

THOUGHT LEADERSHIP

Your AI Is Only as Good as the Units Your Sensors Speak

Why Lab-Equivalent Oil Data Is the Missing Input for Industrial AI



Industrial AI in maintenance is stalling on data, not algorithms. For lubricated assets, the fix is simple to state: measure oil condition inline in the same units the laboratory has always used.

Almost every large industrial company now runs an AI programme for maintenance. Predictive models, digital twins and, increasingly, generative AI copilots that read work orders, manuals and sensor data, then tell a planner what to do next.

The ambition is right. The bottleneck is the input. In a 2026 survey of industrial professionals, 54% named data quality and availability as the single biggest barrier to industrial AI, ahead of integration, skills or cost. Only 7% said AI was embedded in most of their core processes.

Oil condition is a good example of why. Lubricant chemistry is one of the earliest and most direct indicators of machine health: water ingress, oxidation, additive depletion and contamination precede most bearing, gear and engine failures. Yet the oil data most AI teams can get their hands on is either too sparse or in the wrong language.

Two kinds of oil data, neither built for AI

The first kind is the **laboratory report**. It is precise and standardised: oxidation in abs/cm, acid and base number in mg KOH/g, water in ppm, measured to ASTM and DIN methods that OEM limits are

written against. But a typical asset is sampled monthly at best. That gives a model twelve points a year, and one to three per oil drain interval. You cannot learn a degradation rate from three points, and you cannot see an event that starts and ends between samples.

The second kind is the **inline proxy sensor**. Dielectric probes, humidity sensors and single-band optical cells stream data every few seconds and are inexpensive. But they report a dielectric constant, a relative humidity or a vendor-specific “quality index”. Those signals are useful alarms. They have no counterpart in the laboratory archive, the OEM manual or any standard.

That creates three problems for an AI team:

- **No labels.** Decades of lab results cannot be used to train a model on a signal that was never measured in the lab.
- **No specificity.** Water, acids, oxidation, fuel and soot all move a dielectric signal in the same direction. The model must guess which one it is seeing, and each calls for a different action.
- **No portability.** A model trained on one sensor type, at one site, on one oil, has to be relearned when any of those change.

None of this means proxy sensors are bad. It means the AI programme spends most of its effort translating signals rather than improving decisions.

Putting the laboratory method inline

There is a third option: take the laboratory method itself and run it continuously on the machine.

Mid-infrared (MIR) spectroscopy is already the basis of standard in-service oil analysis (ASTM E2412, ASTM D7414). Each chemical change leaves its own fingerprint: oxidation products absorb near $1,700\text{ cm}^{-1}$, nitration near $1,630\text{ cm}^{-1}$, sulphation near $1,150\text{ cm}^{-1}$, water as a broad band around $3,400\text{ cm}^{-1}$. Fuel, soot, glycol and additives each have identifiable signatures too.

At Spectrolytic we built FluidInspectIR around a static MIR spectrometer with no moving parts, mounted in a small bypass loop. Combined with an optical particle counter (ISO 4406) and a viscometer, it reports oxidation, nitration, sulphation, TAN/TBN, water, fuel dilution, soot, additive concentration, particle counts and viscosity every few minutes. Chemometric models convert each spectrum into the same parameters, units and accuracy as a laboratory report, validated against reference methods.

Two things change when that happens.

First, **time resolution**. An asset produces more than 8,000 lab-equivalent results a year instead of twelve. Events that sampling structurally misses become visible.

Second, **the question changes**. Instead of asking whether one result has crossed a fixed limit, you can fit a degradation slope within each drain interval, project when the OEM limit will be reached, identify which parameter is life-determining, and benchmark every asset against the best-performing unit in its fleet. Those slopes and step changes are transparent, explainable features, and a natural input layer for machine learning.

What the field data shows

Three industrial deployments illustrate what continuous, lab-equivalent data makes visible.

Steel rolling: water break-ins. In the gear-oil circuit of a steelworks, three water ingress events between July and October 2026 reached roughly 26,000, 16,000 and 5,000 ppm. Each began abruptly and decayed over several days as water was removed. Weekly or monthly sampling would probably have caught one at most. The continuous ppm signal, validated against Karl Fischer titration, can switch on water-removal equipment automatically. At an earlier rolling installation, the operator estimated savings above USD 100,000 over nine months from avoided damage and oil replacement.

Aluminium foil rolling: additive control. Rolling-oil additives (fatty acids, fatty alcohols, esters, the antioxidant BHT) were previously measured by gas chromatography, with results back in days or weeks. At two European foil rolling plants, inline MIR now tracks them continuously. Against on-site GC, mean deviation was 0.014% for BHT and 0.05% for fatty acid. The data is used to time additive top-ups and is the basis for closed-loop dosing.

