

IT@Intel: From Automation to Autonomy: Agentic AI Is Transforming Intel Facilities Operations

Intel IT is pursuing a vision for self-optimizing, human-aligned facility intelligence

Industrial Automation Authors

Patrick Bradley, PhD
System Software Architect

Manoj Athawale, PhD
Machine Learning Engineer

Subhadra Sampathkumaran
Software Engineering Manager

Table of Contents

Background.....	2
The Opportunity—From Automation to Autonomy	2
Enablers for Predictive Maintenance at Scale	2
Industry 5.0 and AI Technology Convergence	3
Our Facilities Transformation Journey.....	4
Facilities Agentic AI Solution Architecture Overview	4
Architecture Details and Data Flow... 5	
Data Ingestion and Contextualization	6
Feature Engineering and Analytics....	7
Predictive and Anomaly Models	7
Decisioning.....	7
Automation.....	7
Agentic Orchestration.....	7
Integration with Digital Twins	8
Use Cases and Results.....	8
Predictive Maintenance – UPW TOC Monitoring.....	8
Hybrid SPC with AI-Driven Anomaly Detection	8
Digital Twin – Chilled-Water Plant Optimization	8
Unified Facilities Analytics	9
Future Vision: The Path to Lights-Out Facilities.....	9
Conclusion.....	9
Related Content.....	9

Executive Summary

Intel manufacturing facilities have long been the backbone of reliable, safe, and efficient semiconductor manufacturing. As fabs scale in complexity and sustainability expectations rise, traditional automation systems—built around static logic and deterministic control—are insufficient.

To keep up with the pace of business, Intel’s Industrial Automation team, working in tandem with the Facility Operations team, is helping Intel facilities transition from mere automation to complete autonomy using agentic AI. Agentic AI combines machine learning, digital twins, and Human-in-the-Loop (HITL) oversight to orchestrate intelligence across building systems, utilities, and sub-fabs. This distributed, context-aware, multi-agent architecture integrates perception, reasoning, and action across the facilities ecosystem.

Early deployments demonstrate measurable impact, compared to traditional automation systems:

- 30% faster anomaly detection in ultrapure water (UPW) systems.
- 15–20% reduction in technician workload via autonomous work order generation.
- More than 120,000 labor hours saved annually by integrating analytics of asset management data and industrial process data.

Our vision for facilities automation extends far beyond predictive maintenance to achieve self-optimizing, autonomous facilities by the end of the decade, setting a new standard for Industry 5.0 operations.

Contributor

Joe Sartini, IT@Intel EMEA Regional Lead

Acronyms

AI	artificial intelligence
ARIMA	Autoregressive Integrated Moving Average
EWMA	Exponentially Weighted Moving Average
HITL	Human-in-the-Loop
IIoT	Industrial Internet of Things
LSTM	Long Short-Term Memory
MCP	Model Context Protocol
OPC UA	Open Platform Communications Unified Architecture
PCA	Principal Component Analysis
PLC	programmable logic controllers
RUL	remaining useful life
SCADA	supervisory control and data acquisition
SPC	statistical process control
TOC	total organic carbon
UPW	ultrapure water

The Opportunity—From Automation to Autonomy

Historically, automation meant using codified logic with if-then rules that were executed in programmable logic controllers (PLCs) or supervisory control and data acquisition (SCADA) systems. While powerful, this approach is rigid and assumes static operating envelopes.

Agentic AI marks the next evolution for automation, embedding reasoning and learning into the operational fabric. Each agent perceives its environment, reasons for context, and coordinates with other agents to achieve higher-order objectives. Similar to how transistors once scaled exponentially according to Moore’s Law, facilities intelligence is now scaling in depth and adaptability (see Figure 1). In the past, scaling meant adding capacity. In the future, scaling pertains to embedding intelligence into operations, so uptime, efficiency, and sustainability grow exponentially, leading to lower headcount and cost.



Past - Transistor Scaling Future - Self-Optimizing Facilities

Figure 1. Intel’s “Automation to Autonomy” journey mirrors Moore’s Law for operations: As transistors once scaled exponentially, facilities intelligence now scales in depth and adaptability.

Background

Facilities are the unsung enablers of yield, uptime, and cost performance in silicon wafer manufacturing. Yet historically, they have operated through a mosaic of siloed, separate systems for control, telemetry, supervisory layers, advanced process data analytics, dashboards and reporting, and asset management. This approach results in fragmented visibility, manual data correlation, and reactive operations. Technicians often navigate between dashboards, spreadsheets, and maintenance logs to identify root causes or generate work orders.

At the same time, the complexity of facilities is escalating—cleanrooms, ultrapure water (UPW), exhaust treatment, and energy recovery systems all interact dynamically. Sustainability mandates and uptime targets are more demanding than ever. Traditional automation, bound by fixed thresholds, cannot keep pace with contextual reasoning or multi-system optimization. The facilities and automation teams jointly recognized a clear inflection point: a need for adaptive, context-aware intelligence that learns from data, predicts outcomes, and acts autonomously, while keeping human judgment—that is, a Human-in-the-Loop (HITL) strategy—at the center of all decisions.

