



Baština Akademije nauka i umjetnosti Bosne i Hercegovine

SMART MANUFACTURING ON THE HORIZON: A Conceptual Shift from Industry 4.0 to Industry 5.0: Industrija

Karabegović, Isak

2026

Academy of Sciences and Arts of Bosnia and Herzegovina

<https://bastina.anubih.ba/handle/123456789/875>

Preuzeto s Baštine Akademije nauka i umjetnosti Bosne i Hercegovine

<https://bastina.anubih.ba/>



SMART MANUFACTURING ON THE HORIZON:
A Conceptual Shift from Industry 4.0 to
Industry 5.0



AKADEMIJA NAUKA I UMJETNOSTI BOSNE I HERCEGOVINE
АКАДЕМИЈА НАУКА И УМЈЕТНОСТИ БОСНЕ И ХЕРЦЕГОВИНЕ
ACADEMY OF SCIENCES AND ARTS OF BOSNIA AND HERZEGOVINA

Posebno izdanje
Knjiga CCXXV

Odjeljenje tehničkih nauka
Knjiga 26

Međunarodna naučna konferencija
PAMETNA PROIZVODNJA NA HORIZONTU:
Konceptualni pomak od Industrije 4.0 ka Industriji 5.0



Sarajevo, 1–2. oktobra 2026.

Zbornik radova

Urednik
Isak Karabegović

Sarajevo - Bihać, 2026.



AKADEMIJA NAUKA I UMJETNOSTI BOSNE I HERCEGOVINE
АКАДЕМИЈА НАУКА И УМЈЕТНОСТИ БОСНЕ И ХЕРЦЕГОВИНЕ
ACADEMY OF SCIENCES AND ARTS OF BOSNIA AND HERZEGOVINA

Special Editions
Volume CCXXV

Department of Technical Sciences
Volume 26

International Scientific Conference
SMART MANUFACTURING ON THE HORIZON:
A Conceptual Shift from Industry 4.0 to Industry 5.0



Sarajevo, 1st – 2nd October 2026

Proceedings

Editor
Isak Karabegović

Sarajevo - Bihać, 2026

SMART MANUFACTURING ON THE HORIZON:
A Conceptual Shift from Industry 4.0 to Industry 5.0
Sarajevo, 1st – 2nd October 2026

Publisher:

Academy of Sciences and Arts of Bosnia and Herzegovina
Society for Robotics of Bosnia and Herzegovina

For the publisher:

Academician Muris Čičić
Prof. Samir Vojić, PhD

Organizing committee:

Isak Karabegović, PhD
Ahmed Kovačević, PhD
Namik Hrle, PhD
Sead Pašić, PhD
Safet Isić, PhD
Samir Vojić, PhD
Mirha Bičo Ćar, PhD
Mehmed Mahmić, PhD
Ermin Husak, PhD

Editor:

Corresponding member Isak Karabegović

Reviewers:

Prof. Safet Isić, PhD; Prof. Mijodrag Milošević, PhD; Prof. Isak Karabegović, PhD,
Prof. Domenico Guida, PhD; Prof. Slavenko Stojadinović, PhD; Prof. Avdo Voloder,
PhD; Prof. Sead Pašić, PhD; Prof. Radivoje Mitrović, PhD; Prof. Predrag Dašić; Prof.
Ermin Husak, PhD; Prof. Vladimir Tankonogy, PhD; Prof. Savo Stupar, PhD; Prof. Emir
Kurtović, PhD; Prof. Jasmin Suljagić, PhD; Doc. Sanela Nazdrajić, PhD; Prof. Nermin
Goran, PhD; Prof. Aco Antić, PhD; Prof. Aleksandar Živković, PhD.

Translation and lecturer for English:

Mirela Midžić, MSc

DTP

Mehmed Mahmić, PhD

The press

Štamparija Fojnica d.d.
Fojnica

Printing
100

Sarajevo 2026
EBSCO

ISBN 978-9926-574-25-3 Akademija nauka i umjetnosti
Bosne i Hercegovine
ISBN 978-9926-9163-0-5 Udruženje Društvo za robotiku u
Bosni i Hercegovini
CIP zapis dostupan u COBISS sistemu Nacionalne i
univerzitetske biblioteke BiH pod
ID brojem 70679302



CONTENTS

<i>Preface</i>	1
<i>Vidosav Majstorović</i> THE FUTURE OF SMART MANUFACTURING - INTEGRATION, INTELLIGENCE AND INNOVATION	3
<i>Isak Karabegović, Ermin Husak, Mehmed Mahmić, Samir Vojić, Isić Safet, Pjereta Agalliu</i> ARCHITECTURE OF INDUSTRY 4.0 SMART MANUFACTURING SYSTEMS BASED ON MOBILE AND COLLABORATIVE ROBOTICS	27
<i>Ermin Husak</i> APPLICATION OF AUGMENTED REALITY IN COLLABORATIVE ROBOTS	59
<i>Mijodrag Milošević</i> SMART MANUFACTURING ENABLED BY LLM AI AGENTS AND RETRIEVAL-AUGMENTED GENERATION	73
<i>Osman Lindov</i> A HUMAN-CENTRIC DIGITAL TWIN FRAMEWORK FOR INDUSTRY 5.0 IN TRANSPORT AND LOGISTICS	85
<i>Munira Šestić, Mirha Bičo Ćar, Anes Hrnjić, Amila Bijedić</i> DIGITALIZATION AND THE CIRCULAR ECONOMY IN BOSNIA AND HERZEGOVINA: CHALLENGES OF THE TWIN TRANSITION IN A TRANSITIONAL ECONOMY	99
<i>Amra Bratovčić</i> INDUSTRY 4.0 AND GREEN CHEMISTRY INTEGRATION FOR SUSTAINABLE SMART CHEMICAL MANUFACTURING	111
<i>Dušan Janković, Goran Janković, Zora Joković, Aleksandar Supurović, Vidosav Majstorović</i> AN INTEGRATED MODEL OF INDUSTRY 4.0 AND INDUSTRY 5.0 FOR QUALITY MANAGEMENT DURING THE ENTIRE LIFE CYCLE OF PLASTIC INJECTION MOLDS	123
<i>Enes Saletović, Malik Kriještorac, Đulaga Hadžić</i> WIRELESS COMMUNICATION LINK OPTIMIZATION IN CYBER- PHYSICAL SYSTEMS USING SMART SENSOR NETWORKS	153

<i>Milica Tufegdžić</i> AN EXPLAINABLE ARTIFICIAL INTELLIGENCE FRAMEWORK FOR PREDICTIVE MAINTENANCE USING HYBRID MACHINE LEARNING AND MULTI-TASK TRANSFORMER MODELS	163
<i>Nerma Mezetović, Lejla Banjanović-Mehmedović, Isak Karabegović</i> EXPLAINABLE AND TRUSTWORTHY AI FOR STEEL PLATES FAULT DETECTION USING SHAP AND LIME	195
<i>Samir Lemeš</i> SEMANTIC DATA POLLUTION: A NEW CHALLENGE FOR CYBERSECURITY AND STANDARDIZATION OF AI MODELS IN SMART FACTORIES	223
<i>Sonja Dimitrijević, Vidosav D. Majstorović</i> CYBER SECURITY OF SMART MANUFACTURING ACCORDING TO THE INDUSTRY 5.0 MODEL: CHALLENGES, SOLUTIONS AND FUTURE DEVELOPMENT DIRECTIONS	241
<i>Viktorio Malisa</i> DARK FACTORY VS. HUMANOID FACTORY: TWO DEVELOPMENT DIRECTIONS OF INDUSTRIAL AUTOMATION IN COMPARISON	271
<i>Madžida Hundur Hiyari, Nejra Merdović, Jovana Kepčija Pavlović, Adna Softić, Amra Džuho, Ana Lalović</i> DIGITAL COMPETENCE AND TECHNOLOGY ADOPTION IN VOCATIONAL EDUCATION: EVIDENCE FROM BOSNIA AND HERZEGOVINA	291
<i>Nejra Merdović, Madžida Hundur Hiyari, Adna Softić, Lejla Gurbeta Pokvić</i> FROM DIGITAL MATURITY ASSESSMENT TO DIGITAL TRANSFORMATION STRATEGY: EVIDENCE FROM SMES IN BOSNIA AND HERZEGOVINA.....	301
LIST OF AUTHORS	309



FOREWORD

The 5th International Scientific Conference "Smart Manufacturing on the Horizon: A Conceptual Shift from Industry 4.0 to Industry 5.0" continues the international scientific dialogue on the transformation of modern industry, with particular emphasis on the development of intelligent manufacturing systems that integrate digital technologies, sustainability, and the human being as the central driver of innovation. The conference is organized by the Academy of Sciences and Arts of Bosnia and Herzegovina – Department of Technical Sciences, in cooperation with the Robotics Society in Bosnia and Herzegovina and the Foreign Trade Chamber of Bosnia and Herzegovina, with the aim of strengthening collaboration among academia, industry, and policymakers.

Today, the concept of smart manufacturing extends far beyond the digitalization of production processes. The emergence of Industry 5.0 has shifted attention toward intelligent, adaptive, and resilient manufacturing systems in which advanced technologies—including artificial intelligence, mobile and collaborative robotics, digital twins, large language model (LLM) AI agents, the Internet of Things, and cyber-physical systems—serve to augment human capabilities while promoting sustainability, resilience, and responsible industrial development.

The papers presented in this Proceedings reflect this interdisciplinary vision of future manufacturing. The contributions address the foundations of smart manufacturing and Industry 5.0, sustainable manufacturing and digital transformation, digital infrastructure and cyber-physical systems, explainable artificial intelligence for predictive maintenance, cybersecurity and AI governance, as well as emerging directions in intelligent industrial automation. The conference is further enriched by a panel discussion dedicated to the future of smart manufacturing through the integration of intelligence, innovation, and interdisciplinary collaboration.

Modern industry faces increasingly complex challenges requiring higher productivity, energy efficiency, operational resilience, and secure digital infrastructures. Consequently, smart manufacturing has become a strategic framework that combines technological innovation with economic competitiveness, environmental responsibility, and social sustainability. The research presented in this volume introduces new concepts, methodologies, and practical solutions that contribute to the understanding and implementation of these transformative developments.

The international character of the conference is reflected in the participation of distinguished researchers and experts from different countries and scientific disciplines. The exchange of knowledge between academia and industry creates opportunities for new research partnerships, collaborative projects, and faster transfer of scientific achievements into industrial practice.

We express our sincere appreciation to all authors, reviewers, members of the Scientific and Organizing Committees, panelists, and participants whose valuable contributions ensured the success of the conference and the high quality of this Proceedings volume. Their commitment once again demonstrates that scientific collaboration remains one of the most important driving forces behind technological progress and sustainable industrial development.

We believe that these Proceedings will serve as a valuable source of knowledge for researchers, educators, students, engineers, industrial practitioners, and policymakers, while encouraging further innovation, collaboration, and implementation of smart manufacturing principles and the Industry 5.0 vision in Bosnia and Herzegovina and beyond.

Sarajevo, October 2026

*Corresponding member Isak Karabegović, editor
Academy of Sciences and Arts of Bosnia and Herzegovina*



The Future of Smart Manufacturing - Integration, Intelligence and Innovation

Vidosav Majstorović^{*1}

Abstract: *Smart manufacturing represents the confluence of digital transformation and industrial evolution, redefining manufacturing paradigms through seamless integration, artificial intelligence (AI)-driven intelligence, and continuous innovation. This paper explores the future development of smart manufacturing, highlighting the synergistic interaction of Cyber-Physical Systems (CPS), Industrial Internet of Things (IIoT) and Advanced Analytics (BDA) to create adaptive, resilient and sustainable factories. Central to this vision is technology integration, which enables full visibility into supply chains, machines and business systems through standardized protocols and architectures from the edge to the cloud. Intelligence emerges through machine learning (ML), digital twins (DT) and generative artificial intelligence (GenAI), developing predictive maintenance (PrM), autonomous quality control (AQC) and real-time optimization. Innovation is manifested in revolutionary applications – additive manufacturing (AM), human-robot collaboration (cobot) and biologically inspired manufacturing models – that challenge conventional limitations of scalability and customization. Despite progress, significant challenges remain: interoperability gaps, cybersecurity vulnerabilities, workforce skills shortages, and ethical issues related to AI autonomy. The paper critically examines these obstacles while presenting practical ways forward, including modular architectures, federated learning and policy frameworks for responsible adoption. Looking ahead to 2035, the analysis predicts a shift towards Industry 5.0, dominated by human-centric systems, circular economy (CE) and quantum-enhanced systems. Synthesizing current research, industry case studies and emerging trends, this paper provides a strategic roadmap for stakeholders to navigate the next decade of intelligent manufacturing.*

Keywords: *Smart Manufacturing, Industry 4.0, AI Integration, Digital Twins, IIoT, Sustainability, Cyber-Physical Systems*

1. Introductory Notes

The evolution of manufacturing reflects humanity's progressive mastery of tools, energy and information, moving from manual craftsmanship to intelligent, interconnected ecosystems. Industry 1.0 (late 18th century) was created by the mechanization of textile production, driven by water and steam engines, enabling mass production through standardized processes [1]. Industry 2.0 (early 20th

^{*1}The University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia
E-mail: vidosav.majstorovic@gmail.com

century) introduced electricity and assembly lines, epitomized by Ford's conveyor belt system, which dramatically increased throughput and reduced costs through the division of labor and interchangeable parts [2]. The emergence of Industry 3.0 (from the 1970s onwards) marked the era of automation, powered by programmable logic controllers (PLCs), robotics and early computing. This phase moved manufacturing from analog to digital control, enabling flexible manufacturing systems (FMS) and computer integrated manufacturing (CIM) [3]. However, the systems remained largely isolated (islands), with limited real-time data exchange. Industry 4.0, formalized in 2011 by the German government [4], represents a paradigm shift towards cyber-physical systems (CPS) – networks of embedded sensors, actuators and software that enable machines to perceive, decide and act autonomously. Key technologies Industrial Internet of Things (IIoT), cloud computing, big data analytics (BDA) and additive manufacturing (AM) have led to this development. Smart manufacturing, which in practice is often used interchangeably with Industry 4.0, is defined by the American National Institute of Standards and Technology (NIST) as "fully integrated, collaborative manufacturing systems that respond in real time to meet changing demands and conditions in the factory, supply network and customer needs" [5]. The new discourse on Industry 5.0 [6] revises the human role, emphasizing human-artificial intelligence (AI) collaboration, resilience and sustainability. In contrast to the technological focus of Industry 4.0, Industry 5.0 prioritizes social value, circular economies (CE) and personalized manufacturing. Smart manufacturing therefore evolves from connectivity (4.0) to cognitive, adaptive and value-driven systems (5.0), laying the foundations for the integrated, intelligent and innovative future explored in this paper.

Defining the problem: Despite the transformative potential of Industry 4.0 and the new vision of Industry 5.0, modern manufacturing systems are still hampered by systemic fragmentation, operational inefficiencies and scalability bottlenecks, which undermine competitiveness, resilience and sustainability. The main challenge is the existence of silos of data and functionality – disconnected islands of automation in design, manufacturing, logistics and enterprise resource planning (ERP) systems. According to Deloitte research [7], 68% of manufacturers still work with legacy systems that lack real-time interoperability, resulting in delayed decision-making, redundant data entry and error-prone manual interventions. Inefficiencies manifest in many forms: unplanned downtime (800 hours per year on average per facility, according to [8]), energy waste due to suboptimal process control, and quality deficiencies resulting from reactive rather than predictive monitoring. Supply chain volatility—exacerbated by geopolitical disruptions and post-pandemic recovery—further exposes the fragility of linear, "just-in-time" models that lack adaptive intelligence. Scalability issues arise when trying to deploy smart technologies after the pilot stages. High start-up costs (73% of firms

report a shortage of digital talent, [9]) and resistance to change. Furthermore, small and medium-sized enterprises (SMEs), which account for over 85% of the global manufacturing base, have difficulty integrating advanced solutions due to limited infrastructure and standardized frameworks. Without addressing barriers to integration, proactive intelligence scalable innovation, manufacturers risk obsolescence in an era that requires hyper-customization, net-zero operations, and real-time response to dynamic market and environmental pressures. This paper confronts these shortcomings in order to chart a strategic way forward.

Research Objectives and Research Questions: This paper provides a forward-looking, integrative analysis of smart manufacturing to guide industry transformation over the next decade (2025–2035.). It bridges theoretical frameworks with practical roadmaps, emphasizing the triad of integration, intelligence and innovation as the main drivers of the maturation of Industry - Industry 4.0 and the emergence of Industry 5.0.

The primary goal is to articulate a comprehensive vision of next-generation smart manufacturing by synthesizing current technology capabilities, identifying barriers to adoption, and anticipating evolutionary pathways toward resilient, human-centered, and sustainable manufacturing ecosystems. *The partial objectives* are: (a) examine the integration mechanisms that enable seamless connectivity between cyber-physical systems, supply networks, and enterprise layers, (b) assess the role of artificial intelligence in the transition from automation to autonomous, cognitive manufacturing, (c) explore advanced innovations, including additive manufacturing, bio-digital convergence, and circular manufacturing models—and their scalability potential, and (d) assess socioeconomic and environmental challenges (labour reskilling, energy efficiency, ethical artificial intelligence management) and suggest mitigation strategies. *The guiding research questions* are: *RQ1*: How can horizontal and vertical integration eliminate data silos and enable real-time, end-to-end visibility in complex manufacturing environments? *RQ2*: In what ways will AI-driven intelligence—via digital twins, generative design, and autonomous decision systems—transform operational efficiency and product personalization? *RQ3*: What new innovations will redefine manufacturing paradigms and how can they be scaled across different enterprise sizes and regions? *RQ4*: What are the key enablers and barriers to achieving sustainable, resilient and human-centered smart factories within Industry 5.0? *RQ5*: How will integration, intelligence, and innovation come together to shape manufacturing competitiveness, workforce dynamics, and environmental impact over the next decade?

2. Defining the Context of Smart Production

The journey towards smart manufacturing can best be understood through four distinct industrial revolutions. Milestone 1 was Industry 1.0 – mechanization and steam power (late 18th century). The key change was the transition from manual production methods to production on machines powered by water and steam. This led to the creation of the "factory system", as production was centralized, leading to the mass production of goods such as textiles. It was a transition from craft work to mechanized work. Milestone 2 was Industry 2.0 – electrification, mass production and the assembly line (pioneer Henry Ford) [11] (late 19th - early 20th century), which enabled more flexible factory machine layouts and division of labor. Mass production became the gold standard, and the assembly line [12] drastically reduced production costs and made goods like cars affordable to the masses. This era also marked the introduction of concepts such as scientific management (Taylorism) to optimize human work. This model added "nervous system" (electricity) and "process" (assembly line) to the industrial muscle. Milestone 3 was Industry 3.0 – Automation, Computers and Robotics (1970s - 1990s), where the invention of the programmable logic controller (PLC) and the rise of electronics and computers were the main features of this era. A key change came from the integration of digital technology and automation into manufacturing processes [13], where the PLC replaced complex relay systems, facilitating the automation of individual machines and processes. Industrial robots have begun to perform dangerous, repetitive or precise tasks (automotive welding and painting). The development and implementation of IT systems for enterprise resource planning (ERP) and work order management systems (MES) began, which started the digitization of business operations and the management of production plants. Thus, these systems have become "islands of automation". A robot or CNC machine was automated, but could not communicate with other systems. This was "the brain" - the introduction of digital control and computation, but the brain could not easily communicate with other brains or the body. Thus we have reached the dawn of smart manufacturing: the Era of Industry 4.0 (2000s - today), where the concept of "smart" appears. Industry 4.0 is characterized by the fusion of the physical and digital world through cyber-physical systems (CPS) [14]. In this period, there was a convergence and maturation of several basic technologies that form the backbone of smart production, Figure 1 [15]: (a) Industrial Internet of Things (IIoT) - installation of sensors in machines, tools and products to collect data in real time.

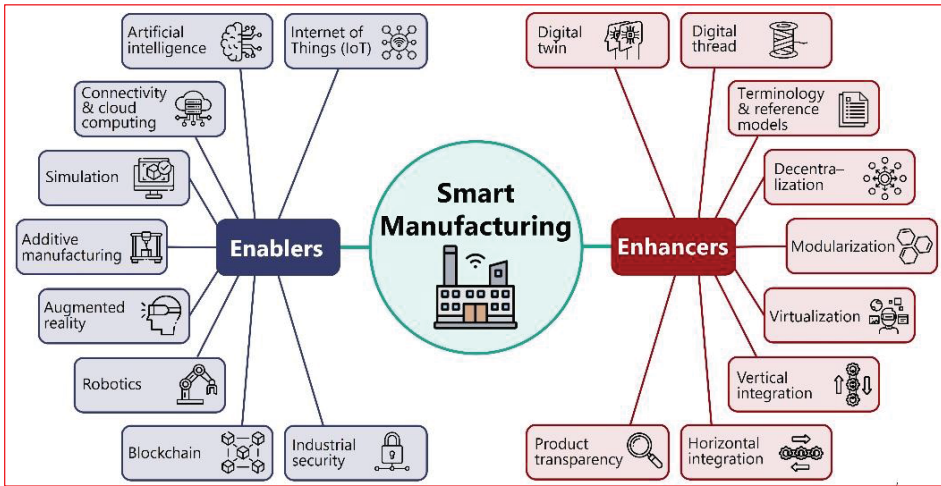


Figure 1. Smart manufacturing model structure [15]

This is the nervous system of the smart factory, (b) big data and advanced analytics (BDA) - the ability to store and process vast amounts of data generated by the IIoT to gain insights, predict failures and optimize processes, (c) cloud/edge computing - providing the scalable, affordable computing power needed to analyze data and connect global operations, in real time, (d) artificial intelligence (AI) and machine learning (ML) - moving from descriptive analytics to predictive and prescriptive actions. AI can optimize production schedules, perform quality control via computer vision and predict maintenance needs (predictive maintenance), (e) additive manufacturing (3D printing) - moving from mass production to mass customization and rapid prototyping, enabling complex geometries and on-demand production, (g) digital twin (DT) - creating a virtual, dynamic replica of a physical asset or process,

3. Defining The Framework of the Smart Production Concept (Integration, Intelligence, Innovation)

Integration (ERP, MES, vertical/horizontal). Integration in smart manufacturing is a strategic and technological practice of seamlessly connecting all systems, data and processes throughout the value chain to create a unified, transparent and fast operation [16]. The central nervous system makes the factory "smart". Without deep integration, technologies like the Internet of Things and artificial intelligence operate in isolation, failing to realize their full transformative potential. The essence of this integration lies in the vertical and horizontal connection of three critical systems: ERP (Enterprise Resource Planning), MES (Work Order Management) and Supply Chain Platform (SCM). The ERP system

is the business brain of the enterprise, managing operations at the business level - finance, human resources, procurement, order management and high-level planning [17]. It answers the question: "What should we produce, at what price and when? It uses low-frequency (cloud computing), high-context business data (customer orders, financial forecasts, raw material orders). MES is the nervous system of the production plant, which monitors, monitors and controls the production process in the production plant [18]. It is the bridge between planning (ERP) and physical operations (OT). It answers the question: "How do we make it, is it according to specification and how far have we come in production? It uses high-frequency (edge computing), operational data (machine status, material consumption, production in progress). And finally, supply chain management (product life cycle management), represents the external connective tissue in this model in warehouses, Product BOMs (BOMs) and engineering change orders. Integration is not just connecting cables, the smart manufacturing model is about creating a continuous, two-way flow of contextualized data. *Vertical integration* is a model that connects the production plant (MES and OT) and all levels up to the top of the enterprise (ERP) [19], Figure 2. The smart integration of production is realized according to the following model - ERP → MES, where a detailed production order is sent, including the specific version of the product, BOM and technological manufacturing process. MES → ERP, provides real-time feedback (edge computing) on order progress, material consumption, produced quantity and work tracking. The moment the planned quantity is produced, the quality check is completed, the work order (WO), the ERP system closes for this WO, with the following facts: (a) real-time costing with actual material and labor consumption, (b) finished goods inventory update is done in real-time, and (c) real-time delivery tracking. *Horizontal integration* is integration throughout the supply chain, from suppliers to customers. Smart manufacturing integration, at this level, starts with connecting to suppliers, so that the manufacturer's MES or ERP, via EDI (Electronic Data Interchange) or API, automatically initiates replenishment orders to suppliers based on real-time consumption data. Suppliers can see manufacturers' inventory levels and production schedules, enabling vendor-managed inventory (VMI) and just-in-time delivery. Connecting with customers is done in such a way that customers can access real-time order status portals or even influence production scheduling. This is the basis of mass customization. PLM integration is performed in such a way that when an engineer finalizes a change to a product design in the PLM system, that change is made automatically transfers to ERP (BOM Update) and MES (Work Order Update), ensuring the factory is always building the latest version. According to the concept presented, integration is the driver of key features of smart manufacturing: (a) end-to-end product visibility and digital thread, because integration creates a "digital thread" - a single, seamless flow of data that connects every aspect of a product's life cycle, from design to disposal.

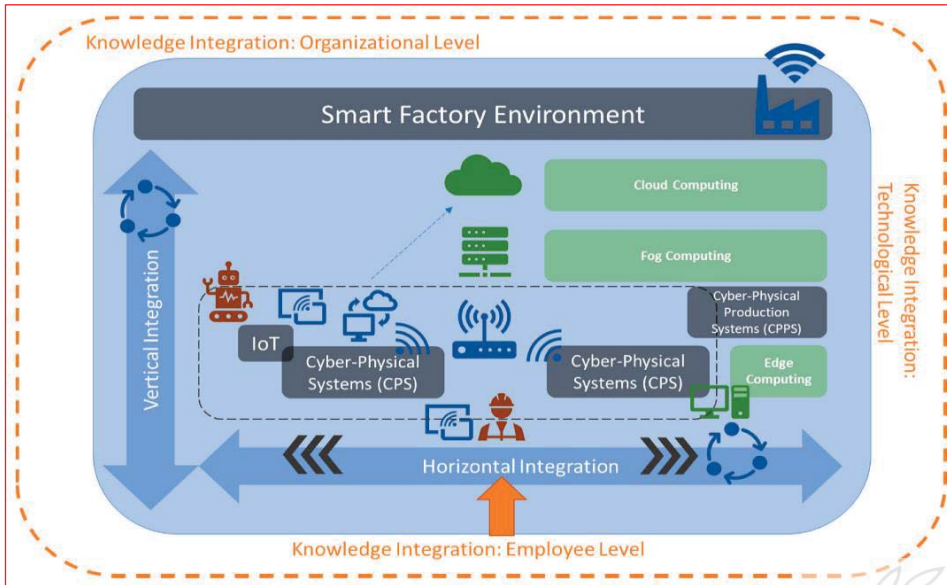


Figure 2. Model and integrations in smart manufacturing [19]

This provides an unprecedented possibility for traceability of all aspects of production, so if a quality problem arises, you can trace it to a specific batch of raw materials, the machine that produced it and the operator on the shift, (b) agile and fast production operations, with real-time data from the production plant (edge computing), the company can immediately react to disruptions. Machine failure automatically triggers order rescheduling in ERP and then informs suppliers of delays in material needs, (c) data-driven decision making, where integration breaks down data silos, creating a single database. This model is now the basis for advanced analytics and artificial intelligence (AI), where an AI algorithm can optimize the entire production plan, if it has access to data from ERP (orders, costs), MES (machine availability, setup time) and SCM (supplier constraints), (d) efficient resource management, where we have real-time material consumption data from the MES, thus preventing both overstocks and stock shortages. Real-time energy monitoring integrated with production scheduling allows optimization of energy consumption, and (e) mass customization, means that in order to profitably produce a "one-piece-sized batch", it also forms a system that must be synchronized. ERP receives a custom order, PLM provides a single BOM and instructions, and MES orchestrates the line to build that particular product non-stop. This is impossible without seamless integration. *The challenges of successful integration* can be defined as [20]: (a) outdated systems, because many factories have decades-old machines and software that were not designed to connect. Retrofitting or bridging these systems is complex and expensive, (b) data

silos and culture - departments may be resistant to sharing data, fearing a loss of control. The cultural shift towards collaboration is as important as the technological one, (c) interoperability - different systems speak different "languages" (protocols). ERP from SAP, MES from Siemens and legacy PLCs from the 1990s all need a common way of communicating, and (d) cyber security, where every new connection point is a potential vulnerability. A secure integration architecture is undeniable and essential. The key technologies that support this concept are: (a) standardized communication protocols - OPC UA (Open Platform Communications Unified Architecture), which provide a secure framework for data exchange, independent of the manufacturer, from the sensor level to the cloud/edge, (b) APIs (application programming interfaces) - modern cloud-based systems, which rely heavily on RESTful APIs, which are the standard method for communicating different software applications over the Internet, (c) midlevel and integration platforms (iPaaS) - tools like MuleSoft or Azure Integration Services act as translators and routers, connecting different systems and transforming data formats between them, (d) cloud/edge - where platforms act as a central hub, facilitating the connection of systems that were previously locked behind corporate protective passwords. They provide scalable computation and storage for the vast amounts of data generated by the integration, and (e) a digital twin - the DT of the production process is both the consumer and the result of the integration. It needs direct transfer of integrated data from ERP, MES and PLC to accurately reflect the physical world. In conclusion, we can say that integration is not a side actor in smart production, it is the main protagonist. It transforms a collection of automated islands into a cohesive, intelligent and self-optimizing organism. While the challenges are significant, the result is a level of operational efficiency, agility and insight that was previously unimaginable. The journey to a truly smart factory does not begin with the purchase of a single advanced robot, but with the strategic and meticulous weaving of the digital thread that connects every asset, process and decision maker.

Intelligence (AI/ML, predictive maintenance, quality control, generative design). Intelligence in smart manufacturing represents a leap from automated to cognitive operations and systems. It is the ability of a production system to learn from data, reason about complex situations, predict future states and make (or recommend) optimal decisions with minimal human intervention. This was developed by bringing together artificial intelligence (AI), machine learning (ML) and predictive analytics (PA), transforming data from a historical record into a proactive, strategic asset. *Predictive analytics* uses historical data and statistical models to predict what is likely to happen. ("based on past vibration patterns, this bearing has an 85% chance of failing within the next 14 days."). *Machine learning* (ML) is a subset of AI, which uses algorithms to analyze data, learn from it, and make decisions or predictions without explicitly programming for that particular

task. The results of learning, and thus of decision-making, are automatically improved through experience. *Artificial intelligence* (AI) is a broad discipline, which as a broader concept of a machine, which has the ability to perform tasks in a way that we would consider "intelligent". ML is the primary way we currently use AI. In manufacturing, this culminates in systems that can perceive, reason and act. Intelligence is not a monolithic system but a set of capabilities applied to specific production domains. *Predictive maintenance (PdM)* is an intelligent approach, as machine learning models analyze sensor data in real time (vibration, temperature, acoustic emissions, wear, power consumption) to detect anomalies and predict failures before they happen, Figure 3 [21].

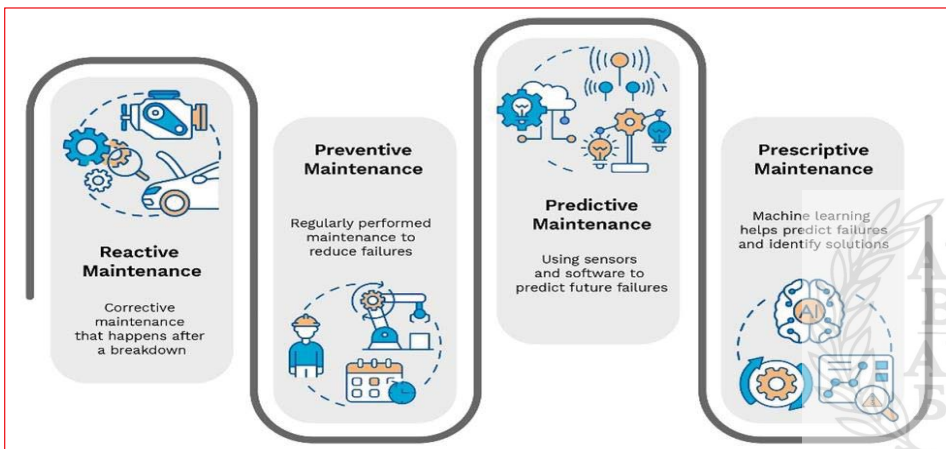


Figure 3. Development of the maintenance model

Using deep data analysis models, such as recurrent neural networks (RNNs), which are extremely good at learning from time series data. They don't just see a high temperature; they learn the complex sequence of events—a slight increase in vibration, followed by a subtle thermal shift, and then a specific increase in energy consumption—that precedes a particular type of failure. In other words, they perform a multifactorial analysis of causes and consequences. This model (PrM) [22], drastically reduces unplanned downtime, extends asset life, optimizes spare parts inventory and shifts maintenance from a cost center to a reliability function. AI-based *quality control* is an intelligent approach that uses computer vision and convolutional neural networks (CNNs) to analyze high-resolution camera images to detect defects that are often invisible to the human eye. The system is trained on thousands of images of "good" and "bad" products. It learns to identify complex, non-linear patterns of defects – microcracks, subtle color changes or misalignments. Unlike a human, it works with 100% consistency, 24/7, without fatigue. Thus, we get "zero scrap" and the ability to trace defects to their root cause

in the process ("95% of defects occur when machine 3 is operating outside parameter set B"), Figure 4 [23].

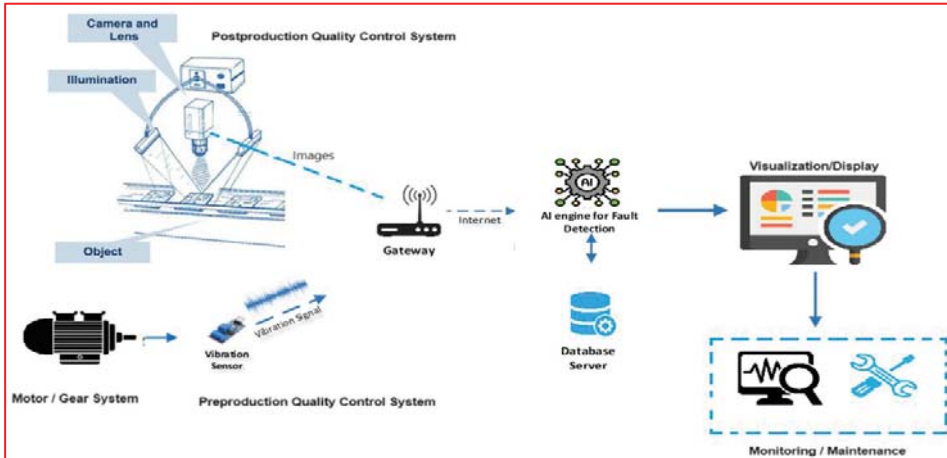


Figure 4. Intelligent Framework for Quality Control in Industry 4.0

Generative design and process optimization guided by artificial intelligence, is based on the application of concepts, where engineers define design goals (weight, strength, material, cost limits), and AI algorithms generate thousands of valid alternative solutions, often giving original, highly efficient forms that a human cannot imagine. Process optimization, uses "reinforcement" learning algorithms, where optimal solutions for the production process are "learned" by generating virtual experiments using digital twins (DT). They continuously adjust the parameters (temperature, pressure, speed) in order to achieve the best result (highest number of produced pieces (productivity), lowest energy consumption), and this knowledge is then applied to the real production line. Radical innovations in product design ensure significant material savings and continuous, autonomous optimization of complex, multivariable production processes. *Self-optimizing supply chains*, based on an intelligent approach, use machine learning models, which analyze a wide range of internal and external data - real-time order flow, weather patterns, geopolitical events, traffic data, supplier risk - to predict disruptions and dynamically reroute logistics, adjust production schedules and reallocate inventory. This creates a resilient, demand-driven supply chain that can anticipate and absorb shocks, minimizing inventory and excess inventory. The basic prerequisites for the development and application of an intelligent model in this area are, Figure 4: (a) a reliable and large enough database, because AI/ML models are big consumers of high-quality, contextualized data, (b) they are reached through integration (previous chapter), because data must flow

seamlessly from machines (IoT sensors), MES, ERP and SCM systems to form a complete picture of the model, (c) data management means that clean, labeled and reliable data is indisputable, inconsistent data must be discarded, (d) cloud/edge computing support is essential, as training complex ML models, especially deep learning networks, requires massive computing resources, which are easily provided by cloud platforms and edge computing devices, and (e) domain expertise is essential, as IT experts must collaborate with production engineers, operators and maintenance technicians who understand the process physics and data context. This partnership is necessary for feature selection and interpretation of results. That is why *the development and application of an intelligent model has several stages* (5 in total) [24], from descriptive analytics to autonomous operation, which at the same time represent the maturity curve of this model. Level 1: Descriptive analytics - "what happened?" (Reporting, dashboards). Level 2: Diagnostic analytics - "why did it happen?" (Analysis of causes). Level 3: Predictive analytics - "what will happen?" (Focus of this analysis). Level 4: Prescriptive analytics - "what should I do about it?" (Automatic intelligence recommends actions). Level 5: Autonomous operations - "the system operates by itself." (Automatic intelligence automatically executes the optimal action, such as a self-adjusting production line - CPS). Smart manufacturing refers to the last two levels.

The challenges and ethical dimensions of this field are: (a) the "black box" problem, where some complex machine learning models (such as deep neural networks) can be incomprehensible, making it difficult to understand why a certain decision was made. This is a major obstacle in safety-critical industries (pharmaceutical, process, military), (b) data quality and availability is the single biggest obstacle, especially for older machines, (d) skills and knowledge shortages, as there is a severe shortage of experts with both data science skills and manufacturing knowledge, (e) change management, as workers may fear that AI will make their jobs obsolete, requiring a thoughtful strategy to retrain and demonstrate how AI augments human capabilities, rather than it replaces them, and (f) security and ethics of AI, because as systems become more autonomous, ensuring their safe operation and making ethical decisions (when allocating resources during a crisis, etc.) becomes paramount. Ultimately, intelligence is the fundamental pillar of smart manufacturing that ensures its full potential. It's moving the industry beyond just doing things faster and cheaper, to doing things significantly better and in entirely new ways. By embedding AI/ML into the very fabric of production, manufacturers gain not only efficiency, but also agility, resilience and a strong new capacity for innovation. The ultimate goal is a symbiotic relationship, where human creativity and strategic oversight are enhanced by tireless, data-driven cognitive systems, creating a future of manufacturing that is not just smart but wise.

Innovations (additive manufacturing, cobots). While the other pillars of smart manufacturing, integration and intelligence, focus on optimizing existing production paradigms, *innovation* is the pillar that fundamentally transforms it. It's about adopting and adapting revolutionary technologies that redefine what's possible in design, manufacturing and the very nature of the factory floor. Innovation in this context goes beyond simply increasing efficiency. It enables new business models, creates unprecedented product capabilities, and reshapes the relationship between workers and machines. Two of the most powerful expressions of this innovative spirit are additive manufacturing (AM) and advanced human-robot collaboration (HRC), Figure 5 [25].

