Porsche Consulting
Strategic Vision. Smart Implementation.

× semi™ SMART MANUFACTURING

From Hype to Impact

Realizing AI's potential in semiconductors manufacturing

INSIGHTS

01 AI-driven transformation constitutes a high-risk, high-reward strategic initiative for semiconductor manufacturers. In the European context, AI can act as a catalyst for competitiveness, helping the sector address persistent structural challenges such as elevated energy costs, talent scarcity, and the complexity of modernizing legacy brownfield sites.

02 Think big, start small - Unlocking the full potential of AI requires a systematic transformation that fosters a robust ecosystem both technologically and regulatorily. While this long-term shift is essential, industry players must begin now by experimenting with targeted use cases, cultivating strategic partnerships, and embedding AI as a company-wide strategy.

03 A key priority lies in establishing the correct data strategy and infrastructure, along with the usage of SEMI standards to transform installed equipment base into data-generating digital replica of physical fab tools, unlocking the potentials that AI-usage will bring forward.

04 Fab parameters such as yield and tool uptime will be boosted by up to 20%, driven by innovative use cases in scheduling, quality control and the MES as the central orchestrator, if common hurdles of implementation can be overcome.

CONTENTS

SEMI AND PORSCHE CONSULTING COLLABORATION

Since 2023 Porsche Consulting has been collaborating with SEMI in various initiatives and provided strategic insights into the End-to-End Smart Manufacturing community. Leading the European chapter for smart manufacturing as a chairman, Porsche Consulting has jointly with the members of the initiative worked on a white paper to uncover the benefits and challenges of Artificial Intelligence in the realm of Semiconductor Manufacturing in a practical and for the reader encouraging, helpful way.

The European Chapter of the Smart Manufacturing Community is dedicated to the vision of an autonomous factory which is end to end vertically integrated along the Semiconductor manufacturing value chain. Contribute to the strategic vision of a competitive Europe in the global arena. Its mission is to unite industry experts, innovators, and thought leaders, catalyzing collaboration that leads to the development and implementation of robust solutions and standards for smart manufacturing across both front-end and back-end Semiconductor production in Europe. It aims to construct best practices across the Semiconductor manufacturing value chain

MEET THE END-TO-END SMART MANUFACTURING COMMUNITY

01

Porsche Consulting
Strategic Vision. Smart Implementation.

INTRODUCTION

AI enabling role for Europe’s semiconductor industry ambitions

For Europe, AI represents not just an enabler, but the true accelerator of competitiveness

The semiconductor industry is undergoing rapid expansion, with annual revenues projected to reach $1 trillion by 2030 and more than 100 new fabs anticipated worldwide by 2027 [1]. However, this growth is accompanied by increasing manufacturing complexity, rising costs, persistent supply chain disruptions, and heightened geopolitical risks—all of which threaten industry competitiveness. Traditional automation and digitalization, while foundational, are no longer sufficient to deliver the efficiency, quality, and agility required at this scale.

Within this global context, European semiconductor manufacturers face additional challenges. Compared to peers in the US and Asia, they must navigate lower levels of public investment and subsidies, greater regulatory complexity, higher utility and labor costs, and smaller average operational scale.

This is where Artificial Intelligence (AI) comes in. AI is already impacting semiconductor manufacturing by analysing vast datasets in real time, enabling early detection and prediction of equipment failures.
Rapidly maturing, AI is predicted to improve yield by up to 30%, reduce unplanned downtime by 40%, reduce cycle times by 20% and even improve time to market by over 30%, through virtual qualification.

For European semiconductors industry then, AI is not merely a promising technology, but potential catalyst and accelerator for transforming the sector, enabling the region to regain competitiveness.

AI AS THE ACCELERATOR, NOT JUST AN ENABLER

01

Efficiency
AI reduces cycle times and cuts downtime

02

Quality
Yield improvement and potential to simplify key processes like qualification

03

Sustainability
Facilitates flexible production and greener operations

04

Talent Bridging
Mitigate skill gap enhancing augmented operations, remote monitoring and automated decision making

01

Technical Challenges
Lack of clean, structured, and comprehensive data. Complex integration with thousands of diverse tools and systems.

02

Organizational inertia
Legacy structures and siloed decision-making. Digital units with no authority to influence core process architectures.

03

Lack of Ecosystem
Limited collaboration and rigid regulatory frameworks between manufacturers, customers, suppliers.

CHALLENGES TO OVERCOME

Figure 1: Benefits of AI and key challenges that must be overcome to achieve them

THE EUROPEAN SEMICONDUCTOR SECTOR STANDS AT A CRITICAL JUNCTURE

Despite the European Union's ambitious goal to capture 20% of global semiconductor market share by 2030, independent audits and industry experts warn that this target is "deeply disconnected from reality" [2]. As of 2022, Europe's share was just 9.8%, with projections suggesting only a modest increase to 11.7% by 2030, far short of what is required to meet strategic autonomy goals.

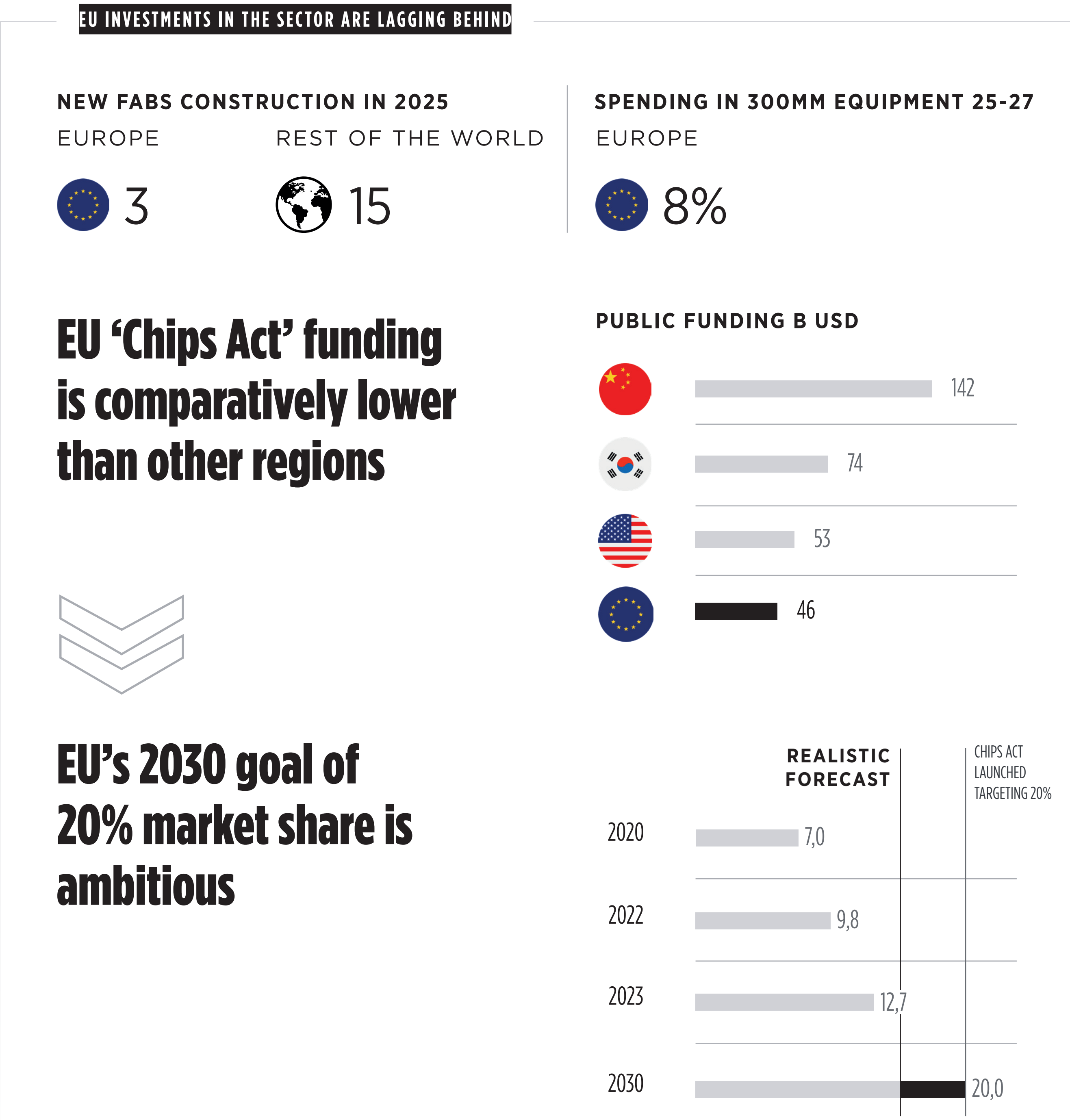


Figure 2: EU investment in Semiconductor sector

Europe's market share gap in semiconductors is closely tied to structural challenges across the region (see Figure 3), which have historically placed it at a disadvantage compared to e.g. Asia and the United States. AI presents a strategic opportunity to overcome these constraints, enabling European manufacturers to do more with less and gain a competitive edge in a rapidly evolving global landscape.

What is holding EU back

01 HIGH UTILITIES COSTS
60 to 150% higher than US or China

02 TALENT SHORTAGES & LABOUR COST
- EU faces a shortfall of up to 350,000 skilled workers by 2030
- Labour cost is 3-4 higher than Asia

03 INDUSTRY FRAGMENTATION
- Regionally dispersed industry
- No global leader in semicon manufacturing
- Lack of ecosystem centralized hubs

04 AUTOMOTIVE SECTOR DEPENDENCY
EU players traditionally exposed to automotive, currently volatile

05 BROWNFIELD SITES AND LEGACY INFRASTRUCTURES
Many EU fabs are older and focused on mature nodes, while leading-edge nodes are dominated by Asia and US

06 RAW MATERIAL SUPPLY CHAIN VULNERABILITIES
EU is heavily dependent on third markets, especially China, for critical semiconductor materials

07 ELEVATED FAB CONSTRUCTION COSTS
Building a fab in Europe costs twice as much as in Asia

08 EXTENDED LEAD TIMES FOR PLANT COMMISSIONING
Fab construction in Europe takes ~34 months, compared to 19 in Taiwan and 23 in Southeast Asia

09 HIGH REGULATORY COMPLEXITY
Fragmented framework with EU directive and national laws result in slow execution speed

10 LACK OF SEMICONDUCTOR INTEGRATED CLUSTER
Development of semiconductor clusters that integrate R&D, fabrication, and equipment manufacturing is significantly less advanced than in Asia

Figure 3: Factors holding back EU Semiconductors industry competitiveness

CONSIDER THE FOLLOWING EXAMPLES:

01 Europe's semiconductor sector faces a critical skill shortage, with industry reports projecting a gap of over 100,000 engineers and technicians annually. This shortfall spans every part of the value chain, from fab operations to chip design and packaging, at a time when AI-driven automation is reshaping global manufacturing. To compensate, European fabs can leverage on AI to automate complex decisions, support remote operations, and accelerate knowledge transfer, helping bridge talent gaps and boost operational resilience [3].

02 AI also plays a pivotal role in reducing manufacturing cycle time and mitigating process variability, key challenges in a field where small deviations can lead to disproportionate losses. Innovations like predictive process control and faster root-cause analysis are now helping fabs both improve yield and speed product qualification, a step that can represent more than 50% of a new product's development timeline. The SEMI Smart Manufacturing Report 2024, developed with Porsche Consulting reveals that facilities adopting smart scheduling with AI have achieved a 9% increase in throughput and a 15% rise in operator value-added time [4].

03 Operating across a landscape dominated by brownfield sites and legacy equipment, European manufacturers depend on dynamic scheduling and predictive maintenance to maximize equipment utilization and optimize yield. According to Applied Materials, fabs deploying AI in predictive maintenance have cut unplanned downtime by up to 40%, while yield optimization has driven multi-percentage-point improvements in output [5].

04 Beyond immediate efficiency gains, European companies embedding AI throughout the manufacturing chain are laying the groundwork for future growth. Enhanced mass customization, resilient supply chains, and improved sustainability are already defining the approaches of global leaders such as TSMC, Intel, Samsung, and NVIDIA. For Europe, accelerating the adoption of AI isn't just a catch-up game, it's a chance to redefine its competitive role in the global semiconductor industry [6]

Given its potential to redefine the paradigm of competitiveness and help Europe regain ground, AI must not be treated merely as an IT tool.

Achieving success will require a structural transformation that spans the entire value chain—from how companies integrate AI into their core strategies, to how policymakers foster a more agile and innovation-friendly ecosystem.

Having explored AI's enabling role in addressing Europe's structural challenges, we now shift focus to its transformative potential, illustrated through ambitious applications, the emerging challenges that accompany this shift, and the strategic priorities the industry must embrace to lead the change.

Redefining possibility: AI's visionary impact on semiconductor design and manufacturing

Each new semiconductor product requires several years to progress from development to qualification and deployment, not accounting for the timelines of downstream technologies. One of the most promising ways Artificial Intelligence (AI) is reshaping the semiconductor industry is by virtualizing key process stages accelerating development cycles, improving yield and reliability, and reducing reliance on physical prototyping.

Already today semiconductor fabrication is advancing thanks to AI. IBM and SCREEN Semiconductor Solutions are collaborating on cleaning technologies for High-NA EUV lithography, addressing sub-2nm contamination risks [7]. With mature development, AI-enabled advanced process control systems are poised to improve yield of up to 30%, potentially saving manufacturers millions by identifying latent defects and minimizing production losses [8].

Cross-industry applications further underscore AI's transformative impact in ways that could soon be applicable even to semiconductors complexity. As an example, Porsche accelerated development of its electric Cayenne SUV by 20% through AI-driven simulation, replacing 120 physical prototypes with virtual models and reducing resource consumption [9]. Mercedes-Benz, in partnership with Nvidia [10], deployed a software-defined architecture based on the NVIDIA DRIVE™ platform, enabling continuous over-the-air updates and shifting automotive design toward software-centric paradigms.

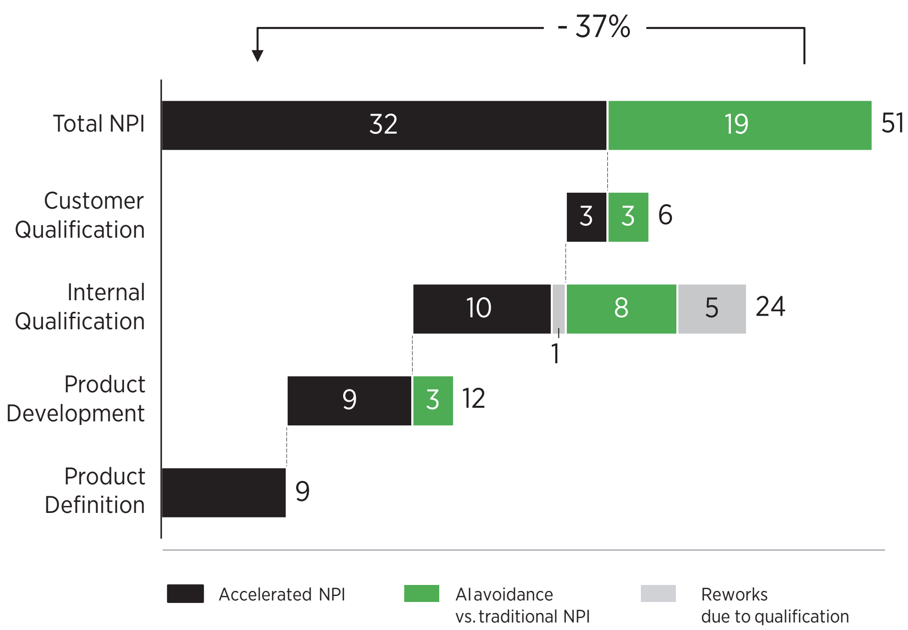
Those who master AI qualification will redefine semiconductor competitiveness, saving up to 1 year and 50M$ per launch

AI has the potential to streamline one of the most resource-intensive phases of semiconductor manufacturing: product and technology qualification. Traditionally consuming up to 50% of the New Product Introduction (NPI) timeline, qualification involves high capital expenditure and operational risk. AI-driven predictive models enable early anomaly detection and real-time deviation forecasting, reducing rework and accelerating time-to-market.

To mitigate challenges from long qualification times, manufacturers have adopted strategies such as similarity-based product architectures, which enable qualification by analogy, and virtual testing of product designs prior to fabrication. Despite these advancements, physical qualification remains indispensable to ensure that theoretical design parameters are accurately replicated in the manufacturing environment. This involves extensive stress testing, often spanning thousands of hours, and meticulous monitoring of hundreds of process steps, where even minor deviations can result in critical device failures. Such failures, typically detected post-production, can incur substantial costs due to redesign, re-manufacturing, and re-testing. The adoption of virtual qualification methodologies presents a significant opportunity to enhance competitive advantage in platform development. By replacing traditional physical testing with digital simulation and validation, we estimate that semiconductor manufacturers can reduce platform qualification timelines by over 12 months. This acceleration not only enables faster market entry but also contributes to substantial financial impact—potentially lowering circulating capital at risk by approximately €50 million per platform. These benefits underscore the strategic value of integrating virtual qualification into core development processes.

Yet, despite the benefits that AI promises to bring, various challenges are still holding organizations back, and this is particularly the case in semiconductors manufacturing.

AI-DRIVEN QUALIFICATION ENABLES NPI EXCELLENCE



-37%

TIME TO MARKET

Due to faster (les tests) and more predictable (expected outcome accuracy) qualification timeline

-53%

PRODUCTION AT RISK

Due to less production material at risk of scrap throughout validation & qualification stages

+30%

INNOVATION CAPABILITIES

Due to decreased costs of early experimentation & validation before production

Figure 4: Estimated impacts of AI-Driven qualification on New Product Introduction (NPI). Source: Porsche Consulting

OVERCOMING BARRIERS REQUIRES MORE THAN TECHNOLOGY, IT DEMANDS A TRANSFORMATION.

AI adoption in semiconductor manufacturing requires more than technological advancement—it demands systemic transformation. Despite proven applications in predictive maintenance and process optimization, industrial AI remains complex due to the need for tailored, use-case-specific models [11]. More specifically, AI implementation in semiconductor manufacturing is hindered by technological, organizational, and process-related constraints. Technological barriers stem from legacy infrastructure and heterogeneous systems that limit data integration and model scalability. Organizational barriers arise from siloed functions and insufficient cross-functional collaboration, which impede the operationalization of AI outputs. Process-related barriers involve rigid production architectures that are not designed for real-time, data-driven workflows, requiring structural redesign for effective AI deployment. Virtual qualification illustrates these challenges, as it often lacks sufficient historical data and requires advanced methods such as synthetic data generation and transfer learning. Without resolving these barriers, AI cannot be institutionalized across the ecosystem, limiting its impact on validation and operational efficiency.