Enablers for Predictive Maintenance at Scale

Scaling agentic AI requires three pillars:

- **Unified data context.** Integrated time-series data and metadata governance enable feature reusability and context sharing across agents.
- **Hybrid autonomy model.** HITL strategy used for judgment and safety; AI executes routine pattern recognition and forecasting.
- **Modular design for replication.** Each agent can be cloned and redeployed across sites with minimal retraining.

Data governance and cross-functional alignment mirror lessons from Intel’s manufacturing data warehouse programs, proving that data context matters more than data volume.

Industry 5.0 and AI Technology Convergence

The agentic AI transformation of autonomy aligns directly with the broader Industry 5.0 paradigm, emphasizing human-AI collaboration, adaptive control, and resilient, sustainable systems.

Our approach operationalizes these principles across five converging AI-enabled technologies, as shown in Table 1. The convergence of AI models, digital twins, and agent orchestration represents the practical realization of Industry 5.0 within Intel’s facilities (see Figure 2). Each system—once isolated—now participates in a shared cognitive ecosystem that is capable of perception, reasoning, and controlled action.

The outcome is living operational intelligence:

- Continuously learning from data
- Augmenting human expertise
- Autonomously driving efficiency, safety, and sustainability across Intel’s Foundry network

Table 1. AI-enabled Technologies for Industry 5.0

Industry 5.0 Dimension	Example of Implementation in Intel Facilities	Core Technologies
Human-Centric Operations	Human-in-the-Loop (HITL) agent governance	<ul style="list-style-type: none"> ▪ Workflow orchestration based on Model Context Protocol (MCP) ▪ Explainable AI dashboards
Autonomous Systems	Self-optimizing sub-fab agents	<ul style="list-style-type: none"> ▪ LangGraph orchestration ▪ Reinforcement learning loops
Resilience and Predictive Intelligence	Anomaly detection across ultrapure water (UPW), chillers, and exhaust systems	<ul style="list-style-type: none"> ▪ Convolutional neural network (CNN) with a Long Short-Term Memory (LSTM) network ▪ Principal Component Analysis (PCA) ▪ Statistical Process Control (SPC) ▪ Wavelet decomposition
Sustainability and Efficiency	Energy optimization in digital twins	<ul style="list-style-type: none"> ▪ Hybrid Multilinear Principal Component Analysis (MPCA) and Sparse Autoencoder (SAE) ▪ Feedback optimization
Interoperability and Integration	Unified process data analytics, asset management, and telemetry systems	<ul style="list-style-type: none"> ▪ MQTT protocol ▪ Open Platform Communications Unified Architecture (OPC UA) ▪ REST APIs ▪ Facilities Unified Tag Model schema



Figure 2. The convergence of AI models, digital twins, and agent orchestration represents the practical realization of Industry 5.0 within Intel’s facilities.

Our Facilities Transformation Journey

Integrating Industry 5.0 capabilities into Intel’s facilities operations is not an overnight process. Our phased approach builds on each success and applies lessons learned as we go along (see sidebar on [page 9](#)). We have completed the first two phases of the transformation and are well-positioned to complete the third phase soon. Here is a summary of each phase:

- **Phase 1 (2020–2022): Digitalization and Automation.** Our initial efforts centered on digitizing existing systems and centralizing telemetry data streams. We deployed Node-RED and MQTT brokers to capture industrial Internet of Things (IIoT) telemetry from sub-fab equipment. Early process data analytics pilot projects provided visualization for utilities and environmental data.
- **Phase 2 (2023–2024): Data Contextualization and Predictive Insight.** We broadly deployed process data analytics capabilities across Fab Facility Operations and Technology Development sites, which unified data across telemetry, asset management, digital twin applications, and water management systems. The unified data enabled users to quickly gather and analyze data, create work orders, and auto-generate condition-based maintenance directly in the analytics application. This milestone marked Intel Facilities’ largest investment in data analytics in a decade, enabling near-real-time insight across global facilities.

- **Phase 3 (2024–2025): Predictive AI Maintenance.** Building on the data foundation established in Phase 2, we launched predictive AI maintenance pilot projects for UPW and chiller water systems. Machine-learning and anomaly detection models reduced false positives by more than 75% and extended forecasting horizons by 3x.
- **Phase 4 (2026 onward): Agentic AI Autonomy.** Now we are integrating agentic AI orchestration—a LangGraph-based multi-agent framework with HITL oversight—to coordinate predictive, prescriptive, and autonomous decisions across the enterprise.

Facilities Agentic AI Solution Architecture Overview

Intel’s agentic AI framework combines control theory, machine learning, and autonomous orchestration through a modular, multi-agent design. The framework spans three execution tiers—perception, reasoning, and action—that are interconnected using a semantic context layer and governed by HITL validation.