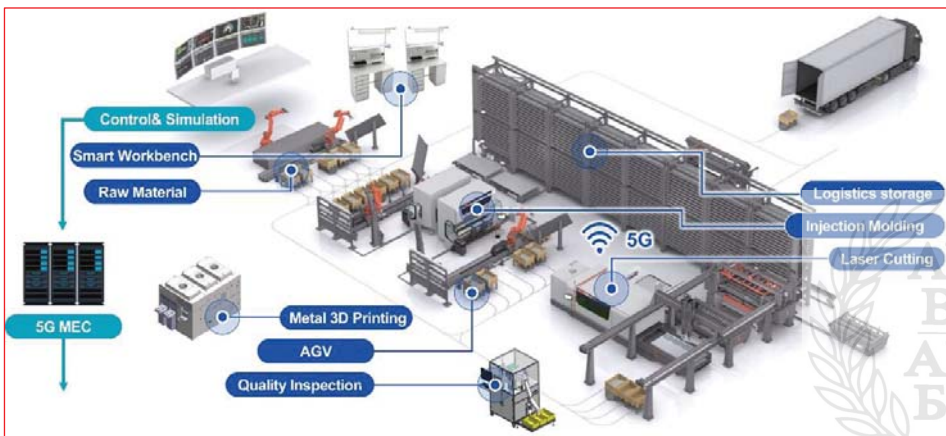


Figure 5. Innovation smart manufacturing [25]

Additive manufacturing is not just a new way of making parts, it is a philosophical shift from subtractive and formative manufacturing (cutting, milling, shaping) to an additive process, building parts layer by layer from digital models. The innovative shift is reflected in the following facts [26]: (a) from design for manufacturing (DFM) to manufacturing for design (MFD). Traditional manufacturing imposes design limitations (inclination angles for molds, tool access paths). AM removes many of these limitations, allowing engineers to design for pure function, (b) from mass production to mass customization, where the economics of AM are decoupled from the scale of production. The production of one unique part is as economically viable as the production of ten identical ones, breaking the centuries-old dominance of the assembly line for certain applications. This concept (AM) has had a profound impact on the manufacturing value chain, product innovation and weight reduction, from the following angles: (a) generative design synergy, where AM physically enabled generative design. AI can create complex, organic and topologically optimized structures that are

traditionally impossible to produce, and AM concept manufacturing can produce them. This is revolutionary in aerospace, automotive and medical implants, where weight reduction leads directly to part performance and efficiency in use, (b) parts consolidation, means that a complex assembly of dozens of parts can be redesigned and printed as a single, integrated component. This reduces assembly time, eliminates potential failure points (fasteners, gaskets) and improves overall operational reliability, (c) digital inventory and make-to-order, where instead of storing physical spare parts in warehouses around the world, companies only need to store digital CAD files and print parts locally and to order. This drastically reduces inventory costs, CO₂ emissions from logistics and delivery time, while mitigating disruptions in the supply chain, and (d) distributed manufacturing, where the concept of a centralized mega-factory is challenged. AM enables a network of smaller, localized "micro-factories" that can quickly respond to regional demand. A particularly important aspect of AM is reflected in tooling and prototyping, from two angles: (a) rapid tooling, where accessories and their components can be printed overnight, speeding up changes and adaptation of production lines, in the morning, at the beginning of the work shift, and (b) functional prototyping, where we have that prototypes are no longer just visual models, they can be full-power functional parts made from final production material, drastically shortening product development cycles. *Human-Robot Collaboration (HRC)* refers to the development and application of cobots [27]. Traditional industrial robots are powerful, precise and dangerous, requiring isolation in safety cages. HRC, primarily through collaborative robots (cobots), breaks this paradigm by creating a shared workspace where humans and robots work together, using their strengths. The innovative shift is reflected in the development of the paradigm - from automation to augmentation. The goal is not to replace the worker, but to increase his opportunities. The system uses human cognition, dexterity and problem-solving skills, along with robotic strength, precision and endurance. Cobots have made a profound impact on factory operations, through three aspects: (a) ergonomics and safety, where cobots are designed with force-limiting sensors and have rounded edges to stop accidental contact, making them inherently safer. Also, they take on "3D" tasks - boring, dirty and dangerous - such as heavy lifting, repetitive work overhead or handling hazardous materials, reducing worker fatigue and injury, (b) improved flexibility and skill enhancement, where we have that unlike fixed automation, cobots are mobile and easily reprogrammed. One cobot can perform multiple tasks during the day – screwing in the morning, packaging in the afternoon and quality control in the evening. Also, they act as a force multiplier for skilled workers, where an experienced technician can guide the cobot through a complex assembly process by demonstration (programming through guidance) and the cobot can then replicate it with perfect consistency, enhancing the technician's expertise, (c) hybrid work cell, which is the physical manifestation of HRC. In a hybrid

workcell, a human can perform complex, variable tasks (final assembly, inspection), while a cobot performs precise, repetitive subtasks (applying glue, holding parts, presenting components). This creates a symbiotic relationship that is more productive and adaptive than if a human or robot were working alone. *The synergistic power of combined innovations* is reflected in the fact that real transformative power is realized when these innovative pillars are connected to each other and to the wider ecosystem of smart production. Here are some examples: (a) AM + HRC, where a cobot is used to post-process a 3D printed part (polishing, cutting) or to load and unload multiple AM assemblies, creating a flexible, seamless cell for the digital production of parts, (b) AM + intelligence, where we have AI algorithms able to monitor the AM process in real time (using computer vision and thermal sensors) to detect anomalies such as voids or bends and correct process parameters in gait, ensuring the quality of the first part, and (c) HRC + Intelligence, where we have that computer vision and AI enable the cobot to perform complex picking from containers (identifying and grabbing disordered parts from containers) and adaptive assembly, where it can adjust its actions based on small variations in the parts that a human presents. *The application of these innovative technologies is not without significant obstacles* : (a) the economy and the required knowledge for application are reflected in the fact that new business models need to be developed, because companies must change their value proposition, from the sale of physical products to the potential sale of "part as a service" or digital files, (b) the workforce needs retraining, engineers need training in AM design, and operators must become programmers and supervisors of cobots, not just physical workers. Technical obstacles relate to: (a) AM limitations, reflected in the fact that speed, material properties, surface finish and multi-material printing are still areas of active development of AM technology in application, (b) HRC complexity, as truly seamless collaboration requires advanced perception and AI, for the robot to understand human intent and operate safely in dynamic, unstructured environments. Finally, there are cultural and organizational changes, as trust must be built between workers and collaborative robots. Management must foster a culture of experimentation and accept that the return on investment for innovation can be in strategic agility and new markets, not just in outright replacement of the workforce. Finally, in conclusion we can say that the pillar of innovation moves smart production from the theory of optimization to the practice of transformation. AM defines new rules of design and supply chains, while human-robot collaboration redefines the factory floor as a dynamic, hybrid space of human and machine intelligence in an advanced form of collaboration. Together, they enable a future of manufacturing that is not only more efficient but also more agile, sustainable and human-centered. These are key technologies that enable companies to respond to hyper-customization demand, build resilient supply chains and create higher-value jobs, ultimately providing a competitive advantage in an evolving global environment.

Factories that master this concept will not only be smarter, they will be fundamentally different.

4. Interoperable Smart Production Models

Interoperable models provide horizontal and vertical data integration in smart manufacturing. *RAMI 4.0* is a three-dimensional model developed in Germany for the standardization and structuring of the application of Industry 4.0 [28]. It provides a reference architecture for the integration of technologies, processes and systems in smart manufacturing. The model is designed to ensure interoperability, scalability and uniform application across industries. It has three axes: hierarchy levels (vertical axis), life cycle and value stream (horizontal axis), and layers (depth axis). *The NIST (National Institute of Standards and Technology)* smart manufacturing ecosystem is a framework developed in the United States of America for the advancement of smart manufacturing through interoperable, secure and scalable systems [29]. Emphasizes standards, reference architectures, and test platforms to enable data-driven manufacturing. It has the following components: product life cycle, cross-cutting issues, supporting technologies and reference architecture. NIST develops standards (NIST SP 800-82 for Cybersecurity) and reference models to ensure interoperability. The focus is on open standards to avoid vendor lock-in. *Cyber-physical production systems (CPPS)* are the backbone of Industry 4.0, which integrate physical production processes with digital systems through cyber-physical systems (CPS) [30]. CPPS refers to networked, smart systems that connect sensors, actuators, control systems and computing systems to enable real-time adaptive manufacturing. CPPS is not a single framework, but a conceptual model with the following components: physical layer, cyber layer, communication layer, human-machine interaction, and control and optimization. A comparison of the basic characteristics of these three models is given in Table 1 [28-30].

The relevance and trends of these models are reflected in the following facts: (a) framework convergence, which is reflected in the fact that there is a growing trend of harmonizing the concepts of RAMI 4.0, NIST and CPPS to create globally accepted standards. For example, efforts by organizations like the Industrial Internet Consortium (IIC) aim to harmonize these models for interoperability. Also, web searches and recent discussions on platforms like X highlight increased collaboration between US and European initiatives to bridge the RAMI 4.0 framework and NIST, (b) DT and AI integration, reflected in CPPS and RAMI 4.0 strongly emphasizing DT, now powered by generative AI and real-time analytics. This enables predictive maintenance (PrM) and process optimization on a large scale. NIST's focus on data analytics aligns with this trend, with test platforms exploring AI-driven manufacturing, (c) the imperative of cyber security,

stemming from the fact that cyber attacks on industrial systems (ransomware targeting IoT devices) are increasing, NIST's focus on cyber security is gaining in importance. RAMI 4.0 and CPPS include stronger security protocols to address vulnerabilities, (d) sustainability and circular economy (CE), where we have all three frameworks adapting to prioritize sustainability.

Table 1. Overview of interoperable models for smart manufacturing

Aspect	RAM 4.0	NIST Smart Manufacturing	CPPS
Origin	Germany (Industry 4.0 Platform)	NIST (USA)	Global, conceptual model
Scope	Comprehensive, life cycle focused	Focused on technology and standards	Integration of technology and processes
Structure	3D model (Hierarchy, Lifecycle, Layers)	An ecosystem with cross-cutting aspects	Layers (Physical, Cybernetic, Communication)
Basic concept	Administrative shell, Interoperability	Data Driven Manufacturing, Cyber Security	Digital twins, Decentralized intelligence
Cyber security	Moderate focus	Strong focus	Limited, depends on application
Suitable for SMEs	Complex, less suitable for SMEs	Moderately, with custom tools	Complex, requires a lot of resources
Standardization	Strong (IEC, ISO compliant)	Strong (open standards)	Weak, lack of universal standards
Applications	Smart factories, supply chains	Optimization, additive manufacturing	Flexible production, predictive maintenance
Maturity	Mature, widely accepted	Development, practical test environments	Conceptually, the implementation varies

The life cycle focus of RAMI 4.0 supports CE principles, while CPPS enables energy efficient manufacturing. NIST is developing standards for sustainable manufacturing metrics, and (e) 5G and edge computing, where we have CPPS making significant use of 5G and edge computing, enabling ultra-low latency communication in smart factories. RAMI 4.0 and NIST are updating their communication layers to include these technologies. *Challenges and future directions* are characterized by the following facts: (a) interoperability shortcomings are reflected in the fact that despite progress, differences in standards (OPC UA vs. MTConnect) create challenges for global adoption. Future efforts will focus on universal protocols, (b) adoption by SMEs, where we have them struggling with the complexity and cost of these frameworks. Simplified, modular implementations and government subsidies could bridge this gap, (c) workforce skills, where we know that implementing these models requires expertise in the Internet of Things (IoT), AI and cyber security. Training programs

and educational partnerships are key, and (d) Ethical and regulatory issues relate to As CPPS enables autonomous systems, ethical issues of human oversight and data privacy arise. Regulatory frameworks are being developed to address these issues. *In conclusion*, we can state the following: (a) RAMI 4.0 is ideal for organizations looking for a comprehensive, standardized approach to Industry 4.0, especially in Europe. Its product lifecycle focus and administrative shell make it a robust framework for large-scale deployment, (b) NIST's smart manufacturing ecosystem excels in practical, data-driven applications with a strong cybersecurity foundation, making it suitable for US-based industries and SMEs, and (c) CPPS offers flexibility and innovation, driving the future of autonomous, smart manufacturing, but requires significant investment and standardization. For companies adopting these models, the choice depends on their industry, scale and priorities (interoperability, cyber security or flexibility). Combining elements from all three - the structure of RAMI 4.0, the practicality of NIST, and the adaptability of CPPS - can result in a hybrid model for the next generation of smart manufacturing.

Research shortcomings and inconsistencies of the analyzed models. In the changing environment of smart manufacturing, theoretical models like RAMI 4.0, NIST's Smart Manufacturing Ecosystem, and Cyber-Physical Manufacturing Systems (CPPS) provide fundamental frameworks for integrating digital technologies into industrial processes. These models emphasize interoperability, data-driven decision making, and system resilience. However, as Industry 4.0 transitions to Industry 5.0, prioritizing human-centeredness, sustainability and ethical considerations, *significant research gaps remain*. These under-researched areas hinder the practical translation of these models into robust, equitable and future-proof systems. This analysis identifies key gaps, drawing on recent literature, including review papers, empirical studies and new papers. It focuses on *the ethical integration of AI, resilience in volatile markets, and additional underexplored domains such as human-AI collaboration, cybersecurity in decentralized systems, and adoption by SMEs*. For each, I discuss the nature of the deficiency, its implications for theoretical models, and opportunities for future research. The aim is to highlight how addressing these shortcomings could bridge theory and practice, achieving more adaptive and inclusive smart manufacturing ecosystems.

Ethical integration of AI in CPPS and RAMI 4.0 [31]. Nature of Deficiencies: While the administrative shell of RAMI 4.0 and the decentralized intelligence of CPPS enable AI-driven automation, research largely neglects ethical dimensions such as mitigating bias, transparency, and human oversight of AI decision-making. Existing studies focus on technical interoperability (OPC UA protocols), but rarely address how AI algorithms in CPPS might maintain biases in predictive

maintenance or resource allocation, leading to inequitable outcomes across different workforces. For example, the NIST ecosystem emphasizes cybersecurity, but provides limited guidance on the ethical governance of AI, with only initial exploration of truth metrics in industrial applications. Recent reviews highlight a "significant lack" of studies on the ethical implications of AI in smart factory education and training, extending to operational ethics such as prioritizing safety in human-robot interactions. *Implications for theoretical models:* In the layered structure of RAMI 4.0, the "business" layer could include ethical property shells, but without research, this risks non-compliance with ever-changing regulations (EU AI Law). Self-organizing CPPS systems amplify the risks, as autonomous decisions can be unaccountable, which undermines trust in volatile multi-stakeholder environments. *Opportunities for future research:* Development of hybrid ethics frameworks that integrate explainable AI (XAI) into CPPS, tested through simulations in NIST test environments. Longitudinal studies could assess bias reduction in real applications, addressing the gap between practitioner concerns (fairness in SMEs) and academic focus on efficiency. New discussions about XAI emphasize the need for ethical revisions in recursive production loops, where AI projects AI, to prevent unintended social harm.

Resilience in volatile markets and supply chains [32]. *Nature of Gaps:* Theoretical models like NIST's Ecosystem and CPPS emphasize real-time adaptability, while empirical research on resilience amid economic policy uncertainty (EPU), geopolitical disruptions, or pandemics remains fragmented. Post-COVID-19 studies call for predictions about Industry 4.0 technologies in supply chain resilience management (SCRM), but comprehensive models linking the RAMI 4.0 life cycle axis to dynamic market volatility are lacking. For example, while green AI and circular economy (CE) ESG show promise within the EPU, the interconnections with traditional energy markets are underexplored, and data from 2017–2023 reveals AI as a net transferor of spillovers, but not a proven stabilizer. This highlights "production resistance" due to bottlenecks in upstream materials, which are not addressed in current CPPS architectures. *Implications for theoretical models:* RAMI 4.0 hierarchy levels support end-to-end visibility, but without resilience metrics for volatile inputs (raw materials), models fail to mitigate "systemic bottlenecks" in global chains. NIST's cross-sector data management is resilient in theory, but lacks blockchain integration for traceable, adaptive logistics in uncertain markets. *Opportunities for future research:* Empirical studies based on the Delphi method could prioritize technologies (IoT-blockchain hybrids) for post-disruption SCRM, expanding the CPPS to include "survival" angles like Ivanov's sustainability models. Explore the role of AI in a predictive supply model, as in Toyota's IoT-blockchain tracking, to quantify resilience gains (34% delay reduction). This could develop NIST into a "first order of resilience" ecosystem for volatile sectors like semiconductors.

Cooperation between man and machine and training of the workforce in the transition to Industry 5.0 [33]. Nature of Gaps: As Industry 5.0 emphasizes human-centricity versus Industry 4.0's focus on automation, there are gaps in how RAMI 4.0 and CPPS accommodate collaborative robotics, emotional AI, and human-digital twins (HDT). The literature calls for in-depth research into the policies and ethical principles of human-AI symbiosis, but the longitudinal impacts on workforce sustainability and resilience are rare. NIST's technologies neglect training for SMEs, where 77% of factory data remains unused due to lack of skills. Discussions of "dark" factories reveal underexplored tensions between autonomy and human oversight in recursive systems. *Implications for theoretical models:* CPPS's cyber layer assumes seamless integration, but without the human factor there is a risk of workers losing qualifications, which reinforces inequalities in volatile labor markets. The RAMI 4.0 functional layer could incorporate empathy models, but current implementations prioritize efficiency over inclusivity. *Future Research Opportunities:* Live lab pilot projects for HDTs in NIST test environments to assess ethical human-machine interactions, with a focus on the role of quantum computing in bioengineering-enhanced collaboration. Case studies on retraining (via AR/VR in CPPS) could address gaps in the routinization stages of adoption.

Cyber security and standardization in decentralized CPPS [34]. Nature of the Defects: While NIST excels in cybersecurity standards, research into quantum intelligence-proof protocols for CPPS under RAMI 4.0 is in its infancy, especially amid growing threats in interconnected ecosystems. Gaps include interoperability between RAMI 4.0 and NIST in multi-agent CPPS architectures, with limited empirical validation of standards like IEC 62264 in volatile, decentralized settings. Only RAMI 4.0 and IIRA had CPPS use in development, leaving the others underexplored. *Implications for Theoretical Models:* Decentralized intelligence in CPPS increases vulnerabilities, undermining NIST's secure data exchange in EPU-prone global chains. *Future research opportunities:* Mapping multi-agent patterns to RAMI 4.0 for secure, X-only CPPS (self-healing networks), including post-quantum cryptography as in NVIDIA's cuPQC. Comparative studies of architectures could standardize resilience.

Application in SMEs and sustainability in green supply chains [33]. Nature of the deficiency: SMEs, key to global manufacturing, face under-researched barriers to adopting RAMI 4.0 or the NIST model, including financial constraints and the integration of ethical sustainability. Reviews point to a lack of links between Industry 4.0 technologies and Green Supply Chain Management (GSCM), with ethical business in SMEs requiring more case studies. China's 31,000 smart factories highlight efficiency gains, but the global gap between SMEs in reducing

carbon emissions remains. *Implications for theoretical models:* Models assume large-scale resources, marginalizing SMEs and limiting sustainable scalability. *Opportunities for future research:* Development of modular, low-cost variants of CPPS for SMEs, assessing the impacts of GSCM through bibliometric analyses. Explore the alchemy of data for the \$900 billion opportunity in untapped factory data.

A comparative overview of the differences between the models is given in Table 2.

Table 2. Model analysis

Area of deficiencies	RAM 4.0	NIST ecosystem	CPPS	A key research opportunity
Ethical AI	Limited in the business tier	Emerging Equity Metrics	Bias in autonomous decisions	XAI integration pilot projects
Market resilience	Life cycle of EPU insufficiently researched	Strong data focus, low volatility	Gaps in self-organization after COVID	Blockchain and IoT hybrids
Human cooperation	Deficit of empathy in the functional layer	Technology enablement ignores skill enhancement	Immature H-CPPS models	Living laboratories for HDTs
Cyber security	Interoperability standards are evolving	Advanced, but without quantum intelligence	Decentralized vulnerabilities	Multi-Agent PQC Mapping
SMEs/Sustainability	Complex for SMEs	Tools for emerging SMEs	A resource-intensive adoption	Case studies of modular GSCM

Broader implications. These shortcomings reveal a technocentric bias in current models, with Industry 5.0's ambitious shift towards ethical, resilient paradigms revealing the need for interdisciplinary research. Implications include stalling adoption (93% of professionals prioritize resilience and face data silos) and risks like ethical lapses in dark factories. *Future directions:* Harmonize models through global consortia, fund grants targeted at SMEs, and use AI for gap-filling simulations. By prioritizing these, research can develop theoretical models in the development of smart, sustainable manufacturing due to uncertainty by 2028.

5. Conclusions and Future Research

The factory of tomorrow is not a place - it is a living system, as integration has dissolved silos, weaving every machine, material and mind into a single, real-time digital thread. Intelligence awakened within that thread, turning data into prediction, and prediction into flawless execution. Innovation now emerges not from isolated breakthroughs, but from the continuous pulse of this connected, cognitive core—giving birth to products that evolve, processes that heal themselves, and business models that scale with human needs. This convergence is irreversible. The lines between design and production, man and machine, factory and customer have disappeared. What remains is a resilient, adaptive, value-creating organism—sustainable by design, personalized by default, and profitable by outcome. The future does not come, it starts in the production plant. Integration built the foundation, intelligence gave it vision, and innovation set it free. With this, smart production has been achieved, and it is not the end of industry but the beginning of industry as intelligence.

For the academic community, the contributions of this theme are: New interdisciplinary fields in manufacturing systems science - combining control theory, AI, materials science and human factors. Open datasets and digital twin benchmarks, as standardized, anonymized lakes of factory data for training next-generation AI models. Real-world simulation frameworks, through validated DT methodologies that bridge physics-based and data-driven modeling. Implications relate to: (a) curriculum revision, where isolated engineering degrees will be replaced by Integrated Manufacturing Engineering (IME) programs, (b) commitment to lifelong learning, as Faculty and students will have to master real-time systems, edge AI and ethical data management, (c) transition to research funding, from component-level grants to convergence grants focused on system-level resilience and sustainability.

For the industrial community, the contributions of this topic are: (a) platform ecosystem, where we have neutral integration layers (industrial data spaces) that allow SMEs to engage in smart value chains, (b) AI as a useful service, where we have pre-trained, sector models (predictive quality for plastic injection molding) implemented via MaaS, (c) standards for CE, where we have industry-driven protocols for modular design, digital product passports and closed-loop recycling. On the other hand, we have implications, related to: (a) the transition from capital to operational costs, which means we have a transition from owning assets to subscription to results ("working time as a service"), (b) retraining of the workforce on a large scale: 70% of operators become process co-pilots - trained in augmented reality (AR) and adaptive decision-making, and (c) a redefinition of competition, because the market leaders will not be the biggest, they will be the most integrated, the most intelligent and the most innovative.

For policy makers and decision makers, the main contributions to the development and application of smart manufacturing should be: (a) global interoperability standards, as harmonized data schemas, cyber security core values and AI ethics for global manufacturing, (b) innovation hubs, as regulatory safe zones for testing autonomous manufacturing and AI-driven supply chains, and (c) sustainability metrics, as KPI parameters of ESG/CE models. The implications of this concept can be defined as: (a) Industrial policy 2.0, means the transition from tax breaks to incentives for integration - financing 5G factory networks, open APIs and shared DT, (d) immigration based on skills, where we have accelerated visas for convergence engineers (AI + manufacturing experts), and (e) geopolitical resilience, through a model of mandatory dual supply through DT, to reduce the risk in supply chains without protectionism.

References

- [1] Rosenberg, Nathan. (2010). The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present. By David S. Landes. Cambridge, England: Cambridge University Press, 1969. Pp. 566. The Journal of Economic History. 31. 497-500. 10.1017/S0022050700091269.
- [2] Carstensen, Fred. (2012). From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States. By David A. Hounshell. (Baltimore; Johns Hopkins University Press, 1984. xxi + 330 pp. Business History Review. 59. 299-301. 10.2307/3114941.
- [3] Groover, Mikell. (2016). Fundamentals in Modern Manufacturing : Material, Processes, and Systems / MP Groover. <https://tusharaneeyrao.wordpress.com/wp-content/uploads/2019/12/fundamentals-of-modern-manufacturing-materials-processes-and-systems.pdf> .
- [4] Kagermann, H., Wahlster, W. and Helbig, J. (2013) Securing the Future of the German Manufacturing Industry: Recommendations for Implementing the Strategic Initiative Industrie 4.0. Final Report of the Industrie 4.0 Working Group, Acatech—National Academy of Science and Engineering, 678 p. https://www.acatech.de/wp-content/uploads/2018/03/Final_report_Industrie_4.0_accessible.pdf .
- [5] AI Wavering, NIST and Smart Manufacturing, NIST, 2020. <https://www.nist.gov/system/files/documents/el/isd/Overview-of-NIST-and-Smart-Mfg-24-Aug-2020.pdf> .
- [6] European Commission (2021), Industry 5.0: Towards a sustainable, human-centric and resilient European industry, Brussels, https://ec.europa.eu/info/publications/industry-50_en .
- [7] Manufacturing industry outlook, Deloitte, 2024, <https://www.deloitte.com/us/en/insights/industry/manufacturing-industry-outlook-2024.html>.

- [8] Identity Security Threat Landscape Report, Vanson Bourne, 2023, <https://www.vansonbourne.com/case-studies/2023-identity-security-threat-landscape-report/>.
- [9] A new talent model for the digital age, PwC, 2025. <https://www.strategyand.pwc.com/ml/en/strategic-foresight/sector-strategies/technology/new-talent-model-for-the-digital-age/strategyand-new-talent-model-for-the-digital-age.pdf>.
- [10] Manufacturing, value added (% of GDP), 2024, <https://data.worldbank.org/indicator/NV.IND.MANF.ZS>.
- [11] Ford, H., & Crowther, S. (1922). *My life and work*. Garden City Publishing Company.
- [12] Taylor, FW (1911). *The principles of scientific management*. Harper & Brothers.
- [13] Dalenogare, LS, Benitez, GB, Ayala, NF, & Frank, AG (2018). The expected contribution of Industry 4.0 technologies for industrial performance. *International Journal of Production Economics*, 204, 383–394. <https://doi.org/10.1016/j.ijpe.2018.08.019>.
- [14] Liao, Y., Deschamps, F., Loures, E. d. FR, & Ramos, LFP (2022). Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal. *International Journal of Production Research*, 55(12), 3609–3629. <https://doi.org/10.1080/00207543.2017.1308576>.
- [15] Smart manufacturing, 2025, <https://www.jeffwinterinsights.com/insights/smart-manufacturing-enablers-and-enhancers>.
- [16] IEC. (2020). *Factory of the future*. International Electrotechnical Commission. <https://www.iec.ch/basecamp/factory-future>.
- [17] Mittal, S., Khan, MA, Romero, D., Wuest, T. (2018). A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). *Journal of Manufacturing Systems*, 49, 194–214. <https://doi.org/10.1016/j.jmsy.2018.10.005>.
- [18] MESA International. (2024). *MESA model: A framework for smart manufacturing*. Manufacturing Enterprise Solutions Association. <https://www.mesa.org/white-papers>.
- [19] Tao, F., Zhang, M., & Nee, AYC (2019). *Digital twin driven smart manufacturing*. Academic Press.
- [20] Zheng, P., Sivabalan, AS (2020). A generic tri-model-based approach for product-level digital twin development in a smart manufacturing environment. *Robotics and Computer-Integrated Manufacturing*, 64, 101958. <https://doi.org/10.1016/j.rcim.2020.101958>.
- [21] Development of Maintenance models, 2025, <https://www.einfochips.com/blog/predictive-maintenance-for-the-enterprise-of-the-future/>.
- [22] Fidma Mohamed Abdelillah, Hamour Nora, Ouchani Samir, Sidi Mohamed Benslimane. Predictive Maintenance Approaches in Industry 4.0: A Systematic Literature Review. 2023 IEEE International Conference on Enabling Technologies:

- Infrastructure for Collaborative Enterprises (WETICE), Dec 2023, Paris, France. pp.1-6, <https://doi.org/10.1109/WETICE57085.2023.10477802> .
- [23] Ali, Y., Shah, SW, Arif, A., Tlija, M., & Siddiqi, MR (2024). Intelligent Framework Design for Quality Control in Industry 4.0. *Applied Sciences*, 14 (17), 7726. <https://doi.org/10.3390/app14177726> .
- [24] Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 18, 20-23. <https://doi.org/10.1016/j.mfglet.2018.09.002> .
- [25] Innovation smart manufacturing, 2025, https://www.smic.kr/SMIC_ENG_120.php.
- [26] Gibson, I., Rosen, D., Stucker, B., & Khorasani, M. (2021). Additive manufacturing technologies (3rd ed.). Springer. <https://doi.org/10.1007/978-3-030-56127-7> .
- [27] Gualtieri, L., Palomba, I., Merati, FA, Rauch, E., Vidoni, R. (2022). Design of a human-centered collaborative framework for assembly and disassembly in smart manufacturing. *Procedia CIRP*, 91, 383-388. <https://doi.org/10.1016/j.procir.2020.02.190> .
- [28] Adolphs, P., Bedenbender, H., Dirzus, D., Ehlich, M., Epple, U., Hankel, M., Heidel, R., Hoffmeister, M., Huhle, H., Kärcher, B., Koziolk, H., Pichler, R., Pollmeier, S., Schewe, F., Walter, A., Waser, B., & Wollschlaeger, M. (2017). Reference architecture model Industrie 4.0 (RAMI 4.0): An introduction. *Platform Industrie 4.0*. <https://www.plattform-i40.de/> .
- [29] Lu, Y., Morris, KC, & Frechette, S. (2018). Current standards landscape for smart manufacturing systems. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.8107> .
- [30] Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., Ueda, K. (2016). Cyber-physical systems in manufacturing. *CIRP Annals*, 65 (2), 621–641. <https://doi.org/10.1016/j.cirp.2016.06.005> .
- [31] Bibri, SE, Allam, Z., & Krogstie, J. (2024). A sociotechnical framework for evaluating Industry 5.0: An integrated human-machine approach to smart manufacturing. *Frontiers in Manufacturing Technology*, 4, 1378418. <https://doi.org/10.3389/fmtec.2024.1378418> .
- [32] Rathore, H., Mohammed, A., & Ali, M. (2023). Industry 4.0 and sustainable manufacturing: A bibliometric analysis. *Sustainable Manufacturing and Service Economics*, 2, 100013. <https://doi.org/10.1016/j.smse.2023.100013> .
- [33] Sony, M., Antony, J., & McDermott, O. (2024). The impact of Industry 4.0 on sustainable manufacturing: A systematic literature review. *Sustainable Manufacturing and Service Economics*, 3, 100026. <https://doi.org/10.1016/j.smse.2024.100026> .
- [34] Zhang, C., Chen, Y., & Liu, Z. (2023). From Industry 4.0 to Industry 5.0: Ethical and technological challenges in smart manufacturing. *International Journal of Production Research*, 61 (12), 3985–4002. <https://doi.org/10.1080/00207543.2022.2091456>.

Architecture of Industry 4.0 Smart Manufacturing Systems Based on Mobile and Collaborative Robotics

Isak Karabegović^{*1}, Ermin Husak², Mehmed Mahmić², Samir Vojić², Isić Safet³, Pjereta Agalliu⁴

Abstract: *Research and development of advanced technologies underpinning the implementation of Industry 4.0 are driving the transformation of traditional manufacturing systems into flexible production systems whose high level of integration and intelligence is based on digital technologies. Mobile and collaborative robotics, supported by artificial intelligence, are becoming central elements of intelligent manufacturing processes through their implementation, complementing other Industry 4.0 technologies. This paper provides a clear and comprehensive overview of the architecture of Industry 4.0 smart manufacturing systems based on the integration of autonomous mobile robots and collaborative robots, with particular emphasis on system hierarchy, data flows, and interoperability with existing industrial information systems. Patent trends over the past decade in robotics and artificial intelligence are analyzed, along with trends in the implementation of mobile and collaborative robots in manufacturing processes. The role of robotic systems in process monitoring, decision-support, and adaptive production control is examined, including their integration with IoT platforms, manufacturing execution systems, and AI-based analytical modules. The paper also discusses the technical and organizational challenges associated with implementing such architectures, as well as their potential benefits in terms of flexibility, scalability, and resilience of manufacturing systems. The proposed architectural approach provides a foundation for the further development of smart manufacturing environments and offers guidelines for the industrial application of mobile and collaborative robotics in accordance with Industry 4.0 principles.*

Keywords: *smart manufacturing systems, Industry 4.0, mobile robotics, collaborative robotics, system architecture, artificial intelligence, systems integration*

1. Introduction

Organizational transformation in modern manufacturing increasingly begins with the adoption of Industry 4.0 technologies, which are characterized by the integration of digital, physical, and cognitive systems into unified smart

^{*1}Academy of Sciences and Arts of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina

²University of Bihać, Bihać, Bosnia and Herzegovina

³University Džemal Bijedić, Mostar, Bosnia and Herzegovina

⁴University of Tirana, Tirana, Albania

E-mail: isak1910@hotmail.com

production environments. The concept of Industry 4.0 emerged within the German industrial initiative “High-Tech Strategy 2020,” and the term itself was first introduced in 2011 at the Hannover Messe trade fair. It was subsequently formalized through reports produced by the Plattform Industrie 4.0 working group, which defined the core objectives of this paradigm, including cyber-physical systems, the Industrial Internet of Things, smart factories, and the horizontal and vertical integration of manufacturing processes [1-3]. The development and implementation of Industry 4.0 rely on the technological convergence of numerous advanced technologies, such as industrial robotics, the Industrial Internet of Things (IIoT), cloud computing, big data analytics, cyber-physical systems (CPS), artificial intelligence (AI), and intelligent sensing platforms. These technologies enable manufacturing systems to become increasingly automated, adaptive, and autonomous. In contrast, traditional production environments—still widely present today—are often characterized by rigid structures, centralized control, and limited flexibility, making it difficult for them to respond to modern market requirements such as short product life cycles, high product variability, and the growing demand for mass customization [4-6]. Through the adoption of Industry 4.0 principles, production systems are transformed into smart and flexible environments capable of autonomous adaptation, real-time resource optimization, and the integration of humans and machines within shared workspaces. Among the enabling technologies of Industry 4.0, mobile and collaborative robotics occupy a central position in the architecture of intelligent manufacturing systems. Autonomous mobile robots (AMRs) provide flexible transportation of materials, tools, and components without the need for fixed infrastructure, while collaborative robots (cobots) facilitate safe and efficient human-robot cooperation in industrial operations. Their integration contributes significantly to increased productivity, reduced downtime, improved workplace ergonomics, and enhanced operational continuity [7-9].

A particularly important trend is that modern robotic systems increasingly rely on artificial intelligence algorithms, enabling advanced perception of the environment, learning from operational data, autonomous planning, and adaptive process control. The incorporation of AI into mobile and collaborative robotics represents a shift from deterministic pre-programmed automation toward intelligent systems capable of real-time decision-making and predictive performance optimization. Combined with smart sensors, vision systems, and analytical platforms, AI-driven robots actively support production process monitoring, decision-making assistance, and the implementation of predictive maintenance strategies [10-12]. In this way, AMRs and cobots become key carriers of manufacturing digitalization and fundamental components of smart factory ecosystems. The aim of this paper is to provide a systematic overview of the architecture of Industry 4.0 smart manufacturing systems based on the integration of autonomous mobile and collaborative robots, with particular

emphasis on system hierarchy, data flows, and interoperability with existing industrial information infrastructures. Special attention is devoted to patent trends in robotics and artificial intelligence over the past decade, as well as to the evolution of industrial implementation practices worldwide. Furthermore, the paper examines the role of mobile and collaborative robotics in monitoring, adaptive control, and predictive maintenance, while addressing the technical and organizational challenges associated with deploying such architectures in real manufacturing environments.

In accordance with these objectives, the study addresses the following research questions:

- How are global patent trends in robotics and artificial intelligence relevant to smart manufacturing evolving?
- What have been the dominant trends in the industrial implementation of mobile and collaborative robots over the past ten years?
- In what ways does the integration of these robotic systems contribute to monitoring, control, and predictive maintenance of manufacturing processes?
- What are the main barriers, opportunities, and long-term implications of such systems within the Industry 4.0 paradigm?

The remainder of this paper is organized as follows. Chapter 2 presents the theoretical framework of Industry 4.0 smart manufacturing systems. Chapter 3 discusses architectures based on mobile and collaborative robotics. Chapter 4 analyzes global patent activity in robotics and AI, while Chapter 5 reviews industrial implementation trends. Chapter 6 focuses on applications in monitoring, adaptive control, and predictive maintenance. Chapter 7 addresses technical and organizational challenges, followed by the final discussion and conclusions.

2. Theoretical Framework of Smart Manufacturing Systems of Industry 4.0

It is well known that the fourth industrial revolution, namely Industry 4.0, represents a modern approach to industrial development based on the application of advanced digital technologies, which aim to transform production processes in order to achieve high flexibility, efficiency, and autonomy. This concept implies the integration of physical production resources with digital technologies, enabling the transition from traditional, hierarchically rigid systems to smart, adaptive, and self-learning manufacturing structures [13-15]. The foundation of smart manufacturing consists of key digital technologies such as the Internet of Things (IoT), cyber-physical systems (CPS), artificial intelligence (AI), and big data analytics, together with other Industry 4.0 technologies. IoT enables the networking of machines, sensors, and devices in real time, while CPS integrate computational and physical components for continuous monitoring and process optimization. AI and big data analytics further enable predictive decision-making,

adaptive planning, and autonomous optimization of production flows, which is essential for the realization of the smart factory concept [16-18]. Intelligent manufacturing systems are most often described through a multilayer hierarchy that includes sensor and actuator levels, control systems, manufacturing execution systems (MES), and enterprise resource planning (ERP) systems. Unlike classical hierarchies, Industry 4.0 introduces horizontal and vertical integration, enabling dynamic data exchange between different system levels and functions. Interconnection is one of the main challenges in intelligent manufacturing systems because it enables communication and data exchange between different systems. In order to achieve efficient integration, standardized communication protocols and common data models are required. Without interoperability, the application of digital technologies is limited, especially in complex manufacturing environments with many different devices and software solutions [19]. Industrial and mobile robotics has a central role in the architecture of the Industry 4.0 smart factory. Industrial, collaborative, and mobile robots act as execution elements of smart systems, directly connected to CPS and AI modules. Their ability to operate autonomously, collaborate with humans, and adapt to changes in the manufacturing environment makes them key enablers of flexible and intelligent production, as well as a bridge between the digital and physical worlds of smart manufacturing systems.

3. Architecture of Smart Manufacturing Systems Based on Mobile and Collaborative Robotics

The architecture of smart manufacturing systems based on mobile and collaborative robotics represents a key element in the implementation of Industry 4.0 in modern factories. Such an architecture is designed as a modular, distributed, and scalable structure that enables flexible integration of various hardware and software components, with support for real-time operation, autonomous decision-making, and adaptive control of manufacturing processes. Fundamentally, it is based on the close integration of physical production resources and digital control layers, thereby creating the foundation for smart and responsive manufacturing. The conceptual system architecture is most often organized through a multilayer model that includes the physical layer, communication layer, control layer, and application layer. The physical layer includes autonomous mobile robots, collaborative robots, sensors, and production equipment, while the communication layer enables data exchange through industrial networks and wireless technologies. The control layer integrates algorithms for coordination, planning, and supervision, while the application layer includes analytical and visualization tools for higher-level decision-making [20]. The integration of autonomous mobile robots (AMRs) and collaborative robots (cobots) enables the realization of highly flexible manufacturing and logistics processes. AMR

systems take over tasks such as internal transport, workstation supply, and dynamic material rerouting, while collaborative robots directly cooperate with humans in assembly, processing, and inspection. Their mutual coordination is based on distributed control mechanisms and shared databases, enabling synchronized operation and adaptation to changes in the manufacturing environment [21]. Data flows and communication structures form the backbone of the architecture of smart manufacturing systems. Data are generated at the level of sensors and robots, processed at the network edge (edge computing) or in cloud environments, and used for real-time process optimization. Standardized communication protocols enable reliable and secure information exchange between AMRs, cobots, control systems, and information platforms, thereby achieving high transparency and traceability of manufacturing processes. Integration with manufacturing execution systems (MES), enterprise resource planning (ERP) systems, and IoT platforms enables vertical connectivity between the operational and business levels. In this way, a continuous flow of information from the shop floor to strategic management is achieved, allowing more efficient planning, performance monitoring, and data-driven decision-making. In this context, AMRs and collaborative robots act as active nodes of the IoT ecosystem, continuously exchanging data with higher-level systems [22]. Security, organizational, and interoperability aspects of the architecture represent particular challenges in smart manufacturing systems. Security involves the protection of people, equipment, and data, especially in environments where humans and robots share the same workspace. Organizational aspects relate to the adaptation of work processes, employee training, and the redefinition of roles in the digitalized factory. Interoperability, as a prerequisite for efficient integration, requires the alignment of technical standards and software solutions to ensure the smooth operation of complex and heterogeneous systems [23,24].

4. Trends in Robotics and Artificial Intelligence Patents

The analysis of patent trends in the fields of robotics and artificial intelligence represents a reliable indicator of technological progress, innovation dynamics, and strategic directions of industrial development within the framework of Industry 4.0. Patents reflect not only the intensity of research and development activities, but also the ability of industries and countries to commercialize knowledge and secure long-term competitive advantage. This chapter examines the methodological aspects of patent analysis, global trends in robotics and artificial intelligence, their geographical and sectoral distribution, as well as the implications of these trends for the development of industrial robotics and smart manufacturing systems.

4.1. Patent Analysis Methodology

The methodology of patent analysis is based on the use of internationally recognized patent and scientific databases that are widely applied in technology and innovation research. In this study, secondary data and search results from databases such as Lens.org and WIPO PATENTSCOPE were used, relying on methodological frameworks from previously published studies indexed in the Scopus and Web of Science databases. The analysis includes a combination of searches based on the Cooperative Patent Classification (CPC) and relevant keywords, ensuring coverage of both hardware and software innovations in robotics and artificial intelligence. As the main search criteria, CPC classes related to industrial robots, manipulators, control systems, machine learning, and intelligent algorithms were used, with a time frame covering the last ten years. The number of patent applications and patent families per year was used as an indicator of innovation activity. This approach enables the identification of long-term trends and reduces the impact of methodological differences between individual databases [25-27].

4.2. Global Trends in Robotics Patents over the Last Decade

Global trends in robotics patents over the last decade indicate a continuous and accelerated growth of innovation activity. Particularly after 2016, a significant increase in the number of patents has been observed, which is associated with the intensive implementation of Industry 4.0 concepts, the development of collaborative robots, and the expansion of autonomous mobile robotic systems.



Note: The presented data represent estimates of the number of patent families, derived from observed growth patterns and publicly available patent analyses (Lens, PatStat, WIPO, CSET, and industry sources). Variations may arise due to differences in the applied keywords, CPC classifications, and filtering criteria.

Figure 1. Global trends in robotics patents worldwide for the period 2014–2024 [28-31]

Robotics is increasingly positioned as a key technology for flexible and reconfigurable manufacturing, moving beyond the boundaries of traditional automation. Over the past ten years, global patent activity in the field of robotics has exhibited a strong and continuous growth, reflecting an increased level of innovation and strategic investment in automation technologies. Between 2005 and 2019, more than 72,000 robotics-related patents were granted worldwide, with a significant acceleration in the number of applications as countries and companies intensified research and development activities in robotics and autonomous systems.

