Driving the Future of Manufacturing With Edge AI: Accelerating the Revolution

intel

In this guide, we'll look at the common reasons AI projects struggle, explore ways to overcome AI challenges, and see how the latest developments in AI are transforming manufacturing.

As existing AI solutions mature and new classes of generative AI prove their value, manufacturers are on the edge of innovations that we are just beginning to imagine. Decision-makers throughout the industry are eager to put AI to work, but putting AI into production isn't as simple as launching an application.

Nine common challenges of deploying AI in manufacturing

01

Figuring out where to start

Developing AI manufacturing solutions and integrating them into existing production lines is a daunting proposition. The underlying infrastructure, data management, model training, and AI application development require skilled specialists—data architects, data scientists, and AI developers—who are in short supply.

Developing production-ready AI applications demands another layer of expertise. The combination of skilled expertise, long development and piloting cycles, plus ongoing governance can add up to major total costs of ownership (TCO).

Organizations can succeed using internal resources. We've found that working with experienced AI systems integrators can shorten the path to production and may even reduce TCO. For even quicker deployment, a turnkey solution or Solution as a Service may be a better choice for many manufacturers.

> The combination of skilled expertise, long development and piloting cycles, plus ongoing governance can add up to major TCO.

02

Getting past single-point solutions

Developing applications and solutions over a series of iterations is sound advice. It helps teams identify problems, learn, and build momentum—all excellent outcomes for rolling out AI. However, single-point AI deployments create small islands of intelligence that aren't part of a larger transformation. They are useful but they can't produce systematic improvement or support new, AI-powered manufacturing methods.

For example, a deep learning computer vision model that detects defects can spot rejects and provide retrospective data for quality control and manufacturing processes. But if the AI application is a single-point function, the data it creates can't be used in real time beyond its single task.

03

Trying to manage the edge without a platform

It's hard to underestimate the impact of DevOps and cloud-native computing on how we build, scale, and maintain applications and services. We can virtualize hardware with code, spawn infrastructure and applications on demand, and manage everything from a single pane of glass. Cloud native computing has been a true revolution; a revolution that laid the foundation for AI and machine learnintg at scale.

A revolution in the cloud, that is. The edge is a different story. Exporting cloud-native practices from the uniform infrastructure of the cloud or a private data center to the mix of servers, devices, and endpoints at the edge is a huge challenge. Doing it in manufacturing, where proprietary devices and supervisory control and data acquisition (SCADA) systems rule operations, is even more difficult.

However, without some sort of edge architecture that can manage data collection, orchestrate applications, and automate system administration and security, AI in manufacturing will struggle to grow beyond single-point solutions. A flexible, cloud-native platform can solve many of these challenges.

04

Starting AI before modernizing data collection and management at the edge

If a factory's IT and OT systems are poorly integrated, accessing, preparing, and sharing data from embedded industrial devices is difficult. Without a modern data architecture, it is practically impossible to serve up edge data in real-time, which means AI can only run at machine speed on specific devices and process steps. Interdependent systems—such as digital twins, autonomous forklifts, or self-configuring production lines—will run into latency issues that can slow or halt processes.

With a robust edge platform in place, data collection, management, and engineering become largely automated processes. Data can be extracted and shared for AI decision-making in low-latency and real-time applications, opening the door to vast new applications for AI.