Figure 3 illustrates the data flow:

1. Real-time telemetry from sensors and SCADA systems is ingested through edge nodes.
2. Next, it is contextualized in our process data analytics application and a PostgreSQL Vector database.
3. Then, it is reasoned on by LangGraph-based agents.
4. The results are executed through asset management, digital twin, and HITL workflows.
5. The loop is closed by sending feedback to the Perception tier, enabling autonomy across facility systems.

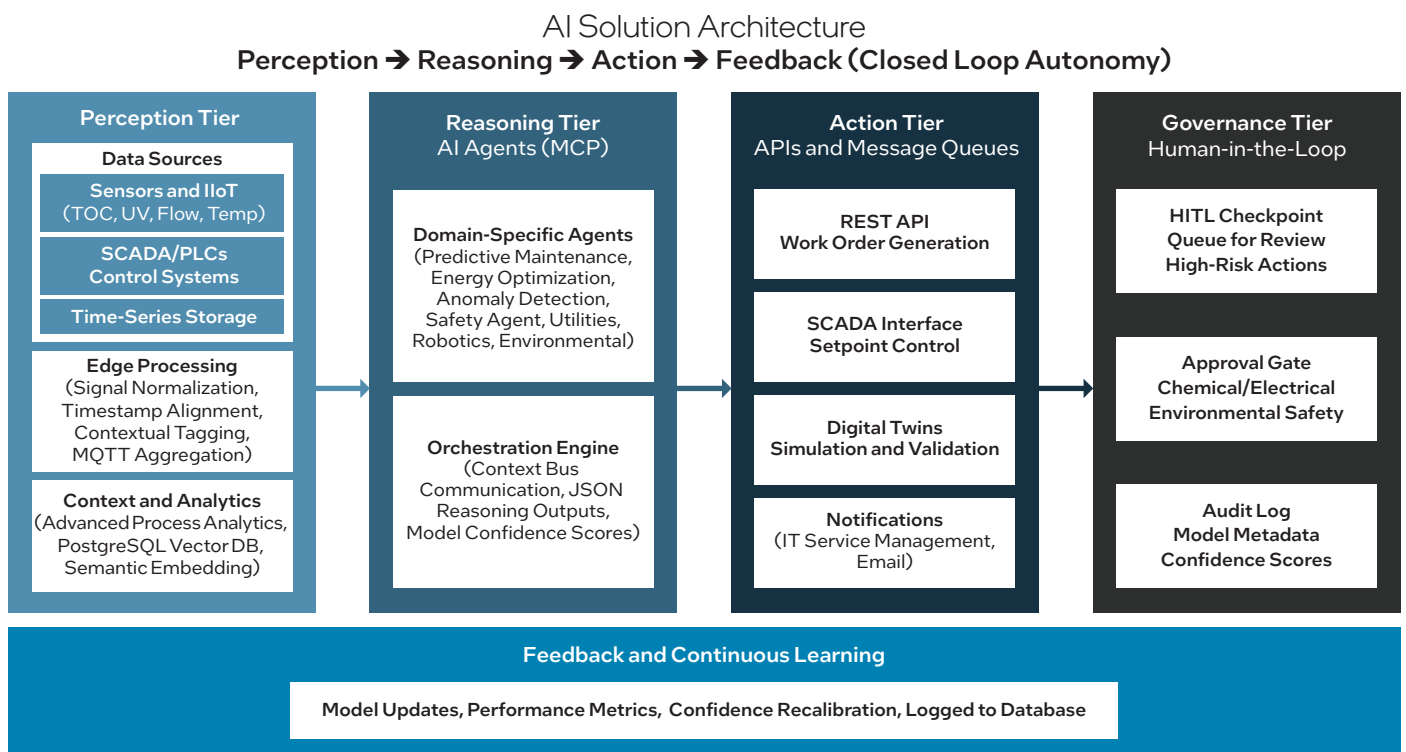


Figure 3. Our Agentic AI facilities architecture combines control theory, machine learning, and autonomous orchestration.

Each facility domain (UPW, Mechanical, Exhaust, IWS, etc.) operates as an agentic cell—a self-contained intelligence node consisting of the following tiers:

- **Perception Tier.** Real-time data ingestion from edge controllers, sensors, and telemetry streams.
 - Protocols: Open Platform Communications Unified Architecture (OPC UA), Modbus, MQTT, HTTP REST
 - Systems: telemetry, digital twin, advanced analytics
 - Edge nodes: Node-RED gateways performing signal normalization, timestamp alignment, and contextual tagging
- **Reasoning Tier.** Cognitive orchestration using LangGraph agents, each with a role-based context; examples of roles include “Predictive Maintenance Agent” or “Energy Optimization Agent.” This tier performs reasoning, planning, and orchestration. Agents specialize by domain—such as utilities, robotics, maintenance, and safety—and share context through embeddings.
 - Agents communicate through the Model Context Protocol (MCP), maintaining semantic awareness of system states and goals.
 - Each agent runs domain-specific models, such as Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA), Principal Component Analysis (PCA), and autoencoders. Agents exchange reasoning outputs as JSON objects over a shared context bus.