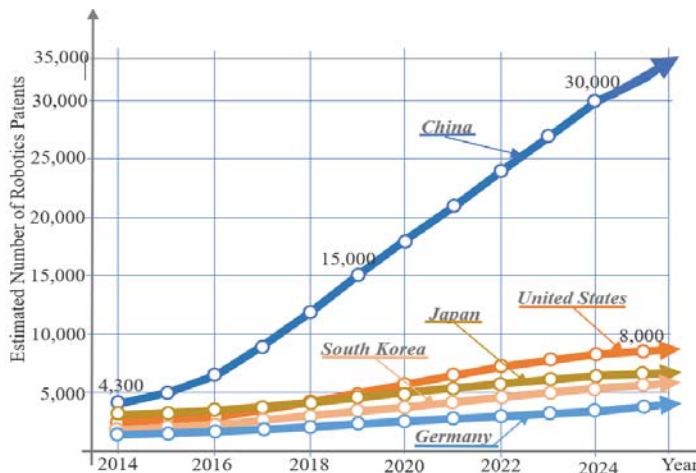


Figure 2. Global trends in robotics patents in the five leading countries: China, the United States, Japan, South Korea, and Germany over the period (2014–2025)

Figure 2 presents the estimated trends in robotics-related patent activity for the five leading countries China, the United States, Japan, South Korea, and Germany over the period 2014–2025 [28–31]. The results indicate a pronounced dominance of China, characterized by accelerated growth, particularly after 2017, followed by steady growth in the United States and Japan, largely driven by industrial and AI-enabled robotics. South Korea demonstrates strong technological continuity and sustained innovation capacity. Germany, ranked fifth globally in terms of the number of robotics patents, exhibits moderate but stable growth, consistent with its strong industrial automation base and the principles of Industry 4.0. Patent filings indicate broader technological advancement in robotics, with high activity in industrial automation, service robots, and systems integrating artificial intelligence. The increasing number of robotics patents further confirms the strategic importance of intellectual property in securing technological leadership and commercial advantage in global markets. Moreover, recent analyses suggest

a continued increase in patent activity, particularly in areas that combine robotics with artificial intelligence and Internet of Things (IoT) technologies. An analysis of patent structures reveals a growing share of solutions related to safety, adaptive control, and human–robot interaction. These trends indicate a shift toward production systems in which robots operate in shared workspaces with humans and other intelligent systems. Robotics patents thus directly reflect industry demands for greater flexibility, cost reduction, and increased productivity [32-33].

4.3. Trends in Artificial Intelligence Patents Relevant to Manufacturing

Compared to robotics, the field of artificial intelligence exhibits an even more intensive growth in patent filings. AI technologies, including machine learning, deep neural networks, and data analytics algorithms, have become key drivers of smart manufacturing. Patents in this domain increasingly address predictive maintenance, optimization of production processes, and intelligent quality control. It is particularly notable that a large share of AI-related patents is directly linked to manufacturing systems, robotics, and industrial IoT platforms. This trend indicates a strong convergence of AI and industrial automation, where algorithms learn from data collected on the shop floor and enable autonomous decisionmaking. The sharp increase in AI patent activity after 2017 confirms that artificial intelligence represents a foundational technology for the realization of smart manufacturing systems [34-35]. Figure 3 illustrates the global trend in artificial intelligence patents relevant to manufacturing processes, that is, global manufacturing, for the period 2014–2024.

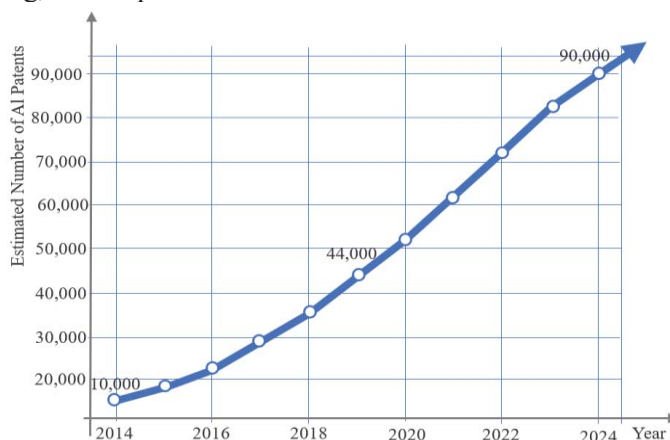


Figure 3. Global trend of AI patents relevant to manufacturing (2014–2024) [28-31]

Over the past decade, a strong and continuous growth in patent activity in the field of artificial intelligence (AI) has been observed, particularly in segments relevant to smart manufacturing and Industry 4.0, as illustrated in Figure 3. The global trend indicates that the number of AI-related patents increased almost multiple times over the period 2014–2024, reflecting the growing role of machine learning algorithms, deep neural networks, computer vision, and autonomous systems in the optimization of manufacturing processes. This growth has been especially pronounced after 2017, when AI technologies became a key factor of competitiveness in industrial automation, predictive maintenance, and intelligent production management.

An analysis of trends in artificial intelligence patents relevant to manufacturing processes, that is, global manufacturing, was conducted for the period 2014–2024 across the five leading countries worldwide: China, the United States, Japan, South Korea, and Germany, and is presented in Figure 4 [28-31].

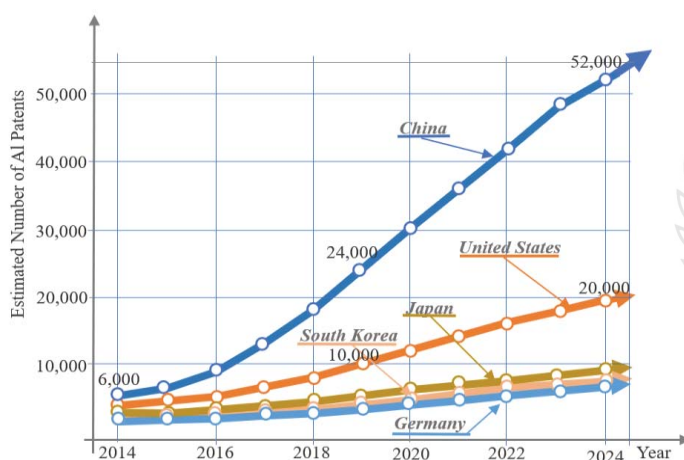


Figure 4. AI patent trends in the top five countries relevant to manufacturing: China, the United States, Japan, South Korea, and Germany (2014–2024)

An analysis of the five leading countries by the number of AI patents reveals a clear dominance of China, which exhibits exponential growth and significantly outperforms the other countries throughout the observed period. This trend is the result of strong state-driven strategies, intensive investments, and the close integration of AI technologies with industrial production. The United States and Japan rank second and third, respectively, showing stable and gradual growth, particularly in the areas of industrial analytics, autonomous systems, and the integration of AI into robotic applications. South Korea and Germany demonstrate moderate but consistent growth, in line with their focus on high-quality industrial automation and smart factories.

Overall, AI patent trends confirm that artificial intelligence represents one of the key technologies of future manufacturing, enabling higher levels of autonomy, flexibility, and efficiency in smart manufacturing systems.

The geographical analysis of patent activity indicates a pronounced concentration of patents in a limited number of technologically leading regions. Asian countries, primarily China, Japan, and South Korea, exhibit a high intensity of patenting in robotics and artificial intelligence, driven by strong government policies and long-term investments in automation. China stands out in particular due to the rapid growth in the number of patents, while Japan and South Korea maintain stable and technologically sophisticated development. The United States plays a dominant role in the development of software- and AI-oriented patents, whereas European countries, especially Germany, retain a strong focus on industrial application, quality, and standardization. From a sectoral perspective, the largest share of patents relates to industrial manufacturing, the automotive industry, and logistics, with a noticeable increase in patents in service robotics and smart energy systems [36-37].

4.4. Impact of Patent Trends on Industrial Robotics and Smart Manufacturing Systems

Patent trends in robotics and artificial intelligence have a strong impact on the development of industrial robotics and smart manufacturing systems. Increased patent activity indicates accelerated innovation and intensified global competition, resulting in a faster transfer of new technologies into industrial practice. The integration of AI into robotic systems enables the development of autonomous and self-learning systems capable of adapting to and optimizing production processes in real time. These trends contribute to the development of more flexible, energy-efficient, and reliable manufacturing systems, while simultaneously imposing the need for standardization and interoperability. Patent analysis thus confirms that robotics and artificial intelligence are key technologies of future industry and form the foundation of the architecture of Industry 4.0 smart manufacturing systems [38-39].

5. Trends in the Implementation of Mobile and Collaborative Robots over the Last Decade

Over the past decade, industrial manufacturing has undergone a profound transformation marked by a shift from rigid, hierarchically organized systems toward flexible, adaptive, and highly digitalized production environments. In this context, mobile robots and collaborative robots (cobots) have become key technological pillars of modern smart manufacturing systems. Their implementation enables efficient integration of physical processes, information

systems, and human labor, significantly increasing the agility and resilience of production systems to market and operational changes. Unlike traditional industrial robots, which are designed to perform predefined tasks in strictly controlled environments, mobile and collaborative robots enable dynamic reorganization of production, rapid reconfiguration of workstations, and safe interaction with human operators. Over the past decade, industrial manufacturing has undergone a profound transformation marked by a shift toward flexible, adaptive, and highly digitalized production environments. Mobile robots and collaborative robots (cobots) have become key technological pillars of smart manufacturing systems, significantly enhancing agility and resilience in line with Industry 4.0 principles [40].

5.1. Evolution of Mobile Robotics in Industry

Industrial mobile robotics has undergone significant evolution, transitioning from conventional automated guided vehicles (AGVs) to autonomous mobile robots (AMRs) that employ advanced sensor systems, real-time computing, and artificial intelligence. While AGV systems relied on fixed infrastructure such as magnetic tapes or reflective markers, modern AMR systems use a combination of lidar sensors, cameras, and simultaneous localization and mapping (SLAM) algorithms to navigate independently in complex and dynamic industrial environments. In contemporary manufacturing systems, mobile robots are primarily used for the autonomous execution of transport and process-support tasks, with their functionality depending on the level of environmental perception, navigation algorithms, and integration with higher-level information systems. Key enabling technologies include localization and mapping techniques, sensor data fusion, and real-time communication, which ensure reliable operation under dynamic industrial conditions. This technological approach makes mobile robots suitable for flexible production concepts and rapid process reconfiguration, regardless of the specific industrial sector, as illustrated in Figure 5.



Figure 5. Autonomous mobile robots perform various tasks in manufacturing processes [43-45]

Numerous studies have shown that the implementation of AMR systems contributes to reduced production downtime, optimized material flows, and increased utilization of manufacturing capacities. In addition, mobile robots are increasingly integrated with manufacturing information systems such as MES and ERP, thereby becoming an integral part of digital twins of production systems. Advances in mobile robotics have also enabled the development of hybrid systems in which mobile robots are combined with robotic manipulators, expanding their functional scope to machine tending, assembly, and inspection. Such systems are particularly suitable for flexible production lines and small-batch manufacturing with frequent product changes [42,46].

5.2. Development and Expansion of Collaborative Robots in Manufacturing Environments

Collaborative robots represent one of the fastest-growing categories of industrial robotics. Their development is based on the integration of force and torque sensors, advanced collision detection algorithms, and functional safety standards, enabling safe operation in shared workspaces with humans. This characteristic makes cobots particularly suitable for applications requiring close human-robot collaboration. In manufacturing environments, cobots are increasingly used for tasks that require precision, repeatability, and ergonomic support for workers. Empirical studies indicate that the use of collaborative robots reduces physical workload, increases productivity, and contributes to improved product quality. They are particularly prevalent in assembly, packaging, palletizing, CNC machine tending, and quality inspection, as shown in Figure 6 [47-51].



Figure 6. Collaborative robots performing CNC machine tending in manufacturing processes

An important factor driving the adoption of cobots is their relatively simple integration and programming. Unlike traditional industrial robots, which require specialized expertise and lengthy setup procedures, collaborative robots often employ intuitive programming methods such as hand-guiding or graphical user interfaces. This enables faster deployment and easier adaptation of production processes.

5.3. Statistical Indicators and Industrial Implementation Examples

Trends in the implementation of professional service robots, logistics service robots, and collaborative robots worldwide were analyzed based on statistical data obtained from the International Federation of Robotics (IFR), the United Nations Economic Commission for Europe (UNECE), and the Organisation for Economic Co-operation and Development (OECD). These data are presented in Figures 7, 8, 9, and 10 [52-60]. Research and development in robotics technology have contributed to the advancement of professional service robots used in manufacturing processes. Currently, service robots for transport and logistics represent the most widespread applications, supporting finished product inspection, monitoring, and tracking of production processes. The global trend in the deployment of professional service robots over the period 2014–2024 is illustrated in Figure 7.

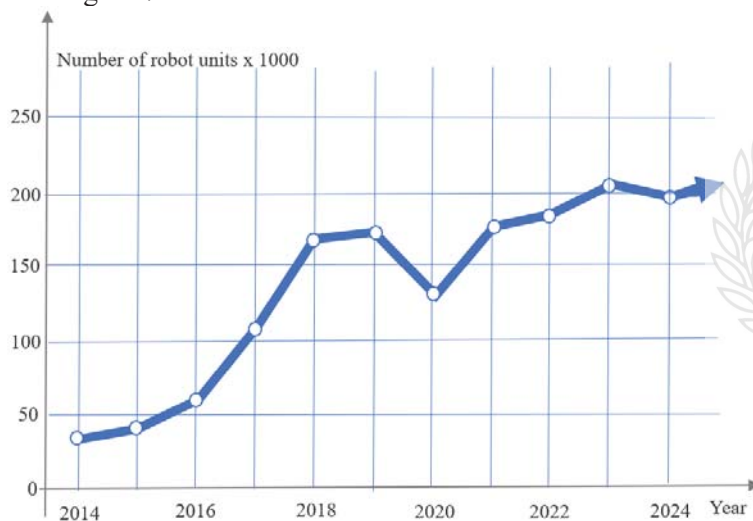


Figure 7. Global trend in the implementation of professional service robots (2014–2024).

The implementation trend of professional service robots shows long-term and structurally stable growth, with occasional short-term fluctuations caused by global economic and societal disruptions. As shown in Figure 7, the number of deployed professional service robots increased significantly from 33,000 units in 2014 to 199,000 units in 2024, representing approximately a sixfold increase and an average annual growth rate (CAGR) of about 19.7%. A notable decline occurred during the pandemic year 2020, followed by a strong recovery in 2021. The data also indicate a peak in 2023, with a slight decline in 2024, suggesting short-term market correction or saturation in certain segments. After a phase of

rapid expansion, the market is entering a more mature stage characterized by gradual stabilization and selective sectoral growth.

The distribution of professional service robot applications across six major application areas in 2024 is shown in Figure 8.

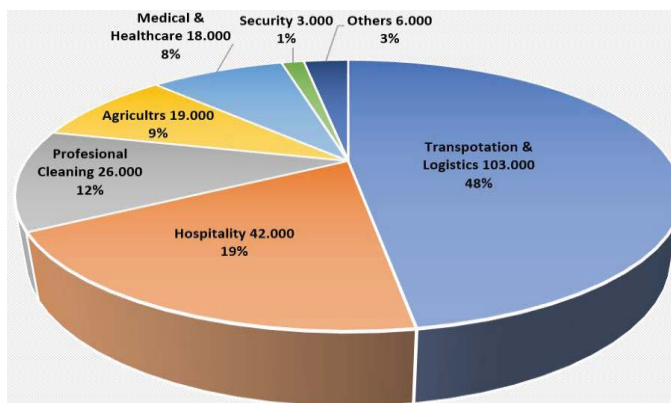


Figure 8. Implementation of professional service robots across six major application areas in 2024.

Based on the data presented in Figure 9, the distribution of service robot applications across sectors can be observed. The largest number of installations is recorded in the transport and logistics sector, with 103,000 units, accounting for 48% of total usage [21]. This is followed by hospitality, with 42,000 units used for food preparation, serving, and delivery in hotels and restaurants. In agriculture, 19,000 units are deployed for harvesting, irrigation, mowing, milking, and crop maintenance. Medical robots account for 8% of total usage, with 18,000 units supporting surgical assistance, rehabilitation, and patient care. Professional cleaning includes 26,000 units, while security, search, and rescue applications represent the smallest share, with 3,000 units, mainly due to high implementation costs and specific operational requirements.

The global trend in the deployment of service robots in logistics and transport over the period 2014–2024 is presented in Figure 9.

The data shown in Figure 9 indicate continuous growth in the number of deployed professional service robots in logistics and transport. After moderate growth until 2018, a period of relative stability followed, with a pronounced increase in installations observed from 2021 onward.

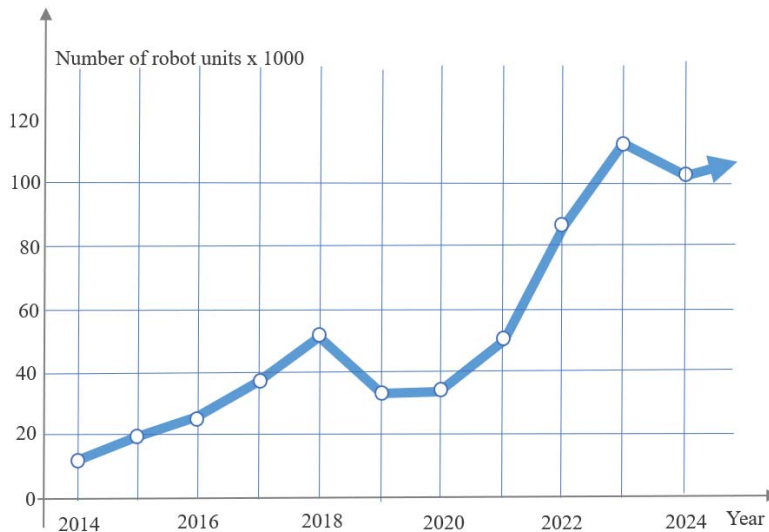


Figure 9. Global trend in the implementation of service robots in logistics and transport (2014–2024).

In 2024, a total of 103,000 units were deployed, representing the highest recorded value during the observed period.

Collaborative robots have become an important component of modern industrial automation between 2017 and 2023. The global trend in their implementation, together with the percentage distribution in 2023, is shown in Figure 10.

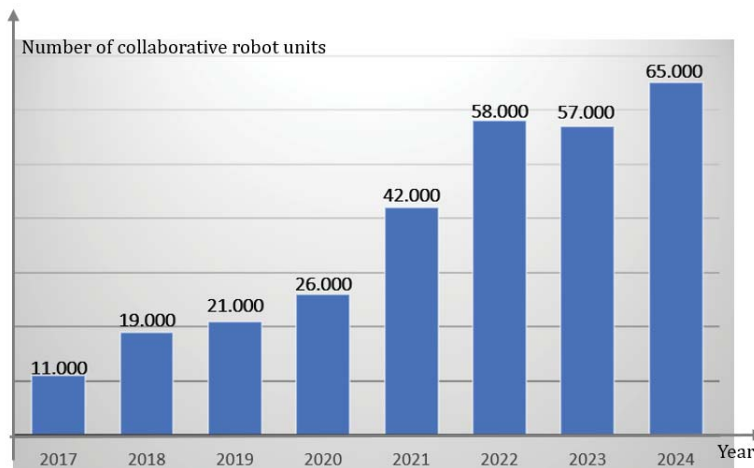


Figure 10. Global implementation of collaborative robots (2017–2023) and percentage distribution in 2023.

According to available data, the global deployment of collaborative robots shows steady growth throughout the observed period. This growth is driven by their flexibility, safety features, and ability to adapt rapidly to diverse production processes. Despite increasing adoption, collaborative robots still account for a relatively small share of the total industrial robot population, while traditional industrial robots continue to dominate automated production facilities.

5.4. Key Factors Influencing Implementation Growth and Sectoral Comparison

The growth in the implementation of mobile and collaborative robots results from the interaction of technological, economic, and organizational factors. Technological advances in sensors, artificial intelligence, edge computing, and wireless communication have enabled reliable robot operation under real industrial conditions. At the same time, declining hardware costs and the development of modular robotic platforms have made these technologies accessible to a broader range of enterprises. Economic factors include rising labor costs, shortages of skilled workers, and the need to enhance competitiveness. In many industrial sectors, robots are not viewed as replacements for human workers but as supportive tools that allow employees to focus on more complex and creative tasks. Organizational factors, such as the adoption of lean and agile production concepts, further stimulate the implementation of flexible robotic systems [42,47]. The application of mobile and collaborative robots varies considerably across industrial sectors, confirming that these technologies represent universal yet adaptable solutions implemented in accordance with the specific requirements of industrial processes.

6. Implementation of Mobile and Collaborative Robots in Manufacturing Processes

The implementation of mobile autonomous robots (AMRs) and collaborative robots (cobots) represents a crucial operational stage in the transition toward smart manufacturing systems aligned with the principles of Industry 4.0. Unlike conventional automation solutions—typically characterized by rigid structures and centralized control—mobile and collaborative robotic systems operate as integral components of cyber-physical production environments. Their functionality extends beyond task execution to encompass continuous process monitoring, adaptive production control, and support for predictive maintenance strategies. This chapter provides an interpretative and integrative discussion of how mobile and collaborative robots are implemented within manufacturing processes, explicitly building on the empirical findings presented in Chapters 4 and 5. Rather than reiterating quantitative results, the focus is placed on explaining

the underlying mechanisms, operational implications, and industrial relevance of the observed performance improvements. The discussion is structured around three interconnected domains: (i) monitoring and real-time supervision of manufacturing processes, (ii) adaptive and data-driven control of production operations, and (iii) robot-supported predictive maintenance aimed at reducing downtime and enhancing system reliability.

6.1. Monitoring and Supervision of Manufacturing Processes

❖ Robotic Systems for Data Acquisition and Visual Inspection

Mobile and collaborative robots increasingly function as distributed sensing and inspection platforms within modern manufacturing environments. Equipped with advanced vision systems—such as 2D and 3D cameras—as well as vibration, temperature, and acoustic sensors, these robots enable continuous acquisition of operational and condition-related data. Compared with fixed inspection solutions, AMRs offer superior spatial flexibility, allowing inspection tasks to be performed across multiple production lines without additional infrastructure investments. The integrated sensing configuration of the AMR–cobot system employed in this study is illustrated in Figure 11.



Figure 11. Autonomous mobile robot and collaborative robot equipped with multiple sensors.

This setup enables continuous visual inspection of components and equipment while simultaneously collecting data relevant to quality assurance and maintenance planning. Such capabilities support early detection of defects, misalignments, and abnormal operating conditions, thereby reducing reliance on manual inspections and mitigating the risk of human error. From an operational perspective, robotic inspection systems enhance consistency and repeatability—two critical factors in high-throughput manufacturing environments. Previous studies have demonstrated that robotic visual inspection systems, particularly

when combined with machine vision and learning-based techniques, outperform manual inspection in terms of accuracy, objectivity, and scalability [61].

❖ Integration of Sensors and AI Algorithms for Real-Time Monitoring

The effectiveness of robotic monitoring systems is substantially enhanced through the integration of sensor networks with artificial intelligence (AI) and machine learning (ML) algorithms. Data acquired by mobile and collaborative robots are processed in real time to identify anomalies, degradation trends, and early indicators of potential failures. This capability enables a shift from passive condition monitoring to active, intelligence-driven process supervision. The real-time data processing and decision-making workflow adopted in this research is illustrated in Figure 12. The figure demonstrates how sensor data collected by the AMR and cobot are analyzed using convolutional neural network (CNN) models and subsequently integrated with Computerized Maintenance Management Systems (CMMS) and Enterprise Resource Planning (ERP) platforms. This architecture ensures that inspection outcomes are directly translated into actionable maintenance and operational decisions.

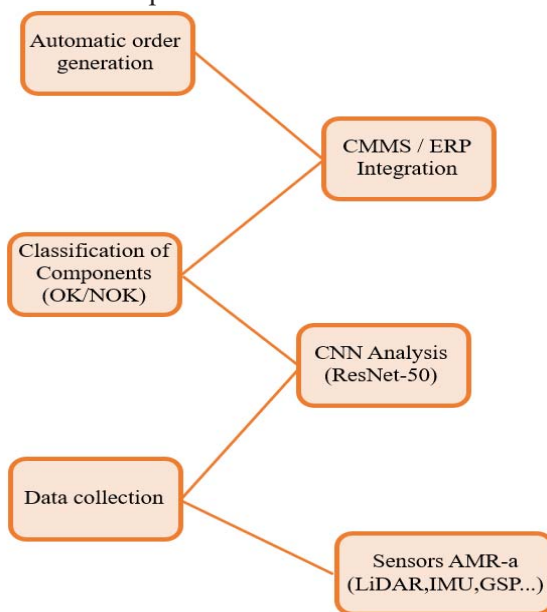


Figure 12. Workflow diagram of the integrated AMR-cobot inspection process.

Such cyber-physical integration is widely recognized as a cornerstone of intelligent manufacturing systems, as it enables continuous learning, rapid response to process deviations, and improved operational transparency [10, 62].

6.2. Control and Adaptive Management of Manufacturing Operations

❖ Role of Robots in Decentralized Process Control

The deployment of mobile and collaborative robots facilitates a transition from centralized control architectures toward decentralized and distributed manufacturing control. In this paradigm, robots operate as autonomous agents capable of local decision-making while remaining interconnected with higher-level manufacturing execution systems. Such decentralization enhances system resilience and flexibility, particularly in production environments subject to frequent changes in demand, product variants, or operating conditions. Through continuous monitoring and bidirectional data exchange, robotic systems dynamically adjust task allocation, inspection frequency, and maintenance priorities. This capability supports adaptive control strategies that respond to real-time system states rather than relying exclusively on predefined schedules. The methodological role of robots in decentralized control frameworks has been extensively discussed in the literature, highlighting improvements in fault tolerance, responsiveness, and overall system robustness [8, 63].

❖ Data-Driven Decision-Making and AI-Supported Control

Data-driven decision-making constitutes a central element of adaptive manufacturing control enabled by mobile and collaborative robots. By continuously collecting, aggregating, and analyzing operational data, robotic systems generate actionable insights that support the optimization of production parameters and predictive maintenance strategies.

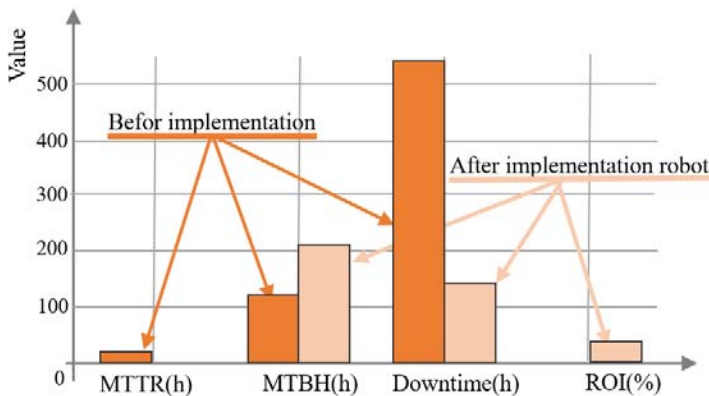


Figure 13. Comparative bar chart of key performance indicators (MTTR, MTBF, downtime, ROI) before and after robotic system implementation.

The quantitative impact of AI-supported decision-making observed in the analyzed case study is summarized in Figure 13, providing a comparative

overview of key performance indicators before and after robotic system implementation. The substantial improvements in mean time to repair (MTTR), mean time between failures (MTBF), and return on investment (ROI) demonstrate the effectiveness of integrating robotic monitoring with AI-driven analytics. Figure 14 presents a comparative analysis of key manufacturing performance indicators before and after the deployment of mobile and collaborative robotic systems supported by AI-driven analytics. The results indicate significant reductions in mean time to repair (MTTR) and overall downtime, reflecting faster recovery from failures and improved operational continuity. At the same time, an increase in mean time between failures (MTBF) highlights enhanced system reliability achieved through predictive monitoring and adaptive maintenance interventions. Furthermore, the observed improvement in return on investment (ROI) demonstrates the economic benefits of combining robotic automation with data-driven decision-making processes. Overall, the figure confirms the positive contribution of decentralized robotic control and intelligent analytics to manufacturing efficiency, resilience, and long-term sustainability, consistent with prior research findings [62].

6.3. Predictive Maintenance Supported by Mobile and Collaborative Robots

❖ Robotic Contribution to Predictive Maintenance Systems

Predictive maintenance (PdM) represents one of the most significant advantages of implementing mobile and collaborative robots in modern manufacturing systems. Acting as mobile inspection and intelligent sensing units, these robots enable continuous condition monitoring of critical equipment while minimizing interruptions to ongoing production activities.

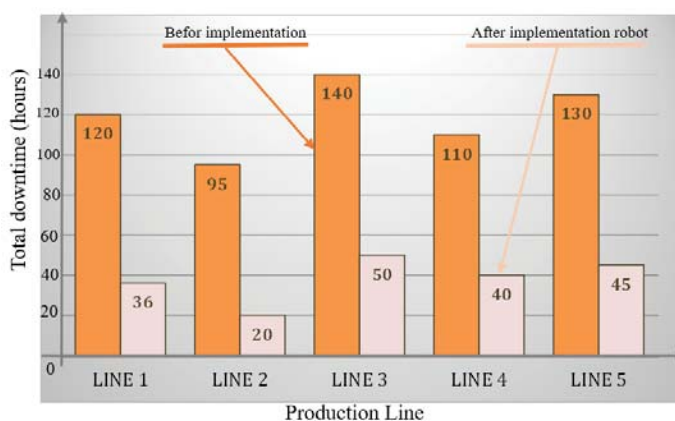


Figure 14. Comparative analysis of total downtime across production lines before and after robotic system implementation

The operational benefits of this approach are reflected in the comparative downtime performance presented in Figure 14, which highlights the measurable impact of robotic support on maintenance efficiency and overall manufacturing stability. Figure 14 presents a comparative analysis of total unplanned downtime across production lines before and after the implementation of mobile and collaborative robotic systems for predictive maintenance [64]. The results indicate a substantial reduction in downtime following the introduction of robotic inspection, autonomous monitoring, and data-driven maintenance interventions. This improvement demonstrates the effectiveness of robot-supported predictive maintenance strategies in identifying early-stage failures, preventing unexpected breakdowns, and enabling timely corrective actions. By providing real-time equipment assessment without disrupting production processes, robotic systems contribute to enhanced operational continuity, increased equipment availability, and improved manufacturing reliability. Overall, the figure emphasizes the critical role of autonomous robotics in minimizing production disruptions and supporting resilient and sustainable smart manufacturing environments.

❖ Integration with Analytical and AI Modules

Robotic data acquisition systems are closely integrated with advanced analytical platforms employing machine learning and deep learning models to forecast equipment failures and degradation trends. This integration allows maintenance activities to be scheduled based on actual equipment condition rather than fixed time intervals, resulting in optimized resource utilization and reduced maintenance costs. The literature consistently reports that AI-enhanced predictive maintenance frameworks outperform traditional preventive maintenance approaches in terms of accuracy and cost efficiency [65,66]. The results discussed in this study further corroborate these findings by demonstrating improved system reliability and sustained reductions in downtime.

❖ Reduction of Downtime and Improvement of System Reliability

The cumulative effect of implementing mobile and collaborative robots is reflected in long-term improvements in production stability and system reliability. The temporal evolution of downtime reduction is presented in Figure 15, which provides a detailed monthly perspective on system performance and operational continuity over the evaluation period. Figure 15 presents the monthly trend of total downtime before and after the implementation of mobile and collaborative robotic systems [64]. The results demonstrate a sustained and progressive decline in downtime throughout the observed timeframe, indicating continuous improvements in production stability and enhanced system reliability. This temporal pattern confirms that robot-supported predictive maintenance offers not only immediate operational advantages but also supports long-term performance optimization through adaptive monitoring strategies.

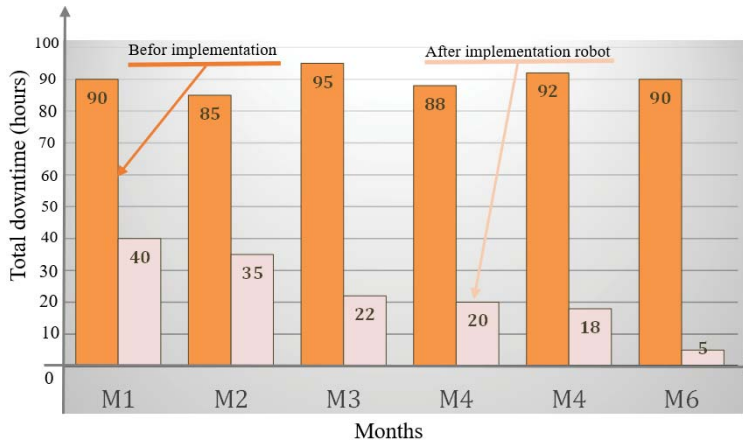


Figure 15. Monthly trend of total downtime before and after robotic system implementation

By enabling proactive interventions, continuous condition assessment, and early-stage fault detection, robotic systems significantly increase equipment availability while reducing unexpected disruptions. Overall, the figure emphasizes the essential role of autonomous robotics in achieving sustainable, resilient, and reliable smart manufacturing operations.

6.4. Critical Discussion and Implementation Implications

While the results demonstrate clear benefits of integrating mobile and collaborative robots into smart manufacturing environments, several practical implementation constraints must be carefully considered. Key challenges include ensuring seamless interaction between robotic platforms and existing industrial infrastructures, addressing cybersecurity vulnerabilities arising from increased digital connectivity, and developing comprehensive workforce reskilling programs to support the operation and maintenance of advanced robotic technologies [67]. In addition, the question of how effectively the proposed architecture can be expanded into less digitally mature manufacturing contexts remains an important area for further investigation. Nevertheless, when appropriately deployed, mobile and collaborative robots should not be viewed merely as automation tools, but rather as strategic enablers of data-driven, cyber-physical production ecosystems. Their combined contribution to real-time monitoring, adaptive process control, and predictive maintenance highlights their role as a technological foundation for resilient, efficient, and sustainable smart manufacturing systems.

7. Technical and Organizational Challenges of Implementing Mobile and Collaborative Robots

Despite the well-documented benefits of mobile and collaborative robots in modern manufacturing systems, their implementation is accompanied by a range of technical, organizational, and economic challenges. The successful integration of these technologies depends not only on the technological maturity of robotic solutions but also on the readiness of organizations to adapt existing processes, structures, and management strategies. This section analyzes the key challenges associated with implementation, with particular emphasis on system interoperability, safety and scalability, organizational and workforce-related issues, economic viability, and standardization and regulatory frameworks.

7.1. Technical Challenges: Interoperability, Safety, and Scalability

From a technical perspective, one of the most demanding aspects of implementing mobile and collaborative robots lies in achieving interoperability with established industrial systems. Manufacturing environments are frequently characterized by heterogeneous machinery, diverse software solutions, and multiple communication protocols, which complicates the integration of robotic technologies into legacy production frameworks. The absence of universally adopted interfaces, together with inconsistent implementation of industrial standards, often results in fragmented system architectures and increased engineering complexity. Safety constitutes another fundamental challenge, particularly in collaborative workspaces where humans and robots operate in close physical proximity. Although modern collaborative robots incorporate advanced safety functionalities—such as force and speed limitation, human presence detection, and emergency stop mechanisms—their effective configuration, validation, and certification in real industrial conditions remains a complex and highly context-dependent task. Moreover, the growing dependence of robotic systems on networked industrial environments and cloud-based infrastructures introduces additional cybersecurity risks, including unauthorized access, data manipulation, and operational disruption [68]. Scalability further intensifies implementation difficulties, as solutions proven in pilot deployments often require major adjustments when extended across multiple production lines or facilities. Without modular system architectures and carefully planned infrastructure upgrades, scaling robotic solutions may lead to disproportionate increases in costs, network demands, and overall technical complexity.

7.2. Organizational and Workforce-Related Challenges

In addition to technical barriers, organizational and workforce-related challenges often play a decisive role in the success or failure of implementing mobile and collaborative robots. The introduction of advanced robotic systems necessitates changes in work organization, task allocation, and the redefinition of employee roles. Resistance to change—particularly among operators and technical staff—can significantly slow down the implementation process. A shortage of appropriate skills represents one of the key challenges. Operating, programming, and maintaining mobile and collaborative robots requires multidisciplinary expertise encompassing robotics, information technologies, and data analytics. Organizations that fail to invest systematically in training and workforce development frequently experience reduced system performance and increased dependence on external vendors and service providers. Moreover, the transformation of human–robot interaction dynamics requires adjustments to safety procedures and organizational culture. Successful implementation relies on building employee trust in robotic systems, which can be achieved through transparent communication, participatory implementation approaches, and continuous education programs.

7.3. Economic Viability and Return on Investment

The economic viability of implementing mobile and collaborative robots is a frequent topic of discussion in industrial practice. Although initial investments in robotic equipment, infrastructure, and training can be substantial, long-term benefits are typically reflected in reduced operating costs, increased productivity, and improved product quality. Return on investment (ROI) depends on multiple factors, including process complexity, robot utilization rates, and the level of integration with other manufacturing systems. Studies indicate that organizations implementing robotics as part of a well-defined digital transformation strategy achieve faster and more stable ROI compared to ad hoc automation initiatives [63]. Nevertheless, economic risks remain, particularly for small and medium-sized enterprises, where limited financial resources and market uncertainty can hinder the justification of initial investments. In such cases, phased implementation strategies and pilot projects represent effective approaches for mitigating financial risk.

7.4. Standardization and Regulatory Frameworks

Standardization and regulatory frameworks play a crucial role in the safe and efficient implementation of mobile and collaborative robots. International standards such as ISO 10218 and ISO/TS 15066 have significantly improved the

safety of collaborative robotic systems; however, their practical application often requires additional interpretation and adaptation to specific industrial contexts. The rapid pace of technological development frequently outstrips the evolution of regulatory frameworks, creating legal and operational uncertainty for organizations deploying advanced robotic solutions. Particular challenges arise with respect to liability in the event of accidents, system certification, and compliance with national and international occupational safety regulations [69]. Further development of harmonized standards and clearer regulatory guidelines is therefore a prerequisite for the broader and safer adoption of mobile and collaborative robots, especially in industries with stringent safety and reliability requirements.

8. Discussion

The results presented in the previous chapters enable a comprehensive discussion of ongoing developments in the industrial deployment of mobile and collaborative robots. By combining patent-based insights with evidence from real implementation cases, the study reveals a clear movement toward robotic systems that are increasingly autonomous, adaptive, and tightly connected with digital manufacturing infrastructures. Patent trends demonstrate an intensified emphasis on artificial intelligence, sensor integration, and autonomous decision support, reflecting the same priorities observed in contemporary industrial applications. This alignment indicates that innovation in robotics is progressively shaped by concrete production requirements and operational challenges rather than purely experimental laboratory advancement. A central outcome of the analysis is the strong linkage between system-level architecture and successful implementation practice. The adoption of modular cyber-physical frameworks provides the structural foundation for integrating mobile and collaborative robots into heterogeneous manufacturing environments. Such architectures enable coordinated data exchange, distributed control functionalities, and improved responsiveness of production systems. The performance improvements discussed in Chapters 6 and 7 suggest that architectural design decisions influence not only technical feasibility but also operational stability and manufacturing continuity. Nevertheless, several limitations must be acknowledged. The study relies on a restricted selection of patent datasets and industrial case implementations, which may limit the broader applicability of the conclusions across all manufacturing sectors. Furthermore, most evaluated examples originate from digitally advanced environments, meaning that the transferability of these findings to less automated or resource-limited contexts remains uncertain. Additionally, longer-term organizational dimensions—including workforce adaptation processes, institutional change, and sustained investment impacts—extend beyond the available observation period. Despite these constraints, the study offers

meaningful implications for industrial stakeholders and future academic inquiry. For industry, the results underline the importance of adopting robotics within an integrated socio-technical framework rather than through isolated automation upgrades. For researchers, the findings emphasize the need to deepen investigations into scalable deployment strategies, harmonized industrial standards, and next-generation intelligent control approaches that support broader adoption across diverse manufacturing settings.

9. Conclusion

This study provides an integrated examination of how mobile and collaborative robots contribute to the evolution of smart manufacturing, drawing together theoretical perspectives, patent development patterns, and practical implementation evidence. The work advances a holistic understanding of robotics adoption by addressing not only technological progress but also the organizational and economic considerations that shape industrial feasibility. The findings indicate that mobile and collaborative robots play a transformative role in enabling new production paradigms characterized by adaptive coordination, predictive capabilities, and increased responsiveness to operational disturbances. Rather than functioning solely as automated machinery, these robotic systems emerge as interconnected elements within intelligent cyber-physical production networks that support more robust manufacturing performance. From an applied perspective, the study suggests that robotics integration should be approached through structured digital transformation planning, with careful consideration of interoperability, safety validation, and workforce qualification. Incremental deployment through pilot-based implementation and continuous evaluation is particularly relevant for manufacturers aiming to balance innovation benefits with investment and operational risks. Future research should focus on expanding empirical validation across broader industrial sectors, advancing AI-enabled autonomy, and developing standardized architectures capable of supporting long-term sustainability objectives. In particular, greater attention is required to understand the socio-economic consequences of robotics-driven transformation as manufacturing systems continue to evolve within dynamic global competitive environments.

