



*An Online CPD Course
brought to you by
CEDengineering.ca*

Predictive Maintenance & Reliability Engineering With AI

Course No: B02-022

Credit: 2 PDH

Brian Lisiewski, P.E.



Continuing Education and Development, Inc.

P: (877) 322-5800

info@cedengineering.ca

Table of Contents

Module 1: Introduction — From Reactive to Predictive, with Safety in Mind	1
1.1 Maintenance Strategies Evolution	1
1.2 Public Safety and Reliability in PdM	2
1.3 Embracing AI Carefully	2
Module 2: Failure Modes, Data Sources & Condition Monitoring Fundamentals	4
2.1 Asset Failure Modes Across Disciplines.....	4
2.2 Condition Monitoring Techniques and Sensor Data	5
2.3 Baselines and Normal Behavior	6
Module 3: Feature Engineering & Anomaly vs. Fault Classification	7
3.1 Feature Engineering for Condition Monitoring Data	7
3.2 Anomaly Detection (Unsupervised) vs. Fault Classification (Supervised)	8
3.3 Example — Bearing Fault Feature and Fault Classification	8
3.4 Avoiding "Black Box" Pitfalls	9
Module 4: Reliability Metrics, RUL Estimation & Validating AI Alarms	10
4.1 Key Reliability Metrics.....	10
4.2 Remaining Useful Life (RUL) Estimation	11
4.3 Validating AI Alarms and Predictions.....	11
4.4 Safety Considerations in Alarm Response.....	12
Module 5: Maintenance Decision Thresholds, Optimization & Economic Trade-offs	13
5.1 Maintenance Thresholds — When to Trigger a Work Order	13
5.2 Basic Maintenance Optimization Trade-off.....	14
5.3 Optimization Using AI Data	14
5.4 Real-World Application — Example Decision.....	15
5.5 Continuous vs. Discrete Maintenance Strategies.....	15
Module 6: Risk-Based Prioritization — Decision Matrices & FMEA	16
6.1 Risk = Likelihood × Consequence.....	16
6.2 Using Decision Matrices.....	16
6.3 Failure Modes and Effects Analysis (FMEA).....	17
6.4 AI Enhancing FMEA and Risk Prioritization	18

Module 7: Communicating Maintenance Insights, Documentation & Professional

Accountability.....	19
7.1 Communication to Stakeholders.....	19
7.2 Documentation and Recordkeeping.....	20
7.3 Professional Defensibility and Ethics.....	20
7.4 Audit Readiness	21
7.5 Continual Improvement	21
References	23

Module 1: Introduction — From Reactive to Predictive, with Safety in Mind

*"If it ain't broke, don't fix it," once used to justify running equipment to failure, is no longer acceptable for critical systems. Unexpected breakdowns are disruptive, dangerous, and expensive.¹ For decades, maintenance evolved from **reactive** (fix when it breaks) to **preventive** (schedule periodic fixes) to now **predictive** (use data to predict and prevent failures).² This module establishes the context for predictive maintenance (PdM): why it is needed, how it differs from past practices, and how it ties to **public safety and reliability**. PEs must implement AI-driven maintenance ethically: validating AI recommendations, ensuring safe operations, and always retaining ultimate **responsibility for decisions**.*

1.1 Maintenance Strategies Evolution

- **Reactive Maintenance (Run-to-Failure):** The oldest approach — fix or replace equipment only after it fails. The **advantage** is minimal upfront cost, but **drawbacks** include high unplanned downtime, costly collateral damage from failures, and potential safety hazards if equipment fails catastrophically.³ Historically common when labor was cheap and machinery resilience low, this approach is now generally avoided for critical systems.
- **Preventive Maintenance (Time-Based):** Scheduled maintenance performed at fixed intervals (e.g., every X operating hours). This reduced random failures but led to **over-maintenance** (replacing parts that still have useful life) and could still miss failures occurring between intervals. Because preventive maintenance is based on statistical averages, it may not hold under unusual stresses — for example, a motor may fail earlier than its scheduled interval under sustained overload.⁴
- **Predictive Maintenance (Condition-Based / AI-Driven):** The modern approach uses continuous **condition monitoring** data to anticipate when maintenance is needed.⁵ Rather than following a fixed schedule, decisions are data-driven: waiting as long as safely possible to maximize component life, while intervening just before failure. AI tools analyze sensor data in real time to detect subtle signs of degradation that humans might miss. This approach aims to deliver the right maintenance at the right time — reducing cost and downtime while improving safety.

AI assists, PEs decide. AI does not replace engineer judgment: even with advanced predictive algorithms, maintenance decisions must be approved by licensed engineers who consider safety and reliability. AI can flag potential issues, but the PE remains accountable for verifying those leads and planning safe, effective maintenance actions.

Smarter maintenance = big benefits. AI-driven predictive maintenance yields tangible ROI and safety gains: upgrading from time-based to predictive strategies has cut maintenance costs by about 25–30%, slashed unplanned downtime by 35–45%, and improved safety by preventing catastrophic failures. These improvements boost asset reliability and production uptime while protecting personnel.