TECHNOLOGICAL BARRIERS

AI systems must be customized to specific processes and integrated with legacy infrastructure, which often lacks standardized interfaces. Data quality and availability remain critical constraints, particularly in virtual qualification where historical data is limited.
These challenges are particularly acute in semiconductor manufacturing. Equipment heterogeneity, strict security protocols, and the absence of historical data—especially in virtual qualification—complicate AI deployment. For instance, qualification processes often lack sufficient data for training, requiring alternative approaches such as synthetic data generation or transfer learning [12].
Techniques such as synthetic data generation and transfer learning offer viable alternatives. Harmonizing heterogeneous data environments—e.g., Intel's fabs with over 2,700 systems [13]—is a prerequisite for effective AI deployment.

ORGANIZATIONAL BARRIERS

From our experience with customers, legacy structures and siloed functions are a major obstacle to efficient AI integration, as use cases are developed in isolation and lack strategic coordination missing the opportunity to scale across the enterprise. Moreover, embedding algorithmic insights into operational workflows requires cultural change, upskilling, and new roles. TSMC's cross-functional teams exemplify successful integration, combining engineering and data science to validate AI outputs. This approach reduces resistance to change and facilitates the integration of AI into standard operating procedures [14]. Qualification processes must shift from sequential handovers to collaborative, data-driven models. Instead, a cross-functional, data-driven approach is required—one in which all stakeholders are involved from the outset and throughout the process lifecycle. This demands a dedicated data analysis function empowered to influence decisions and adapt processes based on algorithmic outputs.

PROCESS BARRIERS

AI can hardly be retrofitted into existing systems; it requires reengineering of production architecture. At Porsche's Leipzig plant, AI-enabled defect detection illustrates how embedded systems improve decision-making. This data-driven feedback loop enables earlier detection of process deviations, allowing for timely corrective actions [15]. In semiconductor fabs, legacy equipment limits real-time data exchange, necessitating a data-centric architecture. Qualification virtualization depends on ecosystem-wide alignment across OEMs, suppliers, and regulators to institutionalize AI-driven validation. The AI-driven approach mirrors the impact of simulation technologies introduced decades ago, which revolutionized chip design by front-loading risk mitigation [16].

Europe's semiconductor AI evolution requires an ecosystem

AI-driven transformation constitutes a high-risk, high-reward strategic initiative for semiconductor manufacturers. The complexity of implementation extends beyond IT infrastructure upgrades, requiring substantial investments in organizational capabilities and process redesign. The potential benefits, however, are equally significant: AI can accelerate innovation cycles and reduce fab capacity absorption by lowering engineering resource requirements. In a sector where production throughput and continuous innovation are critical success factors, AI stands to be a transformative force.

However, the realization of AI-enabled virtualization cannot be achieved in isolation. It demands an enabling ecosystem, both technologically and regulatorily. In particular, regulatory constraints, especially within the European Union, present substantial barriers to full virtualization. For instance, qualification processes are hindered by rigid frameworks that limit the adoption of AI-driven methodologies. This underscores the broader need for harmonized European action to modernize standards, promote cross-border collaboration, and create a regulatory environment conducive to innovation [17].

Achieving technological sovereignty in Europe requires competitive regulatory conditions that balance risk mitigation with innovation support, alongside reductions in energy costs. EU stakeholders must intensify collaboration and foster alliances that enable data sharing and joint development of AI-based business models. While the AI Act represents a foundational step in regulating AI, it is widely perceived as overly restrictive, potentially stifling the dynamism required for growth.

Moreover, the deployment of advanced AI-driven process enhancements will necessitate close cooperation between manufacturers and end-users, as well as a reassessment of legacy regulatory frameworks. Key downstream sectors, such as automotive, impose stringent qualification requirements that constrain development velocity and technological advancement [18].

To maintain competitiveness, Europe must establish a regulatory framework that facilitates, rather than impedes, domestic semiconductor innovation [19]. Addressing these challenges is not merely a technical necessity—it is a strategic imperative to position Europe at the forefront of AI-enabled semiconductor development.

Europe can gain competitive advantage by building a solid ecosystem

POLICY MAKERS

DESIGN

WAFER MANUFACTURING

FRONT END MANUFACTURING

ASSEMBLY TESTING AND PACKAGING

COMPONENTS INTEGRATION AND ECU

FINAL APPLICATION

EQUIPMENT MANUFACTURERS

SHIFT LEFT

Anticipate AI-driven validations and qualifications already in R&D

CONNECT RIGHT

Share data with suppliers, peers, and customers

BREAK OUT OF INTERNAL SILOS

AI initiatives should not be confined within individual companies but designed for broader, cross-organizational impact

FOSTER STRATEGIC COLLABORATION

Partnering with peers, clients, and policy-makers enables shared learning, resource pooling, and accelerates innovation

BUILD SHARED PLATFORMS

Co-developing interoperable platforms creates scalable AI solutions that benefit entire industries and ecosystems

Figure 5: Building a strong semiconductor industry ecosystem is essential for enhancing competitiveness

From strategic vision to practical impact: the experience of SEMI EU E2E Smart manufacturing community

The following chapters, result of the effort of SEMI EU E2E Smart manufacturing community and coordinated by Porsche Consulting as Chairman, present concrete perspectives from leading AI enabled solution providers and technology partners for the semiconductor manufacturing. Each contribution reflects the direct experience of companies actively engaged in developing and deploying AI solutions on the factory floor. By addressing critical domains such as data infrastructure, sensor intelligence, equipment integration, manufacturing execution systems, and production planning, these chapters provide both testimony of the challenges encountered and practical examples of how they are being solved. Together, they illustrate the most relevant applications of AI in today's semiconductor fabs and highlight the pathways toward scalable, impactful adoption.

IN PARTICULAR

ZEISS establishes the foundation by addressing the root cause behind most unsuccessful AI initiatives: insufficient data infrastructure. Their chapter highlights the role of an Industrial Data Platform in breaking down silos, ensuring data quality, and enabling scalable access across all layers of the manufacturing environment. Such structured data provision is the prerequisite for any advanced AI application.

Building on this, **INFICON** focuses on the point where data originates: the sensors within manufacturing equipment. By embedding AI capabilities directly into sensors—edge AI—manufacturers can analyze signals in real time, reduce latency, and lower data transfer costs. This approach enables immediate detection of process deviations and rapid corrective action, thereby enhancing yield and throughput.

While robust data platforms and intelligent sensors are essential, seamless equipment control remains a persistent challenge, **Agileo Automation** demonstrates how AI-driven log analysis transforms equipment integration by automating the review of complex communication and application logs. This shift enables proactive fault detection, faster root cause analysis, and continuous process optimization, reducing downtime and supporting smarter, more reliable manufacturing environments.

PEER Group extends this perspective, by examining the evolution of SEMI standards and the complexity introduced by legacy tools. Their Equipment Automation Framework illustrates how AI can enhance integration layers, ensuring interoperability across diverse equipment generations and creating a consistent operational technology environment.

Critical Manufacturing outlines a three-level AI integration in MES: connecting IoT and data platforms, leveraging GenAI for context-aware insights, and deploying AI agents for self-optimizing operations. This evolution transforms MES from a passive system to an intelligent hub, with AI agents capable of reasoning and self-optimizing operations enabling predictive maintenance, adaptive decision-making, and greater operational autonomy.

Finally, **Flexciton** addresses one of the most complex challenges: production planning and scheduling. Traditional rule-based approaches are no longer sufficient in highly dynamic fabs. By applying advanced optimization and AI, Flexciton demonstrates significant operational benefits, including cycle time reductions of up to 24%, improved throughput, and reduced manual intervention.

Together, these chapters present a coherent journey: from establishing structured data foundations and enabling intelligence at the sensor level, to overcoming integration barriers, transforming MES into an intelligent hub, and optimizing factory-wide planning.

Collectively, they illustrate how semiconductor manufacturing can progress from experimentation to tangible impact, laying the groundwork for resilient, autonomous, and future-ready operations.

FROM VISION TO ACTION:
THINK BIG, START SMALL

INSIDE ORGANIZATIONS: WHAT TO DO

01	02	03	04	05	06
Structured data provision via industrial data platform	**Factory wide equipment log analysis**	**Integration layer for seamless tool connectivity**	**Smart capabilities at the source (edge AI)**	**Adaptive MES supporting autonomous decisions**	**E2E optimized production planning and scheduling**

AI as enabler of talent shortage mitigation
e.g. via augmented operations, remote monitoring, enhanced decision making

Selective partneships with suppliers and clients
e.g. test Virtual qualification in less critical sectors

FOR COMPANIES

Start now, from data structuring and focused use cases, while embedding AI as a corporate wide strategy

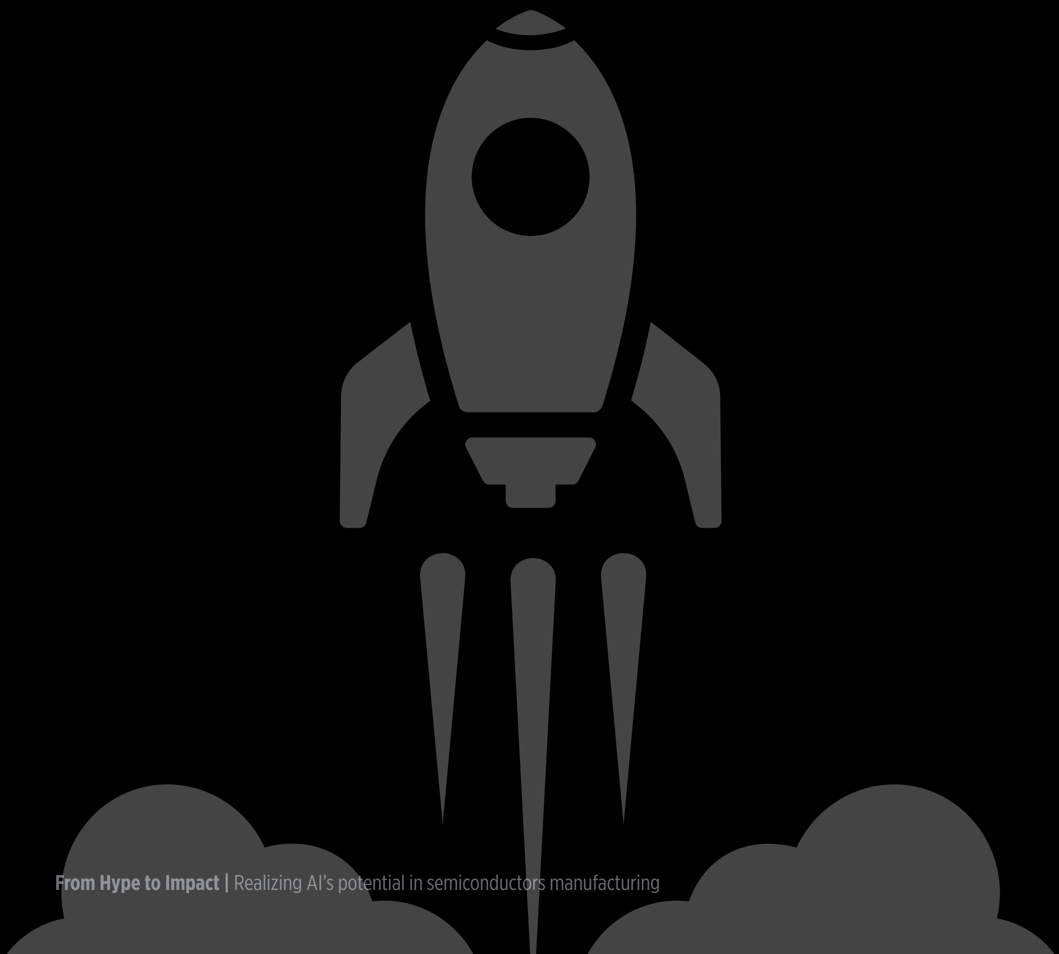
FOR POLICY MAKERS

Deploy regulatory fast-tracks, support European industry alliances

Figure 6: From vision to action

CONCLUSION

AI is our chance to turn limitations into strengths. But without action now, European companies risks staying a follower.



02

BREAKING THE DATA BARRIER

Unlocking AI's true potential in semiconductor manufacturing through structured data enablement

The semiconductor industry is on the verge of a profound transformation, driven by the promise of artificial intelligence (AI) to unlock new levels of productivity, quality, and agility. Yet, despite a wealth of innovative AI approaches, industry experience shows that nearly 90 percent of all AI use cases fail to progress beyond the initial phase [20]. This high failure rate is not primarily due to shortcomings in AI technology itself, but rather to fundamental issues with data availability and data infrastructure. In many fabs, the data required for AI projects is either inaccessible or lacks the necessary quality, and each new IT use case often triggers a separate, complex infrastructure project that is rarely scalable. This fragmented approach creates significant barriers to the integration of new IT technologies, particularly those like AI that depend on data from multiple logical levels within the manufacturing environment [21]. To overcome these obstacles, a structured and centralized approach to data provision is essential.

Only by establishing a robust data foundation can the industry realize the full potential of AI-driven applications, such as autonomous manufacturing, digital twins, and proactive process optimization.

This chapter explores how the concept of an Industrial Data Platform provides a practical and scalable solution to these challenges, enabling greater agility and laying the groundwork for data-driven innovation in semiconductor manufacturing.

Breaking down data silos with industrial data platforms

Traditionally, systems within a semiconductor fab communicate only with adjacent layers in the ISA-95 automation pyramid, with data typically aggregated as it moves upward such as in figure 7. This architecture leads to significant data silos, where information is isolated within specific software systems and is not readily accessible for cross-functional applications.

However, modern AI use cases—such as predictive maintenance, process optimization, and real-time quality control—require simultaneous access to data from all levels of the automation pyramid. The lack of a standardized, cross-layer data access mechanism makes it difficult to scale new digital solutions and severely limits the agility of manufacturing IT.

To address this, the Industrial Data Platform concept has emerged as a key enabler. By providing homogeneous, cloud-based data storage accessible to all relevant applications and technologies, the platform breaks down traditional silos and creates a unified environment for data-driven manufacturing. Cloud-based solutions, however, are optional. Companies can also build their own storage solutions.

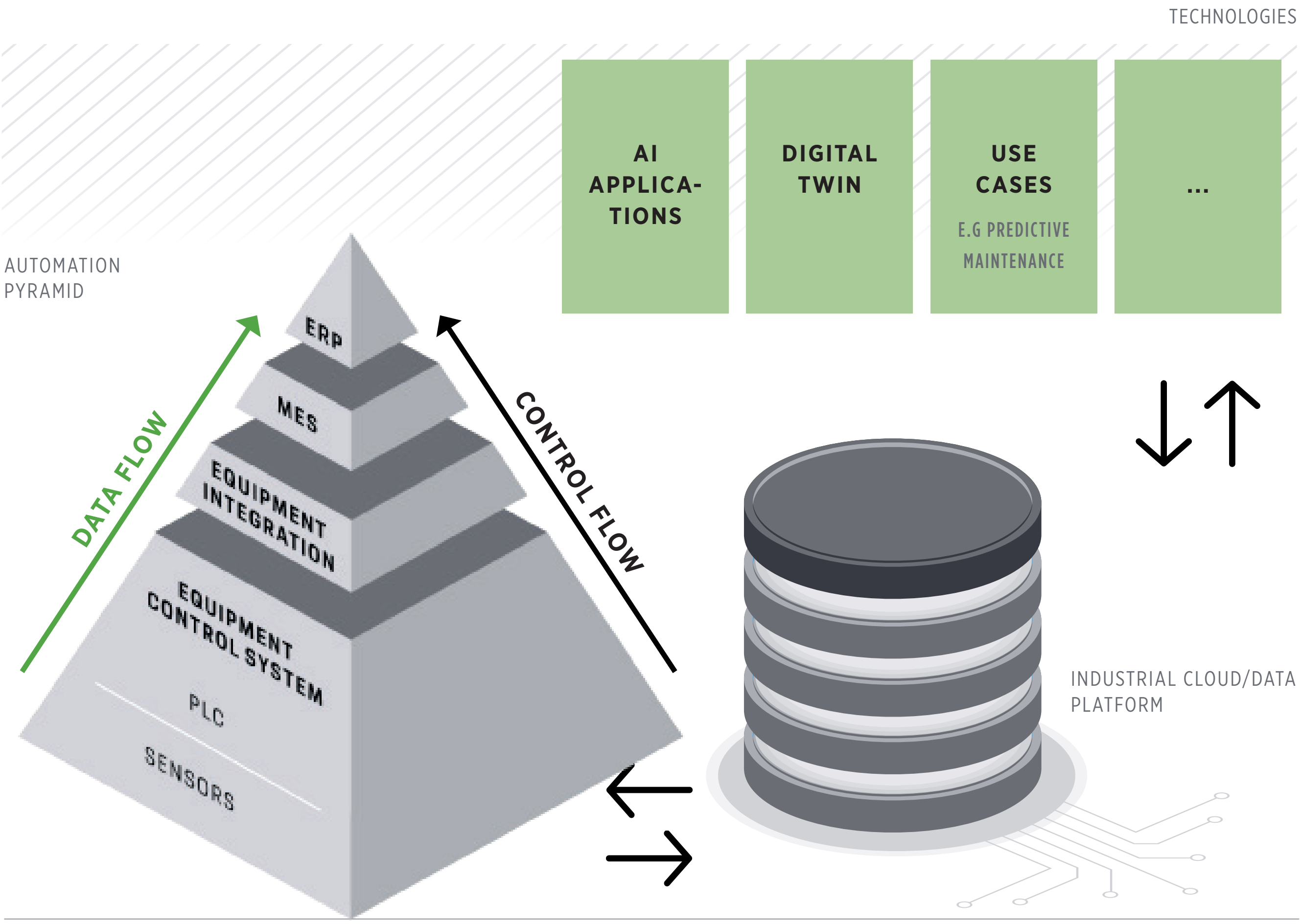


Figure 7: Data Exchange with a Data Platform [DP]

Batch vs. Streaming: the 2 types of data ingestion

A critical aspect of any data platform is the method by which data is ingested and made available for analysis. Two primary approaches are used in modern manufacturing environments: batch data ingestion and streaming data ingestion.