References

- [1] Karabegović, I., Karabegović, E., Mehmić, M., Husak, E. (2021). The Application of Industry 4.0 in production Processes of the Automotive Industry, *Journal Mobility and Vehicle*, Vol.47. No.2. pp. 35-44. DOI: 10.24874/mvm.2021.47.02.02
- [2] Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative Industrie 4.0*. Final report of the Industrie 4.0 Working Group. Acatech – National Academy of Science and Engineering. https://www.acatech.de/wp-content/uploads/2018/03/Final_report__Industrie_4.0_accessible.pdf
- [3] Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the art and future trends. *International Journal of Production Research*, 56(8), 2941–2962. <https://doi.org/10.1080/00207543.2018.1444806>
- [4] Karabegović, I., Kovačević, A. Banjanović-Mehmedović, L. Dašić, P. (2020). *Integration Industry 4.0 in Business and Manufacturing*, IGI Global, Pennsylvania, USA
- [5] Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, 6, 1–10. <https://doi.org/10.1016/j.jii.2017.04.005>
- [6] Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210, 15–26. <https://doi.org/10.1016/j.ijpe.2019.01.004>
- [7] Karabegović, I., (2025). High Robotization as a Foundation for Smart Industry: The Republic of Korea and Singapore as Exemplars of Future Industrial Development. *Digital Technologies Research and Applications* 4(2), 125–141. <https://doi.org/10.54963/dtra.v4i2.1314>
- [8] Villani, V., Pini, F., Leali, F., & Secchi, C. (2018). Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, 55, 248–266. <https://doi.org/10.1016/j.mechatronics.2018.02.009>
- [9] Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, 56(1–2), 508–517. <https://doi.org/10.1080/00207543.2017.1351644>
- [10] Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 18, 20–23. <https://doi.org/10.1016/j.mfglet.2018.09.002>
- [11] Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3, 18–23. <https://doi.org/10.1016/j.mfglet.2014.12.001>
- [12] Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., & Alcalá, S. G. S. (2021). A systematic literature review of machine learning methods

- applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024. <https://doi.org/10.1016/j.cie.2019.106024>
- [13] Karabegović, I., Karabegović, E. (2019). The Role of Collaborative Service Robots in the Implementation of Industry 4.0, *International Journal of Robotics and Automation Technology*, Vol.6., pp.40-46.<https://doi.org/10.31875/2409-9694.2019.06.5>
- [14] Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, 6(4), 239–242. <https://doi.org/10.1007/s12599-014-0334-4>
- [15] Kim, J., Seo, D., Moon, J., Kim, J., Kim, H., & Jeong, J. (2022). Design and Implementation of an HCPS-Based PCB Smart Factory System for Next-Generation Intelligent Manufacturing. *Applied Sciences*, 12(15), 7645. <https://doi.org/10.3390/app12157645>
- [16] Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of Industry 4.0: A review. *International Journal of Production Economics*, 182, 563–579. <https://doi.org/10.1016/j.ijpe.2016.11.005>
- [17] Karabegović, I., Karabegović, E., Mahmić, M., Husak, E. (2018). Innovative Automation of Production Processes in the Automotive Industry, *International Journal of Engineering Works*, Vol.5., Iss.11., Kambohwell Publisher Enterprises, Multan, Pakistan. <https://doi.org/10.5281/zenodo.1486145>
- [18] Karabegović, I., (2022). Sensory Technology is One of the Basic Technologies of Industry 4.0 and the Fourth Industrial Revolution. *ACTA Technica Corviniensis – Bulletin of Engineering*, 15 (4): 33-39. <https://acta.fih.upt.ro/pdf/2022-4/ACTA-2022-4-05.pdf>.
- [19] Wang, S., Wan, J., Li, D., & Zhang, C. (2016). Implementing smart factory of Industrie 4.0: An outlook. *Computers & Industrial Engineering*, 101, 203–216. <https://doi.org/10.1016/j.cie.2016.04.001>
- [20] Monostori, L., Kádár, B., Bauernhansl, T., et al. (2016). Cyber-physical systems in manufacturing. *CIRP Annals*, 65(2), 621–641. <https://doi.org/10.1016/j.cirp.2016.06.005>
- [21] Bogue, R. (2018). What are the prospects for robots in the factories of the future? *Industrial Robot: An International Journal*, 45(1), 1–5. <https://doi.org/10.1108/IR-11-2017-0199>
- [22] Kamble, S. S., Gunasekaran, A., & Gawankar, S. A. (2018). Sustainable Industry 4.0 framework: A systematic literature review. *Journal of Cleaner Production*, 181, 620–634. <https://doi.org/10.1016/j.jclepro.2018.01.131>
- [23] Bahrin, M. A. K., Othman, M. F., Azli, N. H. N., & Talib, M. F. (2016). Industry 4.0: A review on industrial automation and robotic. *Jurnal Teknologi*, 78(6–13), 137–143. <https://doi.org/10.11113/jt.v78.9285>

- [24] Gorecky, D., Schmitt, M., Loskyll, M., & Zühlke, D. (2014). Human-machine-interaction in the Industry 4.0 era. *Procedia CIRP*, 25, 289–294. <https://doi.org/10.1016/j.procir.2014.10.038>
- [25] Griliches, Z. (1990). Patent statistics as economic indicators. *Journal of Economic Literature*, 28(4), 1661–1707. <https://doi.org/10.3386/w3301>
- [26] Karabegović, I., Husak, E., Vojić, S., Karabegović, E., Mahmić, M., (2025). Integration of Robotic Technology and Human Creativity: From Automation to Humanization. In: Karabegović, I. (Ed), *Proc. of the Industry of the Future: From Industry 4.0 to Industry 5.0 - Integration of Technology and People. Special Editions CCXX, Department of Technical Sciences, Volume 24*, ANUBiH, Sarajevo, BiH. 3–28. <https://doi.org/10.5644/PI2025.220.01>
- [27] Ernst, H. (2003). Patent information for strategic technology management. *World Patent Information*, 25(3), 233–242. [https://doi.org/10.1016/S0172-2190\(03\)00077-2](https://doi.org/10.1016/S0172-2190(03)00077-2)
- [28] Lens. (n.d.). *The Lens patent database*. Cambia. Retrieved January 25, 2026, from <https://www.lens.org>
- [29] European Patent Office. (n.d.). *PATSTAT: Worldwide patent statistical database*. Retrieved January 25, 2026, from <https://www.epo.org/searching-for-patents/business/patstat.html>
- [30] World Intellectual Property Organization. (n.d.). *WIPO IP statistics data center*. Retrieved January 25, 2026, from <https://www.wipo.int/ipstats/en/>
- [31] Center for Security and Emerging Technology. (n.d.). *Patent data and emerging technology analysis*. Georgetown University. Retrieved January 25, 2026, from <https://cset.georgetown.edu>
- [32] International Federation of Robotics. (2024). *World Robotics 2024: Industrial Robots*. IFR Publishing. <https://ifr.org/worldrobotics>
- [33] Karabegović, I., Mahmić, M., Karabegović, E., Husak, E., (2024). Advanced robotics as the drive of innovation: The role of the implementation of advanced robotics in Industry 4.0. In: Karabegović, I., Kovačević, A., Mandžuka, S., (Eds), *New Technologies, Development and Application VII - NT 2024. Lecture Notes in Networks and Systems*, 1069. Springer, Heidelberg. https://doi.org/10.1007/978-3-031-66268-3_1
- [34] Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation. *NBER Working Paper*, 24449. <https://doi.org/10.3386/w24449>
- [35] Zhou, Z.-H., Chen, X., & Li, J. (2019). Artificial intelligence: A survey. *Frontiers of Computer Science*, 13(1), 1–17. <https://doi.org/10.1007/s11704-018-8229-1>

- [36] Organisation for Economic Co-operation and Development. (2021). *OECD patent statistics manual*. OECD Publishing. <https://doi.org/10.1787/9789264247310-en>
- [37] World Intellectual Property Organization. (2022). *World Intellectual Property Indicators 2022*. WIPO. <https://www.wipo.int/publications/en/details.jsp?id=4622>
- [38] Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review*, 95(4), 3–11. <https://doi.org/10.2139/ssrn.2901173>
- [39] Karabegović, I., Karabegović, E., (2018) Implementation of Industry 4.0 by Application Robots and Digital Technologies in Production Processes in China. *Tehnika* 73(2): 225-231. <https://doi.org/10.5937/tehnika1802225K>
- [40] Bauer, W., Hämmerle, M., Schlund, S., & Vocke, C. (2018). Transforming to a hyper-connected society and economy—Towards an “Industry 4.0”. *Procedia Manufacturing*, 26, 417–429. <https://doi.org/10.1016/j.promfg.2018.07.056>
- [41] Karabegović, I., (2017). Digital Technology as the key Factor in the Fourth Industrial Revolution - Industry 4.0. *International Journal of Engineering and Advanced Research Technology* 3(3): 17-22. https://www.ijeart.com/download_data/IJEART03313.pdf
- [42] Fragapane, G., Ivanov, D., Peron, M., Sgarbossa, F., & Strandhagen, J. O. (2020). Increasing flexibility and productivity in Industry 4.0 production networks with autonomous mobile robots. *International Journal of Production Research*, 58(21), 6466–6484. <https://doi.org/10.1080/00207543.2020.1753899>
- [43] Quasi AI. (n.d.). *Autonomous mobile robots in material handling: Why manufacturing is finally ready*. Retrieved January 26, 2026, from <https://www.quasi.ai/autonomous-mobile-robots-in-material-handling-why-manufacturing-is-finally-ready/>
- [44] AutoMarket.ro. (n.d.). *Faceți cunoștință cu Spotto, noul câine robot folosit de BMW pentru...* Retrieved January 26, 2026, from <https://www.automarket.ro/stiri/faceti-cunostinta-cu-spotto-noul-caine-robot-folosit-de-bmw-pentru-118223.html>
- [45] ETH Zürich, Department of Management, Technology and Economics (D-MTEC). (n.d.). *MAS+ integrative student-alumni learning format*. Retrieved January 26, 2026, from <https://mas-mtec.ethz.ch/education/special-programmes/mas-plus.html>
- [46] Wurman, P. R., D’Andrea, R., & Mountz, M. (2021). Coordinating hundreds of cooperative, autonomous vehicles in warehouses. *AI Magazine*, 42(1), 20–34. <https://doi.org/10.1609/aimag.v42i1.15650>
- [47] Gualtieri, L., Rauch, E., & Vidoni, R. (2021). Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review. *Robotics and Computer-Integrated Manufacturing*, 67, 102025. <https://doi.org/10.1016/j.rcim.2020.102025>

- [48] Karabegović, E., Karabegović, I., Husak, E., (2019). Collaborative Robots Innovate Automation in metal and Car Industry in the World. *J. of Scientific and Engineering Research* 6(8): 136-145. <https://jsaer.com/download/vol-6-iss-8-2019/JSAER2019-6-8-136-145.pdf>
- [49] Djuric, A. M., Urbanic, R. J., & Rickli, J. L. (2020). A framework for collaborative robot (cobot) integration in advanced manufacturing. *Procedia Manufacturing*, 51, 1310–1317. <https://doi.org/10.1016/j.promfg.2020.10.184>
- [50] Doimak. (n.d.). *RIR+RE 250 grinding machine with robot & artificial intelligence*. Doimak. Retrieved January 26, 2026, from <https://doimak.es/portfolio/rir-series/>
- [51] PNCTECH. (n.d.). *Chi tiết cấu tạo máy cấp phối tự động* [Detailed structure of an automatic feeding machine]. Retrieved January 26, 2026, from <https://pnctech.vn/chi-tiet-cau-tao-may-cap-phoi-tu-dong/>
- [52] Modern Machine Shop. (2018). *Vision-guided cobot doubles daily production*. <https://www.mmsonline.com/articles/vision-guided-cobot-doubles-daily-production>
- [53] International Federation of Robotics. (2015). *Executive summary: World Robotics 2015 – Industrial Robots*. Retrieved January 26, 2026, from https://ifr.org/img/worldrobotics/Executive_Summary_WR_Industrial_Robots_2015.pdf
- [54] International Federation of Robotics. (2023). *World Robotics Report: “All-time high” with half a million robots installed in one year*. Retrieved January 26, 2026, from <https://ifr.org/ifr-press-releases/news/wr-report-all-time-high-with-half-a-million-robots-installed>
- [55] Karabegović, I., (2025). A Comparative Analysis of the Role of Industrial and Humanoid Robots as Drivers of Efficiency and Flexibility in the Automotive Industry. *Journal of Mobility & Vehicle Mechanics*, 51(1), 31–45. <https://doi.org/10.24874/mvm.2025.51.01.03>
- [56] International Federation of Robotics. (2023). *World Robotics 2023 report: Asia ahead of Europe and the Americas*. Retrieved January 26, 2026, from <https://ifr.org/ifr-press-releases/news/world-robotics-2023-report-asia-ahead-of-europe-and-the-americas>
- [57] International Federation of Robotics. (2023). *Executive summary: World Robotics 2023 – Industrial Robots*. Retrieved January 26, 2026, from https://ifr.org/img/worldrobotics/Executive_Summary_WR_Industrial_Robots_2023.pdf
- [58] International Federation of Robotics. (2025). *Executive summary: World Robotics 2025 – Industrial Robots*. Retrieved January 26, 2026, from https://ifr.org/img/worldrobotics/Executive_Summary_WR_2025_Industrial_Robots.pdf

- [59] Bogue, R. (2020). Robots in logistics: A review of industrial applications. *Industrial Robot: The International Journal of Robotics Research and Application*, 47(3), 403–409. <https://doi.org/10.1108/IR-11-2019-0234>
- [60] International Federation of Robotics. (2014). *World Robotics 2014: Service Robots*. IFR Publishing. Retrieved January 26, 2026, from <https://ifr.org/worldrobotics>
- [61] Magalhães, P., & Ferreira, N. (2022). Inspection applications in industrial environments using collaborative robots. *Automation*, 3(2), 258–268. <https://doi.org/10.3390/automation3020013>
- [62] Mourtzis, D., Tsubou, S., & Angelopoulos, J. (2023). Robotic cell reliability optimization based on digital twin and predictive maintenance. *Electronics*, 12(9), 1999. <https://doi.org/10.3390/electronics12091999>
- [63] Kousi, N., et al. (2019). Digital twin-based decision support systems for production planning and control. *International Journal of Computer Integrated Manufacturing*, 32(8), 790–804. <https://doi.org/10.1080/0951192X.2019.1599436>
- [64] Karabegović, I., (2025). AI-Enhanced Mobile and Collaborative Robotics for Autonomous Inspection and Predictive Maintenance in Smart Manufacturing. *Smart Wearable Technology*, 0 (0) 1–12. <https://doi.org/10.47852/bonviewSWT52027397>
- [65] Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment. *IEEE Systems Journal*, 13(3), 2213–2227. <https://doi.org/10.1109/JSYST.2018.2878886>
- [66] Kumar, P., Khalid, S., & Kim, H. S. (2023). Prognostics and health management of industrial robots using deep learning: A review. *Mathematics*, 11(13), 3008. <https://doi.org/10.3390/math11133008>
- [67] Frazzon, E. M., Hartmann, J., Makuschewitz, T., & Scholz-Reiter, B. (2013). Towards socio-cyber-physical systems in production networks. *Procedia CIRP*, 7, 49–54. <https://doi.org/10.1016/j.procir.2013.05.009>
- [68] Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial Internet of Things (IIoT): An analysis framework. *Computers in Industry*, 101, 1–12. <https://doi.org/10.1016/j.compind.2018.04.015>
- [69] Decker, M., Fischer, M., & Ott, I. (2017). Service robotics and human labor: A first technology assessment of substitution and cooperation. *Robotics and Autonomous Systems*, 87, 348–354. <https://doi.org/10.1016/j.robot.2016.09.017>

Application of Augmented Reality in Collaborative Robots

Ermin Husak^{*1}

Abstract: *Augmented Reality (AR) and collaborative robotics represent two interconnected technologies that play a significant role in the development of modern industrial production and the concept of Industry 4.0. Augmented reality enables the display of digital information in a real environment, which improves user interaction with production systems, while collaborative robots enable safe cooperation between humans and robots in a shared workspace. The paper presents the basic working principles of AR technology, its types and areas of application, as well as the characteristics of collaborative robots, their safety requirements and application possibilities in industry. A special emphasis is placed on the integration of these technologies, whereby augmented reality enables visualization of processes in real time, simpler programming of robots, simulation of work operations, increased safety, and more efficient training and technical support for operators. The analysis shows that the combination of AR technology and collaborative robotics contributes to increasing productivity, production quality, flexibility of production processes and safety of the working environment. The further development of these technologies, along with the application of artificial intelligence, machine vision and IoT systems, represents an important step towards the development of smart factories of the future.*

Keyword: *Collaborative Robotics, Cobots, Augmented Reality, Programming, Simulation*

1. Introduction

The development of digital technologies in recent years has significantly changed the way industrial production processes are organized and managed. The concepts of Industry 4.0 and Industry 5.0 are based on the integration of advanced information and communication technologies, automation, robotics and intelligent production systems with the aim of increasing productivity, quality and flexibility of production. In this context, Augmented Reality - AR and collaborative robotics, which represent one of the most promising technologies of contemporary industry, stand out. Augmented reality enables the display of digital content in a real environment, thus providing users with additional information and more intuitive interaction with production systems [1,2,3]. At the same time, collaborative robots enable safe cooperation between humans and robots in a shared workspace, combining human flexibility, creativity and decision-making

^{*1}University of Bihać, Technical Faculty, Bihać, Bosnia and Herzegovina
E-mail: erminhusak@gmail.com

ability with the precision, speed and repeatability of robotic systems[4-7]. The aim of this paper is to present the basic characteristics of augmented reality and collaborative robotics, analyses their application possibilities in industry and point out the advantages of their mutual integration. Special attention is paid to the application of AR technology in the visualization of data in real time, programming of collaborative robots, simulation of production processes, increase of work safety and improvement of training and maintenance processes. The integration of these technologies represents an important step towards the development of intelligent, flexible and safe production systems of the future.

2. Augmented Reality – AR

Augmented Reality (AR) is a technology that allows overlaying digital content with the real environment in real time. Unlike virtual reality (VR), which fully immerses the user in a computer-generated environment, augmented reality maintains a view of the real world and simultaneously displays additional information, three-dimensional models, text, images or animations. The basic working principle of augmented reality is based on the use of cameras, sensors, processors and appropriate software that recognizes objects or the space in which the user is located. After recognizing the environment, the system adds virtual elements to the screen of a mobile phone, tablet, smart glasses or other device that are displayed as an integral part of the real environment. In this way, the user gets additional information without having to leave the real space. The application of augmented reality is very wide today. In education, it is used to display complex processes and phenomena through interactive 3D models, which enables pupils and students to understand the teaching content more easily. In industry and production, AR helps workers during assembly, maintenance and quality control by displaying instructions directly at the workplace. In medicine, it is used to plan operations, educate medical staff, and visualize the anatomical structures of patients. Significant applications of augmented reality are also present in marketing, retail, and the entertainment industry. Shoppers can virtually try on clothes, glasses, or furniture before making a purchase, while video game users can experience interactive content that is connected to the real environment.

Previously, augmented reality was considered complex to use, but today this is no longer the case. Using AR technology generally does not require special devices or special knowledge. Most modern mobile phones and laptops can use AR applications. Users can simply download the appropriate application and use augmented reality. Many believe that using augmented reality requires wearing special glasses or a helmet. It is true that AR glasses have become popular thanks to companies such as Microsoft and Google, but their use is not mandatory. Today, augmented reality content is most often displayed via mobile phones and tablets.

AR visualization technologies can be classified into four main categories: Head-Mounted Displays (HMDs), Hand-Held Displays (HHDs), Fixed or Static Screens, Spatial Displays (Figure1).

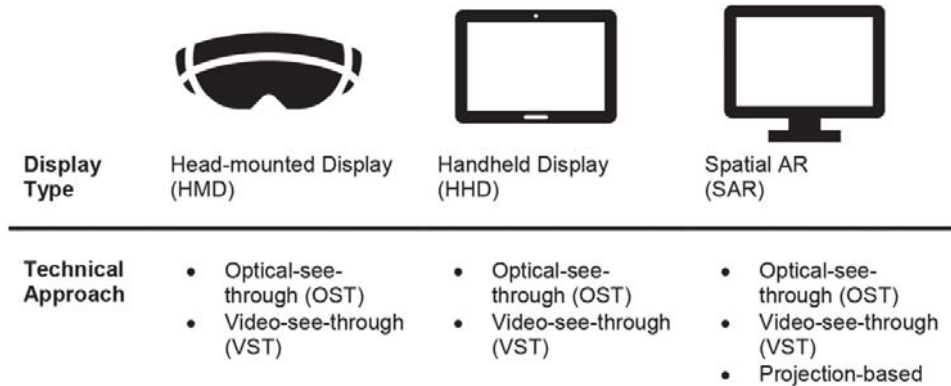


Figure 1: AR visualization technologies [1]

- Head-Mounted Displays (HMDs)

Head-Mounted Display (HMD) is a device that is worn on the head like glasses or a helmet and displays digital content directly in front of the user's eyes. It is used in virtual reality (VR) and augmented reality (AR) systems, enabling the display of three-dimensional objects, information and virtual environments while tracking the movements of the user's head.



Figure 2. Head-Mounted Displays

Examples of HMD devices are: Meta Quest 3, Apple Vision Pro, and HTC Vive.

- Hand-Held Displays (HHDs)

Hand-Held Displays (HHDs) in the context of augmented reality are hand-held devices that enable the display of digital information and virtual objects over an image of the real environment. The most common examples of HHD devices are smartphones and tablets (Figure 3). With the HHD system, the user holds the device in his hand, and the built-in camera records the real environment. The software then displays on the screen a combination of the real image and virtual elements, such as text, graphics, 3D models or animations. The advantages of HHD devices are wide availability because most users own a smartphone, relatively low price compared to AR glasses, simple use, the possibility of using existing cameras, sensors and GPS. The disadvantages of HHD devices are that the user must hold the device in his hand during use, a lower level of immersion compared to AR glasses, a limited field of vision, and hand fatigue during prolonged use.



Figure 3. HD AR application on Apple iPad 10.2"

- Fixed or Static Screens

These are televisions and computer monitors. They display extended content on fixed screens. They require additional recording and processing devices. They are limited to stationary applications and require the user's attention to be diverted.

- Spatial Displays

These are projectors that display content directly on real objects. They do not require user attention. They are suitable for team work, but are limited to

62

stationary use and can have the problem of obscuration. They are also known as SAR. The Spatial Reality Display reproduces spatial images in three dimensions as if they were real and can be viewed by the naked eye without special glasses or headsets (Figure 4).



Figure 4. Spatial Displays

3. Collaborative robotics

The development of industrial automation during the last few decades has significantly changed the way production processes are organized. The introduction of industrial robots increased productivity, improved product quality and reduced the need for people to perform difficult and dangerous jobs. However, despite the rapid development of automation, human workers still play an irreplaceable role in many production activities. Humans have the ability to make decisions, adapt to unforeseen situations, think creatively and perform complex manipulative tasks, while robots provide great precision, speed, power and repeatability of work. Precisely because of the complementarity of their capabilities, the concept of human-robot collaboration (HRC) was developed, which represents one of the most significant trends within Industry 4.0 and 5.0. Instead of completely replacing human workers, modern production systems strive to achieve optimal cooperation between humans and robots in order to maximize the advantages of both participants in the work process. Collaborative robots, also known as cobots, represent a special category of industrial robots designed to work in close proximity to humans. Unlike traditional industrial robots that are mostly housed inside protective enclosures and physically separated from workers, collaborative robots can share the same workspace with

human operators. According to the ISO international standard, defines collaborative robots as robots designed for direct interaction with a human within a limited collaborative workspace. This approach enables a human and a robot to simultaneously perform different or interconnected tasks during the production process.

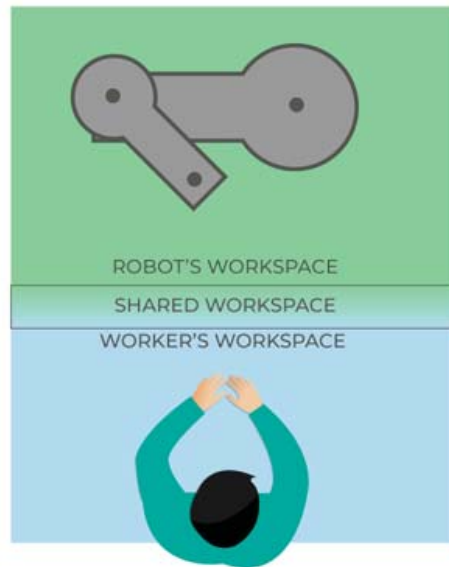


Figure 5 Workers' and robots' workspaces.

The basic idea of collaborative robotics is to combine human intelligence, flexibility and decision-making ability with the precision, strength and durability of robotic systems. In this way, it is possible to achieve higher productivity, better product quality and a safer working environment. Traditional industrial robots are primarily designed to perform repetitive tasks at high speed. Due to the large forces and movement speeds they develop, their use requires the installation of protective fences, safety sensors and other measures that prevent access by workers during robot operation. On the other hand, collaborative robots are designed with special safety functions that enable safe work near people (Figure 5). Their characteristics include:

- lower mass and more compact dimensions,
- limited speed and force of movement,
- built-in force and torque sensors,
- the possibility of detecting human contact,
- simple programming and configuration,
- greater flexibility when changing production tasks.

Thanks to these features, cobots can be relatively quickly integrated into existing production systems without the need for significant changes to the workspace. Safety is the most important aspect of the application of collaborative robots. Although it is often assumed that cobots are completely safe to work with humans, this assumption is not always true. A robot is not necessarily collaborative in itself, but a particular application becomes collaborative only after the implementation of appropriate safety features.

Figure 6 shows different types of cooperation between a human and an industrial robot, arranged according to the level of cooperation and safety requirements.

In the first case, called a cell, the robot works completely separately from the human within an enclosed space. There is a physical barrier that prevents any direct contact, so we cannot speak of real cooperation, but of classical automation. In coexistence, the robot and the human work in the same environment without a physical barrier, but they do not share the same workspace at the same time. Although they are close, their tasks are spatially separated, so the level of interaction is very low. Synchronization implies that the robot and the human use the same workspace, but not simultaneously. Their activities take place alternately, so that one completes its part of the task before the other begins. In cooperation, a robot and a human work simultaneously on the same task or part of the product. Both participants are active at the same time, which requires greater coordination and security measures. The highest level is responsive collaboration, where the robot reacts to human movements in real time. In this case, there is a direct and dynamic interaction, because the robot adjusts its behaviour in accordance with the operator's activities.

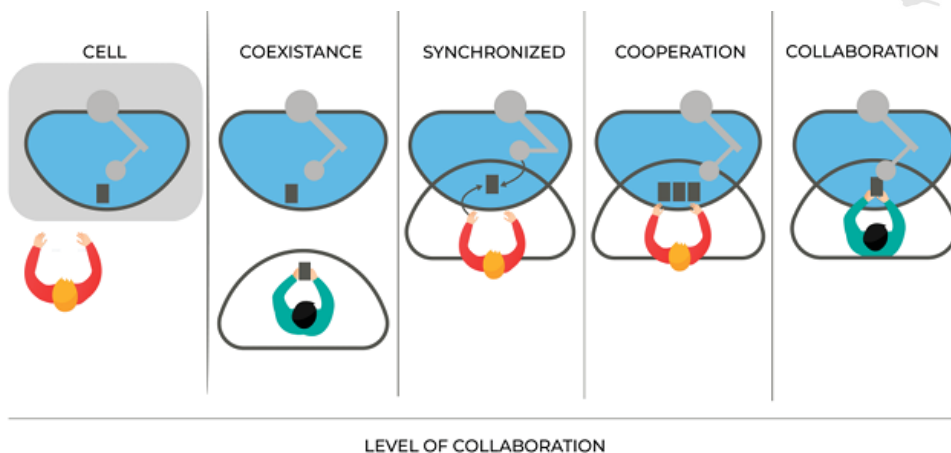


Figure 6. *Levels of Interaction*

Modern collaborative systems use different methods of operator protection. One of the most frequently applied methods is monitoring the speed and separation of man and robot. In this case, the sensors continuously monitor the distance between the operator and the robot and automatically reduce the speed or stop the robot when the human gets too close. The second method involves limiting the force and power of the robot. If there is contact between a human and a robot, the force developed remains within the permissible limits to avoid injuries.

However, safety also depends on the type of tool that the robot uses. For example, a robot equipped with a sharp tool, welding device or other dangerous attachment can pose a risk even when operating at low speeds. This is why a detailed risk assessment is necessary before implementing any collaborative system.

In recent years, the market for collaborative robots has seen extremely rapid growth. A large number of companies develop and produce cobots for various industrial applications. Among the most famous manufacturers are ABB, KUKA, Universal Robots, Yaskawa, COMAU, Franka Emika, TECHMAN Robot and Rethink Robotics. Each of these companies develops its own solutions adapted to the specific needs of users.

Universal Robots is considered one of the pioneers in the field of collaborative robotics and has significantly contributed to the popularization of cobots in small and medium-sized enterprises. ABB and KUKA develop advanced robotic systems for demanding industrial applications, while TECHMAN Robot integrates machine vision systems directly into robotic platforms. Franka Emika is especially known for flexible robotic solutions intended for research and advanced industrial applications.

Today, collaborative robots find application in almost all branches of industry. Their flexibility enables the performance of a large number of different tasks.

One of the most common applications of cobots is product assembly. Robots can precisely turn screws, place components and assemble products with minimal error. The operator performs more complex tasks and process monitoring, while the robot performs repetitive operations [8,9,10].

Collaborative robots are often used to apply adhesives, sealants, paints and other materials. Thanks to high precision, an equal amount of applied material and constant product quality are ensured. Grinding, polishing and surface treatment processes require precise and repeatable movements. Cobots can maintain a constant pressure force and processing speed, which achieves a uniform quality of the final surface.

In many factories, collaborative robots take over the tasks of loading and unloading parts from CNC machines, presses or plastic injection moulding machines. This increases the utilization of the equipment and reduces the burden on the operator. Packing, sorting, palletizing and transporting products are ideal tasks for cobots. Their ability to work in close proximity to people enables efficient organization of logistics processes within production facilities.

Collaborative robots can be equipped with cameras, sensors and machine vision systems that enable automatic product inspection. In this way, it is possible to detect errors in a timely manner and increase the quality of production.

Modern cobots are also used for various welding procedures. The robot ensures precise tool guidance and constant quality of the welded joint, while the operator monitors the process and makes adjustments as needed [11-15].

The use of collaborative robots brings numerous advantages to industrial companies. First of all, the productivity of production processes increases because the robot can continuously perform repetitive tasks without fatigue. At the same time, product quality improves thanks to high precision and repeatability of work. An additional advantage is reflected in increased production flexibility. Unlike traditional automated systems, cobots can be reprogrammed for new tasks and products relatively easily. This is especially important for small and medium-sized enterprises that produce smaller series of products.

Collaborative robots also contribute to improving ergonomic working conditions because they take on physically demanding, monotonous and potentially dangerous tasks. In this way, the workload of workers is reduced and workplace safety is increased.

The further development of collaborative robots is closely related to the progress of artificial intelligence, machine learning, sensor technologies and machine vision systems. It is expected that future robots will be able to better understand human intentions, adapt their behaviour to changes in the environment and learn new tasks independently. Integration of cobots with Industry 4.0 technologies will enable even greater connectivity of production systems, real-time data exchange and optimization of production processes. In this way, human-robot cooperation will become one of the key elements of the smart factories of the future [16-20].

4. The Connection Between Augmented Reality (AR) and Collaborative Robotics

Augmented Reality (AR) and collaborative robotics represent two key technologies of modern industry that together contribute to increasing the productivity, safety and flexibility of production systems. Collaborative robots (cobots) are designed for safe work in close proximity to humans, while AR enables the display of digital information directly in the user's real environment. Their integration creates a more intuitive and efficient working environment [3].

One of the most important advantages of the integration of augmented reality (AR) and collaborative robotics is the visualization of information in real time. This capability allows operators to instantly obtain relevant data on the operation of the robot and the production process without the need to use additional monitors or control panels.

By using AR devices, such as smart glasses, tablets or mobile devices, digital information is displayed directly in a real work environment. The operator can see

the robot's position, movement speed, current task status, actuator load, safety zones and other process parameters at the time of their creation. In this way, an intuitive interaction between man and robot is realized.

Real-time visualization is especially important for collaborative robots because it enables a better understanding of their intentions and future movements. The AR system can display the planned path of the robot's movement before its execution, which reduces the possibility of collisions and increases the safety of the joint work of man and robot.

In production processes, operators can monitor:

- current status of task execution,
- positions and orientations of robotic tools,
- product quality parameters,
- warnings about errors and malfunctions,
- safety zones and areas of possible collision,
- data collected through sensors and IoT devices.

In addition to production, visualization of information in real time is also used during the maintenance of robotic systems. Technicians can instantly see diagnostic data, location of faulty components and recommended troubleshooting steps, significantly reducing system downtime.

Thanks to these possibilities, AR becomes an important tool for the improvement of collaborative robotics, because it increases the transparency of the process, accelerates decision-making, reduces the number of errors and contributes to safer and more efficient cooperation between humans and robots in the Industry 4.0 environment.

Simpler programming of robots with the use of augmented reality (AR) represents a modern approach to managing robotic systems in which complex programming code is replaced by intuitive interaction in real space (Figure 7). Instead of classical programming, the operator uses AR devices (eg smart glasses or a tablet) to define the robot's tasks directly in the physical environment. In the AR environment, the user can visually mark the robot's movement points, define paths, work zones and tool orientation simply by "pointing" or selecting objects in space. The system then automatically converts this information into a robot program. In this way, the need to manually write complex program instructions is eliminated. The key advantage of this approach is intuitiveness – the operator sees the real workspace alongside digital layers of information (eg the planned path of the robot, speed of movement or safety zones). This significantly reduces the probability of errors and speeds up the programming process.



Figure 7. Robot programming using augmented reality

AR also enables visual simulation before execution, where the movement of the robot can be tested in real time without actually starting the machine (Figure 8). This increases security and enables quick corrections. In industrial environments, this way of programming is particularly useful in flexible manufacturing, where frequent task changes must be implemented quickly. Operators without advanced programming knowledge can efficiently adjust robots, which reduces dependence on specialized engineers. Augmented reality (AR) contributes significantly to increasing safety in industrial environments, especially in collaborative robotics systems where humans and robots work in the same workspace. AR allows operators to see in real time additional visual information that is not physically present, but is crucial for safe interaction with the system. One of the most important safety functions of AR is the visualization of safety zones around the robot. Using AR glasses or a tablet, the operator can see marked danger zones, robot reach limits and potential collision areas.



Figure 8. Simulations with the help of AR

In this way, the risk of accidental entry into the danger zone during machine operation is reduced.

Augmented reality (AR) significantly improves training and technical support processes in industrial environments, especially when working with collaborative robots and complex manufacturing systems. Instead of classic manuals and static instructions, AR enables interactive learning directly in the real workspace. In the training process, AR displays step-by-step instructions in real time, which are overlaid on the actual equipment. Trainees can see where certain parts of the robot are located, how to use them properly, and in what order to perform operations. This significantly reduces learning time and increases understanding of complex procedures.

AR also enables simulation of work without real risk, where users can practice programming, operating a robot or maintaining a system in a safe environment. In this way, the possibility of errors in real production is reduced.

5. Conclusion

Augmented reality and collaborative robotics represent complementary technologies that have significant potential for improving modern industrial production systems. While collaborative robots enable safe and efficient cooperation between humans and robots, augmented reality provides an intuitive way of displaying information, managing processes and interacting with the production environment. By integrating these technologies, greater transparency of production processes, simpler programming of robotic systems, more efficient

simulation of work operations, improved work safety and better training and technical support for operators is achieved. At the same time, productivity, production flexibility and product quality increase, while reducing the possibility of errors and downtime in the production process. Further development of artificial intelligence, machine vision, sensor technologies and the Internet of Things will further improve the possibilities of cooperation between humans and robots. It is expected that the integration of augmented reality and collaborative robotics will play an increasingly important role in the development of smart factories, where intelligent systems will enable more efficient, safer and more adaptable production processes in accordance with the requirements of Industry 4.0 and Industry 5.0.

References

- [1] Leins, N., Gonnermann-Müller, J. & Teichmann, M. Comparing head-mounted and handheld augmented reality for guided assembly. *J Multimodal User Interfaces* 18, 313–328 (2024). <https://doi.org/10.1007/s12193-024-00440-1>
- [2] Frederic Z. Zhang, D. C., and Gould, S. Spatially conditioned graphs for detecting human–object interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2021), pp. 13319–13327.
- [3] Costa, G.d.M.; Petry, M.R.; Moreira, A.P. Augmented Reality for Human–Robot Collaboration and Cooperation in Industrial Applications: A Systematic Literature Review. *Sensors* 2022, 22, 2725. <https://doi.org/10.3390/s22072725>
- [4] Bauer, A., Wollherr, D., and Buss, M. Human–robot collaboration: a survey. *International Journal of Humanoid Robotics* 5, 01 (2008), 47–66.
- [5] Subašić, S., Banjanović-Mehmedović, L., Subašić, H., Karabegović, I., Husak, E., 2025, Vision Transformer-Based Multimodal Fusion of Gesture and Object Classification in Human-Robot Collaboration, under exclusive license to Springer Nature Switzerland AG 2025, K. Jovanović et al. (Eds.): RAAD 2025, MMS 190, pp. 505–512, 2025. https://doi.org/10.1007/978-3-032-02106-9_56
- [6] T. Meneweger, D. Wurhofer, V. Fuchsberger, and M. Tscheligi, “Work ing together with industrial robots: Experiencing robots in a production environment,” in *Proceedings of the 2015 IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Kobe, Japan, August 2015.
- [7] Karabegović, I., Husak E., Vojić S., Karabegović E., Mahmić M., 2025, Integration of robotic technology and human creativity: from automation to humanization, *International Scientific Conference, Industry of the future: From industry 4.0 to Industry 5.0 – Integration of Humans and Technology*, 2-3 October 2025, Academy of Sciences and Arts of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina, Special Editions, Vol.CCXX, Department of Technical Sciences, Vol.24. pp:3-28; DOI: 10.5644/PI2025.220.01
- [8] Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 18, 20–23. <https://doi.org/10.1016/j.mfglet.2018.09.002>

- [9] El Zaatari, S., Marei, M., Li, W., & Usman, Z. (2019). Cobot programming for collaborative industrial tasks: An overview. *Robotics and Autonomous Systems*, 116, 162–180. <https://doi.org/10.1016/j.robot.2019.03.003>
- [10] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, “Artificial cognition for social humanrobot interaction: An implementation,” *Artificial Intelligence*, vol. 247, pp. 45–69, 2017.
- [11] K. Darvish, F. Wanderlingh, B. Bruno, E. Simetti, F. Mastrogiovanni, and G. Casalino, “Flexible human-robot cooperation models for assisted shop-floor tasks,” *Mechatronics*, vol. 51, pp. 97–114, 2018.
- [12] F. Chen, K. Sekiyama, F. Cannella, and T. Fukuda, “Optimal subtask allocation for human and robot collaboration within hybrid assembly system,” *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 4, pp. 1065–1075, October 2014.
- [13] B.-J. You, M. Hwangbo, S.-O. Lee, S.-R. Oh, Y. D. Kwon, and S. Lim, Development of a home service robot 'ISSAC', in *IEEE/RSJ International Conference on Intelligent Robots and Systems (2003)* pp. 2630-2635.
- [14] IFR Statistical department, *World Robotics 2019*.
- [15] Arenas, J., Beleno, R., and Moreno, R. Deep convolutional neural network for hand gesture recognition used for human-robot interaction. *Journal of Engineering and Applied Sciences* 12, Specialissue11 (2017), 9278–9285. 1.
- [16] De Luca, A., and Flacco, F. Integrated control for pHRI: Collision avoidance, detection, reaction and collaboration. In *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)* (June 2012), pp. 288–295. 200.
- [17] Karabegović I., Husak E., Vojić S., Karabegović E., Mahmić M., 2024, Trend Innovation of Artificial Intelligence and Robotic Technology: Implementation in Advanced Robotic Systems, 3rd International Scientific Conference, Artificial Intelligence in Industry 4.0: The future that comes true, 3th-4th October 2024, Academy of Sciences and Arts of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina, Special Editions, Vol. CCXXV, Department of Technical Sciences, Vol. 22. pp:179-200; DOI: 10.5644/PI2024.215.10 <https://bastina.anubih.ba/items/064b8e05-7bac-4b30-a717-5cab97b2de2b>
- [18] M. Ito and J. Tani, On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system, *Adaptive Behavior* 12(2) (2004) pp. 93-115.
- [19] Keshvarparast, A., Battaia, D., Battaia, O. et al. (2024). Collaborative robots in manufacturing and assembly systems: literature review and future research agenda. *J Intell Manuf* 35, 2065–2118 <https://doi.org/10.1007/s10845-023-02137-w>
- [20] Karabegović, I., Husak E., Vojić S., 2025, Evolucija robotskih priхватnica u kontekstu Industrije 4.0: Tehnološki napredak i primjena u automatizovanom sortiranju i paletizaciji, Časopis za savremena privredna kretanja „Science Infokom“, Izdavač: Vanjskotrgovinska komora Bosne i Hercegovine, 2025 Nov. Vol. 3., pp. 94-107, <https://komorabih.ba/>

Smart Manufacturing Enabled By LLM AI Agents and Retrieval-Augmented Generation

Mijodrag Milošević*¹

Abstract: *As smart manufacturing evolves beyond automated execution toward knowledge-based decision support, the effective integration of multidisciplinary knowledge has become a key challenge. Retrieval-Augmented Generation (RAG)-based Large Language Models (LLMs) provide a practical approach to incorporating external knowledge into decision-making processes. They are increasingly being applied in on-site manufacturing environments, where decision-making typically requires frequent human–AI interactions and rapid responses. This paper presents a framework for applying RAG-based LLM AI agents in smart manufacturing environments and discusses their potential to support intelligent, context-aware decision-making.*

Keywords: *Smart Manufacturing, Artificial Intelligence, Large Language Model – LLM, Retrieval-Augmented Generation – RAG, AI Agent*

1. Introduction

A Large Language Model (LLM) is a type of artificial intelligence (AI) trained on a large-scale data set. It is designed to understand and generate human language by learning statistical patterns from textual data. At its core, an LLM is based on deep learning techniques. In general, the larger and more diverse the training data set, the more advanced its Natural Language Processing (NLP) capabilities. NLP is a branch of artificial intelligence that enables computers to understand, interpret, and generate human language [1]. It is a multidisciplinary field that combines computer science, artificial intelligence, and linguistics. The relationship between these disciplines illustrated in Figure 1. By bridging the gap between human communication and machine understanding, NLP enables everyday technologies like virtual assistants, translation tools, and automated chatbots. NLP combines computational linguistics (rule-based language modeling) with machine learning and deep learning. Instead of rigidly programming a computer with every grammar rule, NLP models are trained on big data sets using neural networks to learn statistical patterns and contextual meanings. In smart manufacturing environments, NLP-based AI models can

*¹University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
E-mail: mido@uns.ac.rs

automate data collection, extract relevant patterns of system and process behavior, and suggest new strategies and solutions [1].