1.2 Public Safety and Reliability in PdM

Maintenance is not purely a cost concern — it is a major safety issue. **Equipment failures can endanger workers and the public**, especially in critical infrastructure (power grids, aircraft, vehicles, refineries). For example, a bearing seizure in a high-speed turbine could lead to a catastrophic explosion. Predictive maintenance helps mitigate these risks by identifying problems while they are still minor and scheduling repairs before they escalate. Additionally, by reducing unplanned outages, PdM improves reliability of essential services (such as power and water supply), directly benefiting public safety and quality of life.⁶

1.3 Embracing AI Carefully

In predictive maintenance, **AI is used as a tool** to process enormous amounts of sensor data — far more than any person could analyze manually. AI/ML algorithms can find hidden patterns (such as subtle changes in vibration or temperature) that indicate emerging faults. However, regardless of how accurate AI becomes, the **Professional Engineer must oversee** how its outputs are used for maintenance decisions. This means:

- Setting **clear protocols** for how AI-driven alerts are handled — including identifying who is responsible for responding, and verifying alerts with physical inspection.
- The engineer should not defer critical maintenance solely on the basis of an AI recommendation unless they concur that it is safe to do so. The licensed engineer retains ultimate liability for safety.
- Understanding the **limitations and assumptions** of the AI model — for example: What data was it trained on? What failure modes can it predict, and which can it not detect? — and communicating these limitations to stakeholders.

This approach aligns with **quality management practices (ISO 9001)**, where continuous improvement and risk management are central. Engineers must continuously verify that AI predictions are improving maintenance outcomes and adjust thresholds or models as needed. Most importantly, **public safety remains paramount** — predictive maintenance should decrease risk, not inadvertently increase it by fostering complacency or allowing warnings to be missed.⁷

Module 1 Summary

AI-based predictive maintenance shifts maintenance from a calendar-driven chore to a risk-managed process. By anticipating failures, organizations can reduce cost and downtime while enhancing safety.⁸ However, the PE must integrate AI cautiously — human expertise is required to validate and act on AI outputs ethically and effectively. Engineers must incorporate these tools in a way that reinforces reliability and safety, not undermines them.

Knowledge Check — Module 1

1. What is the main difference between preventive maintenance and predictive maintenance?
2. Why does reactive "run-to-failure" maintenance often result in higher costs and risks in the long run?
3. **True/False:** In an AI-driven maintenance program, the engineer's role in approving and overseeing maintenance decisions becomes less important.

Module 2: Failure Modes, Data Sources & Condition Monitoring Fundamentals

Effective predictive maintenance starts with understanding **how assets fail** and **what signals indicate impending failure**. This module examines common **failure modes** for various equipment types and the **condition monitoring** methods used to detect them. Covered topics include typical sensors and data sources — such as vibration, oil analysis, thermography, electric current, and process parameters — and how they apply across mechanical, electrical, civil, and environmental systems.

2.1 Asset Failure Modes Across Disciplines

Different engineering systems have different failure modes:

- **Mechanical equipment** (rotating or moving parts) often fails through wear, fatigue, corrosion, lubrication breakdown, vibration-induced damage, imbalance, misalignment, or overheating. For example, a **pump** might suffer bearing wear (leading to vibration and temperature rise), or a **compressor** might experience gradually declining efficiency due to blade fouling.
- **Electrical equipment** (motors, generators, transformers, switchgear) fails via insulation breakdown (leading to overheating, partial discharge, or arcing), bearing failures, or connection faults. Induction motors can develop rotor bar defects or stator winding short circuits, often detected via changes in current patterns and heat.
- **Civil infrastructure** often incorporates large mechanical components or moving parts (e.g., sluice gates, drawbridges, roller bearings on bridges, pumping stations in water and wastewater systems). Failure modes can be structural (cracking, fatigue in a bridge girder) or mechanical (gear wear in a lift bridge motor). For fixed structures, cracks may propagate slowly and can be detected by strain gauges; moving parts can be monitored like any mechanical asset.
- **Environmental systems** (water and wastewater plants) include pumps, blowers, valves, and mixers — mechanical devices subject to erosion, cavitation, and wear — as well as process performance dependencies (temperature, pressure, chemical levels). Failures can lead to regulatory non-compliance, such as an aeration blower failure causing treatment degradation.

For each asset type, a key part of reliability engineering is identifying its **failure modes and effects** (through FMEA or RCM analysis). Knowing the likely ways an asset can fail and the consequences helps target what to monitor and at what frequency.

2.2 Condition Monitoring Techniques and Sensor Data

AI-driven predictive maintenance relies on data from **sensors and inspections** to gauge equipment health in real time.⁹ Common condition monitoring methods include:

- **Vibration Analysis:** Accelerometers on rotating machinery capture vibration signatures. Healthy machines exhibit a characteristic vibration pattern; developing faults such as **imbalance, misalignment, bearing wear, or looseness** cause distinctive changes in amplitude or frequency content.¹⁰ For instance, an increasing RMS vibration level over time often precedes bearing failure.
- **Oil or Lubricant Analysis:** Periodic sampling of oil to check for contaminants (metal particles indicating wear, water indicating leaks, acidity indicating degradation). Results are trended over time — for example, a rapid rise in iron content signals gear wear and impending gear tooth failure.
- **Thermography (Infrared Imaging):** Detects hotspots in electrical and mechanical equipment. For electrical systems, a rising temperature at a connection may indicate a loose or corroded contact causing high resistance (a potential fire hazard). For mechanical systems, a hot bearing or motor casing is a red flag for friction or electrical inefficiency.
- **Electrical Signature Analysis:** Monitoring motor current and voltage for anomalies such as increased current draw or harmonic distortions. This can reveal motor winding degradation or misalignment (which forces the motor to work harder, increasing current draw).
- **Process Performance Indicators:** In process industries or utility operations, sensors for pressure, flow, temperature, and pipeline vibration can reveal developing problems. For example, increasing pressure drop across a filter indicates clogging, while fluctuating pump discharge pressure may indicate cavitation or control issues.