BATCH DATA INGESTION involves collecting data from various production processes with sampling intervals depending on use case requirements and transmitting it as a package for further analysis. This method is well-suited for use cases where real-time data is not essential, such as periodic reporting, production planning, or providing training data for machine learning models. Batch ingestion keeps infrastructure complexity and operational costs relatively low, making it ideal for overnight evaluations or scheduled analytics.

STREAMING DATA INGESTION, by contrast, transfers production data continuously and almost in real time as soon as it is generated. This approach is indispensable for time-critical applications, such as real-time equipment monitoring, live dashboards, or immediate alerting in the event of process deviations. While streaming imposes higher demands on data platform stability and scalability, it opens up new possibilities for adaptive process control and autonomous manufacturing. For many AI applications that depend on the most current data, streaming ingestion is essential.

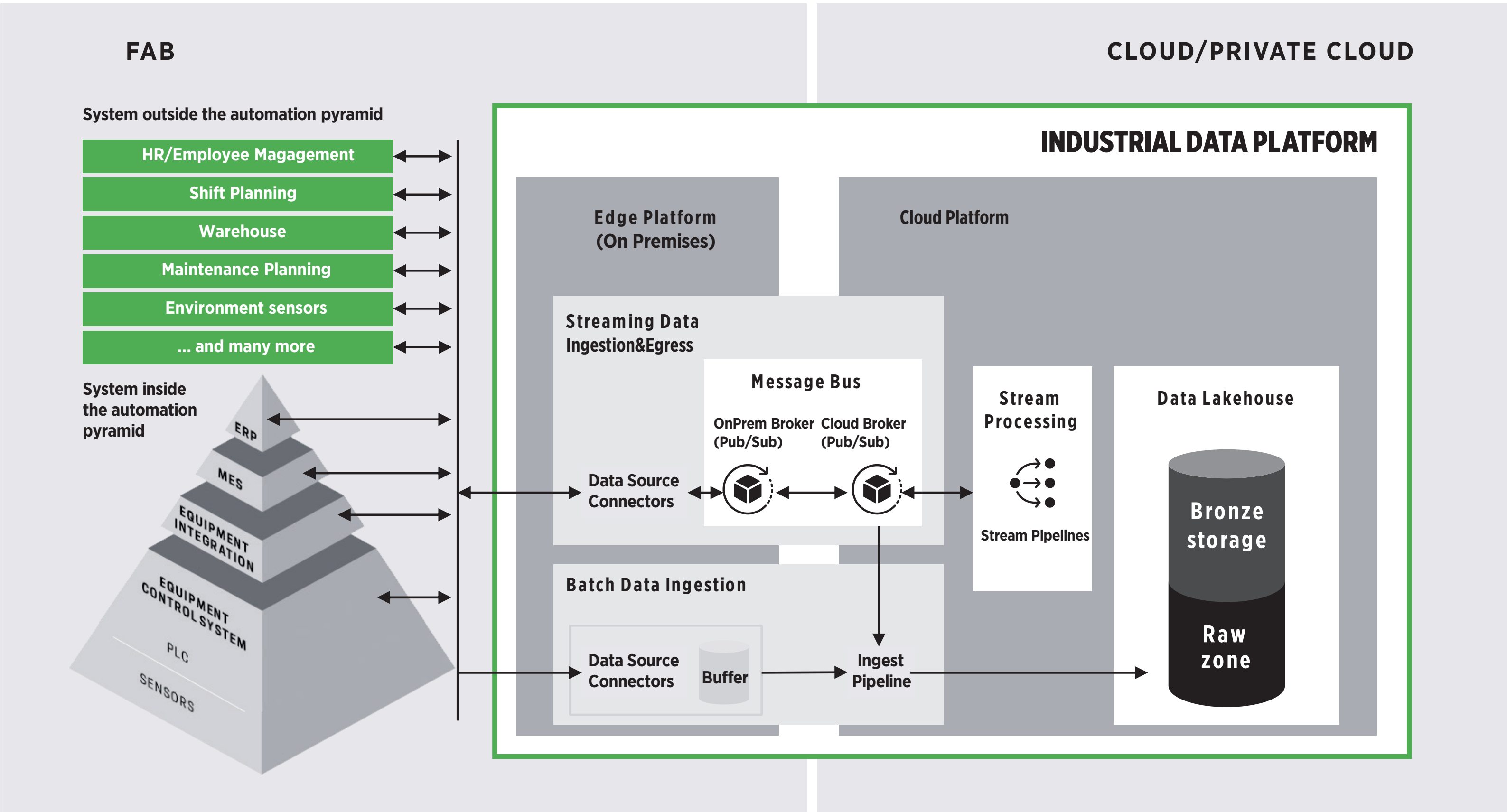


Figure 8: Data flow view Batch Data Ingestion and Streaming

The choice between batch and streaming ingestion should always be guided by the specific requirements of each use case. Batch methods are optimal for periodic, large-scale analyses with low time sensitivity, while streaming is the foundation for dynamic, real-time, and automated production processes (see table 1).

LATENCY RANGE	TYPICAL TIME PERIODS	EXAMPLES OF USE CASES	ADVANTAGE OF DATA INGESTION TECHNOLOGY
True Real-time	<= 10ms	Fast PLC reactions, robotic controls	No data ingestion technology makes sense because use cases are not suitable for a data platform
Near Real-time	10ms … 1s	Warnings for operators in the operation of machines and processes	Data Streaming
Average Latency	1s … 5min	Cycle time-dependent Overall Equipment Efficiency (OEE) reporting, general live dashboards, shift handover protocols, condition monitoring	Data Streaming
High Latency	>= 5min	KPI/reporting applications, training data for AI, planning results for predictive maintenance	Batch – Data Ingestion

Table 1: Data Latency Levels

From raw data to trusted insights: the need for industry standards

Once data is ingested, it must be transformed to ensure quality, consistency, and usability. The Industrial Data Platform employs a multi-stage data preparation process, beginning with the ingestion of raw data into a "Bronze Storage" layer. Here, data may still be incomplete or inconsistently structured. Through a refinement pipeline, data is cleaned, standardized, and enhanced, then stored in "Silver Storage" as quality-assured datasets.

Key transformation steps include filling gaps in time series, removing duplicates, standardizing units and field names, and enriching data with explanatory metadata. These measures ensure that downstream applications—whether for reporting, analytics, or AI—can rely on consistent, accurate, and up-to-date information. For specialized or performance-critical use cases, further condensed and analyzed data is stored in a "Gold Storage" layer, ready for direct consumption by advanced applications as illustrated in Figure 9.

The development of such solutions requires intensive effort due to their specific functionalities and the deep domain expertise needed. Here the development of industry-wide standards is expected to facilitate and accelerate the adoption also for companies with less expertise and capacity.

DATA LAKEHOUSE

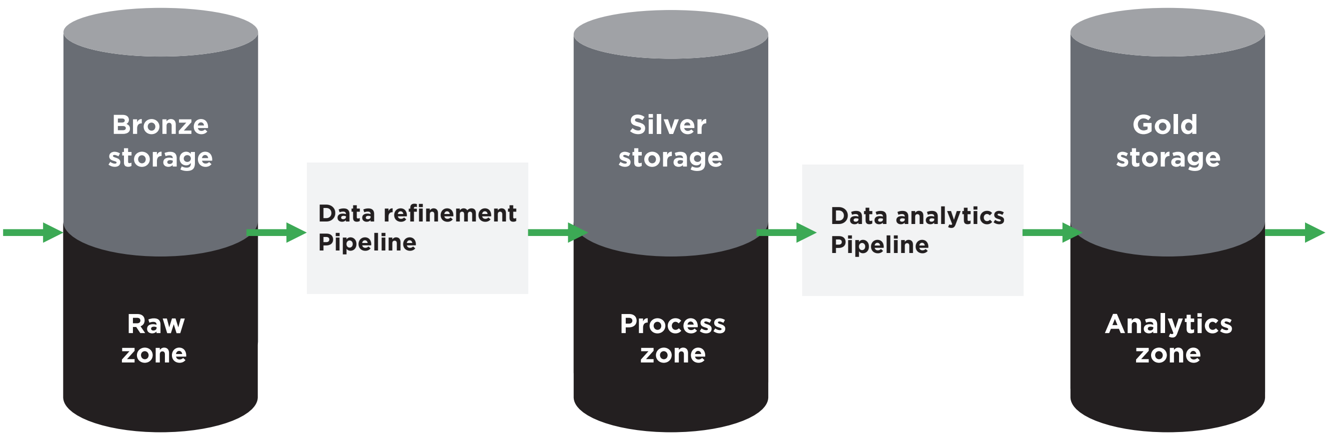


Figure 9: Overview of data storage and processing organization within the data platform

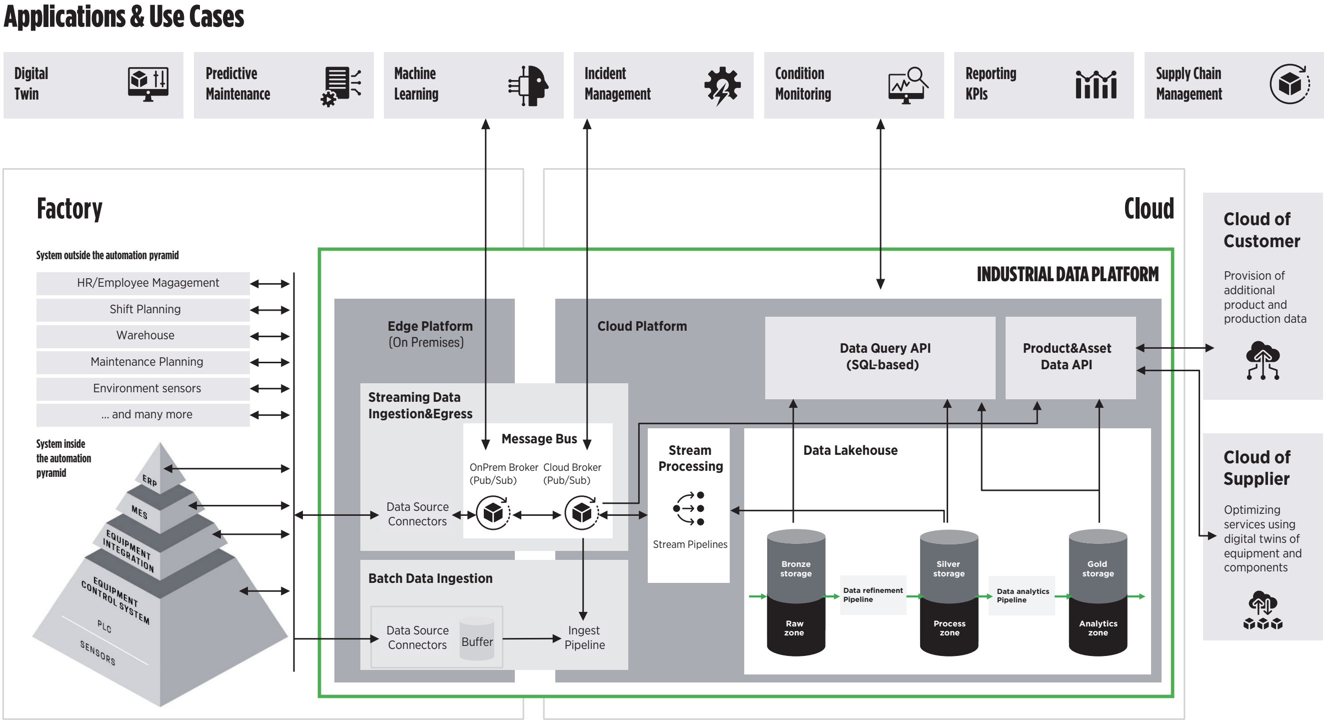


Figure 10: Concept of the Industrial Data Platform, overall overview

Getting flexible data access through APIs and streaming interfaces

Applications interact with the data platform through structured interfaces, most commonly a Data Query API. This API allows for efficient, secure, and controlled access to data at various quality levels (Bronze, Silver, Gold), supporting a wide range of use cases from simple reporting to complex AI-driven analytics. SQL remains the standard query language, offering familiarity and power for data scientists and engineers alike.

For real-time applications, streaming APIs or message broker technologies such as Apache Kafka or MQTT enable applications to subscribe to live data streams. This architecture supports continuous monitoring, instant alerts, and real-time process control, all of which are critical for the next generation of smart manufacturing.

Building the platform step by step allows companies to start with pilot projects

Implementing a comprehensive data platform (see Figure 10) is a journey that must be closely aligned with business priorities and use case requirements. Rather than attempting to build a fully mature platform from the outset, organizations are advised to adopt a phased approach, focusing first on high-impact, quick-win use cases.

Step 1

ANALYTICS & REPORTING

Begin with batch data ingestion and basic analytics to detect anomalies and generate targeted reports. This establishes the foundation for structured data collection and quality assurance as in Figure 11.

Step 2

PREDICTION & OPTIMIZATION

Expand to predictive models as: dynamic production planning, bottleneck analysis and predictive maintenance, leveraging both batch and streaming data ingestion as needed. This phase introduces more advanced analytics and begins to integrate real-time data flows.

Step 3

CLOSED-LOOP PROCESS CONTROL

Ultimately, the goal is to enable closed-loop, autonomous process control, where AI-driven systems can automatically adjust parameters in response to real-time feedback. At this stage, the full capabilities of the data platform are realized, supporting true autonomous manufacturing.

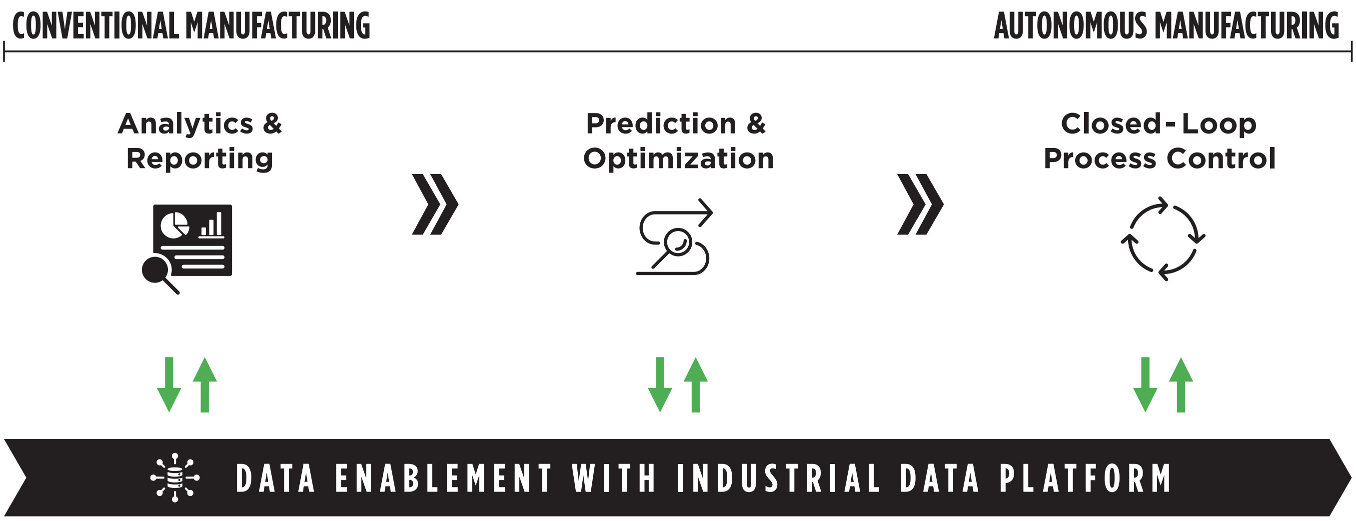


Figure 11: Roadmap towards autonomous production

The entry point and pace of adoption will vary depending on each organization's digital maturity, but the key is to ensure that every step is guided by clear business objectives and measurable outcomes.

The data platform well executed: digital shop floor management at Volkswagen

A leading example of the Industrial Data Platform approach, which could inspire the semiconductors industry, is Volkswagen's Digital Shop Floor Management (DSFM+), developed in collaboration with ZEISS Digital Innovation. DSFM+ is part of Volkswagen's "Digital Product Platform" (DPP), the implementation of a cloud strategy based on Amazon Web Services (AWS) for the entire group. DSFM+ replaces manual, paper-based data collection with a scalable, cloud-based solution that captures, evaluates, and visualizes production data across multiple plants. Core services include real-time machine data transfer, bottleneck analysis, and digital asset management, all built on a robust AWS-based architecture.

The introduction of DSFM+ has enabled Volkswagen to standardize production processes, automate shop floor tasks, and significantly reduce administrative overhead. By providing a unified, digital platform for production data, Volkswagen has increased transparency, comparability, and efficiency across its global manufacturing network. This real-world use case demonstrates the tangible benefits of structured data enablement and serves as a model for other semiconductor manufacturers seeking to modernize their IT [22].

Figure 12: AWS services used in the DSFM+ project at Volkswagen

CONCLUSION

Data enablement as a strategic imperative

The experience of ZEISS Digital Innovation and industry leaders like Volkswagen underscores the urgent need for structured data provision in semiconductor manufacturing.

An Industrial Data Platform is not merely a technical upgrade, but a strategic foundation for realizing the full potential of AI and data-driven innovation. By enabling seamless data access, transformation, and integration, such platforms empower manufacturers to pursue autonomous production, proactive process optimization, and agile business development.

The journey toward data-driven manufacturing is incremental and must be closely tied to business value. By prioritizing use cases, investing in scalable infrastructure, and leveraging proven technologies and standards, semiconductor manufacturers can lay the groundwork for a new era of smart, AI-enabled production.

03

INFICON

AI-DRIVEN DISRUPTION IN FAULT DETECTION

Enabling sensor-level intelligence in semiconductor fabs during wafer production

While a strong data platform is the backbone for AI in semiconductor manufacturing, data ultimately originates from the equipment and sensors within the FAB. Fault detection and classification (FDC) is a critical methodology using the collected data to ensure the reliability, yield, and efficiency of semiconductor manufacturing processes. Since semiconductor fabrication involves hundreds of highly complex, tightly controlled steps, even minor deviations or faults can cause defects, reduce yield, or lead to costly equipment downtime.

Sensors internal to the tool or externally installed by third party suppliers are indispensable for capturing vital process information during each step of the wafer process. These sensors monitor critical parameters such as process pressure, gas flows, gas composition, and plasma characteristics, providing essential insights into the state of the process and the health of the equipment.