- **Action Tier.** Execution through APIs or message queues to external systems.
 - Maintenance uses our asset management application’s REST API to create and update work orders.
 - Operations uses the SCADA command-line interface for setpoint tuning or valve control.
 - Communication uses our IT service management platform, email, and dashboard alerts.
- **Governance Tier.** Governance is implemented using an HITL checkpoint queue for human review of AI recommendations before operational action. All actions are logged with model metadata and confidence scores for traceability. An HITL control plane helps ensure that high-risk actions (such as chemical, electrical, or environmental) require human approval.

Architecture Details and Data Flow

The end-to-end agentic AI data flow follows a closed intelligence loop, merging operational technology (OT) and information technology (IT) environments. Edge nodes process signals locally for latency-sensitive control (less than 100 ms) while higher-level agents run in Intel’s secure IT cloud.

The data flow follows a five-layer design (see Figure 4). Each layer is described in more detail in on the [following page](#).

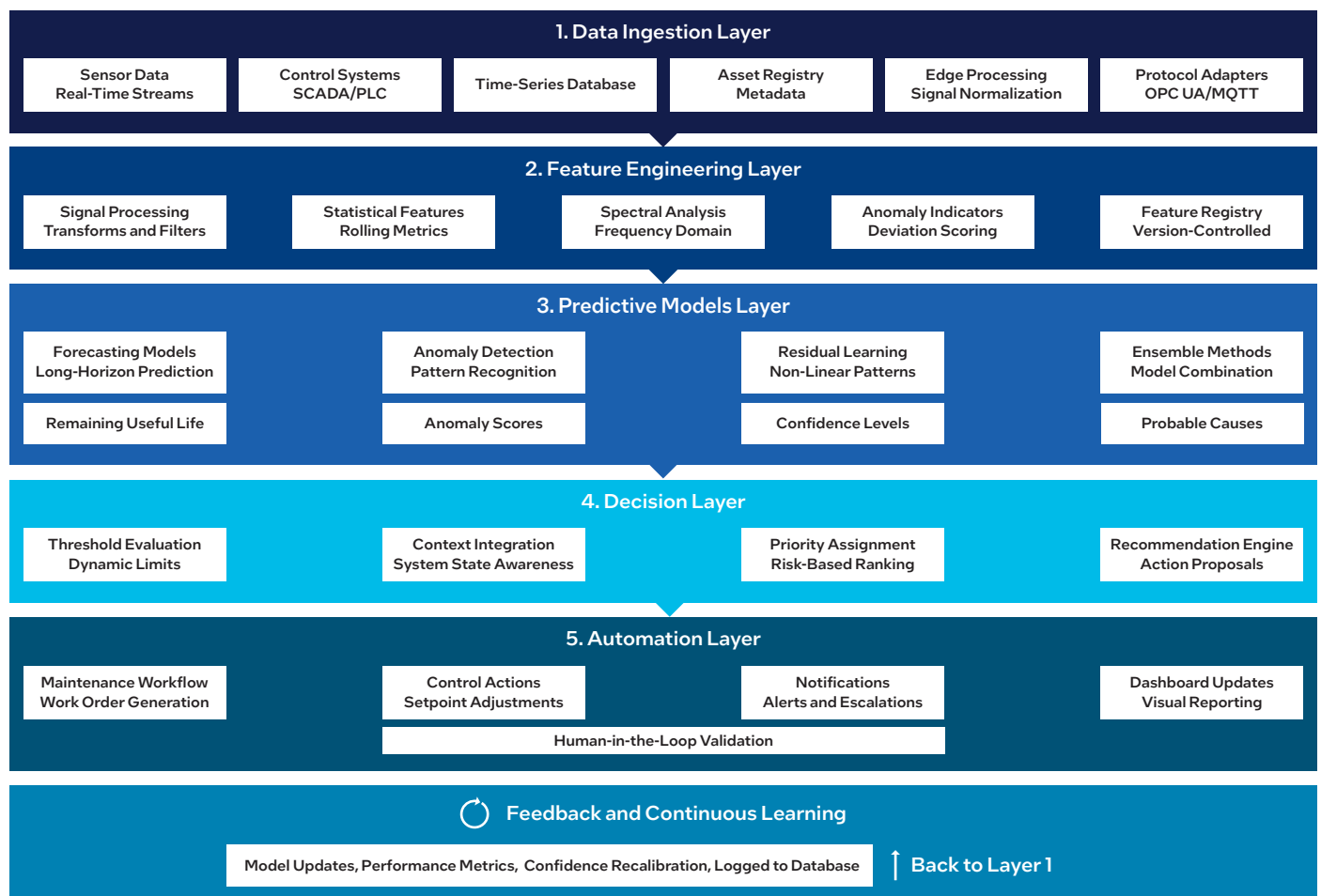


Figure 4. Agentic AI Orchestration: end-to-end data flow.