All these data streams are part of the **Internet of Things (IoT)** environment on modern equipment, feeding into central systems or edge devices. **Data quality is crucial** — sensors must be calibrated and maintained, or AI analysis may be compromised by sensor drift or noise. As ISO 9001 emphasizes, maintaining monitoring equipment is as important as analyzing the data it produces.

2.3 Baselines and Normal Behavior

A hallmark of condition monitoring is establishing what "healthy" operation looks like for each piece of equipment. This means collecting data during normal conditions to form a **baseline** — the AI system's reference for normal behavior.¹¹ For instance, a pump might normally vibrate at 0.1 in/s RMS with certain spectral characteristics. The baseline must account for variations such as load changes or temperature, and multiple operating states may each have their own baseline. Key considerations include:

- **Normalization:** Sensor readings are often normalized against operating conditions. For example, separate baselines should be established at different speeds or loads for a variable-speed pump. This prevents false alarms caused by normal speed-related changes in vibration level.
- **Trend Collection:** Over time, trending data helps distinguish normal drift from abnormal spikes. Some variables naturally change gradually (oil wear metals tend to increase over time — the rate of increase matters more than the absolute value), while others should remain stable (temperature should stay below defined thresholds).
- **Cross-Correlation:** Many systems use multiple indicators in combination. If both vibration and temperature of a bearing are rising simultaneously, that provides stronger evidence of an emerging fault than a single sensor reading alone. AI models ingesting multiple sensor types can build a holistic baseline and reduce false positives by requiring confirmation across data sources.

Module 2 Summary

Effective predictive maintenance combines domain knowledge of failure modes with the right sensors and baseline data. Mechanical, electrical, civil, and environmental systems each produce distinctive signatures when something goes wrong. By deploying appropriate measurement techniques, engineers gather the data AI needs to function. Accurate baselines and quality data are the foundation — a robust baseline enables AI to detect subtle deviations early without being misled by normal operating variations.¹²

Knowledge Check — Module 2

4. Name two condition monitoring techniques and describe what each detects.
5. Why is establishing a "normal" baseline important for AI in predictive maintenance?
6. What type of sensor or monitoring method would most likely detect a loose electrical connection that is overheating?

Module 3: Feature Engineering & Anomaly vs. Fault Classification

AI models do not work directly with raw sensor data — they rely on **feature engineering** to extract meaningful indicators of equipment health. This module explains how engineers derive features from condition monitoring data (such as statistical or frequency features from vibration signals, or trend rates from oil analysis). It then discusses how AI uses these features to detect anomalies and, where possible, classify specific faults. Understanding this step reveals **how AI "thinks"** in a maintenance context and what it can — and cannot — tell you about equipment health.

3.1 Feature Engineering for Condition Monitoring Data

Features are quantifiable characteristics derived from raw data that correlate with asset health. Examples include:

- **Time-domain features:** RMS (root-mean-square) vibration velocity over a period (indicates overall vibration energy), peak amplitude, and statistical measures such as kurtosis (which reveals spikiness from impacts). In lubrication monitoring, the **rate of change** of a contaminant concentration is a useful feature — a rapidly increasing particle count strongly suggests accelerating wear.
- **Frequency-domain features:** Spectral analysis (FFT) of vibration or acoustic data identifies signature frequencies. Bearing defect frequencies are calculated from bearing geometry and running speed, and they typically appear at **non-integer multiples of shaft rotational frequency** (for example, the ball pass frequency outer race, or BPFO, is determined by the number of rolling elements, pitch circle diameter, ball diameter, and contact angle). If an amplitude spike appears at a known bearing defect frequency, that feature directly suggests a developing fault at that bearing element.
- **Model-based features:** Differences between measured performance and expected (physics-model) performance can be powerful features. For example, the difference between a pump's actual flow and the theoretical flow from its pump curve may indicate wear or impeller fouling.
- **Environment and operation context:** Features can also be contextual — for example, number of start-stop cycles, hours in service since last maintenance, or load factor. These help AI models estimate cumulative wear and tear.

PEs must ensure AI models are fed **physically relevant features**. If important indicators are omitted, the model may latch onto spurious correlations. For example, if an AI monitors only motor current but not temperature, it might misinterpret increased current (which could result from a heavier load) as a fault. With temperature included as a feature, the model can distinguish heavy load (higher current, normal temperature) from a true fault (higher current plus rising temperature and vibration).

3.2 Anomaly Detection (Unsupervised) vs. Fault Classification (Supervised)

AI in predictive maintenance typically employs two broad types of ML model:¹³

- **Anomaly Detection** — typically unsupervised learning. The model learns what "normal" looks like and identifies any significant deviation as an anomaly.¹⁴ This is useful when labeled failure data are unavailable for training. The output is typically an **anomaly score** or a simple alert that "something is off." Anomaly detection can catch novel or unexpected issues; however, it may produce **false positives** for benign deviations if not well tuned, and it does not directly identify the cause — only that behavior is not normal.
- **Fault Classification (Diagnosis)** — supervised learning using labeled examples of known fault conditions. If historical data with confirmed faults exist (e.g., known cases of bearing failure vs. misalignment vs. imbalance), an AI classifier can be trained to distinguish between them. For example, a decision tree might learn: "high vibration at $1\times$ running speed frequency $\rightarrow$ likely imbalance; high-frequency noise in a certain spectral band $\rightarrow$ likely bearing defect." These models **identify the type of fault** rather than simply raising a flag. However, they require a library of labeled historical fault data, which may not exist for rare or novel failure modes.