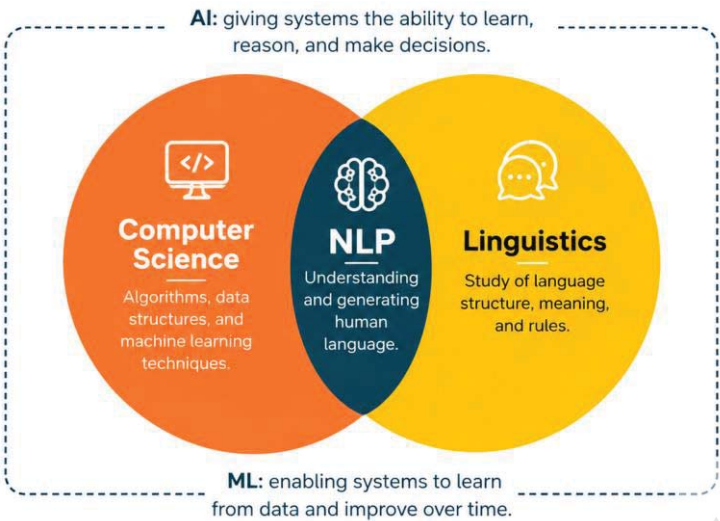


Figure 1. Natural Language Processing (NLP)

In general, AI researchers consider LLMs with billions of parameters to be large language models. Parameters are the internal variables learned during the training process that determine how the model represents and processes information. In general, models with a larger number of parameters exhibit greater representational capacity and are capable of performing more complex language understanding and generation tasks. As LLM architectures continue to evolve, the number of parameters has steadily increased, resulting in significant improvements in model performance and capabilities [2].

2. Context Augmentation Techniques

In today's rapidly evolving AI landscape, organizations are seeking efficient ways to integrate AI solutions that are accurate, secure, and reliable. Although LLMs provide a natural language interface between users and data and are pre-trained on vast amounts of publicly available information, they do not have access to proprietary organizational knowledge [3]. Such knowledge is often organization-specific or domain-specific and is essential for solving specialized tasks. It is typically stored behind APIs, in SQL databases, or within unstructured sources such as PDF documents, technical drawings and tables, as shown in Figure 2.

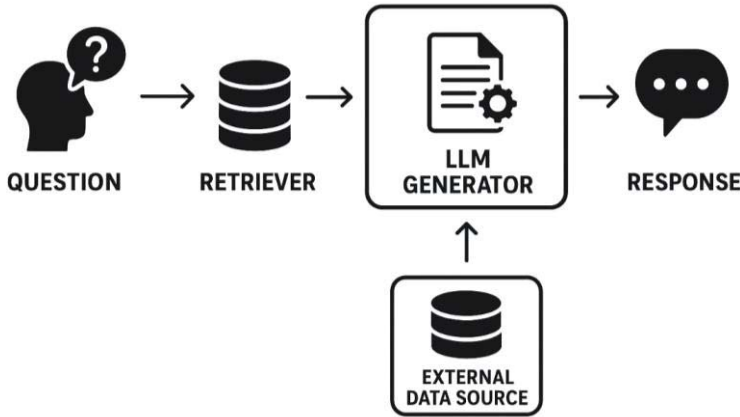


Figure 2. Augmenting the context of an LLM model

Without access to external knowledge, a generative model is limited to producing responses based solely on its parametric knowledge, which is acquired during the training process. Context augmentation enables the integration of external knowledge sources, commonly referred to as non-parametric knowledge, thereby extending the model's effective context and improving the relevance of its responses. Context augmentation is widely used in applications such as:

- Providing more accurate and domain-specific answers to user queries.
- Intelligent chatbots and virtual assistants.
- Document understanding and information extraction.
- Autonomous AI agents capable of reasoning, conducting research, and performing actions.
- Multimodal applications that combine text with images, technical drawings, and other data modalities.

External data sources are typically task-specific and extend beyond the scope of a model's original training data and its parametric knowledge. They are commonly stored in vector databases and may vary considerably in both domain and format. In AI agent systems, this external knowledge base is often regarded as a form of long-term memory that the agent can query whenever additional information is required. Common sources of external knowledge include proprietary organizational databases, legal documents and regulations, scientific literature, technical documentation, and information retrieved from the Web, as shown in Figure 3.

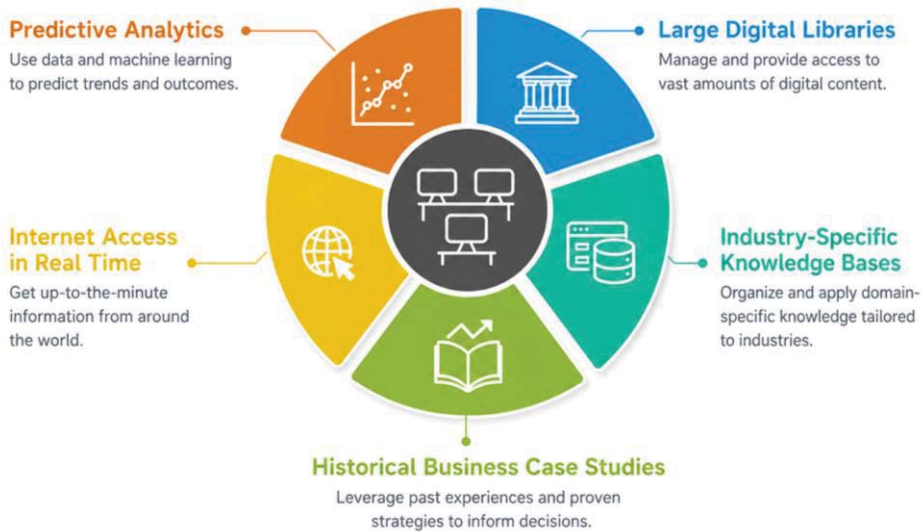


Figure 3. Methods for context augmentation of an LLM model

3. Retrieval-Augmented Generation (RAG)

3.1. RAG Architecture

One of the most important context augmentation methods is Retrieval-Augmented Generation (RAG), which integrates external knowledge with LLMs during the inference phase (Figure 4). This approach enables AI systems to provide up-to-date and accurate information while reducing the risk of hallucinations and incorrect responses. By combining the capabilities of information retrieval with the natural language understanding and generation abilities of LLMs, RAG makes artificial intelligence more reliable, accurate, and practical for real-world applications [4,5]. External data must first be cleaned and transformed into a format that can be processed by the model. Before an LLM can retrieve information from an external knowledge base, the data must be ingested, structured, and indexed. This process is referred to as the indexing (data preparation) phase, which consists of the following steps:

- **Data Loading and Parsing:** Extracting raw data from PDFs, websites, databases, or APIs and converting it into a standardized text format.
- **Chunking:** Dividing large documents into smaller, semantically meaningful segments that fit within the context window of the LLM.
- **Embedding Generation:** Using specialized machine learning models to convert text segments into numerical vector representations (embeddings) that capture their semantic meaning.

- **Storage and Indexing:** Storing these vector representations in a vector database, where they are indexed to enable fast and efficient similarity-based retrieval.

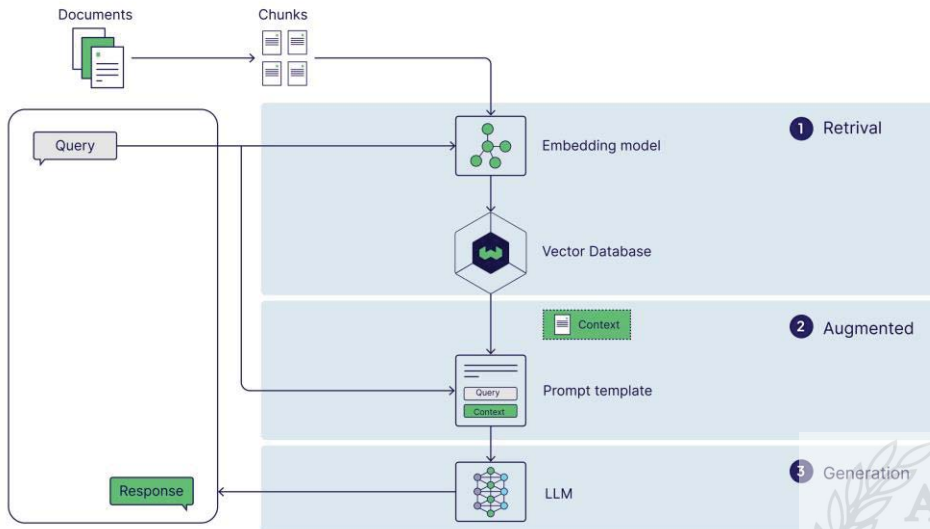


Figure 4. Retrieval-Augmented Generation (RAG) [6]

During the indexing phase, textual or graphical data are transformed from their raw format into vector embeddings through a process known as *vectorization*. Once the embeddings have been generated, they must be stored in a manner that enables efficient retrieval. In most RAG systems, these embeddings are stored in a vector database, which supports fast and efficient similarity-based retrieval for downstream tasks. This external repository serves as a reusable knowledge base from which relevant information is retrieved during the inference phase (Figure 5).

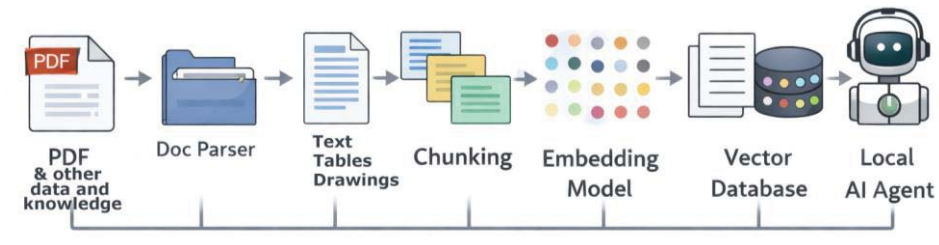


Figure 5. Vectorization of external data and knowledge for context augmentation

After the external data has been encoded and stored, it is ready for retrieval during the inference phase, when the model generates responses to user queries. The inference process consists of three steps: retrieval, augmentation, and generation.

3.2. Retrieval

Based on the user's query, the AI system retrieves relevant information from external knowledge sources, such as documents, databases, or websites. This retrieval process provides the additional context required to generate an accurate response and consists of the following steps:

- **Query Encoding:** The user's query is converted into an embedding within the same vector space as the indexed documents.
- **Vector Search:** The RAG system compares the query embedding with the document embeddings and retrieves the most semantically relevant document chunks.
- **Re-ranking:** An optional specialized model reorders the retrieved chunks according to their semantic relevance, placing the most relevant information at the top of the retrieved context. Although optional, this step is commonly used to improve the overall performance and accuracy of RAG systems.

3.3. Augmentation

The augmentation phase enriches the user's query with relevant information retrieved from external knowledge sources before it is processed by the LLM. During this phase, the retrieved content is selected, filtered, and organized into a structured context that is incorporated into the prompt. This enables the LLM to generate responses based on up-to-date, domain-specific knowledge rather than relying solely on its pre-trained parameters. As a result, the augmentation phase improves response accuracy, reduces hallucinations, and enhances the reliability of the generated output. The original user query is combined with the retrieved context through the following steps:

- **Query Assembly:** The system inserts the retrieved text chunks together with the user's query into a predefined prompt template.
- **Prompt Optimization:** The augmented prompt is refined, compressed if necessary, and formatted to ensure that the LLM effectively interprets and utilizes the provided context when generating the final response.

3.4. Generation

Using the retrieved information, the AI system generates responses that are both accurate and contextually relevant. During the generation phase, the augmented prompt is processed by the LLM, which combines its internal language understanding with the retrieved external knowledge to produce the final response. This enables the model to generate answers that are grounded in up-to-date, domain-specific information while maintaining fluent and coherent natural language generation.

The generation phase focuses on transforming the augmented context into a clear, informative, and well-structured response tailored to the user's query. While the augmentation phase enriches the prompt with relevant external knowledge, the generation phase uses this enriched context to produce the final output. This process typically consists of the following steps:

- **Response Synthesis:** The augmented prompt is submitted to the LLM, which synthesizes the retrieved information with its internal knowledge to generate a coherent and contextually appropriate response.
- **Response Generation:** The LLM produces the final response, leveraging both the retrieved context and its pre-trained knowledge to ensure accuracy, relevance, and fluency.

Figure 6 illustrates an example of an LLM-based AI agent integrated with a RAG architecture for smart manufacturing applications [7].

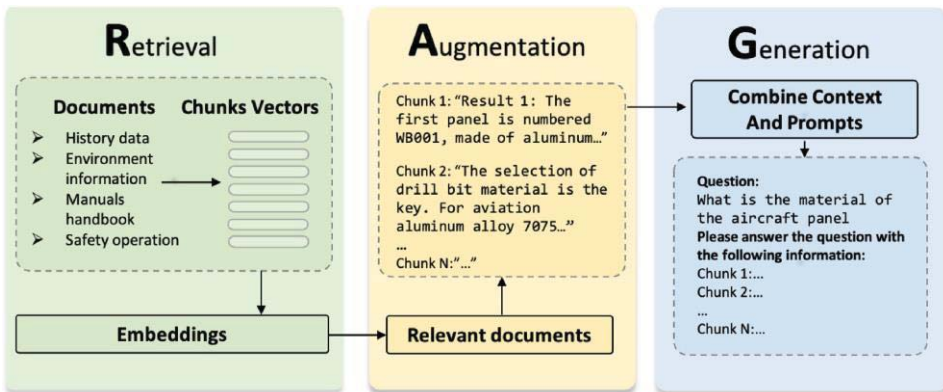


Figure 6. Application of a RAG-based AI agent in Smart Manufacturing framework

4. AI Agentic RAG in Smart Manufacturing

Agentic RAG describes RAG architecture enhanced with AI agents. Unlike conventional RAG systems, AI agents coordinate the retrieval process, manage interactions among system components, and perform additional reasoning and decision-making tasks beyond simple information retrieval and response generation [8]. Rather than relying on a single retrieval step, Agentic RAG employs an iterative workflow in which the agent can retrieve, evaluate, refine, and verify contextual information before generating the final response. This iterative process improves robustness, accuracy, and reliability, making Agentic RAG particularly suitable for complex industrial applications and smart manufacturing environments [9–11].

Figure 7 illustrates the integration of an LLM, a RAG architecture, and an AI agent for the analysis of CAD models and technical documentation. The AI agent orchestrates the entire workflow, beginning with the loading and analysis of the CAD model, followed by the extraction of geometric and technological features [12]. The extracted information is then used to query a vector knowledge base containing embeddings of technical literature, engineering standards, manuals, and previously validated solutions. Finally, the LLM generates a context-aware response that may include recommendations, technical explanations, or references to relevant documentation related to the analyzed CAD model. In this way, the AI agent supports intelligent automation of product analysis and decision-making in a smart manufacturing environment.

As an example of the application of this architecture in smart manufacturing, Figure 8 illustrates a RAG-based anomaly detection system. The system uses RAG to correlate real-time sensor data with historical equipment-specific operational data retrieved from a vector knowledge base. By combining current operating conditions with relevant historical knowledge, the system can identify early signs of anomalies that may not be detected by traditional monitoring methods.

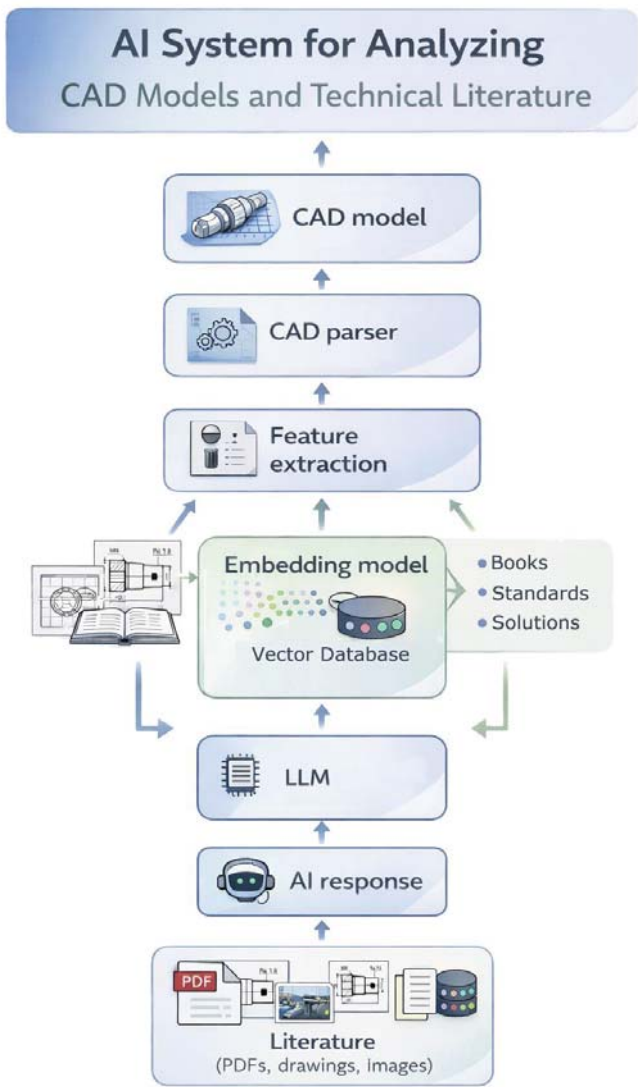


Figure 7. Integration of an LLM–RAG AI agent for CAD model analysis

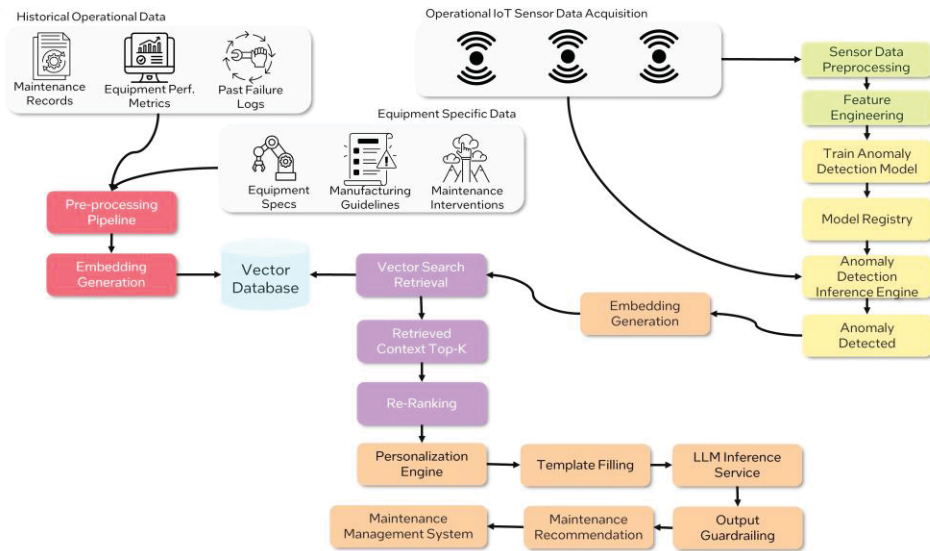


Figure 8. RAG-based Anomaly Detection System [13]

The system continuously monitors equipment data collected from sensors. When an anomaly is detected, it immediately queries the RAG vector knowledge base, which serves as a repository of maintenance records, manufacturer documentation, technical manuals, and historical operational data. This retrieved knowledge is integrated with the LLM to generate context-aware maintenance recommendations and propose an appropriate maintenance strategy.

5. Conclusion

The application of LLMs, enhanced with the RAG context augmentation technique and AI agents, represents a significant step toward the development of intelligent smart manufacturing systems. By integrating advanced natural language understanding, access to relevant domain knowledge, and autonomous decision-making capabilities, these technologies enable more efficient management of manufacturing processes, faster problem solving, improved predictive maintenance, optimized production planning, and enhanced product quality. Although challenges such as reliability, data security, and decision explainability remain, the continuous advancement of LLMs, RAG architectures, and multi-agent systems indicates that these technologies will play a key role in the future of both horizontal and vertical integration in manufacturing, ultimately contributing to the realization of smart manufacturing.

References

- [1] Nabavi, S.F., Garmestani, H., Fekri, F. (2025), AI-powered language models for alloy design and laser-based manufacturing: A review of NLP applications in materials science, *Journal of Manufacturing Processes*, Vol. 156 (Part B), pp. 86–120, <https://doi.org/10.1016/j.jmapro.2025.11.035>
- [2] Annepaka, Y., Pakray, P. (2025), Large language models: a survey of their development, capabilities, and applications, *Knowledge and Information Systems*, Vol. 67, pp. 2967–3022, <https://doi.org/10.1007/s10115-024-02310-4>
- [3] Freire S.K., Wang, C., Foosherian, M., Wellsandt, S., Ruiz-Arenas, S., Niforatos, E. (2024), Knowledge sharing in manufacturing using LLM-powered tools: user study and model benchmarking, *Frontiers in Artificial Intelligence*, Vol. 7, <https://doi.org/10.3389/frai.2024.1293084>
- [4] Zhao, S., Shao, Y., Huang, Y., Song, J., Wang, Z., Wan, C., Ma, L. (2024), Understanding the Fundamental Design Decisions of Retrieval-Augmented Generation Systems, *ACM Transactions on Software Engineering and Methodology*, <https://doi.org/10.1145/3802824>
- [5] Lee, G., Kim, S., Byun, Y., Lee, J., Choi, H.J.(2026), A Review of Knowledge Base Construction Strategies for LLM-based Intelligent Decision Support System in Material Selection and Process Planning, *International Journal of Precision Engineering and Manufacturing-Green Technology*, Vol. 13, pp. 1393-1430, <https://doi.org/10.1007/s40684-026-00890-w>
- [6] <https://weaviate.io/blog/introduction-to-rag>, accessed: June 2026.
- [7] Dong, W., Li, D., Ji, Yuchen, Chen, H., Liu, S., Mab, Z., Hao, F., Ji, Yuqi, Xing, H., Zheng, P. (2025), Towards a next-generation LLM empowered low-code programming industrial robotic system for human-centric smart manufacturing, *Journal of Manufacturing Systems*, Vol. 83, pp. 675–686, <https://doi.org/10.1016/j.jmsy.2025.10.012>
- [8] <https://weaviate.io/blog/what-is-agentic-rag>, accessed: June 2026.
- [9] Li, M., Zhou, Q., Li, W., Qu, T., Yang, M., Jiang, P. (2025), A4PS: Agentic AI-assisted advanced planning and scheduling with large language models for smart manufacturing, *Journal of Manufacturing Systems*, Vol. 85, pp. 207-226, <https://doi.org/10.1016/j.jmsy.2026.01.003>
- [10] Liu, J., Zhou, X., Yu, S., Duan, X., Wang, J., Wen, H., Zhong, J., Dou, Z. (2026), LLM Agent-driven Automated DFA Assessment with Fine-tuning and AAS-based RAG, *Robotics and Computer-Integrated Manufacturing*, Vol. 102, Art. 103325, <https://doi.org/10.1016/j.rcim.2026.103325>
- [11] Zhou, T, Liu, Y., Kumar, M., Zou, L. (2026), Lightweight query-adaptive RAG framework for knowledge support in smart manufacturing,

Journal of Manufacturing Systems, Vol. 86, pp. 774–789,
<https://doi.org/10.1016/j.jmsy.2026.04.016>

- [12] Zhou, J., Camba, J.D. (2025), The status, evolution, and future challenges of multimodal large language models (LLMs) in parametric CAD, *Expert Systems with Applications*, Vol. 282, <https://doi.org/10.1016/j.eswa.2025.127520>
- [13] Alvarez, E., Transforming Manufacturing with RAG: Delivering NextGen Equipment Maintenance, <https://eduand-alvarez.medium.com/>, accessed: June 2026.



A Human-Centric Digital Twin Framework for Industry 5.0 in Transport and Logistics

Osman Lindov*¹

Abstract: *The transition from Industry 4.0 to Industry 5.0 represents a fundamental shift in the design and operation of transport and logistics systems. While Industry 4.0 emphasizes automation, connectivity, and data-driven optimization, Industry 5.0 introduces a human-centric, sustainable, and resilient approach to system development. This paper examines this transition through a conceptual and analytical framework, focusing on the integration of digital twins, artificial intelligence (AI), and human-in-the-loop (HITL) decision-making. A reference architecture is proposed to illustrate a closed-loop cyber-physical system that connects physical logistics operations with sensing technologies, edge/cloud computing, and digital twin-based simulation. The framework incorporates multi-objective optimization, integrating operational efficiency with sustainability metrics and resilience indicators. Furthermore, the study evaluates the role of standardized emissions accounting and circular logistics principles in enabling environmentally responsible transport systems. Selected case studies, including port digital twins, warehouse robotics, and AI-driven route optimization, are analyzed to demonstrate real-world applicability. The results indicate that Industry 5.0 enables more adaptive, transparent, and sustainable logistics systems by aligning technological innovation with human and environmental priorities.*

Keywords: *Industry 5.0, Industry 4.0, smart logistics, digital twin, artificial intelligence sustainable transport, circular logistics, resilience, cyber-physical systems.*

1. Introduction

Industry 4.0 and Industry 5.0 represent complementary paradigms for digital transformation in modern industrial and logistics systems. Industry 4.0 focuses on automation, connectivity, and data-driven optimization enabled by cyber-physical systems, the Internet of Things (IoT), and advanced analytics. In contrast, Industry 5.0 extends this paradigm by emphasizing human-centricity, sustainability, and resilience as core design principles [1], [24].

The transition toward Industry 5.0 introduces a fundamental shift from efficiency-driven optimization to multi-objective system design, where technological performance is balanced with environmental and social considerations. In the context of transport and logistics, this transformation enables the development of

*¹University of Sarajevo, Faculty of Traffic and Communications, Sarajevo, Bosnia and Herzegovina
E-mail: olindov@gmail.com

adaptive, intelligent, and sustainable systems supported by advanced digital technologies such as artificial intelligence (AI), digital twins, and edge/cloud computing.

This paper aims to analyze the transition from Industry 4.0 to Industry 5.0 in transport and logistics and to propose a conceptual reference architecture that integrates technological, human, and sustainability dimensions.

The novelty of this study lies in the integration of Industry 4.0 technologies with Industry 5.0 principles through a unified architecture that incorporates human-in-the-loop governance, sustainability metrics, and resilience-oriented design.

2. Theoretical Background and Literature Review

Industry 4.0 is characterized by the integration of cyber-physical systems, the Internet of Things (IoT), and data-driven optimization to enhance automation and operational efficiency. In transport and logistics, this paradigm enables real-time visibility, automated operations, and optimized decision-making across complex supply chain networks. However, its primary focus on efficiency has revealed limitations related to sustainability, system resilience, and human wellbeing.

In contrast, Industry 5.0 represents a conceptual shift toward human-centric, sustainable, and resilient systems. Rather than focusing solely on efficiency, Industry 5.0 integrates human oversight, environmental responsibility, and long-term system adaptability as fundamental design principles. This transition reflects a broader evolution from purely technological optimization toward a socio-technical approach that balances economic performance with societal and environmental outcomes [1], [24]. Consequently, the transition from Industry 4.0 to Industry 5.0 can be understood not as a technological replacement, but as a strategic reorientation of existing technologies toward human-centered and sustainable objectives.

The evolution from Industry 4.0 to Industry 5.0 has been widely explored in recent academic and policy literature. Industry 4.0 technologies—such as IoT, cyber-physical systems, and data-driven automation—enable real-time monitoring and optimization of industrial and logistics processes, significantly improving operational efficiency, transparency, and responsiveness [7], [8]. However, emerging research highlights the need to complement these capabilities with sustainability-oriented and human-centric approaches.

A key enabling concept in both paradigms is the digital twin, which provides a virtual representation of physical systems for monitoring, simulation, and optimization. Recent studies demonstrate its potential in transport and logistics for improving system performance, predicting disruptions, and supporting real-time

decision-making [22], [25]. In addition, research on supply chain risk and resilience emphasizes the role of digital technologies in mitigating disruption propagation (the ripple effect) and enhancing adaptive system capabilities under uncertainty [23].

Artificial intelligence (AI) and machine learning further extend these capabilities by enabling predictive analytics, dynamic optimization, and autonomous decision-making. However, the increasing reliance on AI introduces challenges related to transparency, accountability, and governance, which are addressed through emerging regulatory frameworks such as the European Union Artificial Intelligence Act [13].

Sustainability has also become a central focus of recent research, particularly in the context of emissions reduction, energy efficiency, and circular logistics. Frameworks such as ISO 14083 and the Global Logistics Emissions Council (GLEC) methodology provide standardized approaches for measuring and managing transport-related emissions, thereby supporting the integration of sustainability into operational decision-making processes [2], [3].

Despite these advancements, existing studies often address technological, environmental, and human dimensions in isolation. This reveals a critical research gap: the lack of integrated frameworks that combine digital technologies with human-centric design, sustainability objectives, and resilience strategies. The present study addresses this gap by proposing a unified reference architecture for Industry 5.0-aligned transport and logistics systems.

3. Methodology

This study adopts a conceptual and analytical methodology to examine the transition from Industry 4.0 to Industry 5.0 within transport and logistics systems. The research approach is structured around three main components: (i) a systematic literature review, (ii) comparative analysis, and (iii) the development of a reference system architecture.

First, a comprehensive review of academic literature, international standards, and policy frameworks was conducted to identify the defining characteristics of Industry 4.0 and Industry 5.0 paradigms. Particular emphasis was placed on sources related to digital transformation, smart logistics, sustainability, and human-centric system design principles [4], ensuring both theoretical and practical relevance.

Second, a comparative analysis was performed to evaluate the differences between Industry 4.0 and Industry 5.0 in terms of technological focus, system objectives, and operational implications for transport and logistics. This analysis enabled the identification of key conceptual shifts, including the transition toward

multi-objective optimization, human–machine collaboration, and resilience-oriented system design.

Third, a reference architecture was developed to illustrate the integration of key enabling technologies, including IoT, edge and cloud computing, digital twins, and artificial intelligence (AI)-driven decision-making within a cyber-physical framework [7]. The proposed architecture incorporates human-in-the-loop (HITL) mechanisms and sustainability considerations, providing a conceptual model for Industry 5.0-aligned transport and logistics systems [23].

Although the study is conceptual in nature, the proposed framework is supported by real-world case studies and established standards, enhancing its applicability to practical transport and logistics systems.

This methodological approach ensures both conceptual rigor and practical relevance for real-world transport and logistics systems.

3.1. System Architecture

The proposed architecture is based on a closed-loop cyber-physical system in which real-world operations continuously inform digital models, while optimized decisions are fed back into the physical system. This bidirectional interaction enables real-time monitoring, adaptive control, and continuous system improvement across transport and logistics processes.

Unlike traditional Industry 4.0 architectures, which primarily focus on efficiency and automation, the proposed model integrates human-in-the-loop (HITL) governance and multi-objective optimization. This ensures that operational performance is balanced with sustainability, resilience, and social impact. The architecture is aligned with Industry 5.0 principles, emphasizing human-centric and environmentally responsible system design. The proposed reference architecture is illustrated in Fig.1.

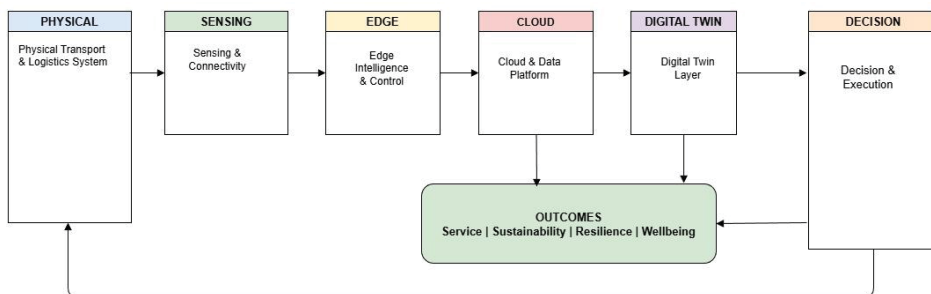


Figure 1. Industry 5.0-oriented digital twin architecture for smart transport and logistics systems integrating physical infrastructure, sensing, edge/cloud computing, AI-based decision-making, and human-in-the-loop governance

The architecture presented in Fig. 1 represents a closed-loop cyber-physical system for transport and logistics aligned with the Industry 5.0 paradigm. It integrates physical assets and infrastructure with sensing technologies, including IoT devices and communication networks, enabling real-time data acquisition and system visibility. Edge intelligence and cloud-based data platforms support distributed processing, data storage, and advanced analytics.

A multi-scale digital twin layer enables real-time synchronization between physical and virtual systems, as well as scenario-based simulation for disruptions, emissions, and policy interventions. Decision-making processes are supported by AI-driven optimization combined with human-in-the-loop oversight, ensuring transparency, safety, and accountability in operational environments [22], [25].

Furthermore, the architecture explicitly links operational layers to outcome-oriented performance indicators, including service efficiency, sustainability (e.g., ISO 14083 and GLEC-based emissions metrics), resilience, and worker wellbeing [2], [3]. This reflects the transition from efficiency-driven Industry 4.0 systems toward human-centric, sustainable, and resilient Industry 5.0 logistics.

The proposed architecture provides a scalable and interoperable framework for implementing Industry 5.0 principles in real-world transport and logistics systems.

4. Results and Discussion

The maturity assessment below is evidence-led (standards maturity + deployment signal from case studies + literature caution where relevant).

Table 1. Technology maturity and adoption signals for transport and logistics

<i>Technology</i>	<i>Typical logistics use cases</i>	<i>Maturity signal</i>	<i>Practical implication</i>
IoT (sensing, telemetry, tracking)	Asset tracking, ETA prediction inputs, cold-chain monitoring	High: strong vocabulary and reference models; widely deployed in ports/vehicles	Focus shifts to data quality, interoperability, and governance rather than “adding sensors”
AI/ML (optimization + prediction)	Dynamic routing, demand forecasting, yard/workforce planning	High in narrow tasks; governance expanding	Requires robust monitoring and human oversight; regulatory expectations growing (EU AI Act [13])

Digital twins (asset + network)	Port/terminal twins, network disruption simulation, urban mobility twins	Medium: standards exist, but SCM deployments often rarer than claimed	Strong value when synchronization, decision integration, and operational ownership exist; otherwise becomes “digital shadow”
Robotics (AMRs/AGVs, cobots)	Goods-to-person picking, pallet moves, yard automation	High in intralogistics; safety standards mature	Best results when designed for augmentation, safety, and job quality; strong change management needed
Edge computing	Low-latency perception, local safety interlocks, V2X-like control	Medium–high: strong standardization; deployment growing	Enables real-time operations and privacy-preserving designs (process data locally)
5G/advanced wireless	Dense sensing, low-latency control, connected mobility	Medium: capability well- defined; rollout uneven by geography	Treat as an enabler (not a prerequisite): architecture should degrade gracefully to LTE/Wi-Fi
Blockchain/DLT	Traceability, document authenticity, multiparty workflows	Mixed: standards vocab exists; network effect risk proven	Consider “minimum blockchain”: align first on standards- based data exchange; use DLT selectively where governance justifies it

(based on [7], [10], [11], [13], [26])

4.1. Human-Centric Design and Human-in-the-Loop Systems

Human-centric design and human-in-the-loop (HITL) systems play a critical role in ensuring transparency, safety, and accountability in AI-driven logistics [4]. By integrating human oversight into automated decision-making processes, these systems enable more reliable, explainable, and ethically aligned operations, particularly in safety-critical and high-impact transport environments [13].

From an Industry 5.0 perspective, human-centricity extends beyond usability toward the design of socio-technical systems that prioritize human wellbeing, trust, and collaboration. In this context, HITL mechanisms support decision validation, system supervision, and intervention in cases of uncertainty or system failure, thereby enhancing operational robustness and governance.

Furthermore, the integration of human expertise with AI-driven systems improves decision quality by combining data-driven insights with contextual knowledge and experience. This collaborative approach enables adaptive and resilient logistics operations while addressing emerging challenges related to ethics, accountability, and workforce transformation.

This approach reflects a shift from automation-centric systems toward collaborative human–AI ecosystems, which are fundamental to the Industry 5.0 paradigm.

4.2. Sustainability and Circular Logistics

Sustainability is integrated as a core operational constraint using standardized emissions metrics and environmental performance indicators. In Industry 5.0-oriented logistics systems, sustainability is not treated as a secondary outcome but as a fundamental design objective, supported by frameworks such as ISO 14083 and the Global Logistics Emissions Council (GLEC) methodology [2], [3].

From a systems perspective, sustainability is operationalized through the incorporation of emissions-aware decision-making, where routing, modal selection, and resource allocation are evaluated based on both efficiency and environmental impact. This enables a shift toward multi-objective optimization, where economic performance is balanced with carbon reduction targets and energy efficiency.

Furthermore, circular logistics principles—such as resource reuse, reverse logistics, and waste reduction—are integrated into supply chain design to enhance environmental efficiency and long-term system viability. These principles extend traditional linear logistics models by incorporating closed-loop flows, thereby supporting sustainable resource management and reducing environmental externalities.

This integration positions sustainability as an active control parameter within logistics systems rather than a passive reporting outcome.

4.3. Resilience and Risk Management

Resilience in transport and logistics systems is addressed through the integration of structured risk management frameworks and digital twin-based simulation approaches. Established standards, such as ISO 31000 and ISO 22301, provide systematic methodologies for identifying, assessing, and mitigating risks across complex supply chain networks [5], [6].

In parallel, digital twin technologies enable scenario-based analysis and real-time evaluation of disruptions, including infrastructure failures, demand fluctuations, and external shocks. By simulating alternative operational strategies and system

responses, digital twins support proactive and data-driven decision-making under uncertainty [23], [25].

The integration of risk management principles with simulation capabilities enhances the system's ability to anticipate disruptions, adapt to dynamic conditions, and recover efficiently from adverse events. As a result, resilience becomes an embedded system property rather than a reactive response, contributing to the stability and long-term reliability of transport and logistics operations.

This approach aligns with the Industry 5.0 paradigm, where resilience is treated as a fundamental design objective alongside efficiency and sustainability.

5. Case Studies and Applications

Case studies and real-world applications demonstrate the practical implementation of Industry 4.0 and Industry 5.0 concepts in transport and logistics. Notable examples include the use of digital twins in port operations, large-scale deployment of warehouse robotics, and AI-driven route optimization systems.

Digital twin implementations in port environments enable real-time monitoring, predictive analytics, and improved coordination of complex operations, contributing to both efficiency and resilience. Similarly, the adoption of warehouse robotics enhances operational productivity while supporting human-machine collaboration, reflecting the human-centric principles of Industry 5.0. In addition, AI-driven route optimization systems improve transport efficiency and reduce environmental impact by minimizing fuel consumption and emissions.

These applications illustrate how advanced technologies can be effectively integrated to enhance operational efficiency, improve sustainability outcomes, and support the development of human-centric and resilient logistics systems. They also demonstrate the transition from isolated technological solutions toward integrated, system-level approaches aligned with Industry 5.0 principles.

Furthermore, these case studies highlight the importance of aligning technological innovation with governance, workforce adaptation, and sustainability objectives to ensure successful implementation.

6. Implementation Roadmap and Governance

A phased implementation approach is essential to ensure scalability, interoperability, and compliance with relevant standards and regulatory requirements in transport and logistics systems. This approach supports the gradual transition from Industry 4.0 capabilities toward Industry 5.0-aligned operations.

The implementation process typically begins with the establishment of a data and interoperability foundation, including standardized data models, integration platforms, and secure data exchange mechanisms. This is followed by pilot deployment of digital twin systems in selected operational contexts, enabling validation of system performance and decision-making processes.

Subsequently, human-in-the-loop (HITL) decision mechanisms are integrated to ensure transparency, accountability, and effective human oversight in AI-driven operations. The final phase involves system scaling, supported by robust cybersecurity frameworks and risk management practices, ensuring system reliability and resilience at scale.

Effective governance frameworks, aligned with standards such as ISO/IEC 42001 [9] and emerging regulatory requirements, play a critical role in managing system complexity, ensuring ethical AI deployment, and maintaining compliance. These frameworks provide the foundation for transparent, accountable, and sustainable system operation in accordance with Industry 5.0 principles.

This phased approach reduces implementation risk and facilitates the alignment of technological deployment with organizational and regulatory readiness.

Industry 5.0 requires KPIs that are **balanced** and **auditable**—not only operational efficiency metrics.

Table 2. KPI framework for Industry 5.0-aligned transport and logistics

<i>KPI category</i>	<i>Example metrics</i>	<i>Why it matters</i>	<i>Notes on measurement sources</i>
Efficiency & service	OTIF, lead time, dwell time at hubs, vehicle utilization, empty-mile ratio	Maintains core logistics performance	Standard ops metrics; link to digital twin interventions
Sustainability & emissions	gCO ₂ e/ton-km, CO ₂ e per shipment/parcel, energy per km, % low/zero-emission miles; circular return rate	Converts sustainability into operational control	ISO 14083 provides common methodology; GLEC offers implementable guidance aligned with ISO 14083

Resilience	Time-to-detect, time-to-recover, service level under disruption, supplier/route concentration indices	Captures network robustness under shocks	Risk and continuity frameworks (ISO 31000/22301) provide process foundations; viability research provides conceptual targets
Human wellbeing	Recordable incident rate, near-miss frequency, ergonomic risk indicators, workload (shift overruns), retention/absenteeism	Aligns with “worker at the center”	ISO 45001 supports OH&S management; ILO highlights OSH policy challenges in digitalized workplaces
Trust & compliance	AI decision audit rate, override rate, model drift incidents, cybersecurity incidents; compliance evidence “completeness”	Prevents unmanaged automation risk	EU AI Act emphasizes oversight, logging, post-market monitoring; NIST AI RMF supports risk framing

(based on [2], [3], [5], [6], [13], [21])

7. Economic Impact and Business Models

Digital transformation fundamentally reshapes cost structures and value creation mechanisms in transport and logistics systems. The integration of advanced technologies, such as artificial intelligence (AI), digital twins, and automation, enables the emergence of new business models based on data-driven services, platform ecosystems, and performance-based logistics.