1. **Data Ingestion Layer.** SCADA, telemetry, and IoT sensors provide real-time data streams, including total organic carbon (TOC) analyzers, UV monitors, flow meters, and chillers.
2. **Feature Engineering Layer.** Derived metrics such as velocity, acceleration, rolling statistics, and spectral entropy are computed in Python.
3. **Predictive Models Layer.** Ensemble machine-learning models—such as Fourier extrapolation, Dynamic Time Warping (DTW), and LSTM—estimate remaining useful life (RUL) and forecast excursions.
4. **Decision Layer.** Thresholding, anomaly scoring, and residual learning integrate statistical and deep-learning outputs.
5. **Automation Layer.** Agents trigger maintenance workflows through the asset management system’s APIs. Our advanced analytics application provides dashboards and cross-platform integration.

Data Ingestion and Contextualization

- **Source systems** include SCADA, PLCs, sensors, IIoT devices, and data analytics application connectors.
- **Edge processing:** MQTT brokers aggregate data every 5–10 seconds.
- **Telemetry:** Our telemetry system stores time-series data with asset-level tagging.
- **Context enrichment:** Metadata (such as equipment ID, location, asset class, and limits) is stored in a PostgreSQL vector store for agent context retrieval.
- **Standardization:** All signals are mapped to our Facilities Unified Tag Model for semantic interoperability.

Figure 5 shows how data flows from various sources—such as PLCs, IIoT sensors, and robots—into the time-series data store. It also illustrates how the analytics platform uses the data store, plus the contextual data from the data lake, to enable AI-powered advanced analytics.

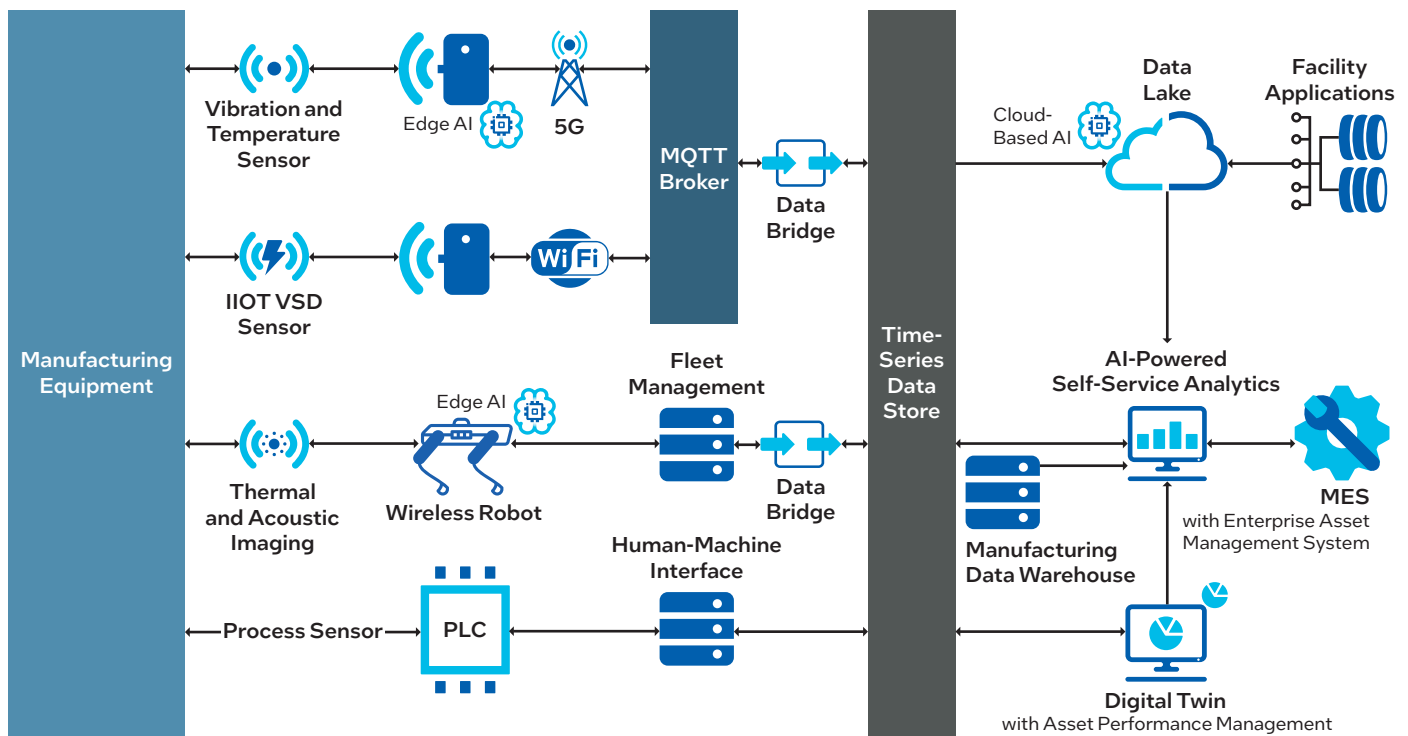


Figure 5. Data from various sources flows into the time-series data store, which the analytics platform combines with contextual data to enable AI-powered advanced analytics.