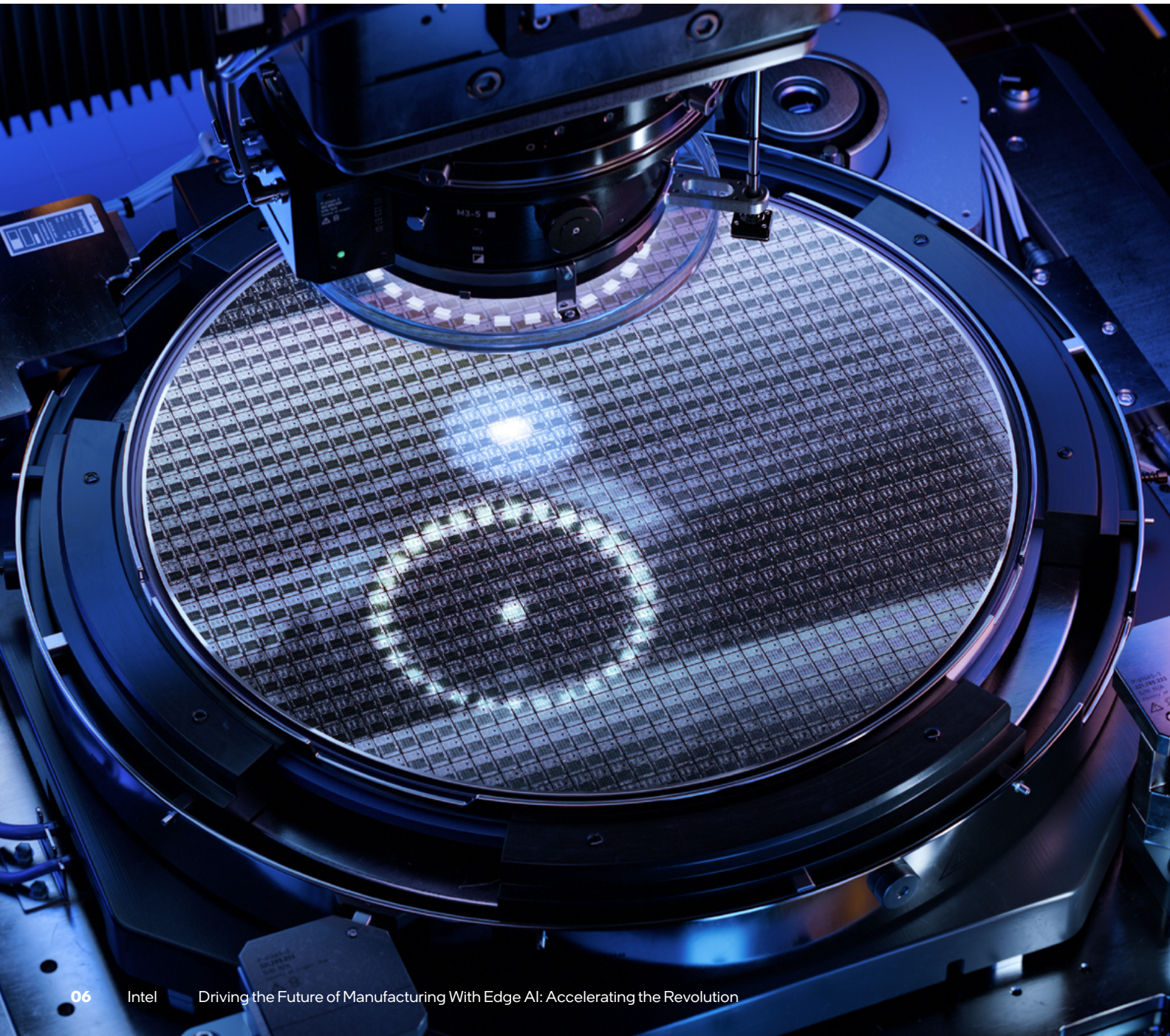
With a robust edge platform in place, data can be extracted and shared for AI decision-making in low-latency and real-time applications, opening the door to vast new applications for AI.

05

Extracting industrial sensor data at scale

Industrial processes and manufacturing lines produce vast amounts of time-series data, including video that monitors everything from machine health to worker safety. All that data is valuable, if it can be extracted through a process that can scale.

With the right platform, manufacturers can ingest video and sensor data, process it, and serve it to AI applications across the organization.