Data can be collected from the tool sensors and any external sensors added to the tool for specific data application needs. Data from external sensors have multiple types of communication protocols requiring an interface to collect the data and make use of it. An FDC system is used in most cases to collect the data, perform the required analysis on the data, adding value and insight to the information collected. The data can then be transferred to the factory systems for further actions including stopping the tool to prevent further wafer scrap as an example.

Traditional FDC systems can be complex to set up due to the manual configuration required to address each FDC requirement. External sensors require specific setup and tool event data is required to synchronise the data with specific wafer process events. Installing a complete FDC system with external sensors can be time consuming and costly.

> Current Fault Detection and Classification (FDC) systems are often difficult to implement due to the need for manual configuration tailored to each specific requirement. AI functionality imbedded in the FDC system add latency, which hinders real-time optimization and quick responses

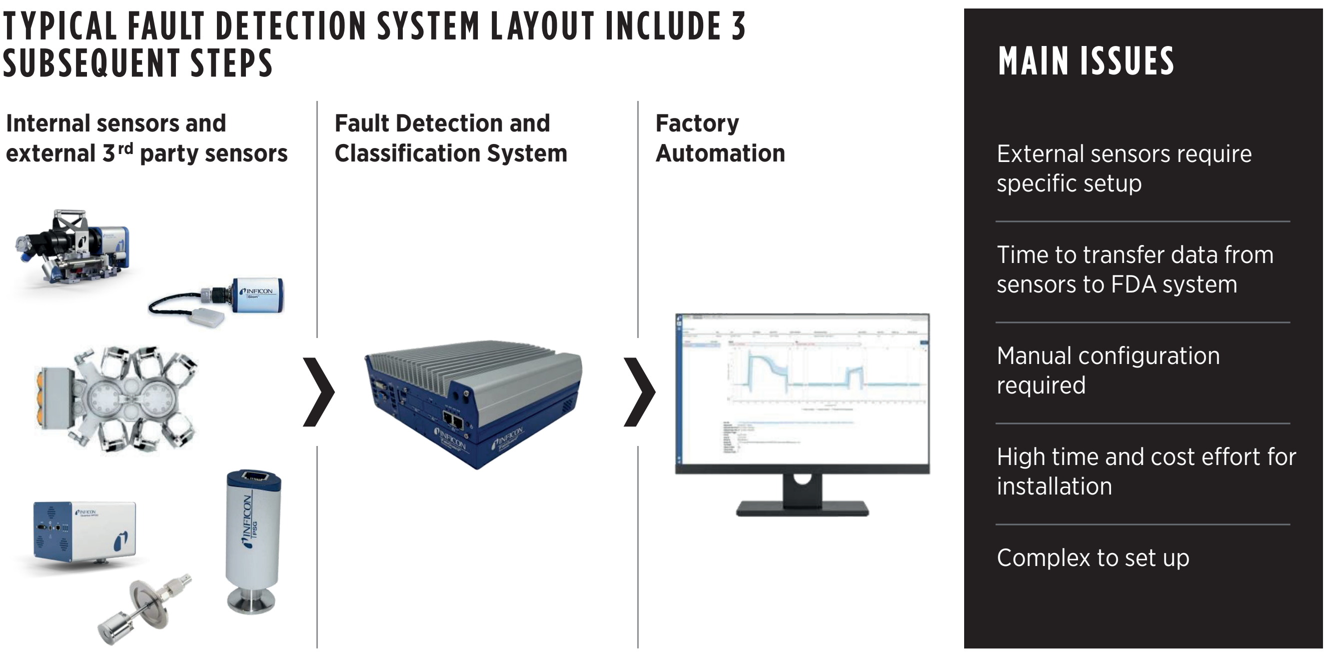


Figure 13: Typical fault detection system process: sensing, fault detection, and factory automation

Smart FDC Systems evolving with the addition of Artificial Intelligence

Artificial intelligence offers some major steps in reducing the time to deploy an FDC system and increasing the capabilities of the system. AI increases the FDC system capabilities by providing:

01 **Anomaly Detection:** AI models (e.g., autoencoders, clustering, LSTMs) learn “normal” equipment and process patterns and flagging subtle deviations earlier than SPC (statistical process control) systems do today

02 **Real-Time Monitoring:** Deep learning can process sensor streams in real-time, detecting outliers instantly and reducing wafer scrap

03 **Noise Reduction:** AI can distinguish between true faults and normal process variation, reducing false alarms

04 **Automated Defect Recognition (ADR):** AI-powered image recognition classifies wafer inspection defects (scratches, pattern bridging, missing vias) far faster and more accurately than manual review

05 **Root Cause Analysis:** Machine learning can correlate equipment conditions with defect signatures to suggest likely fault categories (e.g., chamber contamination vs. process drift)

06 **Adaptive Classification:** AI models can evolve with new defect types without needing constant manual rule updates

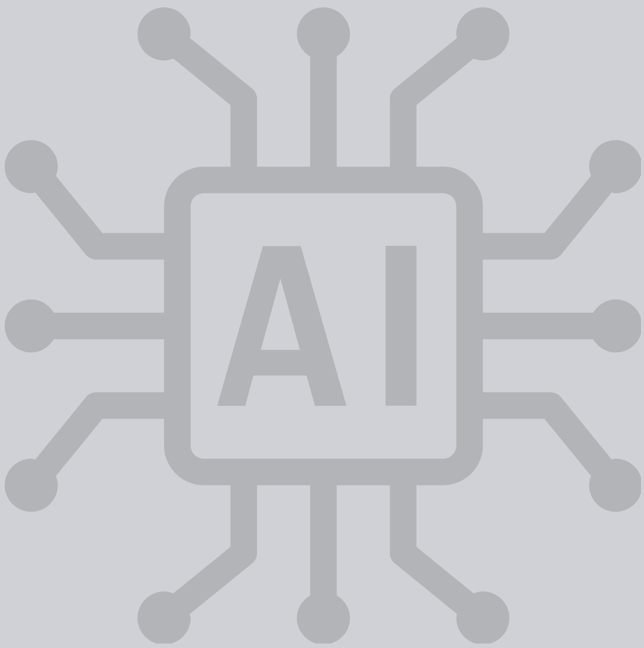
07 **AI predicts equipment failures before they happen** by analysing vibration, thermal, and sensor trends. This reduces unplanned downtime and improves overall equipment effectiveness (OEE)

AI can additionally recommend corrective actions, not just detect problems. For example:

01 **Suggest recipe adjustments** when etch depth drifts

02 **Optimize cleaning cycles** to reduce particle contamination

03 Reinforcement **learning and optimization algorithms** to auto-tune processes for maximum yield



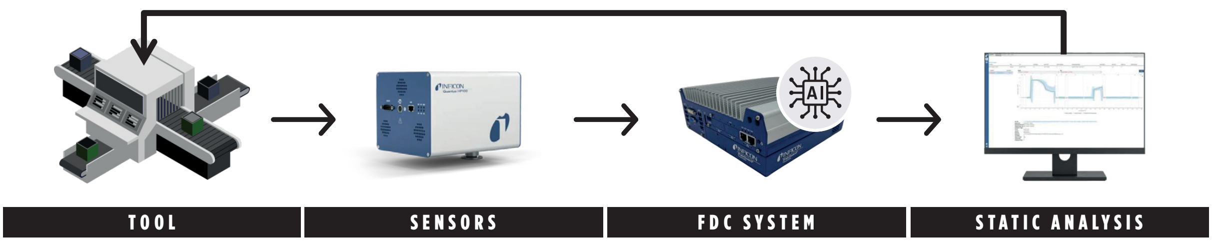
Artificial Intelligence beyond the FDC system

In the future artificial intelligence has the possibility to add value directly at the sensor level of the FDC system. Today, most AI functionality is imbedded in the FDC system via a cloud service adding a level of latency to the real time analysis system. Due to the sheer volume and complexity of the sensor data the overall performance of the system is theoretically slower than having AI directly imbedded into the sensor hardware. This AI implementation approach introduces delays in the detection of process anomalies, limits the speed of response, and incurs additional costs associated with data transfer and the maintenance of edge and cloud computing services. As a result, opportunities for real-time process optimization and rapid intervention are often missed, impacting both productivity and yield. A key constraint in the current range of sensor technologies is that most sensors were designed with a focus on performance, cost, and size, rather than on-board data processing capabilities. Consequently, many sensors lack the hardware necessary for true edge AI computing, and retrofitting existing sensor technologies to enable onboard AI sensor data analysis often requires a complete redesign. This limitation means that, in most fabs today, the potential of advanced sensor data analysis with AI remains limited and real-time event detection and advanced diagnostics are only possible using cloud AI services or server based AI systems.

The future vision, sensors with embedded (Edge) AI

The future of sensor technology in chip manufacturing lies in the integration of artificial intelligence directly into the sensor hardware—an approach commonly referred to as edge AI. By embedding AI-driven analytics within the sensor, it becomes possible to perform high-speed data analysis before the data is ever transferred to an external FDC system. This enables the use of sophisticated mathematical models and advanced statistical techniques to interpret complex data streams in real time, whether the models are trained or untrained. The result could result in a reduction in analysis latency, allowing for immediate detection of process deviations and rapid decision-making. By adding the capability to retrain the model as the data evolves increases the performance of the sensor and analysis detection capabilities over time. Models are updated via an offline server and downloaded to the sensor as a continues improvement step during sensor idle time.

CURRENT FAULT DETECTION SYSTEM PROCESS



FUTURE VISION WITH SENSORS EMBEDDING AI (EDGE)

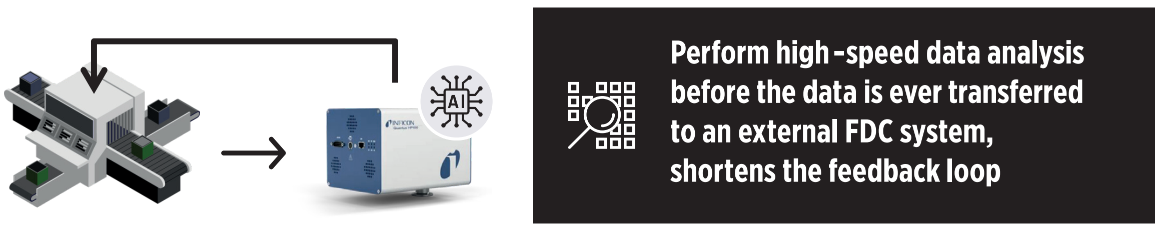


Figure 14: Comparison between the current fault detection system and the future version featuring AI-integrated sensors

Real world example of an AI enabled sensor for real time etch endpoint

Edge AI processing opens new possibilities for process optimization and equipment monitoring. For example, AI models can continuously learn from historical and real-time sensor data, improving their accuracy in detecting faults, predicting maintenance needs, and optimizing process conditions. This continuous improvement loop enhances yield, increases throughput, and reduces downtime, all while lowering the overall cost of analysis.

A compelling real-world application of this approach would be the detection of real-time etch end point in high aspect ratio, low open area processes. A challenge that has grown more acute as devices transition to 3D architectures with deeper trenches and lower light plasma environments. By coupling an electromagnetic antenna with an RF signal processor, a sensor is capable of producing a frequency spectrum of the plasma in applications with less than 10% open area. Edge AI would enable the real-time analysis of this complex frequency data, allowing for precise detection of plasma changes that signal the process end point. Currently, this AI analysis is only possible using offline AI analysis but has proved to be fast enough to detect endpoint signals with less open area structures than ever before. This capability is critical for maintaining tight process control, improving yield, and increasing throughput, all while reducing the cost of ownership in more advanced device designs.

Concept of an AI sensor for low open area endpoint

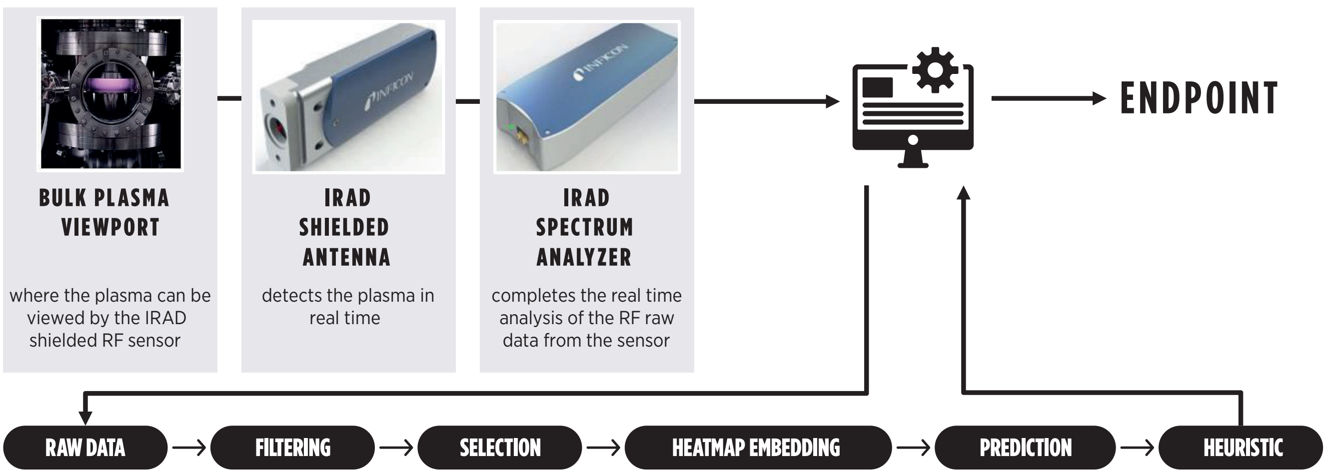
01 **Bulk Plasma Viewpoint** where the plasma can be viewed by the IRAD shielded RF sensor

02 **IRAD Shielded Sensor** detects the plasma in real time

03 **IRAD Spectrum Analyzer** completes the real time analysis of the RF raw data from the sensor

04 **AI Sensor System (Endpoint)** electronics built into the sensor taking the Raw data applying Filtering and Selection to the data and using AI for Prediction and Endpoint analysis

With deep learning techniques, INFICON utilizes the real-time spectral data to identify low open area endpoints that cannot be seen with Optical Emission Spectrometers



AI Sensor System (Endpoint) electronics built into the sensor taking the Raw data, applying Filtering and Selection to the data and using AI for Prediction and Endpoint analysis

Figure 15: Proof of concept using AI to analyze complex data in real time

CONCLUSION

The future is bright for edge AI applications embedded in sensor technologies

The adoption of edge AI in semiconductor sensors is still in its early stages, with most existing sensor technologies requiring significant redesign to support on-board data processing. Nevertheless, pilot projects and early deployments have already demonstrated the transformative potential of this approach. Successful implementation depends on a robust data strategy, investment in next-generation sensor hardware, and a commitment to upskilling teams to leverage AI-driven insights.

In conclusion, integrating edge AI into semiconductor sensors represents a major leap forward in process monitoring and control. By enabling real-time, in-sensor data analysis, manufacturers can detect critical events and process deviations faster and more accurately, reducing the risk of wafer scrap and costly defects. As a technology leader, INFICON is committed to advancing sensor capabilities and empowering semiconductor manufacturers to achieve maximum output and product value with their existing production tools. The future of smart manufacturing in the semiconductor industry will be defined by the ability to harness AI at the FDC system and at the sensor level, unlocking new levels of efficiency, quality, and competitiveness.

04

AGILEO Automation

TURNING DATA LOGS INTO LEVERAGE

Harnessing hidden insights from equipment data to enable AI-supported equipment control

In the fast-paced environment of the semiconductor industry, effective equipment integration is critical for maximizing productivity and minimizing costly downtime. Modern semiconductor fabs generate a wealth of data through equipment communication logs and application logs. While the industry has made significant strides in standardizing equipment communication, a persistent challenge remains: the underutilization of the vast volumes of communication and application logs generated by equipment controllers. These logs, effectively event-based data, capture detailed traces of tool activity, from sensor readings to alarm messages. In practice, only a small fraction of the recorded data ever gets analyzed, meaning potential early-warning signs or optimization insights are often missed. The result is that valuable operational knowledge sits idle in log files instead of driving proactive improvements. This reactive approach leads to increased downtime, reduced yield, and missed opportunities for improvement.

Communication structures still follow vertical principles

The success of semiconductor manufacturing is based on a robust communication and control infrastructure between physical hardware, equipment controllers, and levels of factory systems above them. This infrastructure ensures that the process steps are executed with precision and also guarantees that a continuous stream of right information is transmitted from the equipment to integration layers for monitoring, optimization, and decision-making purposes. Figure 16 illustrates this model of layered communication, from bottom-level physical devices to top-level enterprise-level systems, with historical data and real-time data flowing upward. Physical devices such as sensors, actuators, and subsystems at the lowest level communicate with the process, where sensors generate the raw signals and actuators execute the commands that govern equipment behavior. All complicated manufacturing hardware is made up of various subsystems with local controllers such as a PLC. The equipment controller combines all such subsystems, coordinated in their operations, and presents the entire tool as a single equipment to the fab. These signals and commands, under the control of local controllers, must be integrated into a higher communications infrastructure to enable fabs to achieve consistency and automation at scale. This is where SECS-II (SEMI E5) comes in, defining the actual message format sent between equipment and factory host. For example, a host defines a trace data collection request using S2F23 message, after which the equipment begins sending a stream of data in S6F1 messages. The equipment controller builds this S6F1 message by retrieving the requested data from the subsystems and presenting it in the normalized form. Similarly, status variables are requested through S1F3 (request) and S1F4 (response), so the factory host can poll equipment state in a consistent way. All

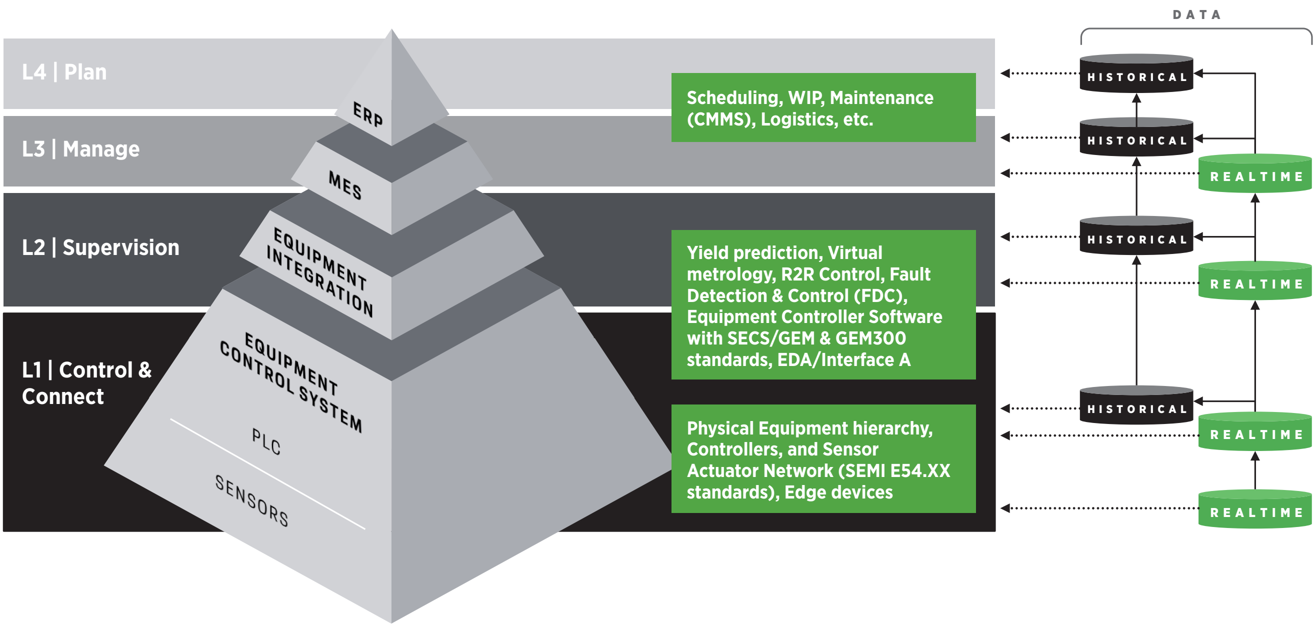


Figure 16: Flow of information & control commands in automation layers of semiconductor manufacturing

such trace and polled data must be logged reliably to enable trend analysis, root cause investigation, and the training of AI models that depend on complete historical records of process variables.