In practice, many PdM systems combine both approaches: unsupervised anomaly detection raises a flag when something unusual is detected, and then — if possible — known pattern matching or classification rules suggest what the issue may be. The **engineer or specialist** then interprets the findings, confirming whether the anomaly matches a known failure mode or planning a direct inspection when the pattern is unclear.

3.3 Example — Bearing Fault Feature and Fault Classification

A centrifugal pump in a water treatment plant has vibration sensors on its motor and pump bearings. The AI system has learned the normal vibration spectrum. It flags an anomaly: increased vibration energy at approximately $5.4\times$ (a non-integer multiple characteristic of rolling element bearing defects) the shaft rotational frequency. The reliability engineer recognizes that for rolling element bearings, high-frequency excitation at a non-integer multiple of running speed is characteristic of a developing crack on a race or rolling element. If the system includes a supervised model trained on historical fault data for similar pumps, it might classify the pattern directly as "bearing inner race defect" with a confidence score.

The maintenance team schedules a bearing replacement during the next planned outage. On removal, the old bearing shows confirmed spalling — validating the AI indication. This outcome builds confidence in the system and generates labeled data to improve the ML model further, closing the improvement loop.

3.4 Avoiding "Black Box" Pitfalls

While ML can reveal complex patterns, there is a risk that a model may respond to irrelevant features. Engineers can mitigate this by applying **explainable AI (XAI)** techniques — tools that highlight which features drove a particular anomaly alert — allowing engineers to verify that the AI is responding to physically meaningful signals. Physics-based or rule-based overlays (such as known alarm thresholds derived from standards) can further constrain AI recommendations within reasonable bounds. This validation process is part of the continuous improvement cycle described in Modules 4 and 7.

Module 3 Summary

Features transform raw sensor data into actionable information, capturing patterns — such as trending vibration levels or spectral frequency spikes — that correlate with faults. AI models leverage these features to detect anomalies and, where possible, diagnose specific fault types.¹⁵ Anomaly detection flags anything unusual; fault classification provides deeper diagnostic insight but requires labeled training data. Engineers must ensure features are physically relevant and interpret AI outputs critically, confirming faults through inspection or additional analysis.

Knowledge Check — Module 3

7. What is an example of a feature that could indicate a developing fault in a rotating machine?
8. How does anomaly detection differ from fault classification in predictive maintenance AI?
9. Why is it important for an engineer to review the features or factors that an AI model is using to flag a machine condition?

Module 4: Reliability Metrics, RUL Estimation & Validating AI Alarms

Maintenance decisions are guided by reliability metrics — numerical indicators of how likely and how soon failures may occur. This module covers fundamental **reliability metrics** and how AI tools estimate them, particularly **Remaining Useful Life (RUL)**. It then addresses how to interpret these metrics and the importance of **validating AI-generated alerts**. Engineers should treat AI-driven alarms like any other condition report: verify their accuracy before taking costly action, finding the right balance between trusting the data and applying appropriate skepticism.

4.1 Key Reliability Metrics

- **MTBF (Mean Time Between Failures) and MTTF (Mean Time To Failure):** Average time an asset operates before failing — a basic measure of reliability. MTBF applies to repairable systems (time between repairs); MTTF applies to non-repairable components. For example, if a pump's historical MTBF is 5 years, one might expect a failure roughly every 5 years on average — but this does not mean a failure cannot occur earlier or later. It is a population average, not a guarantee for any individual asset.
- **Failure Rate and Hazard Function:** The failure rate (λ) is the frequency of failures per unit time for a population of items. The hazard function describes how the failure rate changes over time, as represented in the classic "bathtub curve," which has three distinct phases: an early-life phase of decreasing hazard (elevated failures from manufacturing defects, installation errors, or material flaws that surface quickly and then diminish); a useful-life phase of approximately constant, random hazard (the longest phase, during which failures occur unpredictably rather than from age-related wear); and a wear-out phase of increasing hazard (rising failure probability as components accumulate fatigue, corrosion, and mechanical degradation). Understanding which phase an asset currently occupies is essential to setting appropriate maintenance intervals: assets in the useful-life phase are poor candidates for age-based replacement because failure probability does not increase with time, whereas assets entering the wear-out phase are strong candidates for proactive replacement on a schedule tied to their degradation rate. AI condition monitoring contributes here by tracking real-time degradation indicators — such as rising vibration or increasing wear particle counts — that signal a transition into the wear-out phase earlier and more precisely than calendar age alone can. AI can help update perceived hazard rates by analyzing condition data — for example, a steadily increasing vibration level may indicate a rising effective hazard rate consistent with wear-out behavior.
- **Reliability $R(t)$ and Probability of Failure:** $R(t)$ is the probability that an asset survives to time t . The complementary probability — the probability of failure by a given time — is the risk that maintenance planning seeks to manage. An AI system performing RUL prediction is essentially estimating these probabilities or the expected time to a defined failure threshold.