06

Leaving people out of the AI loop

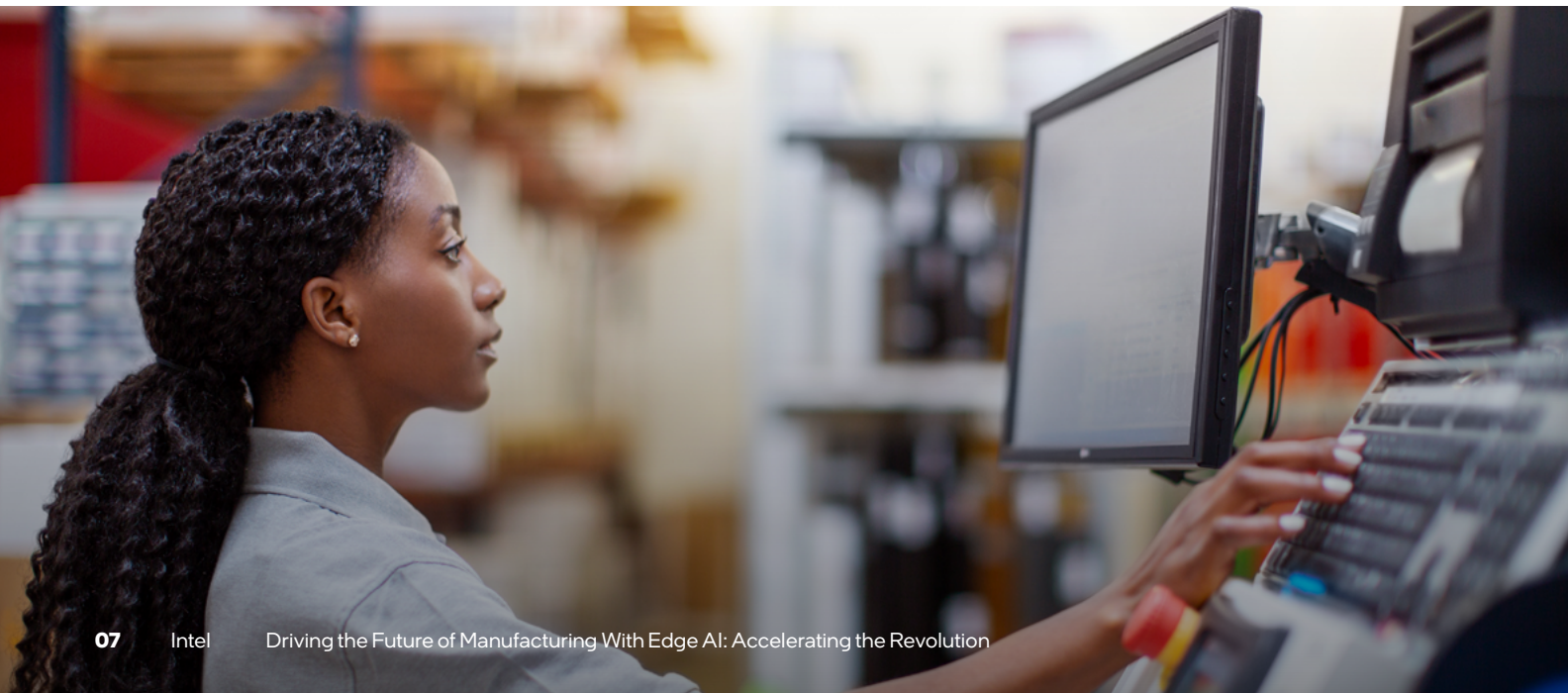
AI model training is time- and computing-intensive work done using specialist tools. Experts like process and quality engineers, production line designers, and safety managers can't be expected to learn arcane AI skills or devote days or weeks to annotating training data.

And yet, if non-AI experts aren't involved in model training and development, the results are invariably poor. Data scientists simply do not have the domain expertise to identify subtle defects in a product or understand what defines safe versus unsafe situations.

Without that expert knowledge, data scientists can't prepare the training data required to produce accurate, trustworthy AI models.

Manufacturers need explainable AI that makes model predictions transparent and keeps humans in the loop during training and deployment. Keeping people with subject area expertise involved is critical for the entire lifespan of an AI application because even well-trained models drift and require continuous retraining and redeployment.

> If non-AI experts aren't involved in model training and development, the results are invariably poor.



07

Unlocking the "black box" of AI

Stable, accurate model behavior is essential in manufacturing AI, but unlocking how deep learning models make predictions and decisions is a major challenge. Techniques like instance-based learning, which can trace decisions back to the original training data, can illuminate a model's process. Keeping humans in the loop as models are trained, deployed, and redeployed adds a critical check on model behavior.