From an economic perspective, these transformations shift value creation from traditional asset utilization toward data-centric and service-oriented models. Logistics providers increasingly leverage real-time data, predictive analytics, and digital platforms to optimize operations, enhance customer experience, and create differentiated service offerings. This evolution supports the transition from transactional business models to integrated, value-added logistics solutions.

Furthermore, the adoption of Industry 5.0 principles introduces additional economic dimensions, including sustainability-driven value creation and resilience-oriented investments. Improvements in energy efficiency, emissions reduction, and operational robustness contribute not only to cost savings but also to long-term competitiveness and regulatory compliance.

Overall, these transformations lead to increased operational efficiency, reduced costs, and the generation of additional value through enhanced service quality,

sustainability performance, and system resilience, thereby redefining the economic landscape of modern transport and logistics systems.

This transformation highlights the growing importance of intangible assets, such as data, algorithms, and digital capabilities, as primary drivers of competitive advantage in logistics systems.

Table 3. Illustrative cost–benefit scenarios

<i>Scenario</i>	<i>Typical capability scope</i>	<i>Cost drivers</i>	<i>Benefit drivers</i>	<i>Major risks</i>
Incremental Industry 4.0 upgrade	IoT tracking, basic analytics, WMS/TMS integration	Sensors, integration work, data platform	Visibility, fewer delays, incremental efficiency	Data quality failure; “dashboard without decisions”
Integrated Industry 4.0 + Industry 5.0	Add emissions accounting, HITL workflows, safety governance, resilience planning	Governance overhead, training, measurement systems	Measurable emissions cuts, safer operations, improved reliability under shocks	KPI conflicts; change fatigue; misaligned incentives
Transformational Industry 5.0 operating model	Digital twins with closed-loop optimization, robotics at scale, advanced edge/5G segments	Capex for automation, advanced modeling, cyber hardening	Step-changes in throughput + sustainability + resilience; new service revenue	Over-automation, workforce pushback, regulatory noncompliance, cyber risk

(author’s conceptual synthesis)

8. Barriers, Ethics, and Workforce Transformation

The adoption of Industry 5.0 concepts in transport and logistics is accompanied by several challenges, including skills gaps, governance complexities, and ethical considerations. The increasing reliance on artificial intelligence (AI)-driven systems raises critical concerns related to transparency, fairness, accountability, and data privacy, particularly in safety-critical and high-impact operational environments.

From an organizational perspective, the transition toward Industry 5.0 requires significant workforce transformation, including the development of new

competencies in digital technologies, data analytics, and human-machine collaboration. This shift also necessitates changes in organizational culture, training strategies, and job design to support effective interaction between human operators and intelligent systems.

Ethical considerations further extend to issues of algorithmic bias, decision accountability, and the potential impact of automation on employment and working conditions. Addressing these challenges requires a multidisciplinary approach that combines technological innovation with robust governance frameworks, regulatory compliance, and continuous workforce development.

Ensuring a balanced integration of technology and human factors is therefore essential for the responsible and sustainable implementation of Industry 5.0 systems.

9. Conclusion

Industry 5.0 represents a significant evolution in the development of transport and logistics systems, enabling a transition toward human-centric, sustainable, and resilient operations. By integrating advanced digital technologies with human oversight and environmental considerations, Industry 5.0 frameworks support more balanced and adaptive decision-making processes across complex logistics networks. This study has demonstrated that the shift from Industry 4.0 to Industry 5.0 is not merely technological but fundamentally conceptual, redefining system objectives from efficiency-driven optimization toward multi-objective performance that includes sustainability, resilience, and human wellbeing. The proposed reference architecture highlights how digital twins, AI-driven decision-making, and human-in-the-loop mechanisms can be integrated within a closed-loop cyber-physical system to support this transformation. Overall, Industry 5.0 enables the development of intelligent and adaptive logistics systems that align technological innovation with societal and environmental objectives, thereby ensuring long-term system viability and sustainable growth in the transport and logistics sector.

Future research should focus on the empirical validation of Industry 5.0 frameworks through real-world implementations and quantitative performance evaluation. These findings contribute to the growing body of research on Industry 5.0 by providing a structured framework for its implementation in complex logistics environments.

References

- [1] European Commission, *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry*, Brussels, 2021.
- [2] International Organization for Standardization, *ISO 14083:2023 – Greenhouse gases – Quantification and reporting of emissions from transport chains*, 2023.
- [3] Global Logistics Emissions Council, *GLEC Framework for Logistics Emissions Accounting and Reporting*, Version 3, 2024.
- [4] International Organization for Standardization, *ISO 9241-210:2019 – Human-centred design for interactive systems*, 2019.
- [5] International Organization for Standardization, *ISO 31000:2018 – Risk management guidelines*, 2018.
- [6] International Organization for Standardization, *ISO 22301:2019 – Business continuity management systems*, 2019.
- [7] International Organization for Standardization, *ISO/IEC 20924:2024 – Internet of Things (IoT) and Digital Twin vocabulary*, 2024.
- [8] International Organization for Standardization, *ISO/IEC 22989:2022 – Artificial intelligence concepts and terminology*, 2022.
- [9] International Organization for Standardization, *ISO/IEC 42001:2023 – Artificial intelligence management systems*, 2023.
- [10] International Organization for Standardization, *ISO 23247-1:2021 – Digital twin framework for manufacturing*, 2021.
- [11] International Organization for Standardization, *ISO/TS 15066:2016 – Collaborative robot systems safety requirements*, 2016.
- [12] International Organization for Standardization, *ISO 3691-4:2020 – Driverless industrial trucks safety requirements*, 2020.
- [13] European Union, *Regulation (EU) 2024/1689 – Artificial Intelligence Act*, 2024.
- [14] Intergovernmental Panel on Climate Change, *Climate Change 2022: Mitigation of Climate Change – Transport Chapter*, 2022.
- [15] International Energy Agency, *Transport Sector Analysis*, 2023.
- [16] European Environment Agency, *Greenhouse Gas Emissions from Transport in Europe*, 2023.
- [17] Port of Rotterdam, *Smart Infrastructure and Digital Twin Initiatives*, 2022.
- [18] DHL, *Robotics and Automation in Warehousing*, 2024.
- [19] United Parcel Service, *ORION Route Optimization System Overview*, 2023.

- [20] World Economic Forum, *Future of Jobs Report 2025*, 2025.
- [21] International Labour Organization, *The Role of AI and Digitalization at Work*, 2025.
- [22] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2405–2415, 2022.
- [23] M. Ivanov, A. Dolgui, and B. Sokolov, "The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics," *International Journal of Production Research*, vol. 60, no. 1, pp. 1–23, 2022.
- [24] European Commission, "Industry 5.0: A Transformative Vision for Europe," 2023.
- [25] S. Ivanov and A. Dolgui, "Digital supply chain twins: Managing the ripple effect, resilience, and disruption risks," *Annals of Operations Research*, 2023.
- [26] K. Demestichas, E. Adamopoulou, and M. Theologou, "5G, AI and IoT convergence for smart logistics and transport systems," *Future Internet*, vol. 14, no. 6, 2022.



Digitalization and the Circular Economy in Bosnia and Herzegovina: Challenges of the Twin Transition in a Transitional Economy

Munira Šestić^{*1}, Mirha Bičo Ćar¹, Anes Hrnjić¹, Amila Bijedić¹

Abstract: *The transition from a linear economic model to a circular economy (CE) is vital for sustainable development, climate mitigation, and resource resilience. Industry 4.0 digital technologies (DTs), as the Internet of Things (IoT), Artificial Intelligence (AI), Big Data analytics, and block-chain are increasingly recognized as structural enablers of this transformation. They facilitate real-time monitoring, resource tracking, and data-driven decision-making across product lifecycles. However, while developed nations rapidly advance toward "intelligent circular systems," transitional economies like Bosnia and Herzegovina undoubtedly face distinct socio-economic, institutional, and technological barriers. This paper evaluates the intersection of entrepreneurship and digital transitions in operationalizing circular business models within Bosnia and Herzegovina. Through a qualitative literature synthesis and contextual analysis, this short study maps the applications of digital tools in resource optimization, identifies systemic issues - including regulatory fragmentation and digital literacy deficits and proposes a strategic pathway for accelerating the twin digital and green transition. The findings indicate that localized entrepreneurial initiatives, supported by targeted institutional incentives, can effectively bridge existing governance gaps and catalyze sustainable economic restructuring.*

Keywords: *Circular Economy, Digitalization, Bosnia and Herzegovina, Transitional Economies, Industry 5.0, Sustainable Entrepreneurship.*

1. Introduction

Today's global economy is facing unprecedented environmental degradation and resource depletion, driven by an unsustainable linear development paradigm [1, 2]. As a response, the concept of the Circular Economy (CE) has emerged as a regenerative alternative aiming at decoupling economic growth from finite resource consumption. CE principles focus on designing out waste, extending product lifecycles, and regenerating natural systems through narrow, slow, and closed resource loops [2, 3].

In international policy frameworks, such as the European Green Deal, the ecological transition is increasingly intertwined with digital transformation and it is conceptually formalized as the "Twin Transition" [3, 4]. Digital technologies

^{*1} University of Sarajevo, School of Economics and Business, Sarajevo, Bosnia and Herzegovina
E-mail: munira.sestic@efsa.unsa.ba

(DTs) serve as the underlying infrastructure and primary catalyst for scaling circular business models. Technologies such as IoT sensors, cloud computing, and big data enable companies to monitor material flows in real time, predict maintenance needs, optimize reverse logistics, and transition toward advanced Product-Service Systems (PSS) [1, 5].

However, the operationalization of this twin transition is heavily contingent upon institutional maturity, domestic infrastructure, and macroeconomic stability. Consequently, the transition dynamic varies widely across geographical and political boundaries. While advanced economies are establishing comprehensive legislative mandates, such as the EU's Digital Product Passport (DPP), transitional post-conflict economies in the Western Balkan and specifically Bosnia and Herzegovina significantly lag behind in both: digital adoption and ecological compliance [6, 7, 8].

Bosnia and Herzegovina face a complex institutional frame characterized by a highly fragmented administrative structure, an export-dependent yet resource-intensive manufacturing base, low recycling rates, and a dominant linear business mindset. Although the country is bound by the Sofia Declaration on the Green Agenda for the Western Balkans, the integration of Industry 5.0 solutions into environmental management is still on its embryonic level.

This paper explores the synergetic potential between digitalization and the circular economy within the specific transitional context of Bosnia and Herzegovina. The primary objective is to analyze how digital tools can enable circular business models, evaluate the profound systemic barriers hindering this integration, and offer base for strategic recommendations for policymakers and entrepreneurial ecosystem in the future [9].

2. Paradigm Shift

The European Commission has defined three core objectives pursued by Industry 5.0: human-centricity, sustainability, and resilience [10]. These three objectives are not isolated; rather, each conditions and reinforces the remaining two, and it is only through their mutual interaction that the full potential of the paradigm is realized.

Human-centricity shifts the focal point of industrial development, whereby the question "what can a human do with new technology?" is transformed into "what can technology do for the human?" [10]. On business perspective, this implies that employees do not adapt to technology - instead, technology is adapted to the employee. In daily activities, this translates to the deployment of collaborative

robots that assume physically demanding, repetitive, and hazardous tasks, while humans remain focused on creative, analytical, and relational work [11]. Employees are actively involved in the design and implementation of new technological solutions to enable technology to fulfill its transformative role. Consequently, humans are placed at the center of all strategic business decisions, while the benefits of Industry 5.0 are manifested through organizational resilience, longevity, and sustainability [12].

Sustainability, as the second objective, focuses on transitioning away from the linear economic model towards a circular economy based on reuse, remanufacturing, and recycling [11]. Organizations leverage digital technologies to monitor energy consumption and optimize resources, thereby aligning their business strategies with the demands of environmental responsibility [13]. This requirement places society as a whole at a turning point, necessitating a shift in both attitudes and behaviors within the workplace, the enterprise, and ultimately, wider society [13].

Industry 5.0 offers a direct response to the lessons learned from previous global crises by building resilience founded on supply chain diversification, modular and adaptable capacities, and reinforced cyber-security - thus, through the processes and practices of crisis-induced learning, organizational routines are established and adaptive capacities are built [14].

3. Digitalization: A Core Circular Enabler

The interplay of Industry 5.0 and circular business practices recently has generated substantial academic attention, and it is moving beyond viewing digital tools merely as functional IT assets – instead, it is conceptualizing them as structural mechanisms for "intelligent circular systems" that actively drive sustainable development goals [5].

According to the Global Alliance on Circular Economy and Resource Efficiency (GACERE), digitalization fundamentally reshapes resource efficiency across macro, mezo, and micro levels [3]. For instance, comprehensive ICT deployments within smart urban ecosystems have the potential to reduce municipal solid waste generation by up to 20% and optimize industrial resource throughput [1].

The literature categorizes the deployment of digital technologies based on their specific roles across the product and material life cycle phases [1, 5]:

1. Internet of Things (IoT) and Remote Sensing: IoT solutions serve as the sensory layer of the circular economy. By embedding sensors into

physical components, enterprises can track the location, health, and status of assets in real time [1]. This forms the operational baseline for reverse logistics - enabling optimized municipal waste collection routes and accurate asset-tracking in construction and heavy manufacturing [15, 16].

2. Artificial Intelligence (AI) and Machine Learning (ML): AI algorithms are critical for processing the massive datasets generated by IoT infrastructure. In circular frameworks, AI is deployed to forecast waste generation patterns, automate the high-speed sorting of recyclable materials, and predict machine component failures, thereby drastically slowing resource loops through preventative maintenance [1, 5].
3. Blockchain and Distributed Ledger Technology (DLT): Block-chain ensures immutable data transparency and cryptographic traceability across fragmented global supply chains [1]. It provides verifiable proof of material provenance, which is an absolute prerequisite for authenticating recycled content and enforcing Extended Producer Responsibility (EPR) mandates [2, 3].
4. Digital Twins and Product Life-Cycle Software: Digital twins create virtual replicas of physical assets or systems. This allows engineers to simulate so called “wear-and-tear” option, to optimize component disassembly, and to design products with eco-design parameters that facilitate straightforward future remanufacturing [1, 16].
5. Digital Matching Platforms and Marketplaces: Online platforms can dramatically lower transaction costs and eliminate information asymmetry by linking waste-generators directly with material recovery facilities - thereby facilitating industrial symbiosis and peer-to-peer (P2P) asset sharing [1, 2].

A vital interest of the theoretical domain is the evaluation of how national macro-environments influence sustainable commercial activities. Authors Fernandes, Veiga, and Ramadani [4] demonstrate that a country's performance in the circular economy is fundamentally moderated by its National System of Entrepreneurship (NSE) [2].

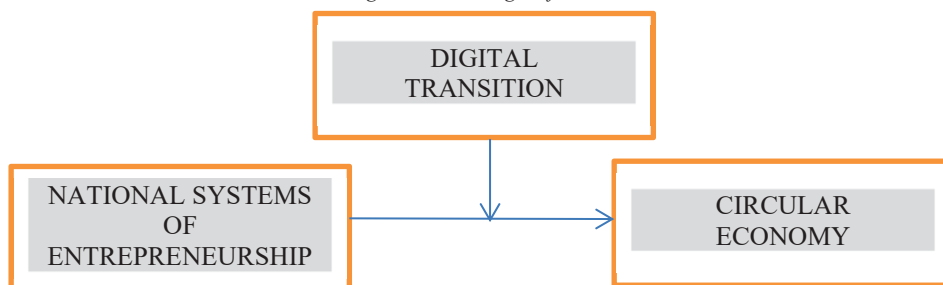


Figure 1. National System of Entrepreneurship [4]

Countries possessing robust entrepreneurial ecosystems, characterized by high-quality institutional setups, deep capital markets, and solid knowledge spillover mechanisms, exhibit significantly higher rates of circular business model innovation. Crucially, their empirical model reveals that the digital transition acts as a powerful moderator that amplifies the positive effects of entrepreneurship on circular outcomes [4].

If digital transition metrics are in advance, entrepreneurial proxies can bypass rigid or inefficient physical institutional infrastructure - utilizing digital platforms to establish agile, resource-efficient supply chains and transition from traditional product-oriented models to use-oriented or result-oriented Product-Service Systems (PSS) [2, 4].

4. Contextual Analysis for Twin Transition: Bosnia and Herzegovina

The paper utilizes an exploratory, qualitative research methodology appropriate for evaluating emerging socio-economic phenomena within constrained data environments. Given the pronounced deficit of localized quantitative metrics regarding digital circularity indexes in the Western Balkans, a multi-step contextual analysis was executed:

1. A review of the literature (2022–2025) - focusing on foundational frameworks regarding intelligent circularity [5], digitalized business typologies [2], and the institutional dynamics of entrepreneurship-led transitions [4].
2. Analysis of institutional reports - strategic development frameworks (including the Federation of BiH and RS environmental and waste management strategies), and commitments under the Western Balkans Green Agenda were analyzed to contrast theoretical digital-circular paradigms against the contemporary socio-economic reality of Bosnia and Herzegovina.

3. Thematic mapping for identify constraints - they were categorized utilizing a modified institutional economics framework to highlight the specific structural friction points inherent in transitional economic systems.

Bosnia and Herzegovina operates within a highly specific socio-economic reality. As a country undergoing protracted economic transition, it simultaneously faces the pressure of harmonizing its legal framework with the EU legislative while managing deep-seated structural inefficiencies.

The application of the "Twin Transition" concept in Bosnia and Herzegovina reveals a significant implementation gap. While the domestic Information and Communications Technology sector (ICT) can be recognizes as highly dynamic, technologically sophisticated, and strongly export-oriented, still it remains fundamentally decoupled from the traditional manufacturing, agricultural, and municipal infrastructure sectors [6, 7, 8, 9, 17].

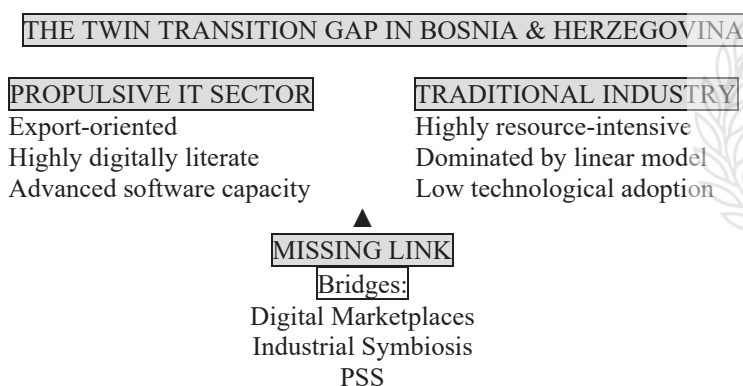


Figure 2. Twin transition gap [6, 7, 8, 9, 17].

The barriers preventing the deployment of digital technologies for circular economy advancement in Bosnia and Herzegovina can be synthesized into four core systemic directions [6, 7, 8, 9, 17]:

A. Institutional and Regulatory Fragmentation

The highly decentralized political governance of Bosnia and Herzegovina (state, entity, cantonal, and district levels) results in an extraordinarily fragmented regulatory environment. Waste management laws and environmental standards are inconsistent across jurisdictions. Crucially, Bosnia and Herzegovina lacks the

penalty economic instruments that make linear business models unviable in European Union, such as high landfill tipping fees, strict enforcement of Extended Producer Responsibility (EPR), and carbon taxes. Furthermore, while digital legislation may exist, operational frameworks for widespread electronic signatures, digital contracting, and cross-institutional open data sharing are either fragmented or under-implemented, stalling the legal deployment of digital material tracking and supply chain transparency.

B. Infrastructure Deficits and Low Industrial Tech Levels

The baseline industrial landscape in Bosnia and Herzegovina is dominated by small and medium businesses (SMBs) engaged in primary or low-value-added secondary processing (e.g., wood processing, metal fabrication, textiles). The vast majority of these firms operate with basic automation profiles. Digital tools are confined to standard administrative or accounting software, with virtually zero penetration of IoT, predictive data analytics, or AI-driven quality control. On a macro-structural level, the delayed rollout of advanced telecommunications infrastructure, such as commercial 5G networks, directly restricts the deployment of distributed sensor networks required for smart utility grids and intelligent municipal waste collection systems.

C. Financial Constraints and Distorted Investment Incentives

Transitioning to circular business models supported by digital infrastructure asks for significant primary capital expenditure. Domestic SMBs in Bosnia and Herzegovina operate within narrow liquidity margins and face high borrowing costs. Local commercial banks possess limited expertise in evaluating "green" or "circular" investments, often classifying digital-circular transformations as high-risk ventures. Furthermore, state-level or entity-level subsidies and grant mechanisms are highly constrained, and they are lacking the scale and strategic continuity needed to de-risk corporate investments in advanced resource-recovery technologies or custom software development.

D. The Interdisciplinary Skills Deficit

The educational architecture in Bosnia and Herzegovina remains partial in its segments. Universities produce technical specialists within isolated domains (e.g., traditional engineering, computer science, or business management), but fail to offer interdisciplinary curricula that merge environmental science with digital systems engineering and entrepreneurship. Consequently, there is an "acute" domestic shortage of industrial data analysts, circular supply chain managers, and

eco-design specialists. This deficit of technical competency is compounded by a modest level of digital awareness among executives, who frequently view sustainability initiatives as regulatory burdens or peripheral public-relations costs rather than core drivers of long-term operational efficiency and global competitiveness.

5. Discussion for Strategic Framework

The intersection of the constraints of digital transition and the circular economy within Bosnia and Herzegovina reveals that while institutional architectures on the macro level stagnate, entrepreneurial activity on the micro-level can catalyze "blastoff" – simply because they have to [3, 4]. By aligning the domestic software engineering capacity with traditional manufacturing, Bosnia and Herzegovina can systematically "bypass" physical institutional gaps through targeted digital interventions.

I - Digital Industrial Symbiosis Platforms

A primary strategic mechanism for Bosnia and Herzegovina may involve the development of automated, localized resource-matching platforms [2, 4]. In key national industrial sectors, such as wood processing and metal fabrication, immense volumes of manufacturing byproducts (e.g., sawdust, timber offcuts, metal scraps) are currently discarded as municipal waste or burned inefficiently for heat.

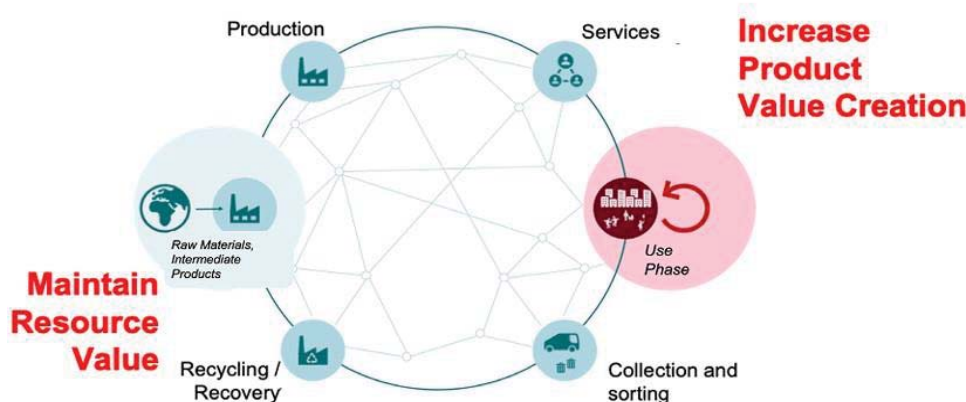


Figure 3. The Circular Economy as a Networked System [2]

By deploying decentralized web-based marketplaces, local software firms can connect these disparate small and medium enterprises (SMEs). This could eliminate localized information asymmetries, also structurally optimizes supply routes, and allows industrial byproducts from one factory to become secondary raw inputs for another, effectively operationalizing industrial symbiosis at a minimal operational cost [1, 5].

II - Product-Service Systems (PSS) and Asset Lifecycle Management

Furthermore, the domestic economy can benefit significantly from transitioning toward use-oriented and result-orientation in Product-Service Systems (PSS) powered by basic digital tracking layers [2]. Rather than engaging in the outright capital-intensive sale of physical machinery, industrial equipment, or eco-efficient lighting installations, domestic manufacturers and distributors can pivot toward leasing and functional performance models [1].

Embedding low-cost microcontrollers and IoT diagnostic sensors into distributed machinery enables manufacturers to maintain ownership and remotely monitor real-time operational efficiency [15, 16]. This operational data-stream facilitates predictive and preventative maintenance protocols, directly extending the functional lifecycle of national industrial assets, reducing localized environmental friction, and aligning commercial incentives with resource conservation rather than raw consumption volume [4, 5].

6. Policy Recommendations

To operationalize the twin transition within Bosnia and Herzegovina, a synchronized multi-tiered policy mix could be implemented across entity and cantonal structures:

1. Regulatory standardization and Open Data frameworks: Align national environmental data frameworks with the EU's Digital Product Passport (DPP) paradigms [3]. Municipalities could standardize electronic waste tracking registries and open up environmental monitoring APIs, enabling external technology entrepreneurs to build commercial solutions as public utility infrastructure [2].
2. Targeted financial risk mitigation: Establish dedicated state-backed and donor-funded green innovation credit lines [4]. Commercial banking institutions require technical assistance to evaluate digital-circular business models, reducing collateral requirements for SMBs that invest explicitly in Industry 5.0 resource-tracking infrastructure [3, 15].

3. Interdisciplinary curriculum integration: Transformation of traditional university silos by establishing structural interdisciplinary education and professional training programs that merge computer science, organization design, environmental engineering, and sustainable entrepreneurship [4, 16].

7. Conclusion

The transition toward a circular economy in Bosnia and Herzegovina cannot be achieved using analog methodologies or isolated ecological mandates [17]. Digital technologies provide the analytical substrate, connectivity, and data transparency required to transform highly fragmented, linear supply systems into resilient, intelligent circular structures [1, 5].

While institutional and infrastructure deficits present ongoing systemic friction in Bosnia and Herzegovina, the national entrepreneurial ecosystem moderated by targeted digital deployment has the strong capacity to spearhead economic “restructuring” [4]. By actively linking its thriving digital sector with legacy manufacturing industries, Bosnia and Herzegovina can forge an autonomous path toward its green future and long-term integration into competitive European value chains [6, 7, 8].

References

- [1] Han, Y., Shevchenko, T., Yannou, B., Ranjbari, M., Shams Esfandabadi, Z., Saidani, M., Bouillass, G., Bliumska-Danko, K., and Li, G. (2023), Exploring How Digital Technologies Enable a Circular Economy of Products, *Sustainability*, 15(3): 2067. <https://doi.org/10.3390/su15032067>
- [2] One Planet Network / Circular Economy Research Division at the Wuppertal Institute .(2024), Digitalization of Circular Business Models: Digital solutions enabling the transition towards circular economy, *Coalition for Digital Environmental Sustainability (CODES)*
- [3] GACERE. (2025), Circular Economy and Digitalization - Working Paper, *Global Alliance on Circular Economy and Resource Efficiency, European Union*
- [4] Fernandes, C.I., Veiga, P.M., and Ramadani, V. (2025), Entrepreneurship as a transition to the circular economy, *Environment, Development and Sustainability*, 27 : 24491–24503. <https://doi.org/10.1007/s10668-023-03513-5>

- [5] Rodriguez Romo, M.F., Siva, V., Papakostas, N., Ghadimi, P., and Hargaden, V. (2025), Circular economy and digital technologies: A research agenda for intelligent circular systems, *Resources, Conservation & Recycling Advances*, 28 : 200292. <https://doi.org/10.1016/j.rcradv.2025.200292>
- [6] Krekić, V. (ur.). (2025), Zbornik radova / Naučno/znanstveno-stručna konferencija “Digitalizacija u zelenoj ekonomiji”, *Federalno ministarstvo obrazovanja i nauke/znanosti; Zaklada za inovacijski i tehnološki razvitak, INTERA TP*, Mostar
- [7] Mastilo, Z., Bosna, J., and Puška, A. (2026), Digitalizacija kao pokretač tranzicije ka cirkularnoj ekonomiji: Studija balkanskih zemalja. In: *Proceedings: MIKET*, pp. 527-535. Bosnia and Herzegovina. <https://doi.org/10.5281/zenodo.18037769>
- [8] Šestić, M. and Bičo Ćar, M. (2022), Izazovi ženskog preduzetništva u kontekstu 4IR: perspektiva Bosne i Hercegovine. In: Dž. Kulović (ur.), *Proceedings: Stanje i perspektive digitalne transformacije poslovanja zemalja Zapadnog Balkana, IX Međunarodna naučno-stručna konferencija Razvoj poslovanja 2021*, pp. 407–421. Ekonomski fakultet, Univerzitet u Zenici, BiH. https://ef.unze.ba/wp-content/dokumenti/Zbornik_BDC_2022.pdf
- [9] Pilav-Velić, A., Mangafić, J., Jahić, H., Hrnjić, A., Hadžiahmetović, N., and Bijedić, A. (2025), Zelena taksonomija: Smjernice za održivu budućnost BiH [Green Taxonomy: Guidelines for a Sustainable Future of BiH]. In: A. Pilav-Velić (ed.), *School of Economics and Business*, University of Sarajevo, Sarajevo
- [10] Breque, M., De Nul, L. and Petridis, A. (2021). *Industry 5.0: Towards a sustainable, human-centric and resilient European industry*, Publications Office of the European Union.
- [11] Guerini, G., and Nelissen, G. (2024). European Economic and Social Committee: *Industry 5.0 – how to make it happen (Initiative)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52024IE1285>
- [12] Saniuk, S., Grabowska, S. and Straka, M. (2022). Identification of Social and Economic Expectations: Contextual Reasons for the Transformation Process of Industry 4.0 into the Industry 5.0 Concept. *Sustainability*, 14(3), 1391. <https://doi.org/10.3390/su14031391>
- [13] Crnjac Žižić, M., Mladineo, M., Gjeldum, N. and Celent, L. (2022). From Industry 4.0 towards Industry 5.0: A Review and Analysis of Paradigm Shift for the People, Organization and Technology. *Energies*, 15(14), 5221. <https://doi.org/10.3390/en15145221>
- [14] Fiksel, J., Croxton, K. L. and Pettit, T. J. (2015). From risk to resilience: learning to deal with disruption. *MIT Sloan Management Review*, 56(2), 78-86.

- [15] Wynn, M. and Jones, P. (2022), Digital Technology Deployment and the Circular Economy, *Sustainability*, 14(15) : 9077. <https://doi.org/10.3390/su14159077>
- [16] Keles, C., Cruz Rios, F., and Hoque, S. (2025), How Can Digital Technologies Enable Circular Economy in the Construction Industry? A Review of Lifecycle Applications, Integrations, Potential, and Limitations, *Preprints.org*. doi:10.20944/preprints202501.0774.v1
- [17] Abaspahić, H., Suljić, V., Garić, M., and Krupić, S. (2022), Bijeli papir cirkularne ekonomije u Bosni i Hercegovini [White paper on circular economy in Bosnia and Herzegovina], *Center for Policy and Governance*, Sarajevo



Industry 4.0 and Green Chemistry Integration for Sustainable Smart Chemical Manufacturing

Amra Bratovčić^{*1}

Abstract: *The chemical industry is undergoing a rapid digital transformation driven by Industry 4.0 technologies and data-driven innovation. Evolving from Cyber-Physical Production Systems, Industry 4.0 integrates Internet of Things (IoT), artificial intelligence (AI), machine learning (ML), big data analytics, robotics, and cloud computing to enable smart manufacturing systems with real-time monitoring, automated control, predictive maintenance, and enhanced decision-making. These capabilities significantly improve process efficiency, safety, and operational flexibility. In parallel, Green Chemistry principles enhance sustainability by promoting pollution prevention, resource efficiency, and the use of safer and renewable feedstocks. The integration of smart manufacturing with Green Chemistry has created sustainable production systems where environmental and economic performance are jointly optimized. Digital twins and process simulation further support virtual optimization, reducing experimental costs and accelerating process development. Beyond manufacturing, AI-, ML-, and modeling-driven approaches are transforming chemical research, particularly in catalysis, through high-throughput screening, predictive modeling, and multiscale analysis. These methods, supported by digital platforms, are advancing applications in hydrogen production, CO₂ conversion, methane reforming, polymer recycling, and photocatalysis, enabling data-centric materials discovery. At the system level, smart manufacturing has evolved into an intelligence-driven paradigm in which continuous data streams enable forecasting, optimization, and anomaly detection for self-adaptive production networks. Despite these advances, barriers such as implementation cost, cybersecurity risks, integration challenges, and skill shortages persist. Overall, the convergence of Industry 4.0, Green Chemistry, and data-driven catalysis is reshaping the chemical industry toward efficient, resilient, and sustainable systems, positioning chemical engineers as key drivers of digital and environmental innovation.*

Keywords: *smart manufacturing; chemical industry; green chemistry; artificial intelligence; machine learning; catalysis; sustainability; process optimization*

1. Introduction

The term *Industry 4.0* was introduced at the Hannover Fair in 2011 and was initially associated with Cyber-Physical Production Systems (CPPS), representing the application of Cyber-Physical Systems (CPS) concepts within manufacturing

^{*1}University of Tuzla, Faculty of Technology, Tuzla, Bosnia and Herzegovina
E-mail: amra.bratovcic@untz.ba

environments [1]. Over the following decade, Industry 4.0 significantly advanced the development of intelligent manufacturing by integrating technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), and Digital Transformation (DT) [2]. A central objective of this transformation has been to minimize human involvement in production processes through increased automation and system optimization. The widespread adoption of automation and digital technologies under Industry 4.0 has not only reshaped manufacturing operations but has also extended its influence into related sectors and everyday life. The emergence of Industry 5.0 has shifted the focus of digital manufacturing toward sustainability, resilience, and human-centered design, creating new expectations for human-machine interaction (HMI) within human-centric smart manufacturing (HCSM). In manufacturing environments, HMI can generally be described through four stages: (1) sensing and hardware components that gather environmental and user-input information, (2) data processing systems that transform signals into meaningful data, (3) communication mechanisms that transfer data to processing units, and (4) interaction and collaboration between humans and machines. Among these stages, sensor technologies and data utilization have become critical areas for advancing HMI performance. This importance stems not only from ongoing technological innovations and emerging research but also from the influence of the human-centered principles introduced by Industry 5.0 [3].

2. Modern Chemical Engineering

Chemical engineering has traditionally been a fundamental driver of industrial progress, continually adapting to scientific advances and societal needs. In the current era, it operates at the intersection of two major paradigms: Industry 4.0, which represents the digital transformation of manufacturing systems, and Green Chemistry, which promotes environmentally responsible design of chemical products and processes. The convergence of these approaches within the framework of smart manufacturing enables the development of highly advanced and sustainable production systems, marking a shift toward more efficient and environmentally conscious industrial practices. Industry 4.0 incorporates a range of enabling technologies, including artificial intelligence (AI), the Internet of Things (IoT), robotics, big data analytics, cloud computing, and cyber-physical systems, all of which contribute to the integration of digital and physical manufacturing environments (Fig. 1). In chemical engineering applications, this transformation is reflected in smart manufacturing systems that connect equipment and processes through real-time data acquisition, advanced analytics, predictive maintenance, and simulation-based optimization. IoT-enabled sensors distributed across chemical plants continuously measure critical process variables such as temperature, pressure, and composition, generating continuous data

streams that support predictive modeling and early detection of process deviations. These data-driven approaches improve operational safety, efficiency, and flexibility while enabling chemical engineers to employ digital twins - high-fidelity virtual representations of physical systems to simulate, analyze, and optimize chemical processes without interrupting actual operations. This capability reduces experimental costs and risks while enhancing process performance, product yield, and responsiveness to market demands. In parallel, Green Chemistry provides a complementary framework focused on minimizing environmental impact by designing chemical processes that reduce or eliminate hazardous substances at their source rather than relying on downstream treatment (Fig 2). While chemists develop safer and more sustainable reaction pathways at the laboratory scale, chemical engineers play a critical role in scaling these innovations by applying principles of transport phenomena, reaction engineering, thermodynamics, and economic analysis. Together, these integrated approaches support the development of chemical manufacturing systems that are not only technologically advanced but also safer, more sustainable, and economically efficient [4].



Figure 1. Industry 4.0 in chemical engineering

2.1 Green Chemistry Principles and Their Integration with Industry 4.0 for Sustainable Chemical Manufacturing

The Twelve Principles of Green Chemistry, introduced by Paul Anastas and John Warner, provide a systematic framework for designing chemical processes that minimize pollution, reduce hazardous substance use, enhance energy efficiency, and promote the use of renewable resources. Unlike traditional end-of-pipe treatment approaches, Green Chemistry emphasizes pollution prevention at the molecular level through strategies such as waste minimization, atom economy,

safer reagents, and renewable feedstocks. This shift from reactive environmental management to proactive process design has led to improved environmental performance, regulatory compliance, and economic benefits via reduced raw material and energy consumption.

The integration of Green Chemistry with Industry 4.0 technologies establishes a synergistic framework for sustainable and intelligent chemical manufacturing. IoT-based sensors enable continuous monitoring of process conditions and rapid detection of inefficiencies, while artificial intelligence supports data-driven optimization of operating parameters to reduce emissions and resource use. Digital twin technologies further enhance sustainability by enabling virtual testing of greener reaction pathways and alternative materials prior to industrial implementation, thereby lowering experimental risk and cost.

At a broader level, Industry 4.0 enhances sustainability across supply chains through improved digital connectivity, traceability, and resource management, supporting circular economy practices and reduced carbon emissions. Industrial applications demonstrate that real-time data analytics and automation can significantly improve efficiency, safety, and environmental performance while maintaining economic viability. Overall, the convergence of smart manufacturing and Green Chemistry delivers substantial technical, environmental, and economic advantages, while also reshaping the role of chemical engineers toward more interdisciplinary, data-driven, and sustainability-focused responsibilities.

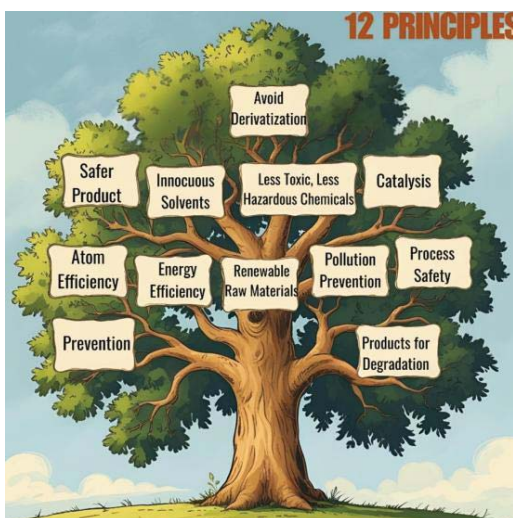


Figure 2. The 12 Principles of Green Chemistry for sustainable chemical design.

Building on this, the evolving role of chemical engineers demands continuous upskilling and adaptation, with increasing emphasis on data analytics, artificial

intelligence, process modeling, and sustainability-driven engineering within both academic curricula and industrial training programs. As digital and green technologies become further integrated, future chemical manufacturing systems are expected to progress toward fully autonomous and self-optimizing plants capable of real-time operational adjustment to maximize efficiency while minimizing environmental impact.

At a broader scale, industrial ecosystems are likely to adopt more collaborative, data-centric frameworks aligned with Industry 5.0 principles, prioritizing human-centric design, sustainability, and ethical responsibility in manufacturing. In parallel, regulatory structures are anticipated to evolve in response to these technological advances, facilitating faster and more structured adoption of smart and sustainable industrial practices. Within this context, chemical engineers will assume increasingly strategic roles, contributing not only technical expertise but also leadership in sustainability governance, digital transformation, and industrial innovation [5].



Figure 3. Framework for smart factory implementation in chemical plants [5].

The chemical manufacturing industry is experiencing a major shift driven by the widespread implementation of smart manufacturing and Industry 4.0 technologies. The deployment of digital systems such as the Internet of Things (IoT), artificial intelligence (AI), machine learning, big data analytics, automation, and robotics has significantly enhanced operational performance by improving process stability, adaptability, and production efficiency (Figure 3). These technologies enable continuous monitoring of industrial operations, advanced data analysis, and automated control strategies that support optimized process conditions, reduced equipment failures, improved product consistency, and lower resource and energy consumption.

In addition, digital transformation strengthens supply chain integration by enabling real-time tracking of materials, inventory, and system performance, which improves coordination and increases resilience against disruptions. However, the implementation of such advanced systems is associated with several challenges, including high investment costs, cybersecurity risks, integration with existing infrastructure, and a shortage of multidisciplinary expertise combining chemical engineering and digital competencies. Addressing these limitations requires targeted investment in workforce development, interdisciplinary training, and strategic talent acquisition.

Overall, the adoption of smart technologies in chemical production not only enhances operational safety, efficiency, and resource management but also accelerates innovation in sustainable manufacturing practices. As industrial digitalization continues to advance, smart factories are expected to become a key driver of competitiveness and sustainability in the chemical sector, enabling companies to meet evolving economic, environmental, and regulatory demands [5].

2.2 Digital Transformation of the Chemical Industry

The chemical industry is experiencing a rapid and multidimensional transformation driven by the adoption of smart manufacturing within the Industry 4.0 framework. As a foundational supplier to sectors such as agriculture, pharmaceuticals, automotive, and construction, it operates under continuous pressure to enhance efficiency, adaptability, and competitiveness while simultaneously meeting increasingly stringent sustainability requirements.