4.2 Remaining Useful Life (RUL) Estimation

RUL is a forward-looking metric — how much time or usage remains before failure is likely. AI can estimate RUL using historical failure data and current condition indicators.¹⁶ For example: if past data show that once vibration exceeds a certain threshold, similar bearings typically fail within 200 hours (with some variance), an ML model might predict that a bearing currently at that vibration level has approximately 180 hours of life remaining, with an associated confidence interval.

Engineers should use RUL estimates as **guidance, not gospel**. Running equipment right up to the predicted failure moment is risky — RUL is a best estimate with inherent uncertainty. Good practice is to schedule maintenance while a comfortable margin remains. For example, if RUL is predicted to be approximately four weeks, planning maintenance in two weeks provides a meaningful safety margin. The appropriate margin correlates with equipment criticality — the more critical the system, the larger the buffer should be.

4.3 Validating AI Alarms and Predictions

When an AI system raises an alarm or forecasts a short RUL, the engineer's immediate question should be: *"Do I trust this? What is the evidence?"* Recommended validation steps include:

- **Cross-check sensor data:** If an AI alarms on a vibration anomaly, check related parameters — is bearing temperature also rising? Is there an audible change in noise? Multiple confirming signals increase confidence that the issue is real. If all other indicators are normal, a sensor malfunction generating a false positive becomes more likely.
- **Physical inspection:** For critical anomalies, a targeted inspection can be decisive. A technician listening to or visually examining a running pump may detect a subtle noise or observable leak confirming a developing problem — or may find nothing, suggesting the AI may be oversensitive or the issue is still at an early internal stage.
- **Use redundancy and comparison:** If parallel measurements exist (e.g., multiple sensors on similar equipment), compare them. If one motor's AI alert triggers while an identical unit does not show the same pattern, this supports the conclusion that one unit is genuinely abnormal. Conversely, if both units show the same anomaly simultaneously, an external common-cause factor (such as a power supply transient) may be responsible.
- **Tune thresholds and models through continuous improvement:** AI systems typically allow sensitivity adjustment. If false positives are frequent, recalibrate the alarm threshold — while being careful not to raise it so high that genuine faults are missed. If a fault was missed — that is, a false negative later identified through inspection or incident review — investigate why and refine the model. This continuous improvement cycle, consistent with ISO quality management principles,¹⁷ is essential to building a trustworthy system. As noted in industry practice, modern systems incorporate feedback from maintenance outcomes to reduce false alarms over time.¹⁸

4.4 Safety Considerations in Alarm Response

If AI predicts imminent failure of a critical or potentially dangerous system (such as a pressure vessel or overhead crane), the response should be escalated quickly — it may be prudent to shut it down for inspection if safety is at stake. For less critical equipment, close monitoring until the next planned downtime window may be appropriate. These decisions require situational judgment: not every alert demands emergency action, but when the consequences of a missed failure are severe, a conservative response is the professional standard.

Module 4 Summary

Reliability engineering metrics quantify failure likelihood over time and frame maintenance scheduling decisions. AI's RUL predictions enhance this by providing real-time estimates of how far along an asset is on its failure curve. Because no prediction is perfect, engineers must validate and interpret AI alarms carefully — verifying abnormal conditions through multiple lines of evidence and ensuring models remain calibrated to minimize both false positives and false negatives.¹⁹ This process enables PEs to integrate AI into maintenance decisions in a trustworthy and professionally defensible manner.

Knowledge Check — Module 4

10. What does RUL stand for, and what does it represent in predictive maintenance?
11. Why should an engineer not blindly trust a single AI alarm without performing cross-checks?
12. If an AI predicts a part has approximately 100 hours of useful life remaining, what considerations should go into deciding when to actually replace it?

Module 5: Maintenance Decision Thresholds, Optimization & Economic Trade-offs

With predictive analytics indicating equipment health, engineers face a practical question: *"When exactly do we perform maintenance?"* This module examines how to set **decision thresholds** for maintenance actions based on AI outputs and reliability goals. It considers the optimization challenge of balancing the cost of performing maintenance too early (lost useful life, unnecessary labor cost) against the cost of acting too late (failure and downtime costs). Simple approaches to deriving a maintenance schedule and incorporating economic reasoning are discussed.

5.1 Maintenance Thresholds — When to Trigger a Work Order

In condition-based maintenance systems, thresholds are set on monitored parameters to trigger maintenance actions. For example:

- **Vibration severity limits:** Standards such as ISO 20816 (which supersedes the earlier ISO 10816) classify vibration severity for rotating machinery in terms of velocity (mm/s RMS). An AI system can incorporate those classifications as initial thresholds — for example, if vibration enters Zone C (short-term operation only), schedule a maintenance check; if it reaches Zone D (damage likely), plan immediate shutdown for repair.
- **Oil contamination limits:** If particle count, water content, or acidity exceeds a predefined safe limit, trigger an oil change or targeted inspection.
- **Anomaly score thresholds:** When using unsupervised anomaly detection, a numerical score above a defined level — historically correlated with actual problems — triggers action. Initial thresholds may be set conservatively (to catch everything), then refined as confidence in the system grows, to reduce unnecessary alarms.

Thresholds are typically established by engineers and reliability specialists, based on manufacturer recommendations, applicable standards, or historical experience. The key principle is making thresholds **risk-informed**: the more critical the equipment, the more conservative (lower) the threshold should be, to ensure time for safe intervention. Non-critical equipment with minor failure consequences can tolerate higher thresholds or even run-to-failure strategies.