08

Making a false choice between cloud and edge

Latency, bandwidth costs, and cloud computing have pushed manufacturing toward processing AI workloads on the edge. In fact, IDC reports that over 56% of enterprises plan to deploy AI at the edge and that over half of all edge deployments will incorporate AI by 2026.[1]

Running AI at the edge has many benefits. Edge AI can analyze data, make decisions, and optimize performance on the fly. With careful model selection and fine-tuning, edge AI can run on existing brownfield infrastructure with limited new hardware and minimal rip and replace.

However, if edge AI isn't integrated with larger systems, its impact is limited to discrete devices, and its data remains siloed. That's why Intel recommends a hybrid, edge-to-cloud approach to AI that combines real-time insight at the edge with deeper context generated in the cloud.

> Edge AI can run on existing brownfield infrastructure

[1] Infrastructure for AI at the Edge, IDC Market Perspective, Dec 2022.

09

Keeping data secure

Manufacturing data can be highly sensitive, especially in industries that work with extremely valuable IP like pharmaceuticals and biotech, semiconductors, and other high-tech manufacturing. Manufacturers with extremely sensitive data often choose on-premises computing over the cloud to reduce exposure, but this can limit AI-driven applications and innovation. With the right platforms in place—at the edge and in the cloud—manufacturers can stand up zero-trust security architectures that combine endpoint security management at the edge with AI-supported monitoring and remediation in the cloud.

Solving the challenges of deploying AI in manufacturing

We've laid out nine major challenges for manufacturers who want to put safe, reliable AI into production. The good news: there are solutions to all of these and you don't have to go it alone. Major firms, systems integrators, hardware manufacturers, and software developers—including Intel and our partners—have pioneered AI in manufacturing. The segment has matured rapidly from pilots and prototypes to turnkey solutions.

We're solving the other challenges with edge computing platforms, model training applications, reference software, AI toolkits, and model-serving solutions. Our approach to AI draws on an open ecosystem that includes solutions from Intel, our partners, and the open-source community.

