capture non-linear interactions and sensitivity to multicollinearity among predictors.

Long Short-Term Memory (LSTM) – A type of recurrent neural network architecture capable of learning long-range dependencies in sequential data. Applied to time-series safety data such as vibration trends or fatigue monitoring. Related terms: gate mechanisms, sequence modeling, back-propagation through time. Example: An LSTM predicts imminent equipment failure by analyzing 30 days of temperature and pressure readings, alerting maintenance before a breakdown. Challenges involve large training data requirements and difficulty in interpreting temporal attention.

Machine Learning (ML) – A subset of AI that enables computers to learn patterns from data without explicit programming. Includes supervised, unsupervised, and reinforcement learning approaches. Related terms: training set, model evaluation, feature space. Example: An ML classifier differentiates between normal and abnormal noise levels in a manufacturing plant, triggering alerts when thresholds are crossed. Challenges include data bias, model drift over time, and ensuring compliance with occupational health regulations.

Model Drift – The degradation of an AI model's performance because the underlying data distribution changes over time. In OHS, new equipment or processes can cause drift. Related terms: concept drift, monitoring, retraining. Example: A risk scoring model trained on legacy machinery underestimates hazards after a plant upgrade, indicating drift that prompts a retraining cycle. Challenges involve detecting drift early and establishing automated retraining pipelines without disrupting operations.

Monte Carlo Simulation – A computational technique that uses random sampling to estimate the probability distribution of outcomes, often for risk quantification. In hazard assessment, it models uncertainty in exposure levels and failure rates. Related terms: probabilistic analysis, stochastic modeling, variance reduction. Example: Monte Carlo runs generate a distribution of possible fire spread scenarios based on variable wind speed, informing emergency response planning. Challenges include selecting appropriate input distributions and ensuring sufficient simulation iterations for convergence.

Natural Language Processing (NLP) – AI methods for analyzing, understanding, and generating human language. Enables extraction of safety insights from free-text incident reports, manuals, and regulations. Related terms: named entity recognition, sentiment analysis, transformer model. Example: An NLP pipeline identifies mentions of "chemical splash" in incident narratives, automatically tagging them for further review. Challenges include handling domain-specific terminology, misspellings, and maintaining confidentiality of personal data.

Neural Network Architecture – The design of layers, connections, and activation functions that define how a neural network processes information. Choices affect model capacity and suitability for specific OHS tasks. Related terms: feed-forward, convolutional, recurrent. Example: A convolutional architecture processes thermal images to locate hot spots on equipment, while a feed-forward network predicts injury probability from tabular data. Challenges involve selecting the right architecture for the data modality and avoiding over-parameterization.

Normalization – Scaling numeric data to a standard range (e.G., 0–1) Or distribution (e.G., Zero mean, unit variance) to improve model training stability. Essential when combining sensor readings of different units. Related terms: standardization, min-max scaling, z-score. Example: Temperature (°C) and pressure (psi) readings are normalized before feeding into a risk classifier, ensuring neither dominates the learning process. Challenges include preserving interpretability after scaling and handling outliers that distort normalization parameters.

Outlier Detection – Identifying data points that deviate markedly from the majority, which may indicate errors or emerging hazards. Techniques include statistical thresholds, clustering, and isolation forests. Related terms: anomaly detection, robust statistics, influence point. Example: An isolation forest flags a sudden surge in airborne dust concentration as an outlier, prompting investigation of a possible ventilation failure. Challenges involve distinguishing true hazards from sensor noise and preventing excessive false alarms.

Parameter Tuning – The process of adjusting hyper-parameters of an AI model (e.G., Tree depth, regularization strength) to achieve optimal performance. Often performed via grid search, random search, or Bayesian optimization. Related terms: hyper-parameter, cross-validation, validation set. Example: Grid search identifies that a random forest with 200 trees and max depth 12 yields the best AUC for injury prediction. Challenges include computational cost and risk of over-fitting to validation data.

Probabilistic Risk Assessment (PRA) – A systematic methodology that quantifies the likelihood and consequences of adverse events using probability theory. AI can enhance PRA by supplying data-driven probability estimates. Related terms: fault tree analysis, event tree, Bayesian network. Example: An AI model estimates failure probabilities for each component in a chemical plant, feeding those values into a fault tree to compute overall plant risk. Challenges involve integrating AI uncertainty with traditional PRA confidence intervals.

Privacy-Preserving Machine Learning – Techniques that protect sensitive data while still enabling model training, such as federated learning, differential privacy, and homomorphic encryption. Critical when handling employee health or biometric data. Related terms: secure aggregation, noise injection, data anonymization. Example: A federated learning framework trains a fatigue prediction model across multiple sites without transferring raw sensor data, preserving worker privacy. Challenges include communication overhead, maintaining model accuracy, and complying with regional privacy legislation.