5.2 Basic Maintenance Optimization Trade-off

Consider a simplified scenario: replacing a part too soon wastes money (because useful life remained), while waiting too long risks a failure with far greater cost. A conceptual trade-off:

- **Cost of planned replacement (C_p):** The cost of replacing a part at a scheduled time — including parts, labor, and downtime within a planned window that minimizes production loss.
- **Cost of failure (C_e):** The cost if failure occurs unexpectedly — typically including higher parts cost, collateral damage, unplanned downtime, and potential safety impact.

A simplified decision rule: replace when the **expected cost of not replacing** equals or exceeds the cost of replacing. For example:

- If a bearing costs \$1,000 to replace and failure would cost \$20,000 (due to downtime and collateral damage), pushing it to the last possible moment is not prudent. If AI indicates risk of failure within two weeks, replacing it at the next convenient planned downtime — perhaps one week out — sacrifices some potential remaining life but avoids any chance of the \$20,000 failure cost.
- When the ratio of failure cost to replacement cost is very large (as in nuclear plants or aviation, where failures are catastrophic), maintenance is performed conservatively early. When the ratio is small and failure is tolerable, intervals may be stretched — an approach sometimes called **risk-based maintenance**.

5.3 Optimization Using AI Data

AI can refine these decisions by providing more accurate estimates of failure probability at a given time. For example, if the probability of failure in the next month is estimated at 5%, a risk-informed calculation might weigh $5\% \times (\text{failure cost})$ against the immediate cost of maintenance. If the expected failure cost ($\$1,000 = 5\% \times \$20,000$) is less than the cost of performing maintenance now (\$2,000), a brief additional wait may be justifiable — *provided* condition monitoring continues and the probability is updated continuously. This highlights the importance of ongoing monitoring, not just a single AI prediction.

Note that the \$2,000 maintenance cost here reflects a broader planned outage scenario rather than the bearing-only replacement cost referenced earlier; both examples illustrate the same principle applied at different scales.

5.4 Real-World Application — Example Decision

A municipal wastewater pump has shown increasing vibration. AI predicts an RUL of approximately six weeks. Replacing the pump requires a one-day controlled shutdown, but a backup pump is available. The maintenance team has a planned window in four weeks. They elect to continue daily monitoring; if any step-change worsening is detected (vibration or temperature exceeding the next threshold level), they will act immediately. Otherwise, they will replace the pump at the four-week window, retaining a two-week buffer before the estimated RUL expiration. This is an acceptable risk given the backup pump and moderate consequences. The decision logic is documented.

5.5 Continuous vs. Discrete Maintenance Strategies

AI often enables **continuous condition monitoring**, but maintenance itself is generally scheduled in discrete intervals (shifts, weekends, planned outages). Engineers should align these by labeling assets for service at the next outage if anomaly severity is moderate, or for immediate action if severity is high. **Maintenance backlogs** can be prioritized by condition severity, using tools such as the risk criticality matrix discussed in Module 6.

Module 5 Summary

Predictive maintenance is not just about predicting — it is about deciding when to act on a prediction. Engineers use thresholds on condition indicators to trigger actions, adjusted to asset criticality and reliability goals. Optimization seeks to minimize total cost and risk: acting neither so early as to waste useful life nor so late as to risk failures. AI enriches these decisions by quantifying conditions and failure probabilities, but human judgment is needed to incorporate real-world constraints such as maintenance windows, staffing, and safety requirements. The goal is the sweet spot where maintenance is efficiently timed to ensure safety and reliability at reasonable cost.

Knowledge Check — Module 5

13. Why might performing maintenance "too late" be more costly than performing it "too early"?
14. What factors would you consider when setting an alarm threshold for vibration on a critical rotating machine?
15. How can cost considerations influence the decision on when to replace a component that an AI model predicts will fail in the near future?

Module 6: Risk-Based Prioritization — Decision Matrices & FMEA

Not all maintenance issues are equally urgent or critical. Risk-based prioritization allows engineers to focus resources on the most important problems first. This module explores **decision matrices** and **Failure Modes and Effects Analysis (FMEA)** to prioritize predictive maintenance actions. These tools are tied to reliability-centered maintenance (RCM) principles, with AI insights (such as condition severity or failure probability) feeding into the risk assessment for each asset.

6.1 Risk = Likelihood × Consequence

Risk is expressed as the combination of:

- **Likelihood** — the probability of a failure or anomaly occurring, and
- **Consequence** — the impact if failure happens, encompassing safety, downtime, repair cost, and environmental or regulatory harm.

AI informs the **likelihood** side — providing a probability or indicator of how close failure may be. Consequence is inherent to the asset's role (a critical hospital pump has higher consequence of failure than a redundant backup pump). **Risk matrices** categorize issues into High, Medium, and Low risk based on these two dimensions. An asset with high failure probability and high consequence is High Risk and should be addressed immediately. An asset with mild anomalies (low near-term failure probability) and low consequence is lower priority.

6.2 Using Decision Matrices

A decision matrix lists assets or maintenance tasks and scores them on criteria such as:

- Current condition severity (from AI condition monitoring)
- Failure probability in the next defined period (if quantifiable)
- Consequence rating (e.g., a 1–5 scale of impact on safety, production, and environment)
- Production impact or regulatory compliance implications

The matrix may use weighted scoring to compute an overall priority, or rely on subject-matter review to rank items. The key is to incorporate AI's factual outputs (e.g., "vibration is in the danger zone" or "estimated 80% probability of failure within one month") alongside human judgment on consequences.