01

Run the edge with cloud-native tools

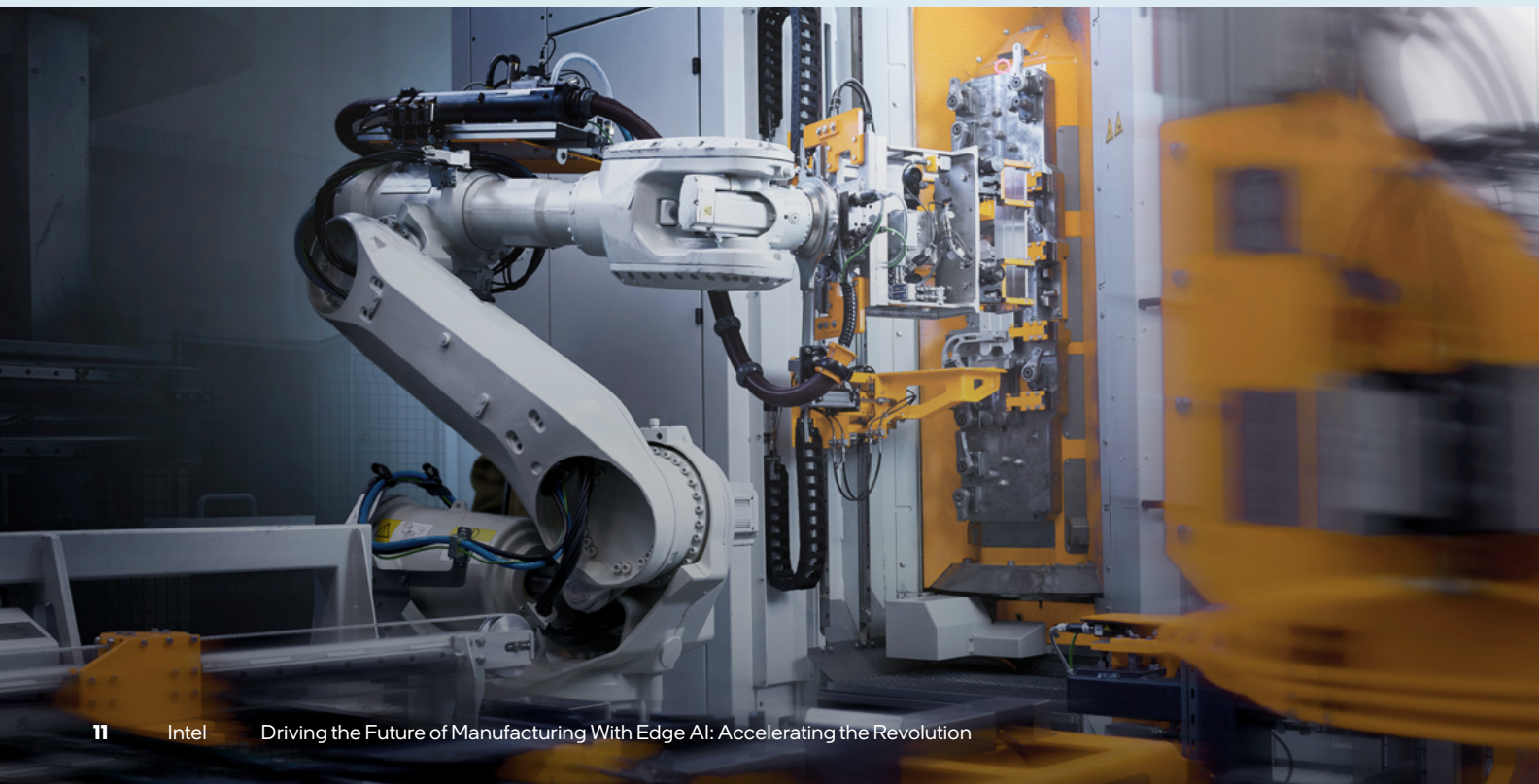
Modernizing edge computing is fundamental to deploying AI in manufacturing. The Intel® Tiber™ Edge Platform provides a cloud-native foundation for the edge that includes data management, infrastructure admin, and application orchestration. The platform gives administrators a single pane for managing complex IT/OT systems, automating operations, and deploying containerized, microservices applications.

We designed the platform to run on existing brownfield hardware so manufacturers don't need to rip-and-replace to field an AI-capable edge platform. The platform puts MLOps components in place for hybrid AI that can run on-premises or in the cloud.

02

Collect, annotate, and govern manufacturing data

Integrating industrial sensors and control systems requires middleware that can extract data, annotate it, and serve it up to downstream applications. Intel® Edge Insights System performs this role—and more—with a portfolio of microservices applications, including applications for ingesting structured and unstructured data, filtering and annotating video, and backing up data to the cloud.



03

Keep subject matter experts in the AI loop

The current tools for training models are not user friendly, but it doesn't have to be that way. We've engineered a training tool called the Intel® GETi™ platform that learns dynamically as the user annotates images with a mouse or a stylus.

For example, a quality control expert might use the Intel® GETi® platform to mark defects in inspection images. The Intel® GETi® platform begins learning immediately and starts identifying defects. The expert then approves or corrects the model's predictions. The process repeats until the model reaches production-ready accuracy. Depending on the complexity of the task, subject matter experts can train an effective model in a few training sessions.

The Intel® GETi® platform can output production-ready models optimized for the OpenVINO™ toolkit. Developers can bring fine-tuned models into their applications and deploy across mixed hardware from edge to cloud.

By opening model training to non-data scientists, the Intel® GETi® platform can help overcome the initial hurdle of getting into production and the long-term challenges of governing model behavior, retraining to correct model drift, and maintaining accuracy.

04

Build, optimize, and deploy learning AI applications edge to cloud

Optimizing AI models and developing applications for the mix of hardware and processor types that span edge to cloud is daunting. With the OpenVINO toolkit, developers can optimize models and convert them to OpenVINO™ format which uses a plug-in architecture to target specific processor types—CPUs, GPUs, and accelerators—for maximum performance on infrastructure from the edge to the cloud. The OpenVINO™ runtime engine is particularly useful for brownfield edge computing because it optimizes inference performance for the low-latency, low-power processors that drive many IoT functions.