When specific events occur in the equipment, such as a process step start or wafer transfer, their occurrence is conveyed through S6F11 messages. Each event has defined data variables linked to it, so that the fabs can receive structured and actionable alerts instead of generic log lines. This way, data collection, polling, and event processing are uniform and standardized across vendors, simplifying integration at factory-level. While SECS-II defines the message syntax, the GEM standard (SEMI E30) defines the behavior and scenarios: it specifies which messages are used to configure event reports on the equipment, how variables and alarms are linked to events, and in what sequence the host and equipment exchange messages to register, enable, and report those events in real operation. The logging of event reports is essential because it provides a chronological record of all critical equipment actions and alarms, making it possible to correlate operational events with process outcomes or failures.

For advanced automation requirements, these capabilities are extended by the GEM300 standards that define explicit state machines governing the automation of wafer transport, job execution, and recipe handling. Events are directly tied to transitions within these state machines. For example, SEMI E87 defines the events that report when a FOUP is docked, clamped, or unclamped at a load port. Each state transition is consistently reported through standardized event reports, giving the host system clear and consistent visibility of the equipment's activities. By leveraging GEM300s event-driven state machines, fabs achieve reliable automation across all compliant semiconductor manufacturing equipment. Agileo Automation's A²ECF framework places these standards to work, providing OEMs with a platform for creating equipment controllers that are intrinsically SECS/GEM, GEM300 and EDA compliant. These equipment controllers then communicate upwards to the Equipment Integration layer to respond to data collection requests and to send event reports. Such logging infrastructure builds the data foundation needed for both real-time monitoring and long-term optimization through AI-based analytics.

Parallel to GEM-based communication, modern fabs also require richer, semantically structured access to data and EDA (Equipment Data Acquisition) standards respond to it. Both provide equipment data, but in complementary ways. While GEM and SECS-II deal with control commands, polling, traces, and event reporting, EDA presents high-rate, model-based data streams to upper-level applications. At the Equipment Integration layer, EDA Clients subscribe to EDA server and gather structured data and events. Comprehensive logging of both GEM-based interactions and EDA data streams ensures that factories retain a consistent historical record, enabling advanced analytics, cross-tool benchmarking, and continuous improvement of equipment performance.

Data governance in this architecture keeps real-time and history data apart. Real-time data (sometimes called streaming data) includes sensor measurements, event messages (e.g., S6F11), and trace responses (S6F1) is the continuous feed of information during operations. Historical data includes logs, traces, and process summaries built up over time, which enable analysis and optimization over time. This data is often stored as historical records (sometimes called training data) in databases or data lakes for offline analysis. Both types of data are important for AI workloads: real-time streams for immediate quick decision-making or anomaly detection, and historical records for training predictive models. For instance, an AI-driven predictive maintenance solution might use streaming data to detect an anomaly (e.g. a vibration spike) and alert staff in real time, while using historical data to train the predictive models and understand long-term trends. Similarly, process optimization AI might analyze batches of historical runs to find patterns, then monitor streaming data from the current run to ensure everything stays on track. By leveraging the SEMI standards (which ensure data is consistently formatted and time-stamped across tools), semiconductor manufacturers can build a data infrastructure that serves both immediate, tactical AI decisions and strategic, long-horizon analyses.

Transforming equipment integration with AI-Driven log analysis

One of the primary constraints in the current landscape is the reliance on manual review of extensive, often unstandardized log files. Engineers must sift through gigabytes of data, spanning multiple files and formats, to identify the root cause of integration issues.

Such issues can also arise during equipment development, Site Acceptance Tests (SAT), and Factory Acceptance Tests (FAT), where the data under review is not production data but test data used to qualify the equipment. After root cause analysis, these problems need to be resolved directly in the equipment controller to ensure that the tool performs reliably once deployed in production. This process can take hours or even days. The diversity of logging conventions used by different equipment suppliers further complicates matters, making it difficult to apply consistent analysis methods or compare performance across tools. Log files may be unstructured or use proprietary formats, such as various versions of SML or may be standardized like SEMI E173 SMN. Despite the availability of standards, automation and cross-tool benchmarking are often hindered because many equipment manufacturers do not consistently adopt or implement these standards. Additionally, the sheer size of these logs often limits their retention, restricting the ability to perform long-term trend analysis or learn from past incidents. Troubleshooting is further hampered by the need for deep domain expertise; deciphering cryptic codes and sequences is a bottleneck if knowledgeable engineers are not immediately available.

AI-driven software can automatically parse and analyze complex equipment logs in real time, instantly flagging anomalies or process inefficiencies suggesting dynamic adjustments to enhance throughput or enabling engineers to address issues proactively before they escalate into downtime

Artificial intelligence offers a transformative solution to these challenges. AI-driven software can automatically parse and analyze complex equipment logs in real time, instantly flagging anomalies or errors that might be overlooked by human inspectors. Machine learning models, trained on historical log data, can recognize early signals of equipment or integration faults, enabling engineers to address issues proactively before they escalate into downtime. Advanced AI systems are capable of correlating events across communication and application logs from multiple sources, providing a holistic view that pinpoints root causes far more efficiently than traditional methods. Natural Language Processing (NLP) techniques further enhance this capability by interpreting obscure log entries and event or alarm identifiers, translating them into clear explanations and actionable recommendations for engineers.

The benefits of AI-driven log analysis extend beyond faster troubleshooting. By continuously monitoring logs, AI systems can identify process inefficiencies and suggest dynamic adjustments to enhance throughput. Over time, these models improve their accuracy in fault detection and process optimization, learning from both historical data and new incidents. This not only shortens the time required for new tool bring-up and qualification but also improves overall production ramp-up and consistency. The process is highlighted in Figure 17.

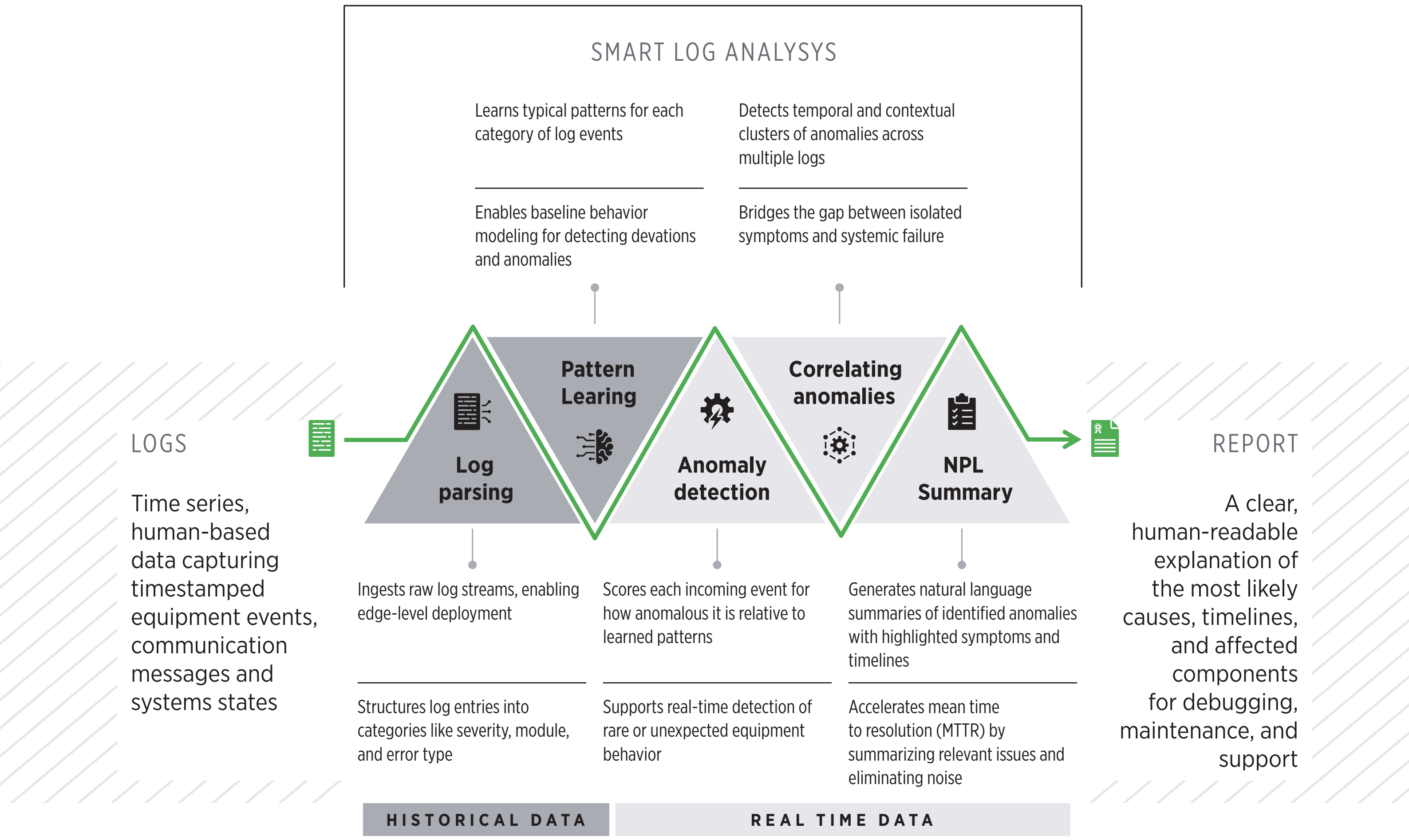


Figure 17: Automated Log Analysis using machine learning

AGILEO AUTOMATION IS LEVERAGING ON AI FOR LOG ANALYSIS

The team at Agileo Automation has been at the forefront of smart factory initiatives, developing software solutions tailored to the unique demands of semiconductor equipment integration. Through direct involvement in edge data collection and log analysis projects within a leading French R&D consortium, Agileo Automation has gained firsthand insight into the challenges and best practices of implementing AI in this context. Our active participation in industry consortia and standardization efforts ensures that we remain at the cutting edge of developments, helping to shape best practices for the application of AI in semiconductor manufacturing.

A practical example of this approach is an ongoing research project focused on metrology equipment, where our AI-powered log analysis solution has significantly accelerated root cause analysis of communication errors between the equipment and the host system. By automatically correlating events and interpreting log data, the AI system has enabled faster resolution of integration issues, reducing downtime and supporting smoother production ramp-up.

CONCLUSION

GenAI is here, right at your disposal

AI-powered log analysis is set to revolutionize equipment integration in semiconductor manufacturing. By shifting from a reactive, labor-intensive troubleshooting process to a proactive and highly efficient practice, semiconductor manufacturers can reduce integration time, improve equipment uptime, and achieve higher yields. While adopting AI in the fab requires investment and cultural change, early successes demonstrate that embracing these technologies is increasingly essential for achieving end-to-end smart manufacturing and maintaining a competitive edge in a demanding market.

Specifically, the use case proposed by Agileo Automation enables equipment manufacturers to enhance the intelligence and reliability of their machines from the design phase through to operation in the field. By embedding standardized logging and AI-ready data interfaces directly into equipment, OEMs can offer advanced diagnostics, predictive maintenance, and continuous performance optimization as built-in features. This approach not only allows manufacturers to identify and resolve issues faster but also enables remote support and faster ramp-up for new tools at customer sites. Leveraging AI, machine learning models can continuously analyze both real-time and historical log data to identify subtle patterns and anomalies that may escape human detection. Over time, these insights contribute to improved machine design, higher quality manufacturing processes, and reduced total cost of ownership for end users. Ultimately, integrating AI-powered analytics at the equipment level empowers both OEMs and fabs to transition from reactive problem-solving to data-driven, predictive decision-making, accelerating the journey toward fully autonomous and self-optimizing manufacturing environments.

05

PeerGroup
Integrating Automation & Information™

REIMAGINING EQUIPMENT INTEGRATION

How AI and Digital Twins are redefining semiconductor automation with AI factory OT equipment integration

SEMI communication standards are constantly evolving to allow new use cases

Semiconductor automation has always been built on standardization. In the early 1980s, SEMI introduced the SECS-I and SECS-II protocols, which for the first time provided a structured way for equipment and factory host systems to exchange messages. Running over RS-232 serial links, these early systems allowed only basic reporting and control, yet they set the stage for the modern fab. In the 1990s, High-Speed SECS Message Services using TCP/IP replaced serial connections, while the Generic Equipment Model introduced a uniform way to describe equipment states, behaviors, and transitions. This represented more than a technical upgrade; it signaled a cultural shift in which automation was no longer optional but an expected foundation for semiconductor production, see Figure 18.

The transition to 300 mm wafers demanded even deeper coordination. Material handling had to be automated, substrates tracked, and process jobs managed seamlessly across hundreds of tools. SEMI responded with the GEM300 suite, which made compliance mandatory for equipment vendors and created the first global baseline for tool automation. But as factories pushed into advanced process control, the volume of data required for analytics grew exponentially. The desire to have multiple data consumers connecting to a tool resulted in the development of Interface A, also known as Equipment Data Acquisition (EDA), to provide a parallel, high-speed data channel capable of streaming context-enriched information to analytical systems. The current Freeze 2 version relies on SOAP/XML over HTTP/1.1, but the forthcoming Freeze 3 will adopt HTTP/2, gRPC, and Protocol Buffers, providing lower latency, secure encrypted communication, and direct compatibility with modern IT and AI infrastructures.

SEMI STANDARD DEVELOPMENT & REVISIONS

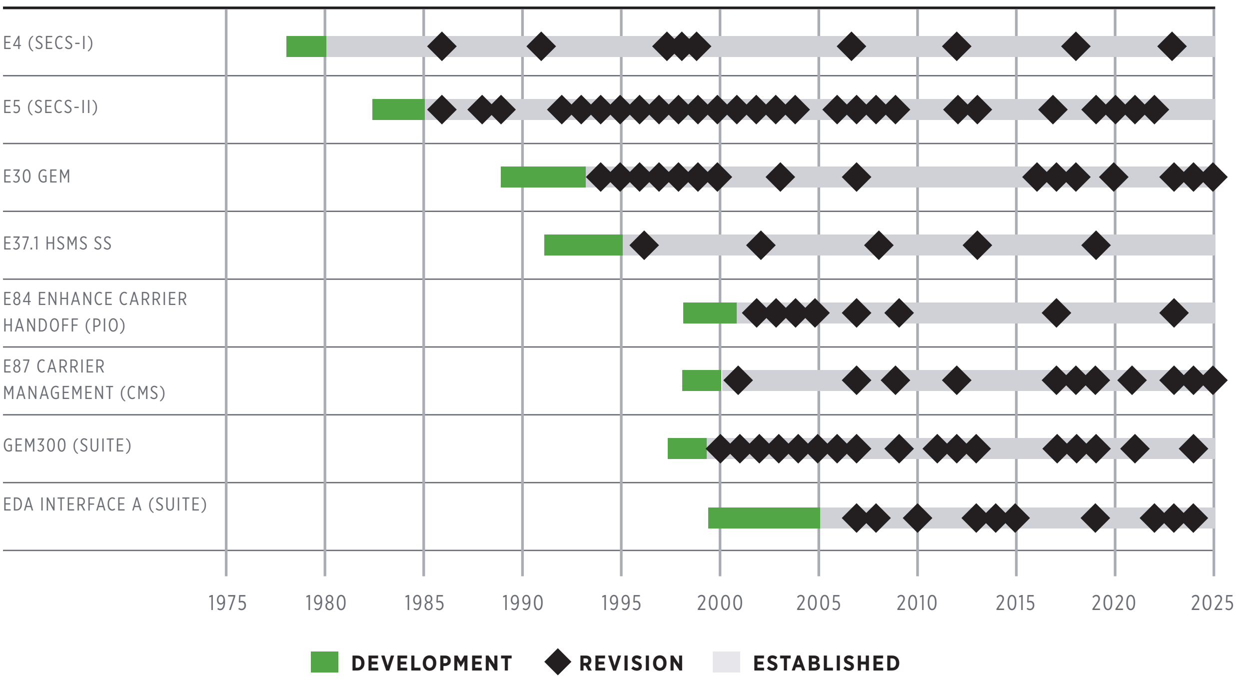


Figure 18: Timeline of SEMI automation standards development and revisions

The Equipment Integration Layer allows fabs to integrate legacy with IOT equipment

Standards provide the language of communication, but they need an interpreter. This role is fulfilled by the equipment integration layer, a dedicated station controller or automation framework that connects every piece of equipment to supervisory factory systems as in Figure 19. The layer ensures that commands flow to the right tools, that data streams are captured accurately, and that context is preserved for advanced analytics. In many ways, it is the nervous system of the fab, connecting the machinery to the brain of the manufacturing execution and control systems. The challenges faced by this layer remain substantial. Despite the maturity of SEMI standards, the promise of plug-and-play interoperability across an entire fab has never been realized. Procurement processes often emphasize throughput and process performance while neglecting automation requirements. As a result, equipment is sometimes delivered with incomplete interfaces or inconsistent implementations. These gaps become visible only at installation, leading to costly engineering effort, integration delays, or tools that cannot achieve the desired automation level. The long service life of semiconductor equipment adds further complexity. It is not unusual to find tools that are decades old still running alongside the latest systems, creating a patchwork of legacy and modern interfaces. In back-end manufacturing this heterogeneity is especially pronounced, with manual and automated steps, old and new hardware, and varying standards coexisting within a single line. Factories increasingly rely on Automation Capability Management to manage these risks. ACM is a governance practice that defines automation requirements up front, ensures they are testable, and validates them before equipment enters production. By shifting responsibility to suppliers and making automation readiness a contractual deliverable, ACM reduces deviations, accelerates ramp-up, and lowers overall integration costs.