6.3 Failure Modes and Effects Analysis (FMEA)

FMEA is a systematic method to evaluate potential failure modes for an asset and prioritize them by risk. For each failure mode, three factors are assigned:

- **Severity (S):** How bad are the consequences if this failure mode occurs? (Scale of 1–10; 10 = worst, such as a safety hazard or major unplanned outage.)
- **Occurrence (O):** How likely is this failure mode to occur? (Scale of 1–10 based on data or expert input; 10 = very frequent.)
- **Detection (D):** How likely is this failure mode to go undetected before causing harm? (Scale of 1–10; a **lower** D value means existing monitoring will catch it early, reducing risk. A **higher** D value means the failure mode is likely to go undetected.)

Risk Priority Number (RPN) = $S \times O \times D$ (maximum 1,000). Maintenance resources should be focused on high-RPN failure modes.

The following table shows a simplified FMEA for a pump system:

Table 1. Simplified FMEA for a pump system.

Failure Mode	S (Severity)	O (Occurrence)	D (Detectability)	RPN	Mitigation / Action
Pump bearing failure	8 — high downtime & repair cost	6 — moderate, given age	3 — good vibration monitoring in place	144	Increase monitoring frequency; schedule bearing replacement at next convenient outage based on trend.
Shaft seal leak	5 — moderate; some downtime, minor env. impact	7 — likely over time	5 — visual inspection monthly	175	Use AI to monitor pump vibration for signs of seal friction; plan seal replacement in maintenance backlog.
Motor winding insulation failure	9 — critical; could cause fire & major outage	3 — low; motor in good condition	2 — IR thermography & electrical monitoring in place	54	Continue regular thermography; low priority for now.
Clogging of pump inlet	6 — downtime if clogged	4 — occasional with debris	4 — flow sensors monitor partially	96	Maintain debris screens; use AI anomaly detection on flow trends to pre-empt clogs.

RPN helps rank which failure modes warrant the most attention in maintenance planning. In the example above, the shaft seal leak has the highest RPN (175) despite a moderate severity rating, because its high occurrence and only adequate detection combine unfavorably. The motor winding insulation, while carrying a very high severity, has a low RPN (54) because occurrence is low and monitoring is strong — reflecting how good detection reduces overall risk.

6.4 AI Enhancing FMEA and Risk Prioritization

AI can assist by providing up-to-date Occurrence data (through real-time condition severity or measured failure probability) and improving Detection (through continuous monitoring, which lowers D values because issues are caught earlier). As shown in the example FMEA above, strong monitoring moves Detection scores down, which lowers RPN — one of the core values of predictive maintenance is precisely this reduction in detection-related risk.

Beyond technical risk, **business risk** should also be considered — a fault in a critical asset just before a major production deadline may carry higher consequence (penalties, lost revenue) than the same fault at a quiet period. Risk prioritization can therefore be time-sensitive, and an engineer ensures that AI's continuous asset health updates feed into an adaptable, current priority plan.

Module 6 Summary

Risk-based maintenance ensures the most critical issues are addressed first. By using risk matrices and FMEA, and integrating AI's condition and probability insights, PEs can logically rank maintenance actions. The result is a reliability strategy that targets high-risk failure modes for proactive intervention, allows benign or low-impact issues to be handled in routine schedules, and optimizes the use of limited maintenance resources. This approach reflects reliability-centered maintenance (RCM) thinking, now enhanced with real-time data.

Knowledge Check — Module 6

16. In an FMEA, what do the letters S, O, and D stand for?
17. How does continuous predictive monitoring (such as vibration sensors) typically change the "Detection" rating in an FMEA, and what effect does this have on RPN?
18. Which failure mode would typically receive a higher maintenance priority: one with a low probability of occurrence but a very high consequence, or one with a moderate probability and a moderate consequence?

Module 7: Communicating Maintenance Insights, Documentation & Professional Accountability

The final module addresses how PEs should **document and communicate** AI-driven maintenance programs. Topics include creating clear communications for management and operations (to gain buy-in for predictive maintenance actions), maintaining thorough records for audit and quality assurance, and upholding professional and ethical standards. These practices ensure that AI adoption in maintenance yields improvements without compromising transparency or compliance, and that engineers retain control of decisions that ultimately affect public safety and reliability.

7.1 Communication to Stakeholders

Transitioning to AI-based predictive maintenance often requires convincing both management and maintenance crews of its value. Key points to communicate:

- **Return on Investment (ROI):** Share concrete metrics — predicted reductions in downtime or maintenance cost savings. Industry data suggest that well-implemented predictive maintenance programs have achieved reductions of 25–40% in unplanned downtime and maintenance costs in various applications.²⁰ Real-world successes (such as early detection of a fault that prevented a costly breakdown) make compelling cases.
- **Safety and Reliability Benefits:** Emphasize that the AI system increases reliability and safety, not merely reduces cost. For example: "By catching faults early, we reduce the risk of catastrophic machine failures that endanger personnel."²¹ For public or critical infrastructure, improved reliability translates directly into better service to the community.
- **Process Integration:** Describe how AI alerts will be integrated into existing maintenance workflows — for example, via CMMS-generated work orders or real-time alerts to field staff.²² This reassures teams that the system is manageable and complements existing processes.