Deploying edge AI unlocks new levels of manufacturing automation

Once AI-capable edge platforms and services are in place, manufacturers can use AI to remake existing operations and deploy entirely new ways of operating. There are two broad computing categories that can benefit from edge computing and AI: real-time functions that must operate at machine speed and less time-critical computing that supports broader insights.

Real-time AI unlocks new levels of automation

Autonomous material handling
Deploying edge AI helps make autonomous, robotic forklifts, pickers/packers, and sorting equipment a reality, accelerating operations and reducing injuries.

Defect detection
AI can detect defects as they happen and automatically halt operations, preventing costly errors and material waste.

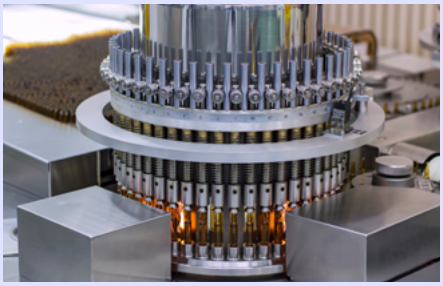
AI-enhanced worker safety
AI can help protect employees by monitoring human-machine interactions and shutting down equipment when it detects a risk of injury.

Continuous process improvement
AI produces real-time insights that autonomous systems can use to optimize efficiency and minimize downtime automatically.

Live digital twins
Transform data from factory cameras, LIDAR, RADAR, and environmental detectors into a simulated environment capable of supporting its own AI. Manufacturers can understand the present, model the future, and simulate changes without interrupting production.

Edge AI can transform manufacturing fundamentals

Flexible manufacturing configuration
Edge AI empowers you to configure and optimize your production lines for maximum efficiency, then reconfigure them to produce a different product.



Post-production defect detection
Ensure the highest quality products by leveraging AI for post-manufacturing defect identification.

Predictive maintenance
Move from reactive maintenance to proactive problem-solving with AI-powered equipment health monitoring.

Deploy edge AI faster with proven solutions from Intel partners

The Intel partner ecosystem is vast, innovative, and filled with ready-to-deploy edge AI solutions, resources, and expertise that simplify AI deployment. Get up and running quickly with pre-built, industry-specific solutions that leverage best-in-class hardware, software, and expertise to address your unique manufacturing needs—from process automation and worker safety to inventory management and predictive maintenance.

Intel works with major cloud service providers (CSPs) to ensure seamless edge-to-cloud performance and interoperability with Intel-based edge solutions. Intel's extensive list of leading systems integrator (SI) partners with manufacturing domain expertise can help deploy edge AI solutions that integrate seamlessly with existing systems and minimize development time.

Explore solutions in the Intel Partner Showcase. →

Open, flexible software solutions for manufacturing

Independent software vendors (ISVs) offer a wide range of AI applications for manufacturing process automation, worker safety, inventory management, and more. Many offerings are built on open-source foundations for greater flexibility in development and integration with other systems.

Leading-edge industrial and IoT hardware

From hardened systems that can handle shock, water, and high temperatures to power-efficient designs for resource-constrained edge environments, Intel's original equipment manufacturer (OEM) and original design manufacturers (ODM) partners deliver high-performance computing that can stand up to industrial use.

Experienced system integrators that specialize in manufacturing

SIs bring the in-depth knowledge of Intel technologies and industry-specific expertise you need to stand up new manufacturing technologies without disrupting production.

Hyperscaler support and collaboration

Intel works with major CSPs to ensure seamless edge-to-cloud performance and interoperability with Intel-based edge solutions. By integrating edge-native and cloud platforms, edge devices can handle real-time data collection and processing tasks while less time-critical processes—like storage, complex computations, and model training— run concurrently in the cloud.