In this context, smart manufacturing integrates advanced digital technologies—including the Internet of Things (IoT), artificial intelligence (AI), machine learning, computer vision, deep learning, and big data analytics—into production systems. This integration enables real-time monitoring, automated process control, and intelligent decision-making. As a result, manufacturers can achieve rapid data acquisition and advanced analytics-driven optimization of process parameters, improved resource utilization, and reduced operational inefficiencies. By reducing dependence on manual data handling and minimizing human error, these technologies significantly improve process reliability and productivity. In addition, AI-supported and automated systems enhance energy efficiency through optimized equipment operation while ensuring consistent product quality. Predictive maintenance strategies, enabled by advanced analytics, further allow early detection of equipment degradation, thereby reducing downtime and lowering maintenance costs. Moreover, data-driven insights into market trends and consumer preferences support continuous innovation and faster adaptation to evolving industrial demands. Collectively, these capabilities enhance production flexibility and strengthen responsiveness to dynamic market and operational

conditions. However, successful deployment depends on the integration of reliable sensing technologies, robust communication networks, and resilient digital infrastructure within existing industrial environments. Overall, smart manufacturing is emerging as a key enabler of sustainability, operational resilience, and long-term competitiveness in the chemical sector [6].

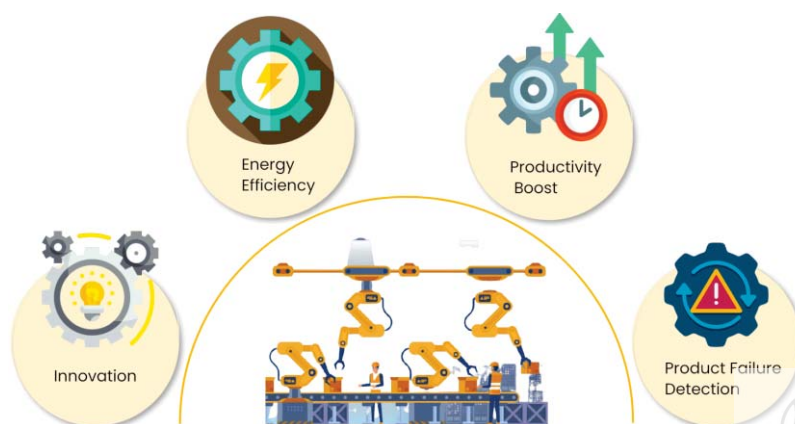


Figure 4. Smart manufacturing in the chemical industry enables productivity gains, energy efficiency, innovation, and early failure detection through AI- and IoT-driven systems.

Building upon these advancements, Industry 4.0 is also being widely implemented across chemical, pharmaceutical, and food production systems, where it enhances automation, process control, and operational intelligence. This results in improved process reliability, repeatability, supervision, fault detection, maintenance efficiency, and product quality (Figure 4). The effectiveness of such systems relies on the integration of high-performance industrial equipment with interconnected digital technologies capable of ensuring seamless data acquisition, integration, and centralized storage within unified platforms. Centralized monitoring systems provide real-time visibility of equipment status, support rapid fault diagnosis, and consolidate operational data for informed decision-making. Furthermore, these technologies improve process monitoring, streamline recipe management, and enable faster product transitions without disrupting ongoing production. As a result, Industry 4.0 adoption leads to enhanced productivity, reduced energy consumption, improved material handling, optimized asset utilization, and lower operational costs, thereby strengthening overall industrial performance [7].

While Industry 4.0 establishes the digital backbone of modern industrial systems through automation, connectivity, and data-driven control, its influence extends beyond manufacturing operations into scientific research and materials

innovation. In particular, the integration of artificial intelligence (AI), machine learning (ML), and computational modeling with experimental catalysis is reshaping the development of advanced materials and chemical processes.

Engineered heterogeneous catalysts, including supported metals, zeolites, defect-engineered structures, and multicomponent systems, exhibit enhanced activity, selectivity, and stability due to precisely controlled active sites and tailored porosity. Data-driven methodologies enable predictive modeling, high-throughput screening, mechanistic insight generation, and rational catalyst design by linking synthesis parameters, structural properties, and performance outcomes across multiple scales. These approaches are widely applied in areas such as carbon dioxide conversion, methane reforming, hydrogen production, polymer recycling, and photocatalysis, supported by digital platforms including PHOTOREAC, QMOF, and PhotoCatDB, which facilitate the transition from laboratory-scale research to industrial-scale applications. Furthermore, hybrid frameworks that combine mechanistic understanding with data-driven learning enhance interpretability, improve predictive accuracy, and optimize process performance. Collectively, these developments represent a shift toward data-centric catalysis, accelerating materials discovery, enabling sustainable chemical technologies, and reinforcing the essential role of human expertise in guiding responsible AI-driven innovation in chemical engineering [8].

Building on this digitally enabled and data-centric foundation, smart manufacturing is evolving toward more integrated, adaptive, and intelligence-driven systems. In this paradigm, sensing technologies, computational analytics, and adaptive control mechanisms operate synergistically to create continuous feedback loops that transform operational data into actionable process improvements.

Within this framework, modern production systems are designed to handle variability in demand, resource availability, and process conditions through dynamic reconfiguration and adaptive control strategies. Data collected from interconnected industrial assets is systematically processed (Fig. 5) to support forecasting, anomaly detection, and performance optimization at both equipment and system levels. This transition toward intelligence-driven manufacturing enables more precise coordination of operations, improved resource allocation, and enhanced stability across production networks.

As industrial systems become increasingly interconnected, emphasis is also shifting toward the integration of optimization models and quantitative decision-support tools that can further improve efficiency and sustainability. These approaches allow manufacturers to evaluate alternative production strategies, minimize operational costs, and ensure consistent product quality under varying conditions. In this way, digital manufacturing evolves into a more analytical and model-driven paradigm that complements existing advances in automation and industrial connectivity [9].



Figure 5. Data gathering and interpretation

Building on these system-level capabilities, the industrial impact of smart manufacturing is most evident in its large-scale adoption across key sectors. The chemical manufacturing industry, in particular, is undergoing a profound transformation driven by increasing demand, regulatory pressure, and global competition.

Within this context, smart manufacturing enables real-time process monitoring, predictive maintenance, automated quality control, improved safety management, and enhanced environmental compliance (Fig. 6).



Figure 6. Intelligent factory automation

AI-based systems analyze large volumes of operational data to optimize reaction conditions, detect anomalies, and improve process efficiency, thereby reducing

waste and operational costs. Meanwhile, IIoT technologies connect sensors, machines, and control systems, enabling continuous monitoring of energy consumption, emissions, and production parameters while supporting remote and automated operation.

Predictive maintenance improves operational reliability by enabling early detection of equipment failures, while smart safety systems reduce risks through real-time monitoring and automated emergency responses. In parallel, sustainability is strengthened through optimized resource utilization, reduced emissions, and the integration of greener feedstocks and chemical pathways.

Beyond production systems, digital technologies enhance supply chain resilience through demand forecasting, real-time logistics tracking, and improved inventory management. Emerging innovations such as digital twins, blockchain, augmented reality, and cloud computing further expand transparency, collaboration, and simulation capabilities. Despite these advancements, challenges such as high implementation costs, cybersecurity risks, workforce skill gaps, and legacy system integration remain significant barriers. Overall, the integration of smart technologies is reshaping the chemical industry toward a more efficient, safe, sustainable, and innovation-driven future [10].

Continuous feedback loops ensure that processes are constantly monitored and refined, enabling self-optimizing workflows that enhance efficiency, resilience, and responsiveness across the smart factory and its extended supply network.

3. Conclusions

The integration of Industry 4.0-enabled smart manufacturing with Green Chemistry represents a major shift in modern chemical engineering toward more intelligent and sustainable industrial systems. The deployment of advanced digital technologies, including artificial intelligence, machine learning, the Internet of Things, robotics, cloud computing, and big data analytics, has transformed industrial operations by enabling real-time monitoring, predictive maintenance, and data-driven process optimization. These developments have improved operational efficiency, safety, product quality, and resource utilization while simultaneously reducing downtime, energy consumption, and production costs.

When combined with Green Chemistry principles, this digital transformation enables the design of chemical processes that prioritize pollution prevention, resource efficiency, and the use of safer inputs from the outset. Environmental performance is increasingly addressed at the design stage rather than through corrective measures, leading to inherently cleaner production systems. In addition, enabling tools such as digital twins, IoT-based sensing, and AI-driven optimization enhance process understanding and support the development of more efficient and environmentally responsible chemical pathways. As a result,

sustainability and economic performance become complementary objectives within modern process design.

Across industrial applications such as catalysis and smart manufacturing systems, data-centric approaches are accelerating innovation, improving process control, and enabling more efficient and scalable production frameworks aligned with low-carbon and circular economy goals. Despite these advancements, challenges such as high implementation costs, cybersecurity risks, integration with legacy systems, and the need for advanced digital skills continue to limit widespread adoption.

Overall, this transformation is driving the chemical industry toward more autonomous, adaptive, and sustainable production systems, positioning chemical engineers as key contributors at the interface of digital innovation and environmental responsibility.

References

- [1] Vogel-Heuser, B., Bauernhansl, T., & ten Hompel, M. (Eds.). (2017). *Handbuch Industrie 4.0 Bd. 4: Allgemeine Grundlagen* (2nd ed.). Springer Vieweg. <https://doi.org/10.1007/978-3-662-53254-6>
- [2] Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production*, 252, Article 119869. <https://doi.org/10.1016/j.jclepro.2019.119869>
- [3] Yang, J., Liu, Y., & Morgan, P. L. (2024). Sensor and data: Key elements of human-machine interaction for human-centric smart manufacturing. *Procedia Computer Science*, 232, 191–200. <https://doi.org/10.1016/j.procs.2024.01.019>
- [4] Kashyap, A. (2025, October 19). *Smart manufacturing in chemical engineering: Integrating Industry 4.0 with green chemistry*. AIChE Executive Student Committee. <https://www.aiche.org/community/sites/committees/executive-student/blog/smart-manufacturing-chemical-engineering-integrating-industry-40-green-chemistry/>
- [5] Rob Wood. (2025, October 29). *Smart factory implementation in chemical plants*. BrightPath Associates. <https://brightpathassociates.com/smart-factory-implementation-in-chemical-plants/>
- [6] Daten & Wissen. (2022, January 19). *Smart manufacturing in chemical industry*. <https://datenwissen.com/blog/smart-manufacturing-in-chemical-industry/>
- [7] Bratovic, A. (2023). *Sophisticated transformation of chemical, pharmaceutical and food industry by implementation of digitalization and automatization*. In *Proceedings of the International Scientific Conference: Basic Technologies and Models for Implementation of Industry 4.0* (pp. 241–253). <https://doi.org/10.5644/PI2023.209.11>
- [8] Bratović, A., & Tomašić, V. (2026). Artificial Intelligence- and Machine Learning-Driven Strategies for Catalyst Design and Sustainable Chemical Processes. *Processes*, 14(12), 1866. <https://doi.org/10.3390/pr14121866>

- [9] Smart Factory Mom. (n.d.). *What is a smart factory?*
<https://www.smartfactorymom.com/what-is-a-smart-factory/>
- [10] Shah, A. (2025). *How smart technology is powering the next era of chemical manufacturing*. GSP Chem. <https://gspchem.com/how-smart-tech-shaping-future-of-chemical-manufacturing/>



An Integrated Model of Industry 4.0 and Industry 5.0 for Quality Management During the Entire Life Cycle of Plastic Injection Molds

Dušan Janković^{*1}, Goran Janković¹, Zora Joković¹, Aleksandar Supurović¹,
Vidosav Majstorović²

Abstract: Plastic injection molding tools are a key component that directly affects product quality, production efficiency, production costs, and sustainability throughout the product's life cycle. Recent advances in Industry 4.0 technologies, including Industrial Internet of Things (IIoT), Cyber-Physical Systems (CPS), Artificial Intelligence (AI), Big Data Analytics (BDA), cloud/edge computing, digital twins (DT) and advanced automation (cobots), have enabled significant improvements in tool performance (quality) monitoring and control. However, most existing approaches primarily focus on technological optimization and operational efficiency, while insufficiently addressing people-centeredness, sustainability and resilience, which are the core principles of Industry 5.0. Consequently, a comprehensive quality management framework that integrates Industry 4.0 and Industry 5.0 paradigms throughout the life cycle of plastic injection molding tools remains largely unexplored. This study proposes an Integrated Industry 4.0 - Industry 5.0 model for quality management throughout the life cycle of plastic injection molding tools, which includes the stages of design, production, inspection, operation, maintenance, renewal (overhaul) and end of life. The research methodology combines an extensive literature review, analysis of current industry practices, conceptual modeling and the development of a multi-layered architecture, which integrates digital technologies with human-centered and sustainable quality management principles. The proposed model consists of interconnected physical, data layer, intelligent, DT, human-centric, sustainability and resilience layers, in an overall multi-layered structure. Industry 4.0 technologies provide real-time monitoring, predictive analytics (quality, maintenance, as well as other KPI parameters), virtual quality assessment and automated decision support, while Industry 5.0 principles enable collaborative human-AI interaction, knowledge-driven quality improvement, resource efficiency, circular economy practices and resilient manufacturing operations. The main contribution of this research is the development of a holistic, lifecycle-oriented quality management framework with a developed set of KPI parameters, which bridges the gap between the technology-driven Quality 4.0 approach and the human-centered Quality 5.0 approach. Expected benefits include improved tool quality, reduced downtime and maintenance costs, improved decision-making capabilities, increased sustainment performance, extended tool life, greater process resilience and improved competitiveness of the injection molding company.

^{*1}Company Inmold Group, Požega, Serbia

²University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia

E-mail: dušan.jankovic@inmold-ltd.com

Keywords: *Industry 4.0, Industry 5.0, Quality Management, Plastic Injection Molding Tools, Smart Manufacturing, Digital Twin, Artificial Intelligence, Machine Learning, Sustainability, Human-Centered Manufacturing.*

1. Introduction

Plastic injection molding is one of the most commonly used manufacturing processes for producing large plastic components with complex geometry, tight dimensional tolerances and high repeatability. Industries such as automotive, aerospace, electronics, medical devices, packaging and consumer goods rely heavily on injection molding technology for its productivity, flexibility and cost-effectiveness. At the core of this process are plastic injection molds, which determine the shape, dimensional accuracy, surface quality and overall functionality of the finished products. Injection molds represent significant capital investments and often make up a significant portion of the total cost of manufacturing plastic products. Their design complexity, production accuracy and operational reliability directly affect production efficiency and business competitiveness. Since molds are exposed to repeated thermal, mechanical and chemical stresses during operation, maintaining their quality throughout the entire life cycle is crucial to ensure stable production and minimize operational risks. That is why effective quality management of injection molding tools, in their lifetime, has become a strategic requirement for modern manufacturing organizations, which are looking for higher productivity, reduced costs and sustainable business, Figure 1.

Industry 4.0 represents the fourth industrial revolution, characterized by the integration of digital technologies, intelligent systems and interconnected production resources. In the production of plastic injection molding tools, Industry 4.0 enables the transformation of conventional production systems into smart, data-driven and highly automated environments [1]. The adoption of digital technologies supports improved quality, productivity, flexibility and lifecycle management of injection molding tools.

Digitization is a fundamental pillar of Industry 4.0 and involves converting physical processes, assets and information into digital formats that can be collected, analyzed and used for decision making. In mold manufacturing, digitization begins during the design phase through the use of computer-aided design (CAD), computer-aided engineering (CAE), and computer-aided manufacturing (CAM) systems.

Digital technologies enable the creation of comprehensive digital records throughout the tool life cycle, including design specifications, machining parameters, dimensional inspection data, maintenance records and operational performance information.

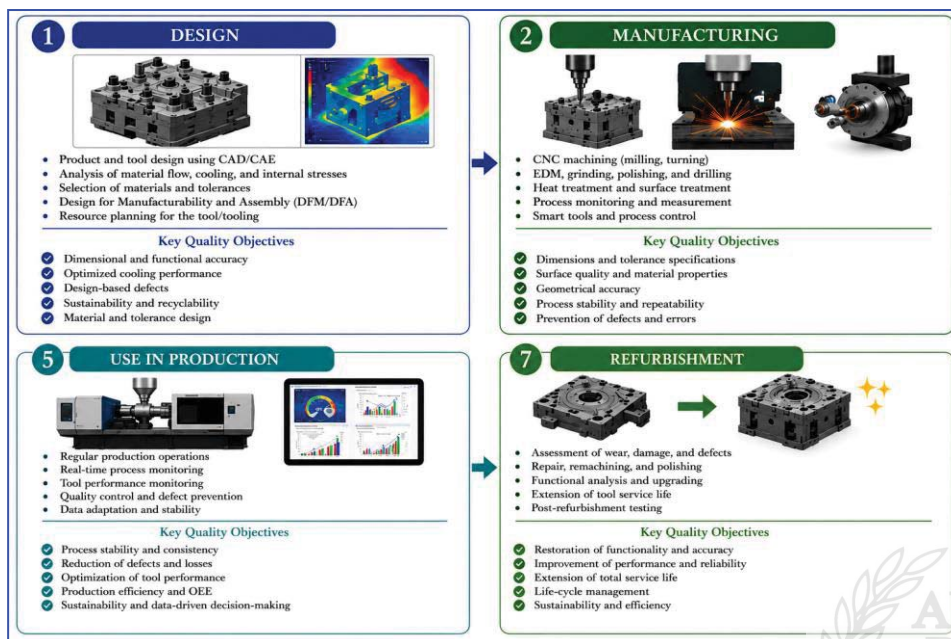


Figure 1. The life cycle of plastic injection molding tools from the perspective of quality management for the two most important phases

Advanced systems for product life cycle management (PLM) enable seamless data exchange between organizational units for design, production, quality and maintenance. The application of Industrial Internet of Things (IIoT) technologies further advances digitization, enabling real-time data collection from machine tools, sensors, coordinate measuring machines (CMM) and production equipment. These digital data streams provide valuable insights into tool condition, machining performance and quality characteristics, supporting data-driven quality management and continuous improvement initiatives.

While Industry 4.0 primarily focuses on digitization, automation and operational efficiency, Industry 5.0 expands this vision by emphasizing collaboration between people and intelligent technologies, while promoting sustainability and resilience. The concept recognizes that technological progress should not only improve productivity, but also create value for people, society and the environment [15]. In the context of plastic injection tooling manufacturing, Industry 5.0 introduces a more holistic approach to quality management, integrating human expertise, environmental responsibility and adaptive manufacturing capabilities throughout the tool life cycle, to increase the resilience of the plastic injection molding system.

1.1 Research Shortcomings

The application of Industry 4.0 technologies in manufacturing has attracted considerable attention from both researchers and practitioners over the past decade. Numerous studies have explored the application of Industrial Internet of Things (IIoT), Artificial Intelligence (AI), Machine Learning (ML), Cyber-Physical Systems (CPS), DT, Cloud Computing, and Big Data Analytics (BDA) to improve manufacturing efficiency, predictive maintenance, process optimization, and product quality. In the field of plastic injection molding, these technologies have been successfully applied to process monitoring, defect prediction, tool condition assessment, and production optimization. However, despite significant progress in the development of manufacturing systems based on Industry 4.0, there are still several important research gaps, especially in the area of quality management throughout the life cycle of plastic injection molding tools [1-28].

Predominant focus on Industry 4.0 technologies. Most existing studies primarily concentrate on the technological aspects of digital transformation, with an emphasis on automation, connectivity, data collection and intelligent analytics. Although these approaches have shown significant improvements in operational performance, they are largely technology driven and focus mainly on productivity, efficiency and cost reduction. Current research often evaluates individual technologies in isolation, such as AI-based quality prediction, IoT-enabled monitoring systems, DT, or predictive maintenance platforms. As a result, there is limited research into how these technologies can be integrated into a comprehensive quality management framework that covers all phases of the life cycle of plastic injection molding tools. Furthermore, the broader goals introduced by Industry 5.0, including human-centeredness, sustainability and resilience, are rarely incorporated into existing quality management models.

Limited integration of human-centered quality management. Human expertise plays a key role in mold design, manufacturing, inspection, maintenance, troubleshooting and quality improvement. However, many Industry 4.0 studies predominantly focus on automation and autonomous decision-making, often neglecting the contribution of engineers, operators, quality experts and maintenance personnel. The new paradigm of Industry 5.0 emphasizes the collaborative interaction between people and intelligent technologies. However, research investigating human-centered quality management within the mold manufacturing environment is still limited. Existing studies rarely address topics such as AI-assisted decision support, operator knowledge integration,

collaborative quality management, human-machine collaboration, workforce empowerment, and organizational learning during the tool life cycle. Consequently, there is a need for quality management frameworks that combine digital intelligence with human expertise to improve decision quality, adaptability, innovation and continuous improvement.

Lack of life-cycle-oriented quality management models. Another significant research gap is related to the lack of comprehensive life cycle oriented quality management models for plastic injection molding tools. Most available studies focus on specific phases of the life cycle, such as tool design, manufacturing, process optimization or maintenance, without considering the entire life cycle as an integrated system. The life cycle of plastic injection molding tools includes design, manufacture, assembly, inspection, commissioning, operation, maintenance, refurbishment and end-of-life management. Quality-related information generated during one phase is often not effectively transferred to subsequent phases, resulting in fragmented decision-making and reduced opportunities for continuous improvement. Although product life cycle management (PLM) and DT systems offer mechanisms for life cycle integration, their application in mold quality management remains relatively limited. Existing research lacks a single framework capable of linking life cycle data, quality indicators, maintenance information, sustainability metrics and human knowledge into a single integrated environment.

The need for an integrated quality management model of Industry 4.0 and Industry 5.0. Based on the identified shortcomings, there is currently no widely accepted model that integrates Industry 4.0 technologies with Industry 5.0 principles for quality management throughout the life cycle of plastic injection molding tools. Existing approaches usually deal separately with either technological innovation or organizational and sustainability aspects. Therefore, there is a clear need for a holistic quality management model, combining digital technologies, cyber-physical systems, DT, AI-driven analytics, human-centered decision-making, sustainability goals and resilience capabilities within a single lifecycle framework. Such a model would enable manufacturers to achieve not only higher quality and operational efficiency, but also greater adaptability, sustainability and long-term competitiveness. This study addresses this research gap by proposing an integrated Industry 4.0-Industry 5.0 model for quality management throughout the life cycle of plastic injection molding tools, providing a comprehensive framework that aligns technological innovation with human, environmental and resilience goals.

1.2 Research Objectives

The growing complexity of the production of plastic injection molding tools, combined with the increasing adoption of Industry 4.0 technologies and the emergence of Industry 5.0 principles, creates the need for a comprehensive approach to quality management that encompasses the entire life cycle of the tool. Although existing studies have shown the advantages of digital technologies for process optimization and quality control, integrated frameworks that simultaneously address technological, human-oriented, sustainable and resilience dimensions are still lacking. Therefore, the primary goal of this research is the development of an Integrated Quality Management Model of Industry 4.0 - Industry 5.0, which supports continuous quality improvement throughout the life cycle of plastic injection tools. To achieve this goal, the study pursues the following research objectives:

O1: Development of a life cycle-based quality management framework.

Development of a comprehensive quality management framework, covering all phases of the life cycle of plastic injection molding tools, including design, manufacture, inspection, commissioning, operation, maintenance, refurbishment and end-of-life management. The aim of the framework is to establish a continuous flow of information, traceability and quality feedback mechanisms in all phases of the life cycle, enabling proactive data-based quality management, with defined KPI parameters.

O2: Integration of Industry 4.0 and Industry 5.0 principles. Design an integrated model that combines the technological capabilities of Industry 4.0, including Industrial Internet of Things (IIoT), Cyber-Physical Systems (CPS), Artificial Intelligence (AI), Machine Learning (ML), Big Data Analytics (BDA), Cloud Computing and Digital Twins (DT), with Industry 5.0 principles of human-centeredness, sustainability and resilience. The goal is to create a balanced approach to quality management that uses advanced digital technologies while preserving the key role of human expertise and organizational adaptability.

O3: Improving tool quality, sustainability and resilience. Improving the overall performance of plastic injection molding tools by improving quality consistency, extending tool life, reducing maintenance costs, minimizing environmental impact and increasing operational resilience. The proposed model seeks to support predictive quality management, resource efficient operations, circular life cycle practices and adaptive responses to production disruptions and emerging risks. Together, these objectives provide the basis for the development of a holistic quality management framework, which aligns technological innovation with

human and organizational priorities, as well as with the environment, thus supporting the transition from traditional quality management approaches to the integrated paradigms of Quality 4.0 and Quality 5.0 in the production of plastic injection tools.

1.3 Organization of Paper

This paper has the following units: Introductory analysis of the basic concept and ideas for quality management of plastic injection tools, in their lifetime. The second part refers to the most important references from this field and their deep analysis, which confirms the need for research, which is the main part of this work. In module three, I present the concept of the Integrated Model with a detailed description of its most important elements. The most important part of this paper is the application of the Integrated Model in the practice of the INMOLD Company, which is shown in detail in point 4. In point 5 of this paper, there was a discussion of the application of the elements of the Integrated Model in the practice of this company, with special reference to the KPI parameters of this model, so that at the end of this paper, point 6, conclusions would be given with directions for future research in the development and application of this model.

2. Literature Review

The life cycle of a plastic injection molding tool includes all activities, from the initial development of the tool concept, to its final disposal or recycling. Effective quality management requires continuous monitoring and control during each phase of the life cycle, because decisions made during one phase directly affect tool performance, product quality, production costs, sustainability and operational reliability in subsequent phases [27,28]. A lifecycle-oriented approach enables manufacturers to improve traceability, optimize resource utilization, extend tool life and support continuous tool quality improvement.

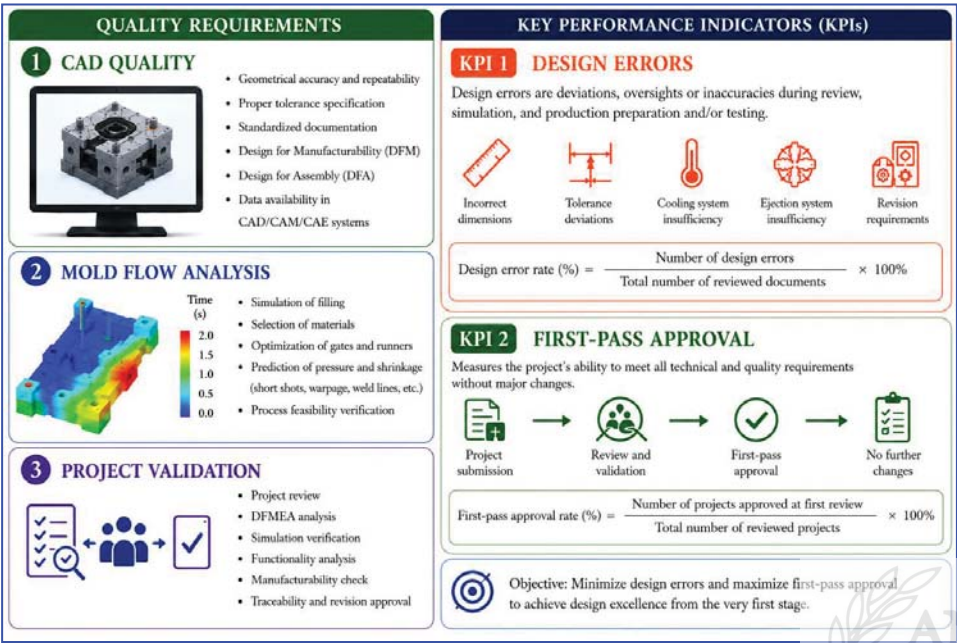


Figure 2. Quality requirements in the design phase of plastic injection molding tools

The design phase is the foundation of the tool life cycle and has a significant impact on future tool performance, Figure 2. During this phase, product requirements are translated into mold design using CAD, CAE and simulation tools. Key activities include cavity design, channel system design, cooling channel optimization, material selection, tolerance specification, and manufacturability analysis, i.e. processing capabilities. Quality objectives in this phase include: ensuring dimensional accuracy and functionality, optimizing cooling performance, reducing potential manufacturing defects, improving maintainability and usability, supporting sustainability through materials-efficient design. Design errors made at this stage can propagate throughout the entire life cycle and significantly increase production and maintenance costs.

A literature review was also performed for the tool life cycle, Table 1 [1-28].

Table 1. Analysis of key references for the development of an Integrated Industry 4.0 and Industry 5.0 model for quality management during the life cycle of plastic injection molding tools

Ref.	Main focus of the work	Industry 4.0 technologies	Industry 5.0 concepts	Tool life cycle	Focus on quality	Research gap
Kagermann et al. (2013) [1]	Industry 4.0 concept	CPS, IoT, Smart Factory	Not processed	No	Digital integration of production	A human-centric approach is lacking
Lasi et al. (2019) [2]	Definition of Industry 4.0	CPS, IoT		Partially	Operational efficiency	No quality 5.0
Xu et al. (2018) [3]	Industry 4.0 Overview	CPS, Cloud, IoT			Smart manufacturing	No sustainability
Frank et al. (2019) [4]	I4.0 implementation	IoT, Cloud, Big Data		No	Operational excellence	The human factor is missing.
Culot et al. (2020) [5]	Analysis of I4.0 technologies	IoT, AI, Cloud		Partially	Process efficiency	No lifecycle approach
Tao et al. (2019) [6]	Digital twin	DT, IoT, Big Data			Product optimization	Sustainability is lacking.
Jones et al. (2020) [7]	DT Architecture	DT			Performance monitoring	Quality 5.0 is missing.
Lang et al. (2021) [8]	DT for smart manufacturing	DT, IoT, AI	Limited	Yes	Quality monitoring	There is no humanocentrism
Fuller et al. (2020) [9]	DT technologies	DT, AI	Not processed	Partially	Predictive analysis	Sustainability is lacking.
Rosen et al. (2019) [10]	CPS and DT	CPS, DT			Process optimization	No quality
Sony et al. (2020) [11]	Quality 4.0	AI, Big Data, IoT	Limited	No	Quality 4.0	No Quality 5.0
Sony et al. (2021) [12]	Implementation of Quality 4.0	AI, Analytics			Digital quality	No sustainability
Jacob (2020) [13]	Quality 4.0 assessment	Big Data, AI			Quality measurement	No resistance
Antony et al. (2022) [14]	Development of Quality 4.0	AI, ML, Analytics			Quality management	Industry 5.0 is missing
Breque et al. (2021) [15]	Industry 5.0	I4.0 technologies	Human-centeredness , Sustainability Resilience	Partially	Indirectly	No quality model
Xu et al. (2021) [16]	Industry 4.0–5.0	AI, DT, CPS	Humanocentricity		Smart manufacturing	No quality tools
Lu et al. (2022) [17]	Human-centric manufacturing	AI, CPS			Decision support	No tool life cycle
Maddikunta et al. (2022) [18]	Industry 5.0 Overview	AI, IoT, Cloud	Human-centeredness , Sustainability	No	Indirectly	No quality
Lee et al. (2018)	Industrial AI	AI, Big Data	Not processed	No	Intelligent quality	No life cycle
Javaid et al. (2021) [19]	AI in manufacturing	AI, ML, IoT	Limited		Production quality	No sustainability
Wuest et al. (2016) [20]	Machine learning	ML	Not processed		Quality prediction	There is no humanocentrism

Kusiak (2018) [21]	Smart manufacturing	AI, Big Data		Partially	Quality optimization	No Industry 5.0
Carvalho et al. (2019) [22]	Predictive maintenance	AI, IoT	Limited	Maintenance	System reliability	No quality 5.0
Zonta et al. (2020) [23]	PdM in I4.0	AI, IoT, Analytics		Maintenance	Equipment reliability	Sustainability is lacking.
Ghobakhloo (2020) [24]	Sustainability and I4.0	IoT, AI, Cloud	Sustainability	Partially	Resource efficiency	No quality
Ghobakhloo et al. (2024) [25]	Sustainable production 5.0	AI, IoT	Sustainability Circular Economy		ESG performance	No quality tools
Rajput & Singh (2020) [26]	Circular economy	IoT, Analytics	Sustainability		Resource efficiency	No QM
Amanov et al. (2019) [27]	Mold life span	Wear monitoring	Not processed	Yes	Mold reliability	No digitization
Mourtzis et al. (2023) [28]	The digital life cycle of tools	DT, AI, IoT	Limited		Digital quality	No Industry 5.0 integration

Based on the analysis of the presented references, the following scientifically relevant conclusions can be drawn: *(a) dominance of Industry 4.0 technologies in the existing literature*. The analyzed literature shows that research in the field of quality management during the last decade has been dominantly focused on Industry 4.0 technologies, such as IoT, DT, AI/ML, cloud environments and CPS. Most authors investigate the possibilities of digitalization, automation and optimization of production processes, while human, social and sustainable aspects are mostly in the background, *(b) Quality 4.0 is well developed, while Quality 5.0 is still in the early stages of development*. Papers [11-13] show that the concept of Quality 4.0 has advanced significantly through the application of AI, Big Data and IoT technologies for quality prediction, process analysis and decision support. However, the concept of Quality 5.0 is still not sufficiently defined, especially regarding the integration of human-centricity, sustainability and resilience in quality management systems, *(c) digital twins represent a key technology for life cycle management*. Works [6-8,28] confirm that DT is one of the most promising technologies for connecting design, production, exploitation and maintenance of products and tools. However, most of the existing solutions are focused on the technical optimization of the process, while the aspects of sustainability, human knowledge and Quality 5.0 are insufficiently researched, *(d) human-centricity represents the biggest drawback of the existing models*. Although Industry 5.0 emphasizes the importance of humans as the central element of the production system, the literature analysis shows that most works do not integrate the expert knowledge of operators, quality engineers and managers into digital decision-making systems. Human-AI collaboration, knowledge management and human decision support remain significantly under-researched areas, *(e) sustainability is largely viewed separately from quality*. Papers [15,24,26] emphasize the importance of sustainability and circular economy in Industry 5.0. However, most research treats sustainability as a separate topic, without directly linking it to

quality management during the life cycle of a product or tool. The integration of quality KPIs and ESG indicators is still an open research question, *(f) the resilience of quality systems is insufficiently researched*. Although predictive maintenance and risk management are active research areas, the literature shows that there are very few works that relate resilience to quality systems. Adaptive quality systems capable of responding to disruptions, failures and changes in market demands represent a significant research opportunity, and *(g) a complete life cycle approach to plastic injection molding tools is lacking*. Most of the analyzed papers focus only on certain phases of the life cycle, most often design, production or maintenance. Very little research covers the complete life cycle of plastic injection molding tools, so this represents one of the biggest gaps in the existing literature.

An Integrated Industry 4.0 and Industry 5.0 model for tool quality management is missing. The most important conclusion of the analysis is that there is currently no single model that simultaneously integrates: Industry 4.0 technologies (IoT, AI, ML, DT, Cloud, Edge), Industry 5.0 principles (human-centricity, sustainability and resilience), Quality 4.0 and Quality 5.0 approaches, and quality management during the entire life cycle of plastic injection tools.

A major research gap. Based on the analyzed literature, it can be concluded that there is a significant lack of an integrated framework that connects Industry 4.0 technologies with Industry 5.0 principles for quality management during the complete life cycle of plastic injection tools. Existing research is mainly focused on the digitalization and optimization of individual processes, while human-centricity, sustainability, resilience and the life cycle of tools are insufficiently integrated into a unique quality model.

3. Integrated Model of Industry 4.0 and Industry 5.0 for Quality Management During the Life Cycle of Plastic Injection Tools

The production of plastic injection molding tools has gone through a significant digital transformation with the adoption of Industry 4.0 technologies. Modern facilities of the INMOLD company for the production of molds use integrated digital production systems (CPS), which connect the functions of design, analysis, production, inspection and business management in a unique information environment, according to the Industry 4.0 model. This integration enables real-time information sharing, improved traceability, improved quality control and data-driven decision making throughout the mold lifecycle.

The existing digital manufacturing system for plastic injection molding tools consists of seven main components: CAD/CAM, CAE, CNC, EDM, CMM, MES and ERP [29-32]. Together, these systems support the planning, execution,

monitoring and optimization of the mold manufacturing process, according to the model in Figure 3.

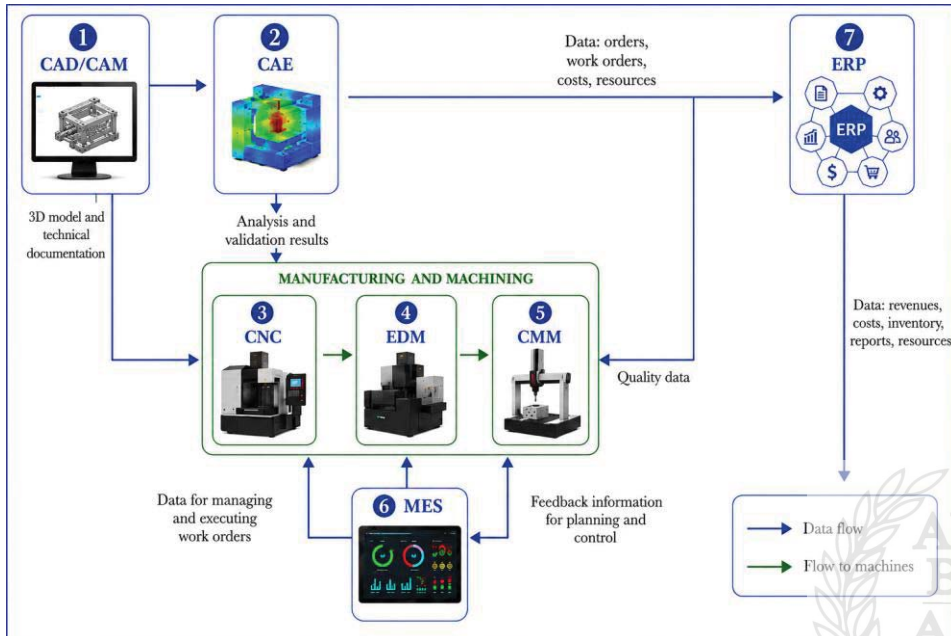


Figure 3. Industry 4.0 model for quality management of plastic injection molding tools at the INMOLD Company

CAD/CAM (1) systems form the basis of digital mold manufacturing. They enable the creation of three-dimensional digital models and the automatic generation of machining programs for production equipment. CAD/CAM improves quality through: accurate geometric modeling, reduced design errors, improved manufacturability, standardized design procedures, and reduced programming errors. The quality data generated are: CAD models, tolerance information, production drawings, CNC programs and process documentation. Coordinate Measuring Machines (CMMs) (5) are the primary metrology equipment used for dimensional verification of mold components. They provide very precise measurements: dimensions, geometric tolerances, surface profiles and shapes and prominent locations. In this Company, the CMM system supports: automated measurement programs, digital reporting, cloud connectivity, DT integration and AI-assisted inspection planning. CMM's contribution to quality management is: production quality verification, process capacity analysis, defect detection and traceability. The quality data generated are: inspection reports, SPC data, process capability indices and certified metrology protocols. A manufacturing

management system (MES) (6), acts as an operational management layer between plant equipment and enterprise-level systems, providing real-time insight into manufacturing operations. Contribution to quality management of MES are: quality monitoring in real time, non-conformity management, traceability and process control. The quality data generated are: production records, quality records, process histories, and performance indicators.

The proposed Integrated Quality Management Model of Industry 4.0 - Industry 5.0 represents the main contribution of this research, Figure 4. The model combines the technological capabilities of Industry 4.0 with the human-oriented, sustainable and resilient principles of Industry 5.0, in order to establish a comprehensive quality management framework oriented towards the life cycle of plastic injection tools.

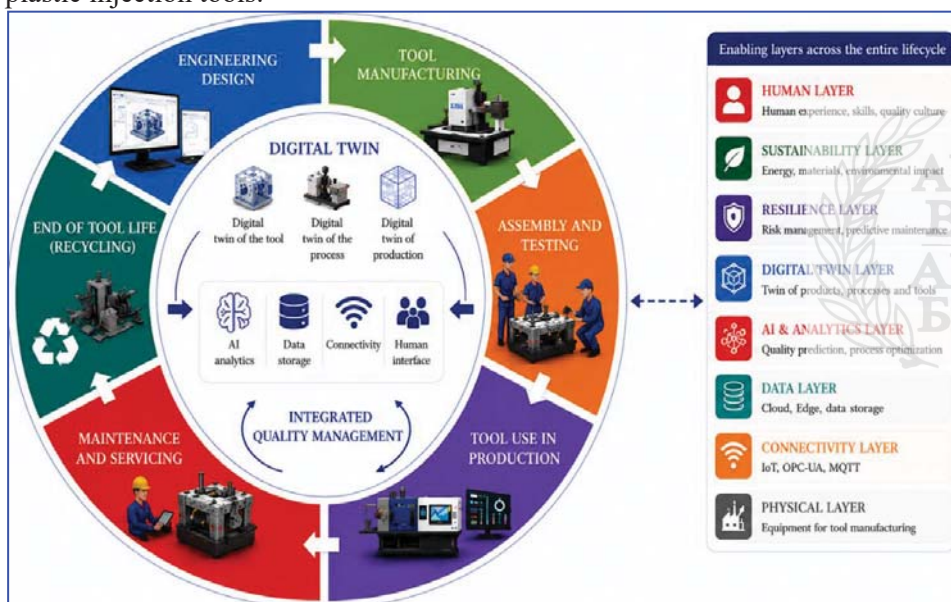


Figure 4. Proposed integrated model of Industry 4.0 and Industry 5.0 for quality management - general model

Existing approaches to quality management primarily focus on isolated phases of the life cycle, specific digital technologies or traditional quality methodologies. In contrast, the proposed framework provides a holistic architecture that integrates physical manufacturing systems, digital technologies, human expertise, sustainability goals, and resilience capabilities throughout the life cycle of a plastic injection tool. The model is defined in seven interconnected layers that

together support data collection, intelligent analytics, virtual representation, decision making, continuous improvement, sustainability improvement and resilience development.

Architecture of the proposed model follows a bottom-up structure, where information originates from physical production assets and moves progressively through connectivity, data management, analytics, DT, resilience mechanisms, sustainability evaluation and people-centered decision-making, Figure 5.