Feature Engineering and Analytics

- Derived metrics: Δ velocity, Δ^2 acceleration, rolling means (6/24/48 hour), Fourier transform features, Z-score anomalies, and spectral entropy.
- Advanced analytics modules execute in our advanced analytics application (Python runtime).
- All processed features are stored in a feature registry, version-controlled by model revision.

Predictive and Anomaly Models

- Forecasting models: An ensemble of Hybrid Fourier, ARIMA, and LSTM for long-horizon predictive maintenance.
- Anomaly detection: Hybrid statistical process control (SPC), including Z-score and Exponentially Weighted Moving Average (EWMA), fused with convolutional neural network (CNN)-LSTM for dynamic thresholds.
- Residual learning framework: Traditional linear models (ARIMA) predict expected trends; CNN-LSTM models discover residual non-linear patterns.
- Thresholding, anomaly scoring, and residual learning integrate statistical and deep-learning outputs.
- Model outputs:
 - Anomaly_score informs model confidence (0–1).
 - RUL provides time-to-failure (measured in days).
 - cause_tags identifies an issue’s probable root cause.

Decisioning

Predictions are probabilities; decisions require thresholds, context, and business rules. This layer transforms model outputs into actionable intelligence. Static thresholds do not work in dynamic systems. Thresholds adapt based on operating context. Thresholds also consider time-to-repair. For example, if ordering resin takes 14 days, the threshold triggers 21 days before predicted failure.

Anomaly scores are calculated using multiple methods: Isolation Forest for detecting multivariate outliers, DBSCAN for identifying density-based anomalies, and Mahalanobis distance for identifying statistical outliers. These scores combine to form a single anomaly index, ranging from 0 to 100. Scores above 70 trigger investigation, above 85 trigger alerts, and above 95 can trigger an automatic action like work order creation.

Decisions optimize multiple objectives: minimize risk, minimize cost, maximize equipment life, and maintain production. The final decision takes into account maintenance crew availability, spare parts inventory, and the production schedule.

Automation

This layer is where decisions become actions, closing the loop from sensing to responding.

APIs expose predictions to downstream systems for custom applications. Predictions become actions through cross-platform integration. For example, predictions and recommendations are pushed to the database, which then provides production schedules and maintenance windows. Work orders are generated automatically when RUL crosses the threshold. Each work order includes the predicted failure date, confidence level, recommended action, and required parts.

Operator dashboards show real-time status and 30- or 60-minute predictions. Engineering dashboards display trends, model performance, and seven-day forecasts. Executive dashboards present fleet-wide health and a 90-day outlook. All dashboards are updated automatically on a set schedule based on the facility’s domain-specific requirements.

Agentic Orchestration

The LangGraph Orchestrator coordinates all agents using an event-driven loop. Here is an example flow (see Figure 6).

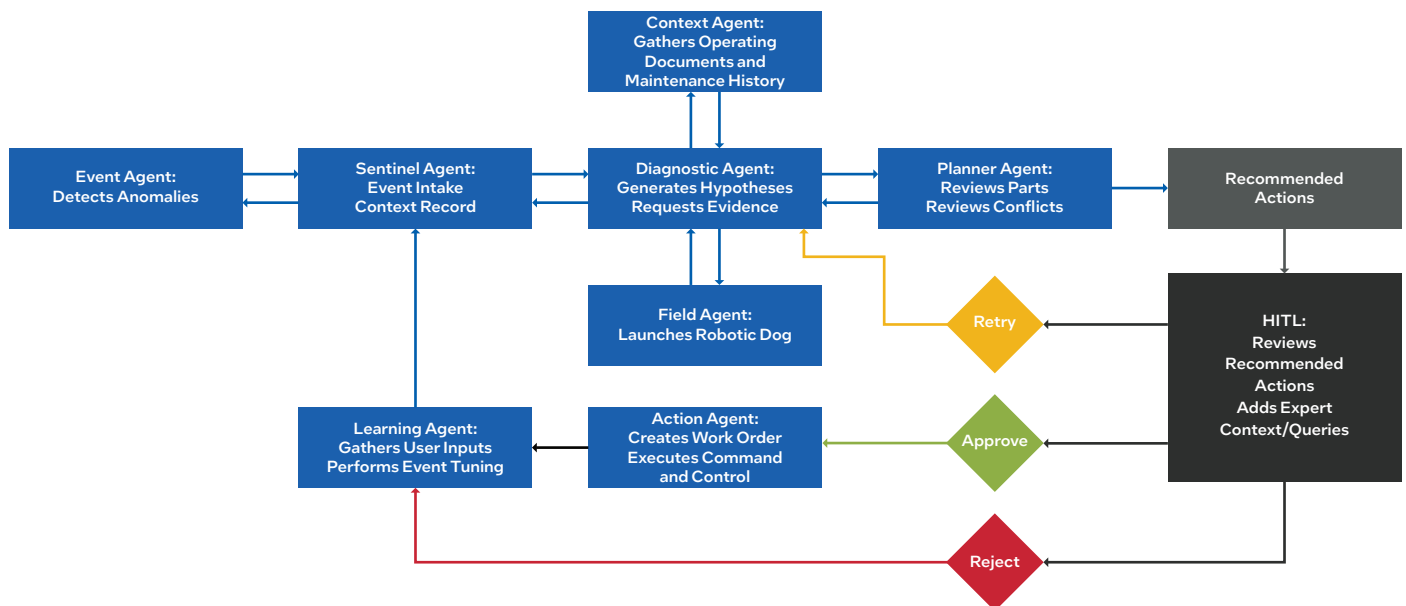


Figure 6. Agentic AI governance through a Human-in-the-Loop (HITL) checkpoint queue.