Intel solutions for every step of your edge AI journey

Intel is a leader in providing compute for both IT and OT environments. We know how to help clients deliver on their core business of manufacturing, distributing, processing, healing, and operating their critical environments. This foundation makes Intel the natural partner for the pragmatic implementation of AI for process and quality improvement.

Intel helps remove the data management and integration blockers that prevent edge AI from delivering on its business promise by ensuring you run AI workloads in the right place—on the edge or in the cloud—with the proper attention to security, cost, power, and results.

Simplify edge development, deployment, and management	The Intel® Tiber™ Edge Platform puts cloud-native infrastructure management, application orchestration, and security to work across greenfield and brownfield edge nodes.
Deploy and manage AI in industrial applications.	Intel® Edge Insights System offers data ingest, model management, registry, data collection, model deployment, and inferencing for video and time series data. It provides services for defect detection and predictive analytics in manufacturing, energy industries, and beyond.
Develop, optimize, and deploy edge AI inference applications	Fine-tune models, develop AI inference applications, and deploy to any mix of Intel hardware from the edge to the cloud with the OpenVINO™ toolkit.
Run AI on existing, low-power edge hardware	OpenVINO™ runtime optimizes performance for low-latency and low-power use of AI inference.
Help subject matter experts train their own AI	The Intel® GETi® platform automates model training with smart annotations and active learning so non-data scientists can train their own expert models.
Track the present and simulate the future with digital twins	Intel® SceneScape goes beyond vision-based AI to create a real-time, 4D digital twin from LIDAR, RADAR, and environmental sensor data. Intel® SceneScape can bring new insight to past analytics, track what is happening in the present, and make predictive decisions for the future.

Unlock the full potential of AI from edge to cloud with Intel® hardware

Intel offers hardware platforms, from edge devices to data centers, which deliver the performance you need to run AI workloads efficiently. Intel® CPUs, GPUs, and dedicated AI accelerators can boost AI performance across the IT/OT spectrum. Our optimized solutions for popular frameworks like PyTorch and TensorFlow ensure you can maximize the value of your AI investment.

Deploy AI in power-constrained edge applications with Intel® Core™ Ultra processors

Intel® Core™ Ultra processors combine CPU, GPU, and a neural processing unit (NPU) in our most power-efficient design to date—up to 2.56x better AI performance per watt than the previous generation.[2] They bring flexible general capabilities plus video and accelerated AI to edge devices.

Accelerate AI, media, and HPC workloads with Intel® Data Center GPUs

Intel® Data Center GPUs accelerate media streaming, HPC workloads, and AI workloads at scale. Manufacturing applications include cloud and on-premises data centers. Each GPU core processes vector and matrix AI workloads faster thanks to Intel® Xe Matrix Extensions (Intel® XMX), a deep systolic array architecture on every Intel® discrete graphics cards.

Train and deploy trillion-parameter AI models with Intel® Gaudi® AI accelerators

Intel® Gaudi® AI accelerators are designed for large-scale deep learning AI training and inference. They deliver superior price per performance on AI workloads and work with ethernet-based systems—no specialized networking required.

Intel hardware platforms feature built-in AI acceleration

Matrix math operations are a key inference workload for natural language processing (NLP), recommendation systems, and image recognition. Intel® Advanced Matrix Extensions (Intel® AMX) and Intel® Deep Learning Boost (Intel® DL Boost) accelerate matrix math on select Intel® CPUs, boosting AI performance. Vector-based computation used for analytics and AI processes can run fast on Intel® Core™ and Intel® Xeon® processors with Intel® Advanced Vector Extensions 512 (Intel® AVX-512).

[2] Performance varies by use, configuration, and other factors. Learn more at intel.com/processorclaims: Intel® Core™ Ultra processors, Edge. Results may vary.

Take the next step in your AI manufacturing journey

AI is evolving rapidly and gaining new capabilities constantly. Establishing a solid, edge-native foundation for data, security, and AI deployment is essential for staying competitive today and being prepared for the innovations on the horizon. Intel and our partners are ready to help you put AI to work and transform your manufacturing operation, no matter where you are in your journey.

Learn more about AI in manufacturing at intel.com/manufacturing.

0524/MN/CAT/PDF 360370-001EN