> Despite the maturity of SEMI standards, the promise of plug-and-play interoperability across an entire fab has never been realized with equipment integration process often stretching into weeks or months

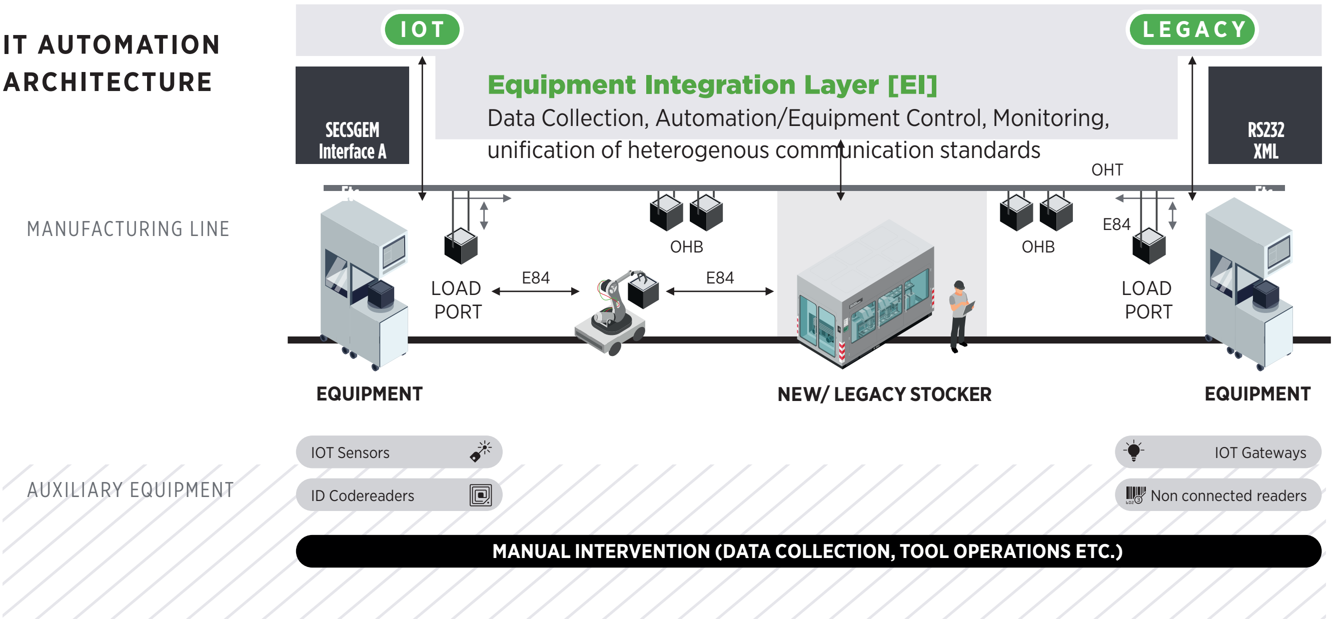


Figure 19: The role of the Equipment Integration (EI) layer as bridging diverse equipment protocols

Against this backdrop, PEER Group has established itself as a leader in equipment integration. The company has spent decades helping fabs automate their production floors and reduce the complexity of tool connectivity. Its Equipment Automation Framework, introduced in 2010 and illustrated in Figure 20, provides a modular and configurable platform that supports both SEMI and non-SEMI protocols.

The EAF architecture is organized in three layers. At the base, the connectivity layer supports a broad spectrum of interfaces, from SECS/GEM and GEM300 to Interface A, OPC-UA, and PLC protocols. Above it, the automation layer implements fab scenarios as a digital twin of automation, allowing compliant tools to be integrated by configuration and non-compliant ones to be adapted through software extensions. This layer also serves as the execution environment for business logic and edge AI functions. The integration layer then connects the framework to the broader OT and IT systems, including MES advanced process control, and factory-wide AI applications.

In practical deployments, the EAF has delivered substantial benefits. First-of-a-kind tools can be brought into production faster, replications across multiple tools are easier, and maintenance effort is reduced by more than half through centralized server-based configuration. Out-of-the-box adapters shorten deployment times, while consulting services and the DiDaCT testing tool support automation readiness and enforce standard conformance. By participating actively in SEMI task forces, PEER Group ensures that its solutions evolve in step with the global standards on which the industry relies.

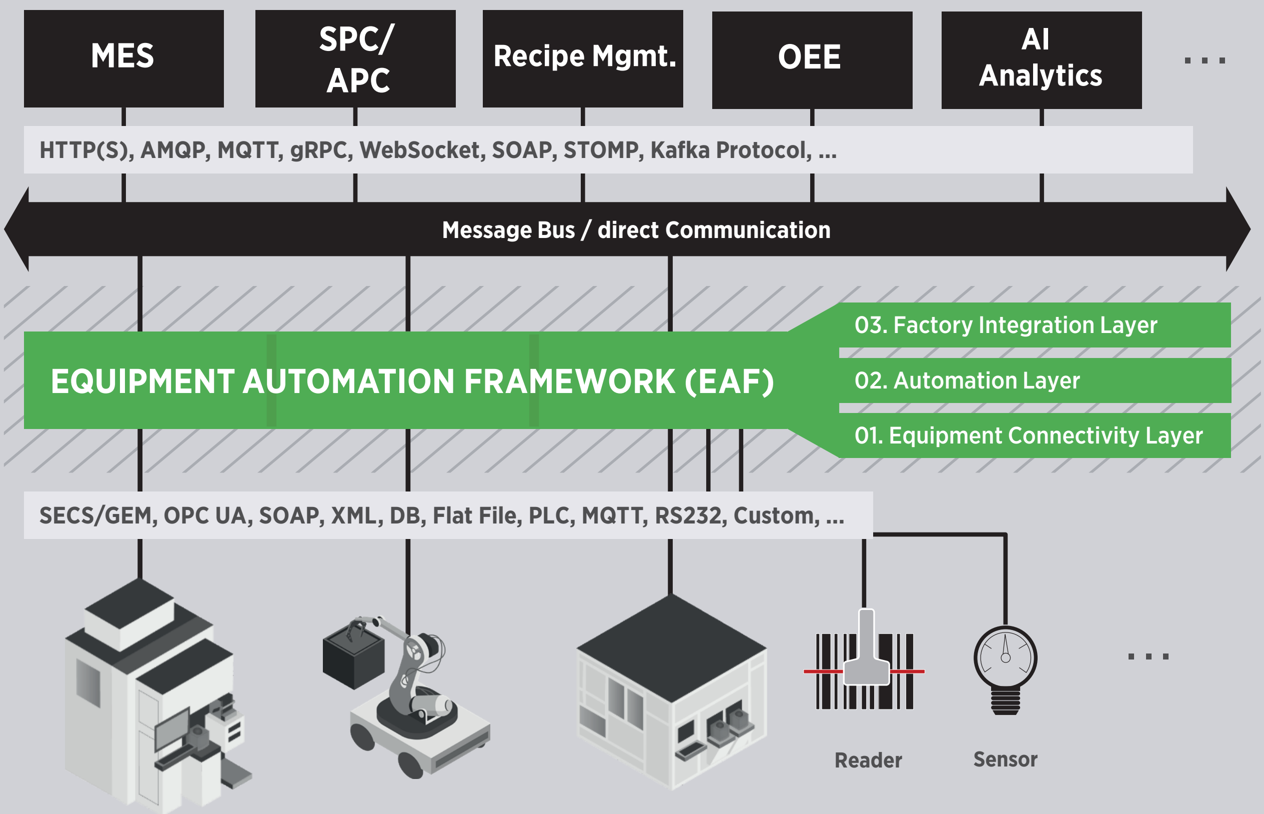


Figure 20: Three-layer architecture of the Equipment Automation Framework (EAF), jointly developed with Infineon Technologies

AI in equipment integration: automation and adaptation

Equipment integration is one of the most complex and resource-intensive tasks in semiconductor manufacturing. Each new tool must be connected to the factory host using SEMI protocols such as SECS/GEM, HSMS, or EDA (Interface A), and then adapted to fab-specific automation conventions. Traditionally, this required engineers to build dictionaries and event reports from scratch, configure middleware, and debug extensive test cases. The process often stretched into weeks or months, delaying production ramp-up and raising engineering costs. Artificial intelligence is now changing this paradigm by automating core integration steps and enabling adaptive, self-maintaining connections between tools and fab systems.

AI models trained on historical SECS/GEM logs and the EI code base can automatically generate preliminary mappings of events, variables, and alarms. Instead of creating dictionaries manually, engineers receive AI-generated drafts that align with factory schemas, providing a validated baseline for refinement rather than starting from zero. Especially for non-standard tools this could reduce setup time up to 40% according to initial estimation shifting to higher-value optimization tasks. Testing and validation are also enhanced: AI-driven harnesses analyze schemas and prior test results to generate comprehensive test cases automatically. These include not only standard operation flows such as recipe start, stop, and carrier handling but also historically problematic scenarios like alarm floods, parameter drifts, and conflicting state transitions. By embedding regression testing, AI ensures that any update to mappings or logic is validated against baselines without requiring engineers to redesign test suites.

> AI, by automating core integration steps, enables adaptive, self-maintaining connections between tools and fab systems. Engineers can receive AI-generated drafts that align with factory schemas, reducing manual effort and setup time up to 40%

Adaptation is another critical advantage. Vendor updates whether firmware revisions, new recipe formats, or modified message sets can disrupt existing integrations. Without automation, these changes demand time-consuming manual re-engineering. AI addresses this by continuously monitoring communication logs and specifications, identifying deltas, and proposing corrective updates. The system validates these changes automatically against historical baselines, ensuring that integrations remain synchronized with vendor software while minimizing downtime. For example, if a tool introduces a new alarm code or modifies a process event, AI highlights the change, updates the dictionary, and generates validation test cases. This adaptive capability reduces integration drift and ensures long-term stability across production environments.

CONCLUSION

The future development of equipment integration: Digital Twins and scalable standardization

The most significant evolution for AI in equipment integration lies in the rise of digital twins -virtual models of a tool's automation interface and behavior. Constructed from vendor specifications, dictionaries, and historical logs, digital twins allow engineers to begin integration activities before the physical tool arrives. Full host-tool communication sessions, including start-up routines, message exchanges, and alarm conditions, can be simulated in a virtual environment. This enables engineers to detect integration issues early, accelerate installation, and validate robustness under controlled conditions. AI enhances these twins by modelling atypical behaviors and stress scenarios that would be difficult to reproduce physically, such as simultaneous alarms or rapid recipe adjustments. This expands test coverage and strengthens system resilience. Digital twins also adapt dynamically. When firmware changes or new SECS/GEM dictionaries are released, AI-enabled twins learn the revised schema, update their mappings, and generate regression tests. Over time, they accumulate integration experience across multiple projects, effectively serving as knowledge bases that accelerate future deployments. This capability transforms integration from a one-off task into a continuous improvement process. Beyond testing, digital twins serve as training platforms. Engineers can rehearse complex or high-risk scenarios in a safe environment, gaining experience with error conditions without jeopardizing live production. AI-generated scenarios further enrich this training by introducing rare but plausible conditions, improving preparedness for real-world failures. Standardization and scalability are equally important. Centralized integration servers, coupled with AI-driven consistency checks, ensure that alarm mappings, state models, and parameter configurations remain aligned across all tools and fabs. Version control and rollback mechanisms provide auditability and regulatory compliance, while automated deployment reduces engineering overhead. This approach enables manufacturers to replicate integration solutions globally without reinventing the process for each site. For example, once an AI-enhanced twin validates a new mapping in one fab, the same validated logic can be deployed across other facilities with minimal adaptation.

06

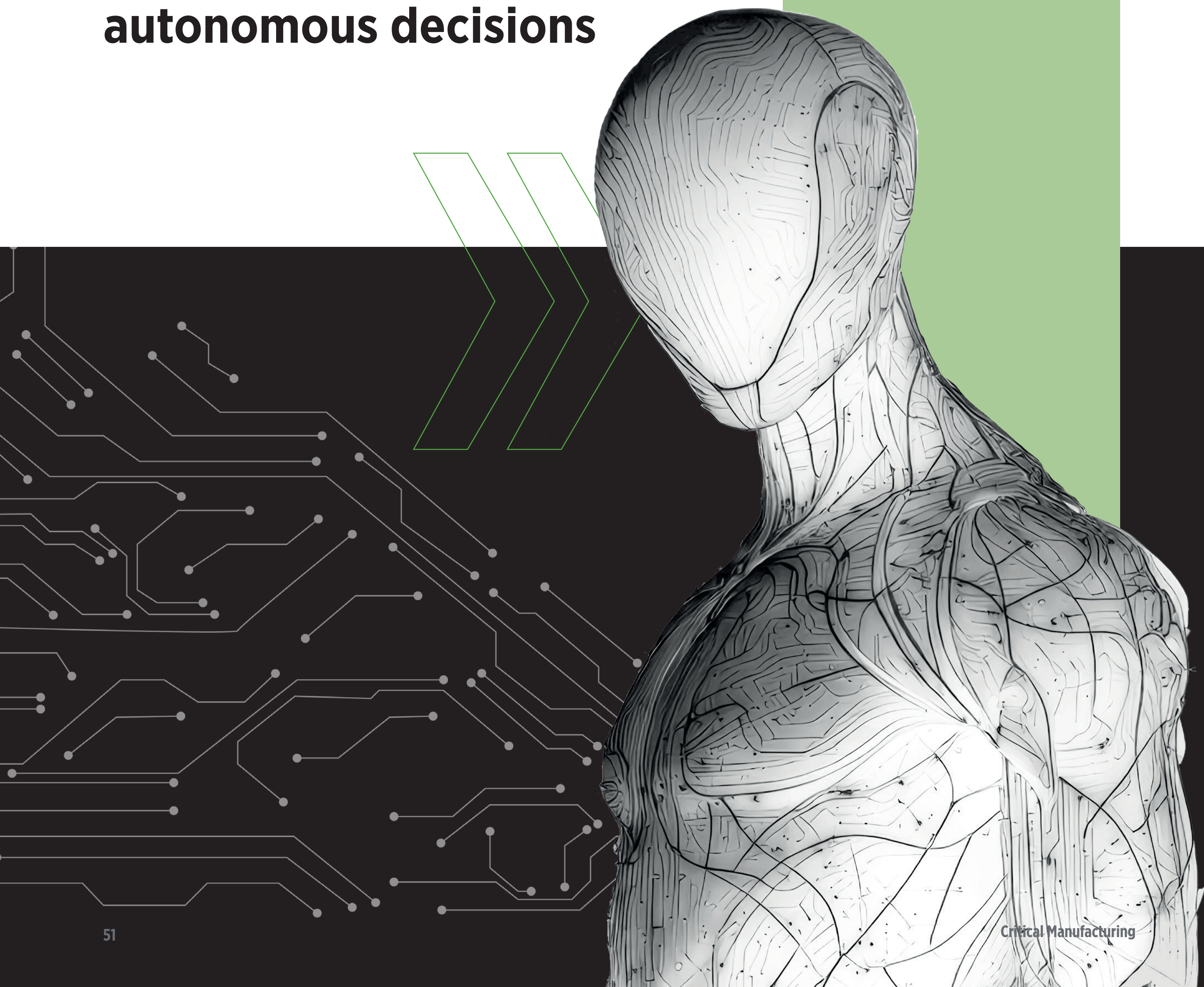
Critical manufacturing

REINVENTING MES

How AI is transforming manufacturing execution systems into adaptive, intelligent decision platforms

Artificial Intelligence (AI) is reshaping the role of Manufacturing Execution Systems (MES), transforming them from structured execution platforms into adaptive systems capable of supporting real-time insights, recommendations, and increasingly autonomous decisions. Traditionally positioned between enterprise planning and the shop floor, MES platforms are now evolving into intelligent hubs that interpret data, suggest actions, and adapt to changing conditions. This chapter explores this transformation in three levels, each representing a deeper level of AI integration, and outlines the operational and strategic value at each step.

(AI) is reshaping the role of Manufacturing Execution Systems (MES), transforming them from structured execution platforms into adaptive systems capable of supporting real-time insights, recommendations, and increasingly autonomous decisions



Integrating IoT, MES, and data platforms

BUILDING THE FOUNDATIONS FOR INTELLIGENT MANUFACTURING

At the first level of AI integration, the focus is on establishing a unified data environment by integrating equipment integration, MES, and data platforms. These systems, often siloed in traditional manufacturing, must be tightly connected to contextualize real-time data and enable intelligent analysis.

Level 1 integration can be understood through three interconnected layers:

CONNECT (EDGE)

Equipment integration at the edge captures real-time data from sensors, machines, and PLCs. These systems preprocess data locally, reducing latency and supporting timely decision-making and by allowing actions to be performed, that may be triggered by external systems (e.g. AI systems). In industries like semiconductors, bidirectional communication is crucial, allowing the MES to both read equipment conditions and issue commands (e.g. recipes).

EXECUTE (MES OPERATIONS)

MES is the operational backbone being responsible for executing production orders, managing work-in-progress, ensuring traceability, and enforcing quality procedures. However, with real-time IoT data now available, MES can begin to operate with a higher level of responsiveness, triggering Statistical Process Control (SPC), exception handling, or process holds, enabling more responsive and data-informed execution flows.

ANALYSE (DATA PLATFORM)

Data platforms ingest, store, and analyse data from both MES and IoT systems. Structured and unstructured data is contextualized into unified semantic models. This allows for more advanced analysis such as root cause analysis, anomaly detection, and yield prediction.

The convergence of these three layers enables foundational use cases where data is not only collected but actively interpreted in real-time. Machine learning models can be trained on historical data, then deployed to operate on live streams, providing alerts, recommendations, and triggers for MES actions.