Predictive Maintenance – The use of AI to forecast equipment failures before they occur, allowing scheduled repairs that reduce downtime and associated hazards. Related terms: remaining useful life, condition monitoring, prognostics. Example: Vibration analysis combined with a gradient boosting model predicts bearing failure 48 hours in advance, prompting pre-emptive replacement and avoiding a potential mechanical injury. Challenges include sensor reliability, model interpretability for maintenance crews, and integration with existing maintenance management systems.

Probabilistic Graphical Model – A framework that represents random variables and their conditional dependencies via graphs, such as Bayesian networks or Markov random fields. Enables reasoning under uncertainty in safety contexts. Related terms: nodes, edges, inference algorithm. Example: A Bayesian network models the relationship between equipment age, maintenance frequency, and incident probability, allowing scenario analysis. Challenges include eliciting accurate conditional probabilities and computational complexity for large networks.

Quality Assurance (QA) for AI – Systematic processes to ensure AI models meet performance, reliability, and safety standards before deployment. Includes testing, validation, and documentation. Related terms: model audit, verification, validation. Example: QA procedures involve stress-testing a hazard detection model against simulated extreme conditions to confirm robustness. Challenges involve defining appropriate acceptance criteria for safety-critical applications and maintaining traceability of model changes.

Quantile Regression – A statistical technique that estimates conditional quantiles of a response variable, providing a more complete view of outcome distribution than mean-based methods. Useful for assessing worst-case risk scenarios. Related terms: prediction interval, asymmetric loss, tau. Example: Quantile regression predicts the 95th percentile of exposure levels for a volatile chemical, informing the design of protective ventilation systems. Challenges include selecting appropriate quantile levels and ensuring sufficient data coverage in the tails of the distribution.

Random Forest – An ensemble learning method that builds multiple decision trees on random subsets of data and aggregates their predictions, improving accuracy and reducing over-fitting. Related terms: bagging, feature importance, out-of-bag error. Example: A random forest model predicts injury severity using variables such as task type, PPE compliance, and environmental conditions, delivering higher AUC than a single decision tree. Challenges include interpretability of many trees and increased memory usage for large forests.

Reinforcement Learning (RL) – A learning paradigm where an agent interacts with an environment, taking actions to maximize cumulative reward. In OHS, RL can optimize safety protocols by simulating interventions and learning which actions reduce risk most effectively. Related terms: policy, reward shaping, exploration. Example: An RL agent learns to schedule maintenance checks in a way that minimizes downtime while keeping equipment failure probability below a safety threshold. Challenges involve defining safe reward structures, avoiding risky exploratory actions, and ensuring the simulated environment reflects real-world dynamics.

Risk Matrix – A visual tool that plots risk severity against likelihood, helping prioritize hazards for mitigation. AI can automate the placement of hazards within the matrix based on data-driven scores. Related terms: risk priority number, color coding, decision grid. Example: An AI system assigns a likelihood score of 0.7 And severity score of 4 to a chemical spill scenario, placing it in the “high” region of the matrix. Challenges include subjectivity in severity definitions and potential oversimplification of multi-dimensional risk factors.

Safety Culture Assessment – Evaluation of organizational attitudes, values, and behaviors related to safety. AI can analyze survey text, communication logs, and incident trends to gauge cultural health. Related terms: sentiment analysis, climate survey, leading indicators. Example: NLP sentiment scores from internal safety newsletters reveal a decline in positive safety language, prompting leadership to reinforce safety messaging. Challenges include capturing nuanced cultural cues and avoiding misinterpretation of linguistic sarcasm.

Scenario Analysis – The systematic examination of possible future events by modeling different combinations of variables. AI can generate numerous plausible scenarios rapidly. Related terms: what-if analysis, Monte Carlo, sensitivity analysis. Example: An AI tool simulates fire spread under varying wind speeds, fuel loads, and suppression response times, providing a range of outcomes for emergency planning. Challenges involve selecting realistic parameter ranges and communicating probabilistic results to non-technical stakeholders.

Semantic Segmentation – A computer vision technique that classifies each pixel in an image into a category, enabling precise identification of hazardous objects or zones. Related terms: pixel-wise labeling, U-Net, fully convolutional network. Example: Semantic segmentation maps a live feed of a construction site, highlighting “open trenches” in red, allowing real-time exclusion zones for workers. Challenges include the need for extensive pixel-level annotations and computational demands for high-resolution imagery.

Sensor Fusion – The integration of data from multiple sensors to produce more accurate, reliable, and comprehensive information than any single sensor could provide. Related terms: Kalman filter, data aggregation, multimodal sensing. Example: Combining temperature, humidity, and gas concentration sensors yields a robust index of fire risk, which feeds an AI model that predicts ignition probability. Challenges involve synchronizing disparate data rates, handling sensor failures, and calibrating weighting schemes.