Engine fleets: drain intervals. On a large mining excavator, inline oxidation rose almost linearly within each drain interval. Projecting each slope to the limit showed oil being changed with 250–550 hours of usable life remaining. Draining at 75% of projected life would remove about a third of oil changes, each of which costs around six hours of production. On a biogas engine, the same method extended the drain interval from 600 to 1,400 hours, and a sulphation spike exposed failing H₂S filtration before it caused damage.

In each case the value came less from any single reading than from rates and events that only continuous data reveals.

Why units matter to machine learning and LLMs

Here is the part most AI roadmaps miss. When an inline sensor reports in laboratory units, three things follow.

Your lab archive becomes training data. Most operators hold years, often decades, of oil analysis results. Models trained on that history can be applied directly to the inline stream, because both speak abs/cm, mg KOH/g and ppm. No re-labelling, no translation layer.

General-purpose LLMs can reason about it. A large language model already knows what TBN is, what an oxidation limit means and why water in a gearbox matters. Give a maintenance copilot “oxidation 18 abs/cm, rising 12 per 1,000 h, OEM limit 25” and it can retrieve the OEM manual, compare against the limit, estimate remaining hours and draft a work order. Give it “quality index 63, down 4 points” and it can only report that something changed. The model should interpret the numbers, not invent them, and physically defined inputs make that interpretation reliable.

Models travel. Because the inputs are physically defined rather than instrument-specific, a model built on one fleet is far more portable to another site or asset type, provided calibrations are consistent.

The economics of building a training dataset show the scale of the difference. For an illustrative 50-asset fleet needing 100,000 labelled oil-condition records:

	Monthly lab sampling	Inline lab-equivalent sensing
Records per asset-year	12	up to ~8,000
Records per drain interval	1–3	several hundred
Time to 100,000 records	> 160 years	≈ 3 months
Cost of 100,000 records	≈ USD 5 million	< USD 0.1 million

Table 1: Assumes USD 50 per lab sample and about USD 7,000 per asset-year for an inline system. Figures are site-specific and illustrative.

To be precise about the comparison: any inline sensor closes the volume gap with the lab. What lab-equivalent sensing adds is data that can be joined to the lab archive, read by an LLM and reused across sites. Volume without meaning is just more data to clean, and data preparation is already where most industrial AI effort goes.

Where it isn't the answer

Inline MIR is not the right tool for every machine, and claiming otherwise would not help anyone.

- **Cost per point.** A lab-equivalent analyser costs more than a screw-in probe. It pays back on assets where an unplanned stop is expensive: rolling-mill gearboxes, large engines, turbines, compressors. For a water alarm on a small hydraulic unit, a humidity sensor is cheaper and good enough.
- **Installation.** The system needs a small bypass flow. Some machines have no practical place for one.
- **Calibration.** Unlike proxy sensors that can be plugged in and used to provide arbitrary units, inline MIR analysis requires an appropriate model to interpret the measurement. Developing and validating these models takes expertise.
- **Wear metals.** Inline MIR measures chemistry, not elemental wear debris. The laboratory remains essential for elemental analysis, for calibration reference and for failure investigation.

The right architecture uses each method for what it does best: inline chemistry for continuous condition and early warning, the laboratory for reference and root cause, and vibration or thermography to localise mechanical faults. The inline system also tells you when a lab sample is actually worth taking.

Where to start

If your organisation is investing in predictive maintenance or AI copilots, three questions are worth asking before the next pilot:

1. **What units does our oil data arrive in?** If the answer is an index or a relative signal, your models are learning a translation, not a machine.
2. **Can our lab archive train our models?** If not, you are leaving your largest labelled dataset unused.
3. **Which assets justify lab-grade data?** Start with the one or two where an unplanned stop costs most, stream the data into your historian or AI platform, and compare it with your lab reports for three to six months.

AI will not fix poor inputs. But give it the same language your engineers, OEMs and laboratories have used for decades, and it can start doing what it promises: turning oil chemistry into decisions, early enough to matter.

References

- [The 2026 Industrial AI Readiness Report](#), IIoT World / HiveMQ (survey of 272 industrial professionals)
- ASTM E2412-10 (2018), Condition Monitoring of In-Service Lubricants by Trend Analysis Using FT-IR Spectrometry
- ASTM D7414-22, Condition Monitoring of Oxidation in In-Service Lubricants by Trend Analysis Using FT-IR Spectrometry
- ASTM D6304-20, Water in Petroleum Products by Coulometric Karl Fischer Titration
- ISO 4406:2021, Hydraulic fluid power: coding the level of contamination by solid particles