Strategic value of triangular integration

Fully integrating MES, IoT, and data platforms enables real-time quality control, exception handling, equipment readiness validation, and unified visibility across production. While not fully autonomous, this level creates a rich, contextualized data environment essential for advanced intelligence.

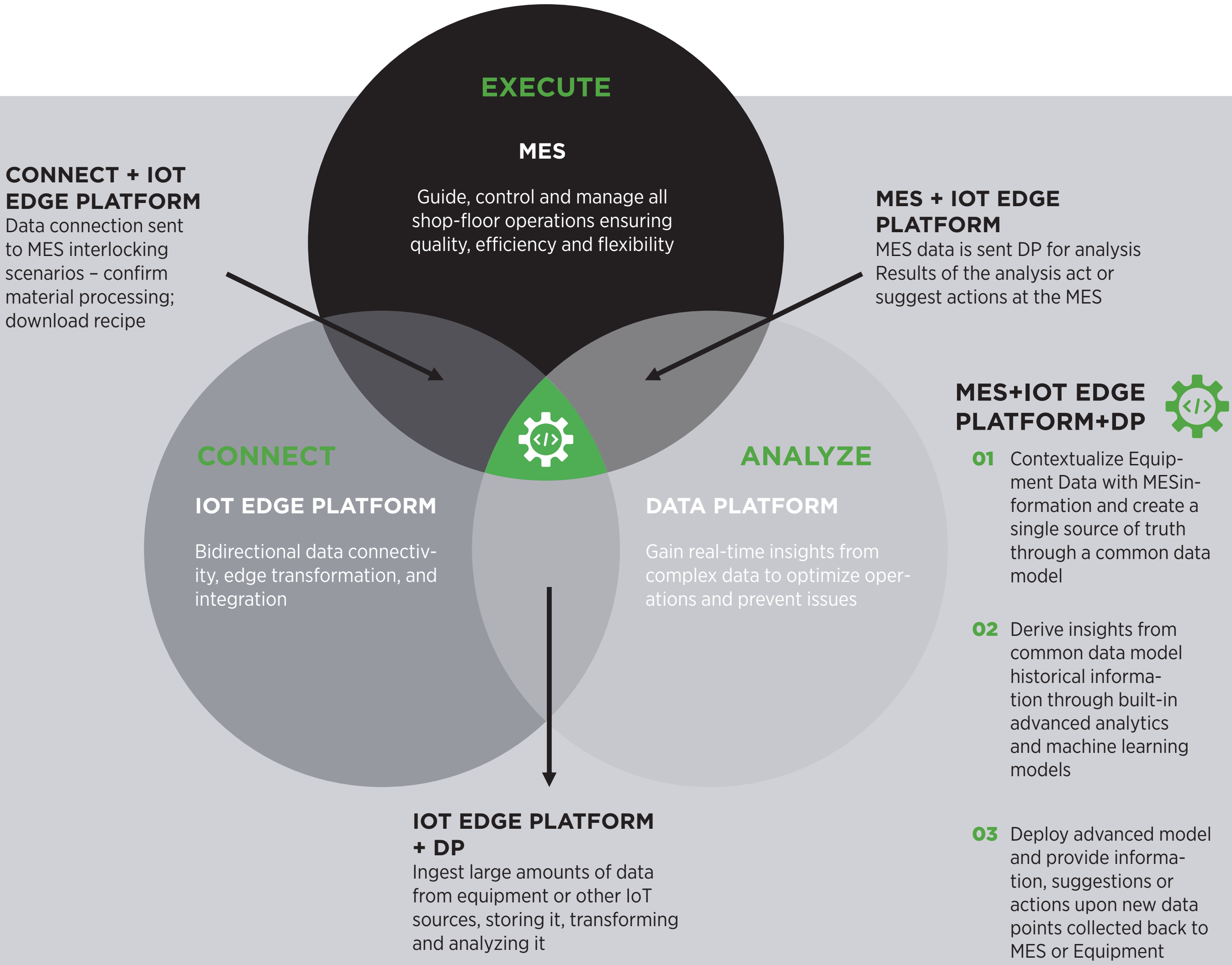


Figure 21: Convergence of MES, IoT and Data Platform

Representative use case: predictive maintenance using Time-to-failure models

Predictive maintenance is one of the most practical applications of AI in manufacturing, and at Level 1, it becomes feasible through the structured integration of MES, equipment sensor data, and cloud-based modelling. This use case uses a supervised learning model to estimate Remaining Useful Life (RUL) and anticipate failures before they occur.

How it works:
- MES data provides context such as historical failure records, known root causes, and tool usage metrics (e.g., number of cycles).
- Sensor data from the equipment is continuously collected via IoT connectors.
- This combined dataset is used to train a prediction model capable of estimating the time to failure for specific machines or components.

Model output is integrated into an operator dashboard, showing the estimated time remaining until failure, with thresholds set to trigger alarms or actions in the MES, as in Figure 22.

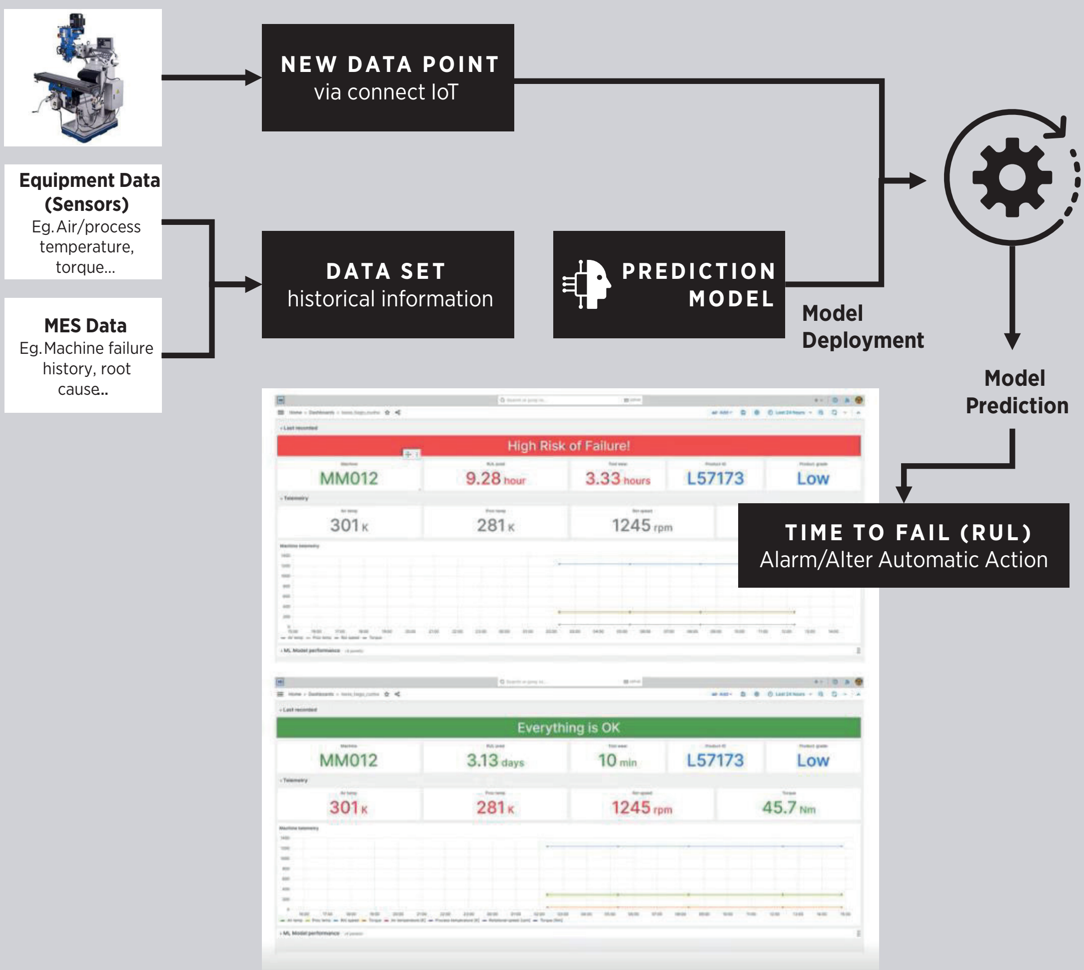


Figure 22: Predictive Maintenance using Supervised Learning

Context-aware intelligence

GENAI IN MES THROUGH RAG AND FINE-TUNING

With foundational integration in place, Level 2 shifts the focus from connectivity to accessibility. AI becomes a conversational partner—interpreting, summarizing, and explaining complex data using natural language.

This is made possible through generative AI (GenAI), particularly using two techniques:

RAG
RETIREVAL-AUGMENTED GENERATION

Combines a pre-trained language model with domain-specific databases (e.g., SOPs, MES logs, tool manuals). When queried, the model retrieves relevant documents and generates accurate, context-specific responses.

FINE-TUNING

General-purpose AI models are further trained on plant-specific data, enhancing their understanding of terminology, workflows, tools, and local performance metrics.

These technologies allow engineers, supervisors, and operators to interact with MES systems naturally—asking questions, summarizing operations, and retrieving specific metrics—without needing programming or database expertise.

Representative use case: end-of-shift summary generation

A shift supervisor managing a critical area needs a concise yet comprehensive report summarizing what occurred during the last production cycle. The GenAI assistant receives a query like: "Summarize the key events and KPIs for the last shift in area XYZ."

The response may look like:
- **Production Lots:** 4 completed; 1 delayed at metrology
- **Tool Events:** Tool ABC-3000 had two 15-minute pressure alarms
- **KPIs:** OEE = 81% (Availability: 87%, Performance: 95%, Quality: 98%)
- **Anomalies:** Slight increase in rework rate (+1.2%)
- **Recommendation:** Monitor chamber clean intervals on ABC-3000

This functionality improves shift handovers, enhances traceability, and ensures critical issues are not lost documentation gaps.

Agent-based MES

TOWARD SELF-OPTIMIZING OPERATIONS

At Level 3, AI transitions from being a supportive tool to becoming an active participant. AI agents, powered by LLMs and enriched with real-time production context, begin to interact with MES data and business logic, making and executing decisions within defined bounds.

AI agents are systems that interact with APIs, retain memory, reason through complex tasks, and learn from outcomes to improve future decisions. Rather than simply following rules, they adapt over time. Initially, they suggest actions for human review, but in familiar domains, they gradually gain autonomy.

Integration with MES and the role of MCP

MES platforms are built on three layers: data models (e.g., lots, tools), transactions (e.g., move, hold), and business logic (rules and constraints). At Level 3, AI agents enhance the business logic layer by responding to patterns and exceptions using adaptive reasoning, see Figure 23.

Model Context Protocol (MCP) is standard which gives agents structured access to MES data, available actions, and live system context. MCP allows agents to interpret and interact with MES autonomously (calling API's) while maintaining transparency and alignment with enterprise rules.

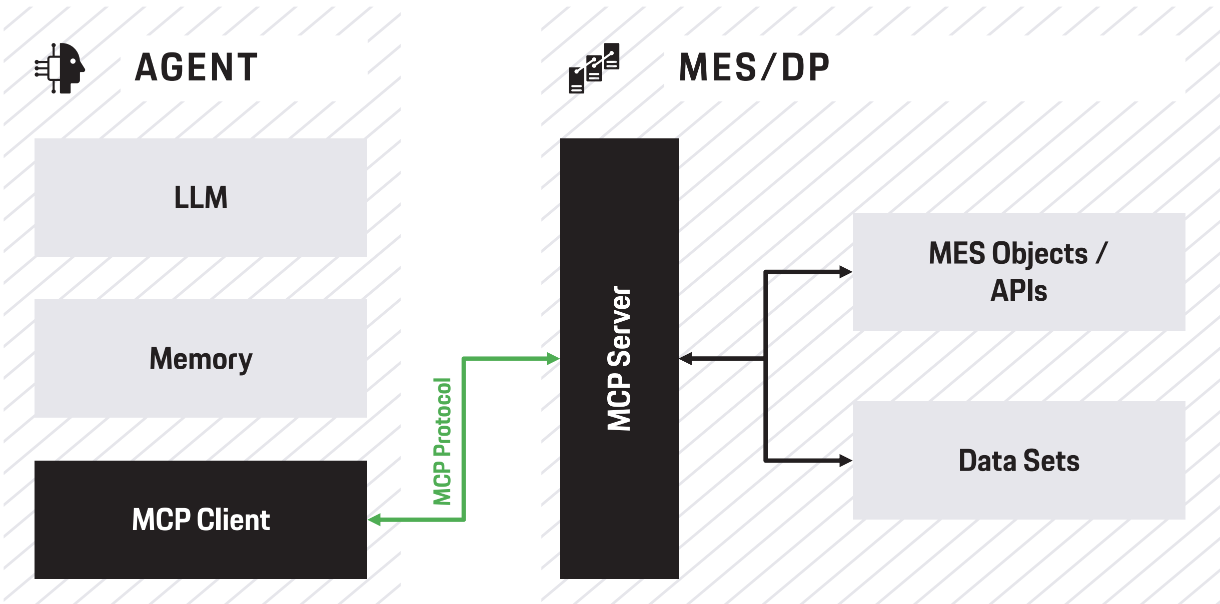


Figure 21: Convergence of MES, IoT and Data Platform

AI risks and how to mitigate them

AI agents don't come without risks, especially in the early stages of adoption, but there are three key aspects to mitigate them. Guardrails define operational boundaries to prevent unsafe or non-compliant actions. Human-in-the-loop oversight ensures accountability by requiring operator review before execution. And decision trails provide transparency into the reasoning behind actions, helping troubleshoot issues and build trust. Together, these mechanisms significantly reduce potential headaches while enabling safer progress toward autonomy.

CONCLUSION

Evolution of AI within MES unfolds in three progressive levels:
Level 1 establishes the data foundation through IoT, MES and analytics integration.
Level 2 adds GenAI for context-aware insights.
Level 3 introduces agents that learn, reason, and act within MES environments

The evolution of AI within MES unfolds in three progressive levels. Level 1 establishes the data foundation through IoT, MES, and analytics integration. Level 2 adds GenAI for context-aware insights and natural language interaction. Level 3 introduces agents that learn, reason, and act within MES environments. Each level builds upon the previous, and skipping foundational steps risks failure in higher-level deployments.
Today, Level 1 delivers the most tangible results, powering predictive maintenance, quality control, and early fault detection. Meanwhile, Level 3 technologies are maturing quickly and will redefine operations in the coming years.

Looking ahead to 2030, MES will likely become a more autonomous, intelligent environment. AI agents will play a larger operational role, decision-making will become more adaptive and collaborative, and real-time data will drive continuous optimization. MES will remain central, but increasingly as a platform for intelligent coordination between humans and AI [23].

07

Flexciton

FROM COMPLEXITY TO CLARITY

AI-powered production scheduling and planning for high-performance semiconductor manufacturing

The challenge of planning and scheduling in modern fabs

Effective production planning and scheduling lie at the heart of semiconductor manufacturing success. Their influence on on-time delivery, tool utilization, and cycle time is profound. The extraordinary complexity of semiconductor processes—with high product variability, long cycle times, and tightly constrained tool capacities—means that maintaining efficient operations is a highly important task. Yet, many factories still rely on legacy solutions that are simply inadequate for the sector's evolving needs.

Many factories still rely on legacy solutions for planning and scheduling which are highly labour-intensive and simply inadequate for the sector's evolving needs.

Traditionally, wafer fabs have depended on rule-based dispatching and scheduling systems. While these approaches offer some structure, they are highly labour-intensive: frequent manual updates are required to keep rules relevant as factory conditions change. As a result, suboptimal decisions are common, with inefficiencies appearing when the human factor cannot cope with operational complexity.

Beyond scheduling, production planning typically relies on simplistic line balancing algorithms. Such tools often use average cycle times to prioritize work-in-progress (WIP), but they overlook the nuanced constraints of real factories, therefore they are generating bottlenecks. Manual wafer start management, mostly handled by planners, struggles to keep up with the optimal product mix—leading to uneven flow and missed opportunities to maximize capacity.

Capacity planning, meanwhile, is often carried out using inflexible spreadsheets that use steady, linear WIP movement—an unrealistic assumption in dynamic fab environments. Even preventive maintenance is too often planned in isolation, resulting in timed interventions that ignore actual tool utilization patterns and production bottlenecks. This fragmented approach limits the ability to manage key performance indicators effectively and consumes significant manual effort, see Figure 24.

FACTORY OBJECTIVES
(E.G., THROUGHPUT, ON-TIME-DELIVERY, ETC)

How do we run production to fulfill customer orders and achieve our business objectives efficiently?

	CAPACITY PLANNING	PRODUCTION PLANNING	PRODUCTION SCHEDULING
TIMING	Long-term horizon, typically 3 months – 5 years	Mid-term horizon, typically 1 – 4 weeks	Short-term horizon, typically 6 – 12 hours
DESCRIPTION	Assessing manufacturing capabilities to meet customer orders. Decisions about commits, new equipment, facility expansions, workforce planning	Translating customer orders into executable manufacturing plans. Determines wafer starts and lot prioritisation weekly.	Detailed allocation of lots to machines. Determines exact sequence and timing to optimise throughput, cycle time, etc.
COMMON APPROACH	Static excel-based capacity models or highly complex simulation software	Manual excel-based wafer starts plans. Simplistic line balancing with avg. cycle time to prioritise WIP.	Rule-based dispatching and scheduling requiring frequent manual tweaks
AI TRANSFORMATION	AI systems build live models of factories to simulate and evaluate scenarios with unprecedented speed	AI-powered planning systems model real factory constraints to generate optimised wafer start plans and prioritise WIP	AI-enhanced models autonomously generate and adjust schedules in real time, reducing human interventions

Figure 24: Factory objectives for different planning horizons

How AI is revolutionizing scheduling and planning

Artificial intelligence is redefining what's possible in fab operations. By tackling the chronic computational complexity of scheduling and planning, AI unlocks a range of benefits that boost factory performance.

AI-driven planning and scheduling systems evaluate immense sets of possible scenarios to deliver truly optimal decisions—far exceeding the coverage of rule-based systems. This doesn't just improve outcomes (such as on-time delivery and throughput), it dramatically reduces the human effort required for setup and ongoing maintenance. Where older systems depend on frequent manual interventions, AI models autonomously create, tune, and run themselves in real-time.

The power of AI extends to transparency and accuracy. Advanced solutions now leverage large language models (LLMs) as schedule explainers. These can distil highly complex scheduling decisions into clear, human-readable explanations. As a result, production staff can query the system in natural language—asking, for example, "Why is lot X delayed?"—and receive instantly understandable responses grounded in the factory's real-time state.