When presenting AI findings, keep communications clear and actionable. Rather than presenting raw data or complex model internals, provide **dashboard summaries** (such as traffic-light status indicators for asset health) and specific recommendations (e.g., "Replace Pump A within two weeks"). When asked, be prepared to explain the basis — have trend charts or underlying sensor data available to support recommendations.

7.2 Documentation and Recordkeeping

Adopting AI in predictive maintenance must be accompanied by rigorous documentation, consistent with ISO 9001's quality management emphasis on records and traceability. Good practices include:

- **Data logs and model records:** Maintain a database of sensor data history, anomalies detected, and actions taken. If an AI model is updated or retrained, log the date, the new data used, and any changes to thresholds. This ensures reproducibility and auditability.
- **Maintenance decisions and rationale:** For significant decisions (such as shutting down a critical system due to an AI warning), document why the decision was made, who approved it, and what the outcome was. For example: "On [date], AI predicted impending bearing failure on Compressor B; maintenance executed bearing replacement on [date] during controlled shutdown. Removed bearing confirmed spalling, validating prediction." Such records provide professional defensibility if decisions are questioned.
- **Alignment with standards:** Note compliance with applicable guidelines (e.g., ISO 17359 for condition monitoring, or RCM analysis conducted per SAE JA1011). Referencing established standards strengthens the credibility and defensibility of the program.

7.3 Professional Defensibility and Ethics

Under the NSPE Code of Ethics, engineers must hold public safety paramount. In the context of AI-driven maintenance, this means:

- Never allow AI recommendations to override safety codes or professional judgment in a way that jeopardizes safety. If AI suggests extending maintenance intervals in a manner that conflicts with applicable codes or accepted engineering practice, the engineer must override to maintain compliance.
- The engineer should understand — at least qualitatively — which factors an AI model uses in its analysis, and verify that these align with engineering principles. This guards against blind trust in a model with unseen biases or failure mode blind spots.
- Where limitations or scope restrictions exist (e.g., "this model was not trained on failure modes associated with process chemistry changes"), communicate these clearly to relevant personnel. Clarity about scope prevents false confidence in coverage the system does not actually provide.

7.4 Audit Readiness

In the event of a failure and subsequent investigation, an engineer who used AI in maintenance decisions should be able to demonstrate:

- The data being monitored at the time of the failure,
- The AI system's output prior to the failure,
- The actions taken or not taken, with documented rationale.

For example: "The AI indicated an anomaly on the bearing on [date]. We inspected and assessed the indication as minor at that time (inspection notes attached). We continued monitoring and scheduled maintenance for the following week. The bearing failed earlier than anticipated; given all information available at the time, our decision was consistent with standard engineering practice, and we have since adjusted the monitoring threshold for similar bearings." A documented decision trail is essential to professionally defend the process followed.

7.5 Continual Improvement

Building an AI-assisted maintenance program is iterative. Use **lessons learned from each event** — near-misses, false alarms, confirmed predictions, and missed detections — to refine thresholds, retrain models, and improve staff training. Document improvements as part of a formal reliability program. Many organizations hold quarterly reliability reviews covering all events from the preceding period; this practice builds maturity and aligns with the ISO Plan-Do-Check-Act (PDCA) improvement philosophy.

Module 7 Summary

Introduce AI into maintenance with the same rigor and transparency applied to any engineering process. Communicate the value — safety, cost, and reliability improvements — clearly, and document everything, from data to decisions to outcomes, to maintain trust and auditability. Always keep PE ethical obligations in focus: use AI as a tool to help make safe, effective decisions, but never cede professional responsibility. With proper documentation and communication, predictive maintenance becomes not just an internal engineering tool but a value proposition recognized across the organization, delivered in a professionally accountable manner.

Knowledge Check — Module 7

19. Why is thorough documentation of predictive maintenance actions important for later audits or investigations?
20. How should an engineer address the known limitations of an AI maintenance system when communicating with stakeholders?
21. What key professional and ethical obligation must always guide a PE's use of AI in maintenance planning?

References

The following sources were consulted in the preparation of this course. Superscript numbers in the course text correspond to the numbered references below.

22. IBM. (2024). *AI in predictive maintenance*. IBM Think Insights. Retrieved from <https://www.ibm.com/think/insights/ai-in-predictive-maintenance>
23. Cryotos. (2024). *Vibration analysis in predictive maintenance*. Cryotos Blog. Retrieved from <https://www.cryotos.com/blog/vibration-analysis-in-predictive-maintenance>
24. International Organization for Standardization. (2016). *ISO 20816: Mechanical vibration — Measurement and evaluation of machine vibration*. ISO. (Note: ISO 20816 supersedes the earlier ISO 10816 series.)
25. International Organization for Standardization. (2015). *ISO 9001:2015 — Quality management systems: Requirements*. ISO.
26. International Organization for Standardization. (2018). *ISO 17359:2018 — Condition monitoring and diagnostics of machines: General guidelines*. ISO.
27. SAE International. (2009). *SAE JA1011 — Evaluation criteria for reliability-centered maintenance (RCM) processes*. SAE International.
28. National Society of Professional Engineers (NSPE). (2019). *NSPE Code of Ethics for Engineers*. NSPE. Retrieved from <https://www.nspe.org/resources/ethics/code-ethics>
29. Moubray, J. (1997). *Reliability-centered maintenance* (2nd ed.). Industrial Press.
30. Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510. <https://doi.org/10.1016/j.ymssp.2005.09.012>
31. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799–834. <https://doi.org/10.1016/j.ymssp.2017.11.016>