1. The perception agent detects an anomaly pattern and pushes the event to the sentinel and context bus.
2. The reasoning agent evaluates the historical context, then correlates it with trend capsules in the data analytics application.
3. The planner agent decides on the course of action, and the action agent prepares a work-order ticket, which is auto-drafted; however, human approval is required.
4. The HITL workflow validates and executes an action, and the confirmation is logged.

This orchestration process enables a “perception → cognition → action → feedback” loop within milliseconds for non-critical processes, or minutes for supervisory actions.

Integration with Digital Twins

The architecture supports bidirectional digital-twin coupling, where AI agents continuously update and learn from virtual plant models. Digital twins enable simulation, scenario testing, and predictive optimization of utilities (see Figure 7). For example, each asset domain (such as chilled water or UPW) has a Level-4 “Living Digital Twin” (real-time model and machine learning coupling). Digital twins run predictive simulations to validate AI recommendations before action execution. Agents consume simulated versus real data differentials to update model confidence in real time.

Use Cases and Results

Agentic AI-driven automation is delivering measurable results in our facilities across various use cases, including

UPW monitoring, anomaly detection, chilled water plant optimization, and unified facilities analytics.

Predictive Maintenance – UPW TOC Monitoring

The UPW system must provide water with a TOC concentration of less than 0.7 parts per billion. Any excursion can halt lithography and increase costly downtime. Our AI-enhanced monitoring solution predicts resin exhaustion and TOC breakthroughs before they occur, with a forecasting horizon of 90 days (compared to a 30-day baseline). The model accuracy, as measured by Root Mean Square Error (RMSE), has improved prediction accuracy by 68% and the false positive rate has fallen by 82%. We estimate that this AI-enhanced monitoring solution will deliver about USD 7.5 million in annual savings by preventing excursions.

Hybrid SPC with AI-Driven Anomaly Detection

Combining SPC (EWMA and Hotelling T²) with deep learning (CNN-LSTM and ARIMA) delivers an early warning of multivariate anomalies in facility data. Our hybrid residual-learning models detected anomalies 2–5 days earlier, with a low false positive rate (<5%) and a true positive rate greater than 97%.

Digital Twin – Chilled-Water Plant Optimization

The integration of digital twins with analytics and sparse autoencoders has enabled adaptive optimization of chilled-water systems. We’ve increased energy efficiency by 15% with automated drift detection via Moving Mean PCA. Our models also continuously learn baseline conditions without manual recalibration.