Looking to the medium-to-long term, AI-powered planning systems dynamically model real factory constraints and prioritize WIP for maximum on-time delivery. They also inform crucial decisions, such as when to start wafers and how best to coordinate equipment and process teams for maximum performance.

Dynamic maintenance planning is another breakthrough: AI can optimize preventive maintenance schedules based not just on fixed intervals but on actual WIP flow, tool utilization, and the need to minimize disruption. In capacity modelling, AI systems construct live digital models of factories, enabling managers to simulate and evaluate diverse scenarios with unprecedented speed and fidelity by simply describing desired "what-if" situations in plain language.

The current maturity of those solutions is summarized in Table 2.

By tackling the chronic computational complexity of scheduling and planning, AI-driven systems deliver truly optimal decisions reducing cycle times up to 24%

SOLUTION	READINESS	MATURE	EARLY STAGE	PROTOTYPING
Autonomous Scheduling	Proven since 2019; demonstrated 5–10% increase in lot moves, 3–4x reduction in manual overrides. AI schedule explainer prototyped.	✓		
AI-powered Planning Systems	Deployed since mid-2024; co-developed with Tier 1 & Tier 2 fabs. ~10% cycle time reduction, improved on-time delivery		✓	
Capacity Modelling Using AI	Prototype stage; production readiness expected by 2025. Direct extension of AI production planning			✓
Dynamic Maintenance Planning	Prototype stage; production readiness expected by 2025. Direct extension of AI production planning			✓

Table 2: Maturity of different use cases leveraging AI

Real-world impact: notable use cases

At Seagate Technology's Springtown wafer fab, Flexciton's toolset delivered remarkable performance improvements. Fully deployed across both local and fab-wide operations, the solution achieved a 24% reduction in cycle time for key toolsets and a 9% increase in photolithography lot movements while reducing reticle moves by 10%. Manual interventions dropped three- to four-fold, and overall throughput for the fab increased between 6.7% and 8.6% [24], as displayed in Figure 25.

Another pivotal project, at a leading US foundry, demonstrated that piloting AI for WIP flow optimization could yield a 5% reduction in cycle time, a 4.5% increase in lot moves, a 14.6% reduction in lot lateness, and a 2.5% improvement in tool utilization.

Critical for the deployment of these AI-driven solutions in fabs are some baseline enablers that need to be tackled first.

Rich and consistent historical as well as real-time manufacturing data—backed by robust MES and recipe management systems—is essential for effective AI adaptation.

Cross-functional teams and thoughtful change management ensure technical feasibility and operational acceptance, particularly as autonomy advances.

Furthermore, standardized data schemas promote seamless integration, easier scaling, and unified analytics.

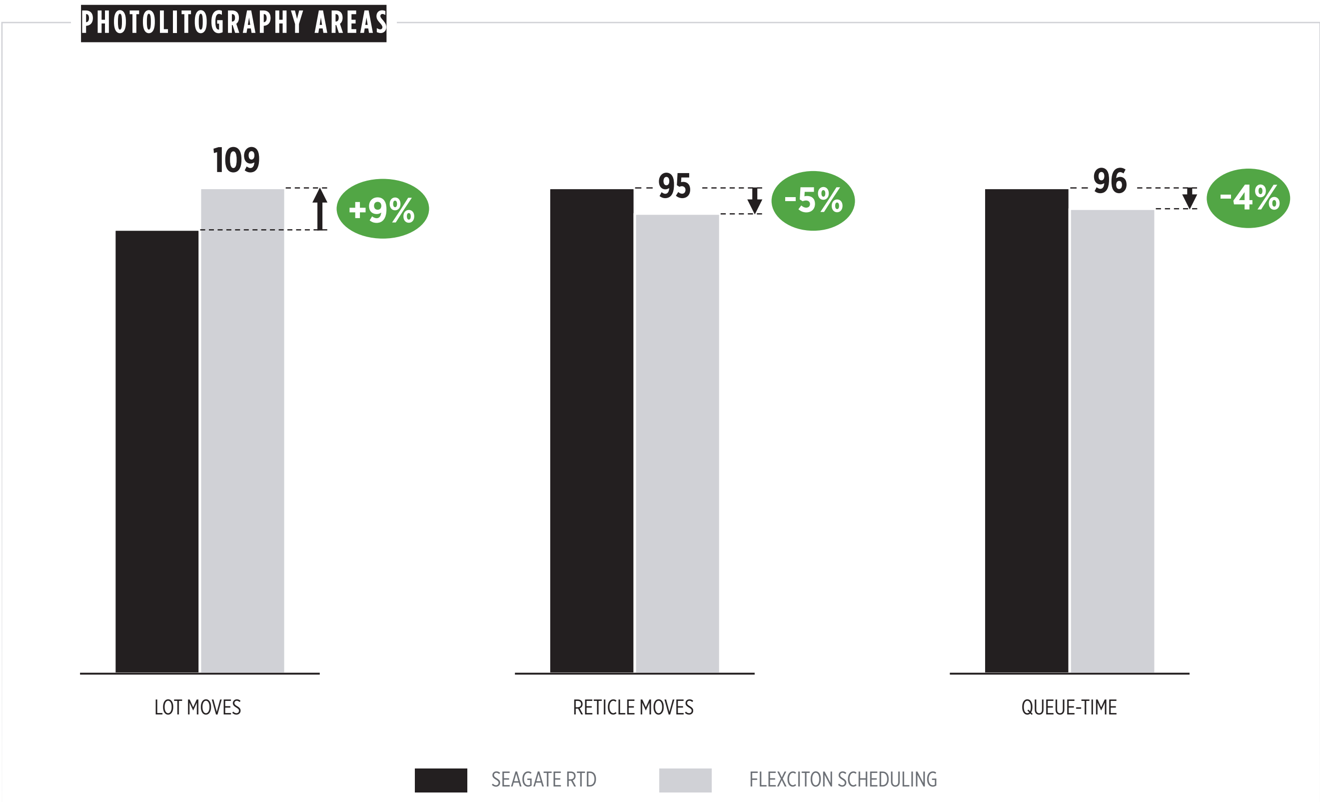


Figure 25: Published results from deploying AI Scheduling at Seagate

Key learnings from AI deployment in fabs

Experience in deploying AI across semiconductor fabs reveals several key learnings.

01 Strong collaboration between fabs and technology vendors is vital, with internal champions required to contextualize change and drive adoption

02 Data integrity emerges as non-negotiable—clean, validated, and pipeline-ready data power AI performance and trust

03 Many operational constraints within the “hidden factory” surface only during rollout, emphasizing the need for iterative deployments and continuous feedback

- **Effective change management, including communication, training, and demonstration of KPI improvements, helps overcome resistance and fosters confidence in AI-driven decisions**
- **Additionally, intuitive visualizations – clear UI designs that illuminate AI logic – significantly boost user trust and operational clarity**
- **Defining success from the outset, with clear KPIs and rigorous analytics, remains a cornerstone of AI project effectiveness**

CONCLUSION

A new essential standard, AI integration in production planning

As traditional approaches struggle under escalating fab complexity, the integration of AI into production planning and scheduling has become not just advantageous but essential. AI's ability to optimize, explain, and autonomously improve manufacturing KPIs is setting a new standard for the industry. Fabs that embrace these technologies are securing far more than process efficiency – they are laying the foundation for adaptive, learning-driven manufacturing where the factory itself becomes a continuously optimizing system.

Authors

Marco Grafe
Solution Specialist for Smart Manufacturing, ZEISS

marco.grafe@zeiss.com

As part of the **ZEISS Group,** ZEISS Digital Innovation drives digital transformation in manufacturing and the semiconductor industry. Combining deep technology expertise with agile software development, the company delivers cutting-edge, future-proof solutions that enable smart fabs, holistic data integration, and measurable improvements in productivity, quality, and efficiency—helping customers achieve sustainable competitive advantage.

Mark Yeates
Sales Segment Manager, Inficon

mark.yeates@inficon.com

INFICON is a global leader in sensor technology for the semiconductor industry, with expertise ranging from precision pressure measurement to advanced gas analysis. Its solutions enable real-time leak detection, contamination monitoring, and higher process reliability. By combining decades of sensor innovation with edge AI development, INFICON helps fabs maximize yield, throughput, and overall equipment performance.

Fahad Golra
Director Product Innovation, Agileo Automation

fahad.golra@agileo.com

Agileo Automation is a French software company specializing in equipment control and connectivity solutions for the semiconductor industry. It supports equipment manufacturers worldwide in automation, digitalization, and factory integration, with a focus on SEMI standards compliance. It helps OEMs accelerate equipment development, ensure compliance, and transform complex technologies into reliable production systems.

Michael Arnold
Managing Director, PEER Group

michael.arnold@peergroup.com

PEER Group is the leading global supplier of factory and tool automation and connectivity software for the semiconductor industry. Founded in 1992, the company delivers best-in-class solutions that help OEMs and fabs reduce time to market, risk, and costs by addressing complex challenges in automation, connectivity, and SEMI standards compliance.

Francisco Almada Lobo
CEO, Critical Manufacturing

franciscolobo@criticalmanufacturing.com

Critical Manufacturing, part of ASMPT, stands at the forefront of digital transformation with its Industry 4.0-ready MES platform. Headquartered in Portugal and present worldwide, the company supports high-tech industries such as semiconductors, electronics, and medical devices in scaling smart factory operations. Its platform combines data integration, equipment engineering, quality, and logistics, enabling manufacturers to rapidly adapt and optimize. Recognized by Gartner, IDC, and Frost & Sullivan, Critical Manufacturing continues to shape the future of intelligent manufacturing.

Jamie Potter
CEO & Co-Founder, Flexciton

jamie.potter@flexciton.com

Flexciton is a leader in AI-powered optimization for fab scheduling and planning. Founded in 2016, it has deployed patented technologies in fabs worldwide, proving its ability to manage highly complex, constraint-rich environments. With one of the largest optimization teams in the sector, Flexciton drives innovation at Tier 1 fabs and advances dynamic capacity modelling for smarter, AI-driven production planning.

Authors

Porsche Consulting
Strategic Vision. Smart Implementation.

Giovanni Notarnicola
Partner & Chairman of SEMI E2E Smart manufacturing community
giovanni.notarnicola@porsche-consulting.com

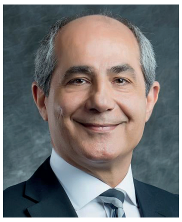
Marco De Appollonia
Senior Manager
marco.de_appollonia@porsche-consulting.com

Marco Paul
Industry Expert
marco.paul@porsche-consulting.com

We extend our sincere appreciation to the following colleagues for their valuable contributions to the preparation of this white paper:
Edoardo Menghini, Senior Consultant
Manuel Micaroni, Consultant
Emma Di Liberto, Associate Consultant

Porsche Consulting

Porsche Consulting GmbH is a leading German management consultancy and employs 950 people worldwide. The company, leading in putting strategies into practice, is a subsidiary of the sports car manufacturer Porsche AG. In addition to its headquarters in Stuttgart, the consultancy has offices in Hamburg, Munich, Berlin, Frankfurt am Main, Milan, Paris, São Paulo, Shanghai, Beijing, Atlanta, and Palo Alto. The experts in the areas of strategy and organization, brand and sales, development and technology, as well as operations, advise clients globally in the automotive, consumer goods, aerospace, industrial goods, life sciences, energy, construction, and transportation sectors. Following the principle of "Strategic vision. Smart implementation," the management consultancy has a clear mission: generating competitive advantage based on measurable results – always with a focus on people.



Laith Altimime
President, SEMI Europe

Ana Bernardo
Senior Manager, Technology Programs, SEMI Europe

About SEMI

SEMI® is the global industry association connecting over 3,000 member companies and 1.5 million professionals worldwide across the semiconductor and electronics design and manufacturing supply chain. We accelerate member collaboration on solutions to top industry challenges through Advocacy, Workforce Development, Sustainability, Supply Chain Management and other programs. Our SEMICON® expositions and events, technology communities, standards and market intelligence help advance our members' business growth and innovations in design, devices, equipment, materials, services and software, enabling smarter, faster, more secure electronics.

Appendix

References

[1] SEMI , "Global Semiconductors industry plans to invest $400 billion in 300mm Fab equipment over next 3 years," SEMI, 26 September 2024. [Online]. Available: https://www.semi.org/en/semi-press-releases/global-semiconductor-industry-plans-to-invest-%24400-billion-in-300mm-fab-equipment-over-next-three-years-semi-reports.

[2] News Desk, "EU Faces Setback on 2030 Semiconductor Production Target, Auditors Reveal," SEMICONDUCTORS INSIGHTS, 29 April 2025. [Online]. Available: https://semiconductorsinsight.com/eu-faces-setback-on-2030-semiconductor-production-target-auditors-reveal/.

[3] J.-B. Smits, "The Semiconductor Talent Crisis: Why Growing Demand Can't Find Leaders," SEMI, 23 June 2025. [Online]. Available: https://www.semi.org/eu/blogs/the-semiconductor-talent-crisis-why-growing-demand-cant-find-leaders.

[4] Porsche Consulting, "SMART MANUFACTURING REPORT 2024," SEMI, November 2024. [Online]. Available: https://www.semi.org/en/form/smart-manufacturing-report.

[5] N. S. a. J. M. Lucas Vann, "Are You Running Your Equipment or is Your Equipment Running You?," APPLIED Materials, 2021. [Online]. Available: https://appliedsmartfactory.com/wp-content/uploads/2021/10/Applied_Are_You_Running_Your_Equipment_White_Paper.pdf.

[6] F. Hicks, "AI In Semiconductor Industry : Innovations Ahead," Aegis, 31 December 2024. [Online]. Available: https://www.aegissofttech.com/insights/ai-in-semiconductor-industry/.

[7] W. Hahn, "SCREEN and IBM Sign Agreement for Next-Generation EUV Lithography Cleaning Process Development," IBM, 24 09 2025. [Online]. Available: https://newsroom.ibm.com/2025-09-24-screen-and-ibm-sign-agreement-for-next-generation-euv-lithography-cleaning-process-development.

[8] Averroes, "How AI is Radically Advancing Semiconductor Manufacturing," 13 02 2025. [Online]. Available: https://averroes.ai/blog/artificial-intelligence-in-semiconductor-manufacturing.

[9] "Cayenne Testing under extreme conditions - real and virtual," Motor Illustrated, 18 09 2025. [Online]. Available: https://motorillustrated.com/porsche-cuts-cayenne-electric-development-time-by-20-through-digital-twin-testing/163023/.

[10] D. Shapiro, "Mercedes-Benz Taking Vehicle Product Lifecycle Digital With NVIDIA AI and Omniverse," Nvidia, 23 02 2023. [Online]. Available: https://www.nvidia.com/en-us/solutions/autonomous-vehicles/partners/mercedes/.

[11] A. L. S.-C. M. S. a. R. S. C. R. Rakholia, "Advancing Manufacturing Through Artificial Intelligence: Current Landscape, Perspectives, Best Practices, Challenges, and Future Direction," IEEE Access, Bd. 12, pp. 131621-131637, 2024.

[12] N. Chen, "Integrating AI into Semiconductor Design and Fabrication: Methodologies, Challenges and Future Prospects," in ITM Web conf. Vol. 78, International Conference on Computer Science and Electronic Information Technology (CSEIT 2025), 2025.

[13] Intel, "How a Semiconductor Factory Works," Intel, 09 06 2023. [Online]. Available: https://newsroom.intel.com/tech101/how-a-semiconductor-factory-works.

Appendix

References

[14] TSMC, “Agile and Intelligent Operations,” TSMC, [Online]. Available: https://www.tsmc.com/english/dedicatedFoundry/manufacturing/intelligent_operations. [Zugriff am 29 09 2025].

[15] Metrology.news, “Porsche Automates Paint Inspection with AI-Powered Robotic System,” Porsche, 21 04 2025. [Online]. Available: https://metrology.news/porsche-automates-paint-inspection-with-ai-powered-robotic-system/.

[16] S. Woo, “How AI Is Driving Changes Throughout The Semiconductor Industry,” Forbes, 05 09 2025. [Online]. Available: https://www.forbes.com/councils/forbestechcouncil/2025/09/05/how-ai-is-driving-changes-throughout-the-semiconductor-industry/.

[17] “Europe, a laggard in AI, seizes the lead in its regulation,” The Economist, 10 12 2023. [Online]. Available: https://www.economist.com/europe/2023/12/10/europe-a-laggard-in-ai-seizes-the-lead-in-its-regulation.

[18] AECouncil, “AECouncil,” [Online]. Available: http://www.aecouncil.com/. [Zugriff am 29 09 2025].

[19] M. Draghi, “The Draghi Report: A Competitiveness Strategy for Europe,” 2024.

[20] P. Bendor-Samuel, “Reasons why generative AI pilots fail to move into production,” Forbes, 2024. [Online]. Available: https://www.forbes.com/sites/peterbendorsamuel/2024/01/08/reasons-why-generative-ai-pilots-fail-to-move-into-production/.

[21] Z. D. Innovation, “Data Enablement,” Zeiss Digital Innovation, 2025. [Online]. Available: https://www.zeiss.com/digital-innovation/manufacturing-solutions/data-enablement.html.

[22] Z. D. Innovation, “VW AG expands production platform with digital shop floor management,” Zeiss Digital Innovation, 2024. [Online]. Available: https://www.zeiss.com/digital-innovation/references/digital-shop-floor-management.html.

[23] C. Manufacturing, “Welcome to the thinking Factory,” 2025. [Online]. Available: https://www.criticalmanufacturing.com/campaign/welcome-to-the-thinking-factory/.

[24] Flexciton, “Seagate Case Study,” Flexciton, [Online]. Available: https://www.flexciton.com/resources/seagate-case-study-2-0.

[25] P. Singer, “Building Fabs in the U.S. vs Taiwan: Twice as Long, Twice as Much,” SEMICONDUCTORS DIGEST, 18 February 2025. [Online]. Available: https://www.semiconductor-digest.com/building-fabs-in-the-u-s-vs-taiwan-twice-as-long-twice-as-much/.

[26] P. C. EU, “Comparison of European and non-European region clusters in KETs,” A study prepared for the European Commission, [Online]. Available: https://www.ipcei-me.eu/wp-content/uploads/2020/11/ComparisonofEuropeanandnon-EuropeanregionalclustersinKETs-thecaseofsemiconductors_2013.pdf

Porsche Consulting
Strategic Vision. Smart Implementation.

semi™

SMART MANUFACTURING