Shapley Additive Explanations (SHAP) – A game-theoretic method for interpreting individual predictions by assigning each feature an importance value. Provides transparent explanations for complex models. Related terms: feature attribution, local interpretability, model-agnostic. Example: SHAP values reveal that “lack of lockout-tagout” contributed 0.35 To a high-risk score for a maintenance task, guiding targeted training. Challenges include computational overhead for large datasets and communicating probabilistic contributions to end users.

Supervised Learning – A machine-learning paradigm where models are trained on labeled data, learning a mapping from inputs to known outputs. Core for classification and regression tasks in hazard detection. Related terms: training label, loss function, over-fitting. Example: A supervised model learns to classify images of safety signs as “compliant” or “non-compliant” using a labeled dataset of 10 000 images. Challenges include acquiring high-quality labels and ensuring the model generalizes to new sign designs.

Support Vector Machine (SVM) – A classification algorithm that finds the hyperplane maximizing the margin between classes, effective in high-dimensional spaces. Related terms: kernel trick, soft margin, support vectors. Example: An SVM separates safe versus unsafe operating conditions based on vibration frequency and amplitude, achieving high precision with limited training data. Challenges include selecting appropriate kernels and scaling to large datasets.

Temporal Drift – Changes in data patterns over time due to evolving processes, regulations, or equipment, leading to reduced model relevance. Related terms: concept drift, time-series analysis, model monitoring. Example: After a new PPE policy is introduced, the distribution of injury types shifts, causing a previously accurate model to misclassify certain incidents. Challenges involve detecting drift promptly and establishing retraining pipelines without interrupting safety operations.

Transfer Learning – Leveraging a pre-trained model on a related task to improve performance on a target task with limited data. Reduces training time and data requirements. Related terms: fine-tuning, domain adaptation, pretrained weights. Example: A model trained on general industrial images is fine-tuned on a specific plant’s safety camera footage to detect blocked exits, achieving high accuracy with only a few hundred labeled frames. Challenges include negative transfer when source and target domains differ significantly.

Uncertainty Quantification (UQ) – The process of characterizing and reporting the confidence or uncertainty associated with model predictions. Essential for risk-based decision making. Related terms: confidence interval, Bayesian inference, Monte Carlo dropout. Example: An AI risk score is presented with a 95 %

confidence interval, informing managers about the reliability of the prediction before allocating resources. Challenges include computational cost of generating uncertainty estimates and communicating them effectively to non-technical audiences.

Unsupervised Learning – Techniques that identify patterns or structure in data without labeled outcomes, such as clustering or dimensionality reduction. Useful for discovering unknown hazard groupings. Related terms: k-means, hierarchical clustering, autoencoder. Example: An unsupervised clustering algorithm groups incident reports into latent categories, revealing a previously unnoticed cluster of ergonomic injuries linked to a specific workstation layout. Challenges include interpreting clusters meaningfully and selecting appropriate distance metrics.

Validation Set – A subset of data used during model development to tune hyper-parameters and prevent over-fitting before final testing on a separate test set. Related terms: hold-out, cross-validation, model selection. Example: After training on 70 % of incident data, the model's hyper-parameters are optimized on a 15 % validation set, ensuring balanced performance before final evaluation. Challenges involve ensuring the validation set remains representative of future data distributions.

Variable Importance – Metrics that rank input features based on their contribution to a model's predictive power. Guides OHS professionals in focusing on key risk drivers. Related terms: feature importance, permutation importance, Gini importance. Example: Variable importance analysis shows that "equipment age" and "maintenance frequency" are the top contributors to machinery-related injury predictions, prompting targeted inspection schedules. Challenges include bias toward variables with more levels or higher cardinality and interpreting interactions among variables.

Vision Transformer (ViT) – A deep learning architecture that applies transformer concepts to image data, achieving competitive performance with reduced convolutional layers. Related terms: self-attention, patch embedding, positional encoding. Example: A ViT model processes high-resolution thermal images to detect overheating components, providing precise bounding boxes for maintenance crews. Challenges include large data requirements for pre-training and higher memory consumption during inference.

Weighted Loss Function – A loss function that assigns different penalties to errors on various classes, useful for imbalanced OHS datasets where hazardous events are rare. Related terms: class weighting, focal loss, cost-sensitive learning. Example: In a safety incident classifier, a weighted cross-entropy loss penalizes false negatives (missed hazards) more heavily than false positives, improving recall for rare accident classes. Challenges involve selecting appropriate weight ratios and avoiding over-compensation that harms overall model balance.

Zero-Shot Learning – The ability of a model to recognize classes it has never seen during training by leveraging semantic relationships. In safety, it can identify novel hazard types based on textual descriptions. Related terms: semantic embedding, attribute learning, transfer learning. Example: A zero-shot model identifies a newly introduced chemical as hazardous by linking its safety data sheet to known toxicant categories, even though the model was never trained on that specific substance. Challenges include ensuring reliable semantic mappings and managing uncertainty when extrapolating to truly unknown hazards.