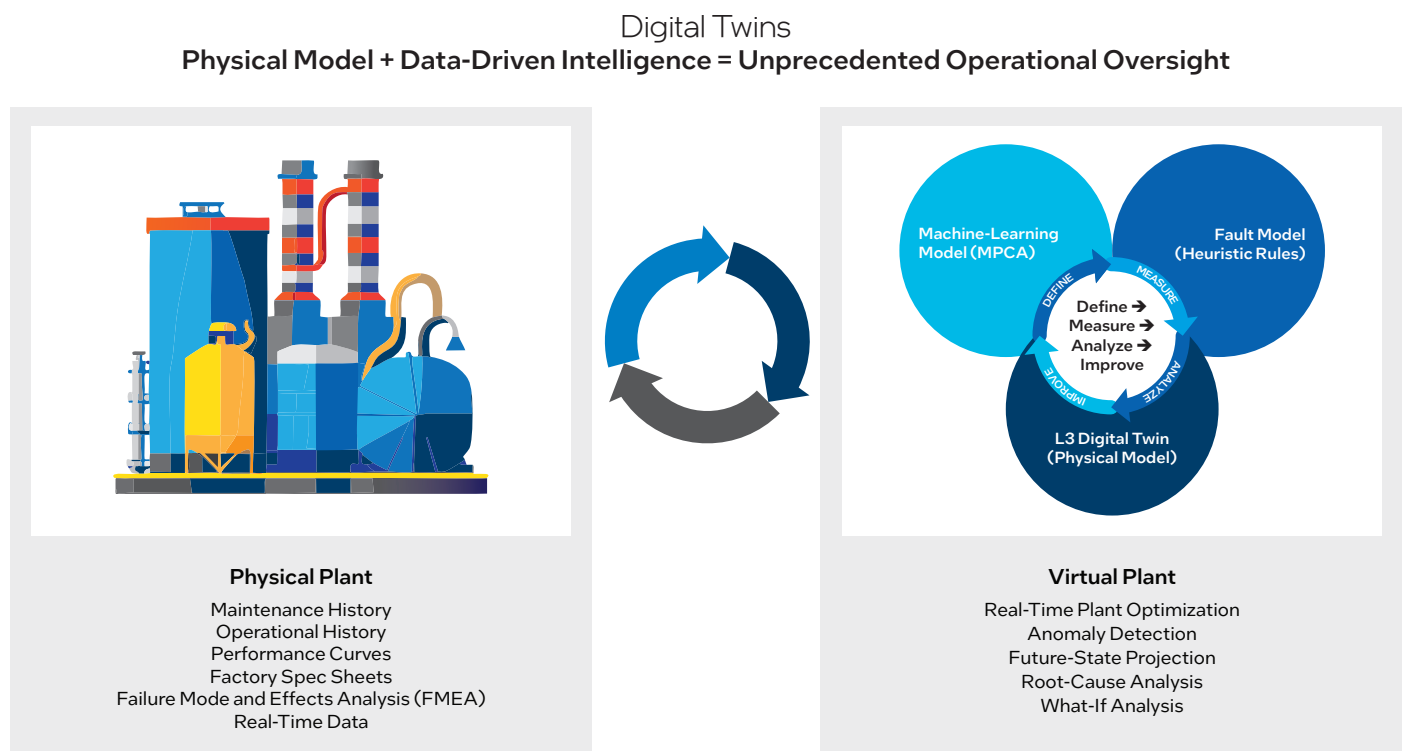


Figure 7. Digital twins enable simulation, scenario testing, and predictive optimization of utilities like chilled water and ultrapure water (UPW).

Lessons Learned

Our agentic AI journey toward autonomous facilities has revealed the following truths:

- **Hybrid autonomy builds trust.** Operator confidence grew through incremental automation, where AI proposes actions that humans can approve.
- **Integration trumps innovation.** Value emerged not from new algorithms alone, but from the seamless integration of facility systems, such as asset management, historical data, and advanced analytics systems.
- **Data context is king.** Standardizing signal naming and contextual metadata was our largest accelerator for AI adoption.
- **Cross-team collaboration is vital.** Tight coordination between Facilities and Automation teams enabled fast iteration and safe deployment.

Unified Facilities Analytics

Launching our new advanced data analytics application created a unified data ecosystem connecting asset management data, water systems, IIoT sensors, and robotics. Operators can create and track work orders within the data analytics application, saving up to 120,000 hours through automation and reduced data lookup time.

Future Vision: The Path to Lights-Out Facilities

Our long-term goal for AI-driven, intelligent automation is to establish a network of autonomous facilities that operate as self-optimizing systems. Our plans include the following:

- **2025–2026: Predictive Expansion.** Agentic AI pilot projects will be scaled across global fabs and support systems. Integration with the IT service management platform and email agents will enable end-to-end incident management.
- **2027–2028: Prescriptive Coordination.** Agents will collaborate across domains—energy, water, and safety—to optimize holistically for cost and carbon reduction. Digital twins will simulate optimal states before real-world execution.
- **2029–2030: Autonomous Facilities.** Facilities will achieve “lights-out” operation through continuous learning, goal-driven autonomy, and HITL governance, helping to ensure optimal safety and ethical operations. Facilities will become self-healing organisms characterized by interacting AI agents that anticipate needs, balance resources, and enhance sustainability.

Conclusion

Intel’s facilities transformation embodies the convergence of human expertise and machine intelligence. By combining agentic AI with robust data infrastructure and human oversight, we are defining the next era of operational excellence. The journey from automation to autonomy is not a single leap, but a continuum—each phase delivering tangible value while laying the foundation for the next. Through collaboration and innovation, Intel facilities are evolving from reactive to predictive to self-optimizing systems, pioneering the future of industrial operations in the age of agentic AI. Through the IT@Intel program, the solution development team is open to consultation or pilot requests and is eager to help other manufacturers continue their AI and automation journey.

Related Content

If you liked this paper, you may also be interested in these related stories:

- IT@Intel: Training Machine-Learning Models on Intel® Gaudi® Accelerators
- IT@Intel: Transforming Siloed Manufacturing Data into Unified Insights
- IT@Intel: Innovations in Confidential Manufacturing Improve Security and Scalability
- IT@Intel: Expanding Low-Cost IIoT Manufacturing Use Cases
- IT@Intel: Delivering Operational Efficiencies Using a 5G Private Network
- IT@Intel: Agentic AI in the Enterprise

For more information on Intel IT best practices, visit intel.com/IT

IT@Intel

We connect IT professionals with their IT peers inside Intel. Our IT department solves some of today’s most demanding and complex technology issues, and we want to share these lessons directly with our fellow IT professionals in an open peer-to-peer forum.

Our goal is simple: improve efficiency throughout the organization and enhance the business value of IT investments.

Visit us today at intel.com/IT or contact your local Intel representative if you would like to learn more.