 HUMAN LAYER	Integrate design with quality	Qualified operators and technicians	Integrated quality validation and approval	Operational performance and measurement	Structure for decision-making and actions	Developing employees and knowledge transfer	Market awareness and customer satisfaction	Cross-functional teamwork	Human-centered leadership
 SUSTAINABILITY LAYER	Include human input and experience	Energy efficiency awareness	Optimize resources through integration	Efficient use of energy and materials	Continuous learning and improving skills	Share knowledge and best practices	Understand customer needs and create value	Improve internal communication and information flow	Create a healthy working environment
 RESILIENCE LAYER	Eco-friendly design and material selection	Reduce energy consumption	Optimize resources through integration	Efficient use of energy and materials	Reduce waste and emissions	Recycle, reduce waste and water usage	Reduce emissions and carbon footprint	Promote circular economy	Create long-term sustainable value
 DIGITAL TWIN LAYER	Risk assessment during design	Manage risks in processes and quality	Quick detection of defects and risk mitigation	Reaction to changes and uncertainties	Predictive maintenance and prevention	Improve the organization's resilience	End-to-end risk management and control	Ensure continuity and stability of processes	Adaptability and rapid recovery
 AI & ANALYTICS LAYER	Digital twin of the product	Digital twin of the process	Digital twin of the data	Digital twin linked to the real-time system	Digital twin for monitoring status and performance	Digital twin for simulation and prediction	Real-time digital twin updates and feedback	Data integration for better information	Transparency and traceability in real time
 DATA LAYER	AI for project and defect prediction	AI for process and tool optimization	AI for supply chain optimization	AI for process monitoring and alerting	AI for predictive maintenance and servicing	AI for quality control and improvement	AI for material selection and smart actions	AI for failure prediction	Intelligence for better decisions
 CONNECTIVITY LAYER	Data for design and simulation	Data for manufacturing and control	Data for testing and measurement	Data from IoT sensors and machines	Data for monitoring and system status	Data for analysis and reports	Data for traceability and life cycle	Data for change and initiatives	Secure and integrated data sources
 IoT Sensors  Wireless Connectivity  OPC UA  MQTT Enable seamless data acquisition and communication across all phases of the lifecycle.									
 PHYSICAL LAYER	Workstations for design and simulation	CNC, EDM, machines and other equipment	Equipment for assembly, CMM and testing	Machines for injection and testing	Machines for initial machining and texturing	Devices for sensing and measuring	Equipment for finishing and inspection	Systems for maintenance and repairs	Automation and physical safety

Figure 5. Integrated model of Industry 4.0 and Industry 5.0 for quality management of plastic injection molding tools - detailed structure

Layer 1 – Physical layer, represents the operational production environment in which quality-related information is generated. The component layers are *CPS processing systems* (Cutting, EDM, Grinding, Heat Treatment, Polishing and Plastic Injection Molding), *CMM, Mold Tools and Smart Sensors*. Primary functions include: production execution, inspection activities, process monitoring, tool condition monitoring, and product quality generation. This layer serves as the source of all quality data throughout the life cycle.

Layer 2 – Connection Layer, enables communication between physical devices, machines, sensors and higher-level digital systems. Its components are: *Internet of Things (IoT)* - enables device connectivity and real-time data collection, *OPC-UA* - supports secure and interoperable communication between industrial

systems, *MQTT* - enables easy and efficient communication between sensors, edge devices and cloud platforms. The functions of this layer are: real-time data transmission, interoperability, remote monitoring and system integration.

Layer 3 - The data layer stores, organizes and manages the production and quality information generated during the life cycle. The components are: *databases* - data on design, production, inspection, maintenance history and quality reports, *edge computing* provides local processing, low-latency analytics, and provides immediate decision-making support, *cloud computing*, provides central database generation, provides enterprise-wide accessibility, and scalable computing resources. The functions are: data management, traceability, historical analysis and preservation of knowledge about the life cycle of the tool.

Layer 4 – AI Analytics Layer, transforms raw data into actionable knowledge and predictive insights. Its components are: quality prediction using AI, related to - *dimensional deviations*, surface defects, process instability, and product inconsistencies, *process optimization, with the support of AI - processing parameters and modes, injection molding conditions, maintenance plans and energy* consumption monitoring. The technologies that support this concept are: AI, ML, BDA, DML, predictive analytics. This layer supports defect prevention, intelligent decision support, continuous process improvement, and predictive quality management.

Layer 5 – Digital Twin Layer (DT), creates virtual representations of physical entities and production processes, and its components are: *product digital twin* - represents the shaped product and its quality characteristics, which is applied for product validation, quality prediction, and life cycle analysis, *process digital twin* - represents production operations, and it is applied for process simulation and optimization and virtual commissioning, *digital twin tool* - represents the mold during its entire life cycle, and it is applied for: monitoring of wear, assessment of remaining useful life, predictive maintenance and overhaul planning. Features include virtual quality management, life cycle simulation, predictive maintenance, and continuous optimization.

Layer 6 – Resilience layer, introduces Industry 5.0 capabilities that improve system robustness, adaptability and business continuity. Its components are: *risk management* - risk identification, risk assessment, failure mode analysis, and quality risk prediction, *predictive maintenance* - preventing breakdowns, improving tool availability, reducing downtime, and optimizing maintenance. The functions of this layer are: enabling operational continuity, adaptive

manufacturing, cyber resilience, and supply chain resilience. The benefits are reduced downtime, improved reliability, and faster recovery from disruptions.

Layer 7 - Layer of sustainability, integrates aspects of environmental protection and resource efficiency in quality management. Its components are: *energy consumption monitoring* - machine energy consumption, process energy efficiency, and life cycle energy performance, *materials* - utilization and recycling of materials and waste generation, *carbon footprint - production-related CO₂ emissions*, life-cycle environmental impacts, and performance sustainability. The functions of this layer are: optimization of resources, reduction of environmental impact, application of circular economy, and sustainable quality management. The advantages are: reduced waste, lower energy consumption, and improved environmental performance.

Layer 8 – Human layer, represents the highest level of architecture and embodies the human-centric philosophy of Industry 5.0. People remain the final decision makers within the system. Its components are: *engineers* - responsible for design decisions, process optimization, and solving technical problems, *operators* - responsible for process monitoring, operational feedback, and participation in continuous improvement, *quality managers* - responsible for quality strategy, quality assurance and performance evaluation. The functions are: decision-making with the help of AI, integration of expert knowledge, human-robot collaboration and continuous learning. The benefits are better decision quality, increased innovation, improved organizational learning, and improved workforce engagement.

4. Application of the Integrated Model in the INMOLD Company

Here are three examples from the digital manufacturing model of the company Inmold, which illustrate the practical application of the elements of the Integrated Model. They relate to digital quality management (CAI/QMS), work order (ERP/MES) and new tool commissioning (MES).

The first example is related to the CMM, and the shown Figure 6 is the CMM measurement report for the control of the machined part M-975 NF #1 from the rear after the milling operation.

The report was created by integrating the 3D CAD model and CAI (PC DMIS model), and the control includes a total of **42 measurement locations (LOC1–LOC42)**. Most of the 42 controlled locations show very small deviations, usually in the range of a few micrometers to approximately 0.010 mm. Special attention is required for LOC5 and LOC6 with a deviation of **0.495 mm**, LOC34 with a diameter deviation of **+0.200 mm**, as well as three depth features LOC40–LOC42 with uniform deviations of approximately **-0.055 mm**.

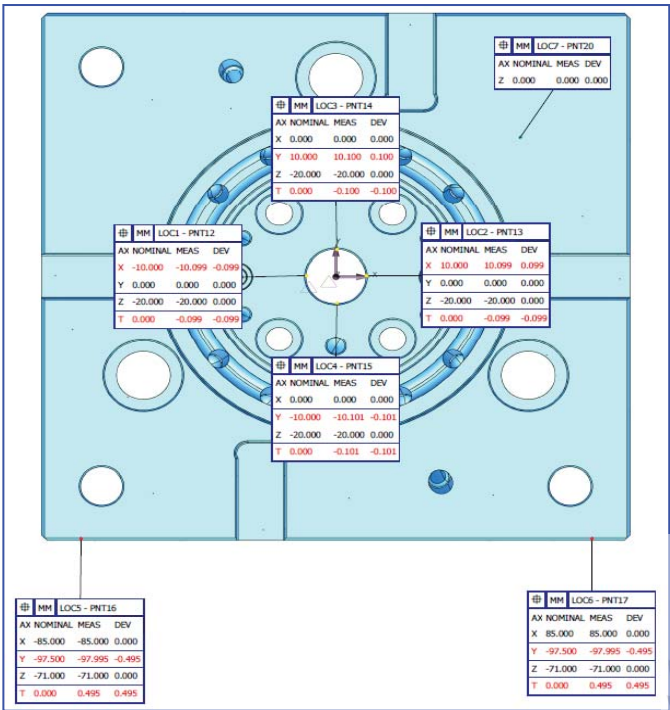


Figure 6. CMM report for part M- M-975 NF #1

Without the displayed tolerance limits, the conformity status cannot be definitively determined, but the image clearly enables the identification of critical features and is the basis for milling process correction, additional CMM checking and making a decision on part acceptance.

Another example refers to Work Order management, as part of the ERP/MES model, whose project is now being implemented, Figure 7.

In this company, the work order is given as a production work card, which contains a detailed work plan for the production of the part of the plastic injection **tool**, made within the project **M-990**, and it includes the complete technological flow of the production of the part (CAPP), from the initial preparation of the raw material, to the final CMM control. In the header of the first page, the basic information about the project is listed: project type is *M Packaging Mold M-XXX*, drawing number **prd-27373255-01857268**, article number **M-990-F-1.1-01-01**, and the name of part **33 – Cavity (NF)**. It is a part of a quadruple tool for the production of a bucket, where the planned quantity is four pieces. The project

implementation period is defined as of **March 25, 2026. until 14.08.2026.** The central part shows a 3D CAD model of the controlled part of the tool.

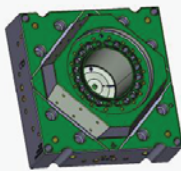
		Radna karta & Plan rada		Br. Projekta: M-990	
Vrsta projekta: M_Packaging_Mold M-XXX			Jedinična cena: NG/03		
Br. Crteža: prd-27373255-01857268		Datum dodavanja: 25.03.2026		Datum završetka: 14.08.2026	
Količina: 4 Piece		Opis: Four cavity bucket JST13-9 mold (I02773)			
Br. artikla: M-990-F-1.1-01-01			Naziv: 33 - Cavity (NF)		
					
10	Aktivnost: M_Rough milling preparation Komethnar: Obrada spoljašnjih gabarita Po visini ostaviti 1mm dodatka. Raspodela dodatka je pola/pola.			Datum: 06.05.2026 Datum / ime: 30.04.2026 ✓ DDomanović	
P846354					

Figure 7. Front page of the Work Order (Work Card) for part M-990

The work card includes **25 numbered activities**, from 10 to 250, and shows a complete digitally traceable flow: **preparation → CAPP/CAM programming → CNC milling → drilling cooling channels → five-axis drilling → control → finishing → heat treatment → grinding → finishing → EDM → CMM control → polishing → final CMM control (CAI inspection).** The document shows a high level of organization and traceability of production in the ERP/MES model. Each activity is associated with a corresponding code and barcode, which enables the connection of the Work Order with production data, operators, execution dates (scheduling) and process status.

The picture represents a practical example of **digitally supported quality management in the life cycle - the manufacturing phase of a complex part of a plastic injection molding tool.** The most important feature of the document is the integration of design (CAD/CAE), CAM programming, CNC and five-axis processing, EDM processes, thermal processing, grinding, polishing and CMM control into a single technological sequence. The existence of control and corrective loops is particularly important: the part is controlled before heat treatment, refined and re-controlled if necessary, while intermediate and final CMM measurements (CAI inspection) are carried out in the final phase. This structure of the Work Order is the basis for connection with **ERP, MES,**

nozzle temperatures; tool closing and opening parameters; parameters of injection molding and subsequent pressure; as well as the parameters of dosing and movement of the auger. Cylinder temperatures range from approx **260 to 270 °C**, while the temperatures of the hot block and active nozzles are set at around **270 °C**. A height of **878.2 mm**, an opening stroke of **500 mm**, a closing time of **1.02 s**, an opening time of **0.75 s** and a locking force of **2700 kN** were recorded for the tool. In the injection molding section, a cooling time of **1.5 s**, a switching pressure of **1654.1 bar**, and a maximum subsequent pressure of **1033.2 bar** are shown. and a clutch position of **72.13 cm³**. The document represents an important record for **tool validation, process reproducibility, cycle optimization and product quality control**. It can be used as a reference process sheet for future tests, batch production and process parameter deviation analysis.

Figure 8b shows a **technical drawing of the control points of the plastic product "Lid D116"**, made within the **M-975 project**, belonging to the Inmold Group company and serving as the basis for dimensional control of the product after the injection molding process. The mass of the product is **5.2 g**. The drawing shows the basic projections of the product, a side view, **section A-A in a scale of 5:1** and enlarged **detail B in a scale of 20:1**. The product is circular in shape, with a nominal diameter of **Ø116.53 mm** and a total height of approximately **8 mm**. **The 13 numbered control positions (pos. 1–13)** are of particular importance. Wall thicknesses and local geometric characteristics of the product are checked on them. Nominal values shown include **0.778 mm** at position 1; **0.39 mm** in positions 2, 3 and 4; **0.44 mm** in positions 5, 6 and 9; **0.378 mm** at position 7; **0.349 mm** at position 8; **0.42 mm** at position 10; **0.872 mm** at position 11; **0.513 mm** at position 12 and **0.74 mm** at position 13. Enlarged detail B clearly shows the complex geometry of the product rim: thin walls, transition radii, stepped surfaces and local thickness changes. This display enables precise control of critical zones that may be subject to deformation, shrinkage or uneven filling during injection molding.

Figure 8c shows a **screenshot of the parameters of the first injection stage** on a plastic injection molding machine, recorded **on April 4, 2026**. The document represents an operational record of actual and set process parameters during the test operation of the **M-975 tool**. The upper part shows the **injection profile**, with graphical representations of pressure and speed as a function of time and volume. The specified **injection volume is 211.66 cm³**, while the displayed injection speed is **85.0 cm³/s**. The graphics enable visual monitoring of changes in process parameters during filling of mold cavities. The central part shows the time parameters of the process. **The cooling time is 1.50 s**, the actual registered value is **1.51 s**, **the subsequent pressure time is 0.10 s**, and **the spraying or injection time is 0.14 s**. These very short time intervals indicate a highly productive process

with a short overall cycle. **The monitoring of the mass cushion**, whose value is **37.22 cm³**, is shown in particular. In the section for switching from the filling phase to the afterpressure phase, the actual switching volume of **73.10 cm³**, the set capacity value of **73.12 cm³** and the specific switching pressure of **1654.1 bar** are shown. In the section **Injection pressure**, the limit value of the specific pressure of **1859.7 bar** is specified, while the actual specific value of the injection pressure is **1654.1 bar**. It shows that the process is taking place below the defined maximum limit. The lower part of the picture shows a plasticizing unit with a screw diameter of **70.0 mm** and a screw geometry of **the Barrier type**. Overall, the figure represents an important process record for **test injection validation, parameter optimization, process stability control, and linking of injection parameters with the results of dimensional control of the Lid D116 product**. The drawing represents **the dimensional control plan of the product** and is used for CMM measurement, section control, tool validation and comparison of actual results with nominal values. It is especially important for connecting the process parameters of injection molding with product quality and proving the stability of the production process with the M-975 tool.

5. Discussion

For the Integrated Industry 4.0 - Industry 5.0 model for quality management throughout the life cycle of plastic injection molding tools, key performance indicators (KPIs) should cover not only the traditional quality, productivity and cost metrics, but also the three basic pillars of Industry 5.0: human focus, sustainability and resilience. The framework of KPIs can be structured according to the stages of the life cycle of the tool, Figure 9. In the design area, seven parameters are defined, from the design error rate (%), to the knowledge usage rate, also in %. For the area of tool production, ten KPI parameters were defined, from dimensional accuracy analysis in (μm), to equipment downtime (h/month). For assembly and test work activities, six parameters are defined, from the assembly defect rate (%), to the accuracy of digital balance validation (%). Using the tool in production is the next activity in the life of this product. A total of ten KPI parameters were defined for this phase, from the rate of defective products (ppm), to the time of starting work after repair (h). The next stage is tool maintenance. Eight KPI parameters are defined here, from the tool failure rate (%), to the failure prediction time horizon (days). Repairs and overhaul is one of the very important phases in the tool life, for which five KPI parameters are defined. The first is the overhaul success rate (%), to the accuracy of decisions with expert support (%). The last stage in the life of a tool is retirement or end of life. There are five KPI parameters defined for this activity, from the degree of

recycling (%) to the circularity index (%). Summarizing this analysis, we can say that a total of 51 KPI parameters were developed for this model.

Data sources for these parameters are the elements of Industry 4.0 in application, namely: IoT sensors, MES/ERP systems, SCADA systems, quality control, big data platform and cloud computing. The key element of this data hub is the DT tool, and the added value to this concept is given by the elements of Industry 5.0, and they generate KPI parameters related to: human-centeredness, sustainability, resilience, human-AI cooperation and knowledge and experience sharing. This concept is implemented in the KPI feedback loop of continuous quality improvement, which includes: real-time data collection, KPI monitoring at each stage, analysis and diagnosis (AI & expertise), actions and improvements (corrective / preventive), and validation of results, where KPI improvements are measured. Based on all fifty-one KPI parameters, its integral framework was defined, which has five dimensions: quality (Q), productivity (P), sustainability (S), human-centeredness (H), and resilience (R). On the axis of these elements and weight coefficients (Wi), the Total Quality Index during the life cycle of the tool is defined - OLQI. It should be emphasized here that as part of the developed Integrated Model, there is also a set of strategic KPI parameters for the management level, six basic ones (from first time correct $\geq 90\%$, to energy per tool < 1400 kWh/tool). The six additional parameters are, from the parameter related to the carbon footprint < 350 kg CO₂ / tool, to the OLQI index $\geq 85\%$.



Figure 9. Structure of KPI parameters in the lifetime of the tool

The left module contains (not shown in Figure 10) four layers of the architecture, namely: data collection, Industry 4.0 analytics, KPI calculation and decision support. This last layer is particularly interesting, because it contains the functions of online monitoring of KPI parameters, predictive warnings, determined based on the trends of KPI parameters and recommendations for action, with the aim of improving the Integrated Tool Quality Management Model according to the Industry 4.0 and Industry 5.0 models for plastic injection molding. The right module has the following overviews: (a) for management (KPI parameters for the strategic level - OLQI and so on), (b) KPI panel by stages of the tool life, seven stages in total, with an indicator - good (and this indicator can be: critical, weak, moderate, good and excellent), (c) the performance panel of Industry 5.0, for its three pillars, and (d) the OLQI panel. Particularly important is the Trend Panel, which for the six indices of the most important KPI parameters, shows their real-time In this way, the second output of this project is completed for the Integrated Management Model of Plastic Injection Molding.



Figure 10. Integrated KPI dashboard for a plastic injection molding tool

The values of the weighting coefficients of the Integrated Quality Index are specifically stated here, namely: w_1 - quality (30%), w_2 - productivity (25%), w_3 - sustainability (20%), w_4 - human moderation (15%), and w_5 - resistance (10%). From the analysis of these weight coefficients, we can conclude that tool quality and productivity (application of Industry 4.0 elements) have the greatest impact on the Integrated Index.

These KPIs provide a comprehensive measurement system, which simultaneously evaluates Quality 4.0 (digitalization, AI, DT, IoT) and Quality 5.0 (human-centeredness, sustainability and resilience), throughout the life cycle of a plastic injection tool, and they are well aligned with the model defined in this paper.

In the conclusion of this discussion, we can confirm that the basic framework of the life cycle KPI architecture is: design → production → assembly and trial operation → production → maintenance → overhaul → end of life, which provide a flow of key performance indicators (KPI) between phases, as well as feedback loops for continuous improvement. The hierarchical KPI model is defined over three levels: strategic KPIs (management level), tactical KPIs (at department level), and operational KPIs (at production plant level). The key performance indicators (KPI) layer of Industry 4.0 are: IoT connectivity index, DT accuracy, AI prediction accuracy, data availability, and cyber security index. Key Performance Indicators (KPI) for Industry 5.0: and human-centeredness index, sustainability index, and resilience index, and especially as an aggregate value - Total Life Cycle Quality Index (OLQI), Figure 11.



Figure 11. Table of the overall quality index of plastic injection molding tools

The proposed Integrated Key Performance Indicator (KPI) dashboard enables: quality support (reduced defects, improved dimensional accuracy, and improved process capacity), productivity support (higher OEE, reduced downtime, and faster maintenance), sustainability support (lower energy consumption, reduced waste, and increased tool life), human-centric support (better decision support, improved employee engagement, and improved knowledge sharing), and resilience support (faster recovery from disruptions, improved predictive maintenance, and better preparedness for cyber security).

The dashboard therefore acts as **the digital nervous system of the entire tool life cycle**, integrating Industry 4.0 technologies and Industry 5.0 values into a unique lifecycle-oriented quality management platform, focused on the **Integrated Life Cycle Quality Index (OLQI)** of plastic injection molding tools.

6. Conclusions and Future Research

This paper defined an Industry 4.0 - Industry 5.0 Integrated Quality Management Model for plastic injection molding tools, addressing the growing need for an intelligent, sustainable, resilient and human-centered quality management model throughout the tool life cycle. The research is motivated by the increasing complexity of the mold manufacturing process and the limitations of existing quality management approaches, which often focus either on traditional quality methods or on isolated Industry 4.0 technologies, without fully considering human-centeredness, sustainability, resilience and life cycle integration.

The research objectives were successfully achieved. First, a comprehensive life-cycle-based quality management framework was developed, covering all major phases of the life cycle, including design, manufacturing, assembly and testing, production use, maintenance and overhaul, and end-of-life management. Second, Industry 4.0 technologies such as IoT, AI, ML, BDA, cloud/edge computing, DT and cyber security are systematically integrated with the principles of Industry 5.0, which relate to human-centeredness, sustainability and resilience. Third, the proposed framework has shown potential for improving tool quality, operational performance, environmental sustainability, and organizational adaptability.

The main contribution of this research is the development of a new multi-layered architecture, which combines technological intelligence with human expertise and sustainability-oriented decision-making. The proposed framework consists of physical, connectivity, data, AI analytics, DT, resilience, sustainability and human

layers that together support predictive quality management and continuous lifecycle improvement. Unlike conventional quality management systems, the proposed model promotes proactive defect prevention, predictive maintenance, lifecycle traceability, resource efficiency, and collaborative decision-making between humans and AI.

From a quality management perspective, the framework extends the existing concepts of Quality 4.0 towards a more comprehensive approach, Quality 5.0. Integrating DT, AI-driven analytics, predictive quality management and human knowledge, enabling more efficient quality control, faster root cause identification, improved dimensional accuracy, reduced defect rates and extended tool life. Furthermore, sustainability-oriented quality management contributes to reduced energy consumption, less material waste, improved use of resources and improved application of the circular economy. The resilience layer strengthens operational continuity, through predictive maintenance, risk prediction, cyber resilience and supply chain resilience.

The practical implications of the proposed model are significant for tool manufacturers, who want to improve competitiveness in increasingly digitized and sustainably oriented industrial environments. The framework can support the implementation of smart manufacturing strategies, improve life cycle management practices, reduce operational costs, improve product quality and facilitate compliance with new sustainability and traceability requirements. In addition, the model provides a practical roadmap for organizations transitioning from traditional quality management systems to manufacturing ecosystems oriented towards Industry 5.0.

Future development, Figure 12, should focus on the integration of new technologies, such as generative artificial intelligence (GenAI) for tool design, autonomous quality systems, explainable artificial intelligence (XAI), federated learning, green digital twins and digital product passports [14,15,17,18].

These technologies have the potential to further improve lifecycle intelligence, sustainability and decision-making transparency. Furthermore, future research should explore the development of a 5.0 tool quality maturity model, quantitative validation methodology, and industrial case studies, to evaluate the practical effectiveness of the proposed framework in real manufacturing environments. In conclusion, the proposed Integrated Quality Management Model of Industry 4.0 - Industry 5.0, provides a comprehensive basis for the future of quality management of plastic injection molding tools.

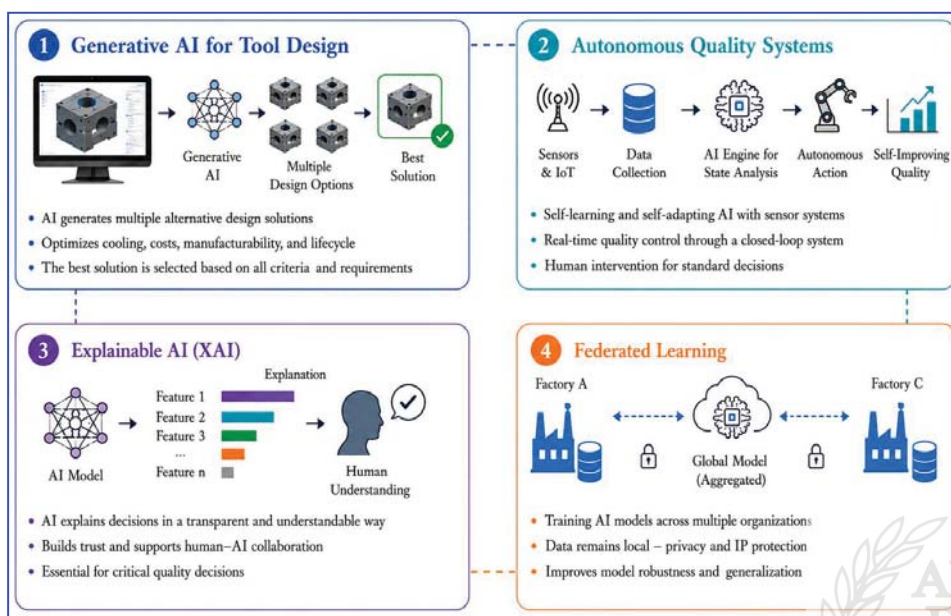


Figure 12. Future directions of research and development of the Integrated Model

Combining advanced digital technologies with human expertise, sustainability principles and resilience capabilities, the framework supports the development of intelligent, adaptive, sustainable and human-centered manufacturing systems, capable of achieving long-term excellence in quality, throughout the mold life cycle.

References

- [1] Kagermann, H., Wahlster, W., Helbig, J. (2013). Recommendations for implementing the strategic initiative Industrie 4.0. Acatech. <https://en.acatech.de/publication/recommendations-for-implementing-the-strategic-initiative-industrie-4-0-final-report-of-the-industrie-4-0-working-group/> .
- [2] Lassi, H., et al. (2019). Industry 4.0. Business & Information Systems Engineering, 6 (4), 239–242. <https://doi.org/10.1007/s12599-014-0334-4> .
- [3] Xu, LD, Xu, EL, Li, L. (2018). Industry 4.0: State of the art and future trends. International Journal of Production Research, 56 (8), 2941–2962. <https://doi.org/10.1080/00207543.2018.1444806> .
- [4] Frank, AG, Dalenogare, LS, Ayala, NF (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. International Journal of Production Economics, 210, 15–26. <https://doi.org/10.1016/j.ijpe.2019.01.004> .
- [5] Culot, G., et al. (2020). Behind the definition of Industry 4.0. Computers in Industry, 113. <https://doi.org/10.1016/j.compind.2019.103122> .
- [6] Tao, F., et al. (2019). Digital Twin in Industry. IEEE Transactions on Industrial Informatics, 15 (4), 2405–2415. <https://doi.org/10.1109/TII.2018.2873186> .
- [7] Jones, D., et al. (2020). Characterizing the Digital Twin. CIRP Journal of Manufacturing Science and Technology, 29, 36–52. <https://doi.org/10.1016/j.cirpj.2020.02.002> .
- [8] Leng, J., et al. (2021). Digital twins-based smart manufacturing system design. Journal of Manufacturing Systems, 60, 119–137. <https://doi.org/10.1016/j.jmsy.2021.05.007> .
- [9] Fuller, A., et al. (2020). Digital Twin: Enabling technologies. IEEE Access, 8, 108952–108971. <https://doi.org/10.1109/ACCESS.2020.2998358> .
- [10] Rosen, R., et al. (2019). About the importance of autonomy and Digital Twins. IFAC-PapersOnLine, 48 (3), 567–572. <https://doi.org/10.1016/j.ifacol.2015.06.141> .
- [11] Sony, M., Antony, J., Douglas, JA (2020). Essential ingredients for Quality 4.0. The TQM Journal, 32 (4), 779–793. <https://doi.org/10.1108/TQM-12-2019-0275> .
- [12] Sony, M., et al. (2021). An empirical examination of Quality 4.0. The TQM Journal, 33 (6), 1241–1262. <https://doi.org/10.1108/TQM-05-2020-0118> .
- [13] Jacob, D. (2020). Quality 4.0 impact and assessment. International Journal of Quality & Reliability Management, 37 (9), 1165–1185. <https://doi.org/10.1108/IJQRM-07-2019-0210> .
- [14] Antony, J., et al. (2022). An exploration of Quality 4.0. International Journal of Quality & Reliability Management, 39 (5), 1249–1273. <https://doi.org/10.1108/IJQRM-10-2020-0362> .

- [15] Breque, M., De Nul, L., Petridis, A. (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. European Commission. <https://doi.org/10.2777/308407> .
- [16] Xu, X., et al. (2021). Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems*, 61, 530–535. <https://doi.org/10.1016/j.jmsy.2021.10.006> .
- [17] Lu, Y., et al. (2022). Outlook on human-centric manufacturing towards Industry 5.0. *Journal of Manufacturing Systems*, 62, 612–627. <https://doi.org/10.1016/j.jmsy.2022.02.001> .
- [18] Maddikunta, PKR, et al. (2022). Industry 5.0: A survey. *Information Systems Frontiers*, 24, 123–147. <https://doi.org/10.1007/s10796-021-10143-z> .
- [19] Javaid, M., et al. (2021). AI applications in manufacturing quality. *Materials Today: Proceedings*, 44, 5086–5091. <https://doi.org/10.1016/j.matpr.2021.01.585> .
- [20] Wuest, T., et al. (2019). Machine learning in manufacturing. *Engineering Applications of Artificial Intelligence*, 48, 118–133. <https://doi.org/10.1016/j.engappai.2015.10.008> .
- [21] Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, 56 (1–2), 508–517. <https://doi.org/10.1080/00207543.2017.1351644> .
- [22] Carvalho, TP, Soares, FAAMN, et al. (2019). Predictive maintenance review. *Computers & Industrial Engineering*, 137. <https://doi.org/10.1016/j.cie.2019.106024> .
- [23] Zonta, T., et al. (2020). Predictive maintenance in Industry 4.0. *Computers & Industrial Engineering*, 150. <https://doi.org/10.1016/j.cie.2020.106889> .
- [24] Ghobakhloo, M. (2020). Industry 4.0 and sustainability. *Journal of Cleaner Production*, 252. <https://doi.org/10.1016/j.jclepro.2019.119869> .
- [25] Ghobakhloo, M., et al. (2024). Industry 5.0 and sustainable manufacturing. *Journal of Cleaner Production*, 435. <https://doi.org/10.1016/j.jclepro.2023.140461> .
- [26] Rajput, S., Singh, SP (2020). Industry 4.0 and circular economy. *Business Strategy and the Environment*, 29 (6), 2380–2396. <https://doi.org/10.1002/bse.2517> .
- [27] Amanov, A., Pyun, YS, Sasaki, S. (2019). Lifecycle improvement of injection molds. *Wear*, 426–427, 1141–1149. <https://doi.org/10.1016/j.wear.2018.12.073> .
- [28] Mourtzis, D., Angelopoulos, J., Panopoulos, N. (2023). Digital technologies in tooling lifecycle management. *Journal of Manufacturing Systems*, 69, 639–657. <https://doi.org/10.1016/j.jmsy.2023.08.009> .
- [29] Vojin Vukadinović, Jovan Živković, Dragan Đurđanović and Vidosav Majstorović, Management of tools in digital manufacturing - A case study, *MATEC Web Conf.*, 368 (2022) 01014, <https://doi.org/10.1051/mateconf/202236801014> .

- [30] Jokovic, Z., Jankovic, G., Jankovic, S., Supurovic, A., & Majstorovic, V. (2023). Quality 4.0 in Digital Manufacturing – One Example. Preprints. <https://doi.org/10.20944/preprints202305.1156.v1> .
- [31] Karabegović, I., Majstorović, V. (Eds), 2024, *Industry 4.0: Digital Transformation Shaping the Future* (Industry 4.0: Digital Transformation Shaping the Future), *Society for Robotics in Bosnia and Herzegovina, Bihać, Academy of Sciences and Arts of Bosnia and Herzegovina, Sarajevo*. ISBN 978-9958-9262-8-0, ISBN 978-9926-410-94-0, <https://bastina.anubih.ba/handle/123456789/789> .
- [32] Majstorovic, VD, Vukadinovic, VS, & Živković, J. (2023). Towards the Digital Model of Tool Lifecycle Management in Sheet Metal Forming. *Journal of Machine Engineering*. <https://doi.org/10.36897/jme/171664>.



Wireless Communication Link Optimization in Cyber-Physical Systems Using Smart Sensor Networks

Enes Saletović^{*1}, Malik Kriještorac², Đulaga Hadžić¹

Abstract: *This paper examines the optimization of wireless communication link performance for controlling unmanned aerial vehicles (UAVs), mobile robots, and remote industrial systems using a smart sensor network. Based on the analysis of sensor integration and data fusion algorithms, the proposed approach enables adaptive antenna control to maintain a stable and reliable communication link. The system utilizes measurements of position, distance, velocity, and other relevant physical parameters to improve signal quality and extend communication range. The paper also discusses the applicability of the proposed solution in cyber-physical systems and industrial environments, highlighting its potential to enhance communication reliability and operational efficiency.*

Keywords: *Wireless communication, smart sensor networks, cyber-physical systems, UAV systems, sensor fusion, adaptive control.*

1. Introduction

The field of remote control of UAVs, robots, and industrial systems has been studied for decades. Continuous advancements in telecommunications, sensors, and microelectronics have enabled the development of increasingly sophisticated RC (Remote Control) systems that provide improved control of Cyber-physical Systems (CPS), which rely on a strong synergy between computational and physical components. The primary performance parameters of wireless remote control systems are stability, range, bandwidth, and link latency, while secondary parameters include power consumption, interference immunity, scalability, compatibility, and support. Significant progress has been observed in the development of communication protocols aimed at achieving high-performance communication links. Advanced MEMS (Micro-Electro-Mechanical Sensors) enable straightforward integration into various control systems, while the application of AI (Artificial Intelligence) makes it possible to implement complex algorithms based on the fusion of different sensor measurements. High precision, compact dimensions, low power consumption, availability, and affordable cost contribute to the widespread use of various sensors delivered with standardized communication interfaces.

^{*1}University of Zenica, Zenica, Bosna i Hercegovina

²Sarajevo School of Science and Technology, Sarajevo, Bosnia and Herzegovina

E-mail: saletovic.enes@gmail.com

Within the scope of this work, the possibility of improving RC link performance through optimization of antenna spatial orientation on both the transmitter (Tx) and receiver (Rx) sides was investigated, using an adaptive algorithm and GPS (Global Positioning System) modules for position measurement.

2. Theoretical Background / Literature Review

In the existing literature, various solutions aimed at improving the performance of remote control systems have been presented. Papers [1] and [2] discuss the application of CPS systems in Industry 4.0 and robotics. In paper [3], the authors address specific problems related to spatially distributed measurement points and sensor networks in the oil industry. In paper [4], the authors analyze modern trends in the field of sensor networks and discuss contemporary solutions developed for healthcare, navigation, rescue, intelligent transportation, social networking, and gaming applications. In paper [5], the authors address current open issues related to enabling seamless integration of WSNs (Wireless Sensor Network) with CPS designs.

In paper [6], the authors solve the problem of integrating WSNs into a spatially distributed system by implementing a WMN (Wireless Mesh Network). The focus of this article is on evaluating and improving WMN performance by increasing network throughput. A regression model is utilized as a mathematical tool to predict the number of possible links and their lengths in each proposed network topology. In paper [7], the authors study the problem of jointly optimizing BS (Base Station) placement and power control in buildings in order to prolong the battery lifetime of sensors within CPS network infrastructures, and they present an appropriate algorithm suitable for application in smart buildings and other large-scale sensor networks.

In paper [8], the authors deal with Fractional Programming (FP), a class of optimization problems in which the objective function is expressed as a ratio of two functions. Examples of such objectives include the signal-to-interference-plus-noise ratio (SINR) and Fisher information. The authors present a novel approach for accelerating the quadratic transform using Nesterov's extrapolation scheme, which is relevant for next-generation wireless networks. In paper [9], the application of quantum search algorithms for accelerating optimization tasks in wireless systems is considered, with the aim of enhancing system performance and reducing complexity. In paper [10], the authors investigate the network connectivity of wireless sensor networks using directional antennas. A new analytical directional antenna model, called the Iris model, is proposed in order to balance accuracy and complexity.

2.1. Research Objective

- Development of an appropriate method and a compact algorithm for optimal antenna alignment in a point-to-point link
- Selection of MEMS sensors enabling optimal spatial orientation of antennas on both nodes of the point-to-point link
- Selection of optimal antennas with respect to directivity and radiation pattern
- Experimental validation of the proposed method
- Analysis of the influence of sensor accuracy on link performance
- Analysis of the influence of interference on the performance of the developed point-to-point link optimization method

3. Methodology

Within the scope of this research, several scientific methods were employed. A mathematical method was used for the analytical description of the antenna spatial orientation problem. Based on the established mathematical model, a method for calculating the antenna orientation angles (Yaw and Pitch) at both ends of the point-to-point RC link was developed. Based on the proposed method, a suitable algorithm for calculating optimal angles, appropriate for software implementation, was designed. Through experimental methods, the developed approach was tested and verified using an implemented RC link prototype.

3.1 Derivation of The Method for Optimization of Antenna Spatial Orientation

In a real-world scenario, the remote controller Tx and the controlled object (or process) Rx are located at different positions, separated by a distance D that can be calculated based on the GPS positions of both nodes, as shown in Fig. 1. In the standard three-dimensional GPS coordinate system WGS84, the position of the transmitter Tx is uniquely determined by the coordinates Lat_T , Lon_T , and HT , while the position of the receiver Rx is uniquely determined by the coordinates Lat_R , Lon_R , and HR . These positions are obtained using standard GPS modules (NEO-6M) located in both the transmitter and receiver. If the transmitter position is approximately constant, it is sufficient to measure only the position of the receiver Rx while using the known or initially measured transmitter position. For easier mathematical analysis, the origin of the reference three-dimensional xyz coordinate system is placed at the Tx position, where the x-axis is directed toward the east, the y-axis toward the north, and the z-axis vertically upward. In the general case, Tx and Rx may be located at different altitudes. Therefore, in order to orient the antenna of the Tx node directly toward the antenna of the Rx node, it is necessary to calculate the Yaw angle in the horizontal x-y plane and the Pitch

angle, which represents the inclination relative to the horizontal x-y plane. In spatial geolocation systems, the Yaw angle is commonly measured relative to the north direction (0 degrees) in the clockwise direction.

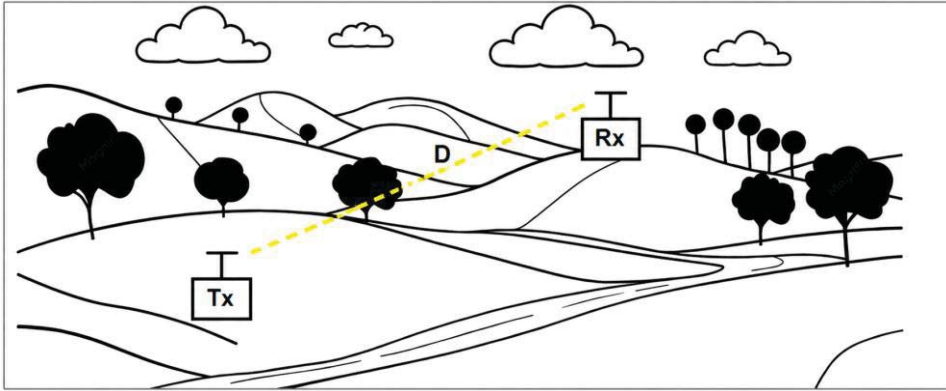


Figure 1. Positions of the transmitter Tx and receiver Rx in a real-world scenario

For the adopted coordinate-axis orientation, and based on the measured and filtered GPS coordinates of the transmitter (Tx) and receiver (Rx) shown in Fig. 2a, it is possible to calculate the horizontal Yaw angle at which the Tx node “sees” the Rx node using the following known parameters:

- φ_T - transmitter latitude in radians
- λ_T - transmitter longitude in radians
- H_T - transmitter altitude above sea level
- φ_R - receiver latitude in radians
- λ_R - receiver longitude in radians
- H_R - receiver altitude above sea level
- Yaw_{TR} - horizontal angle from Tx toward Rx in radians
- Yaw_{RT} - horizontal angle from Rx toward Tx in radians
- $Pitch_{TR}$ - vertical angle from Tx toward Rx in radians
- $\Delta\lambda_{TR}$ - longitude difference between Tx and Rx
- P_{TR}, Q_{TR} - intermediate variables used in the Yaw calculation
- D_{TR} - distance between Tx and Rx
- ΔH_{TR} - difference in altitude between Rx and Tx
- R - radius of the Earth

The horizontal angle from Tx toward Rx can be calculated using the following expression for the Yaw angle in radians:

$$Yaw_{TR} = \text{atan2}(P_{TR}, Q_{TR}) \quad (1)$$

where:

$$\Delta\lambda_{TR} = \lambda_R - \lambda_T \quad (2)$$

$$P_{TR} = \sin(\Delta\lambda_{TR}) \cos(\varphi_R) \quad (3)$$

$$Q_{TR} = \cos(\varphi_T) \sin(\varphi_R) - \sin(\varphi_T) \cos(\varphi_R) \cos(\Delta\lambda_{TR}) \quad (4)$$

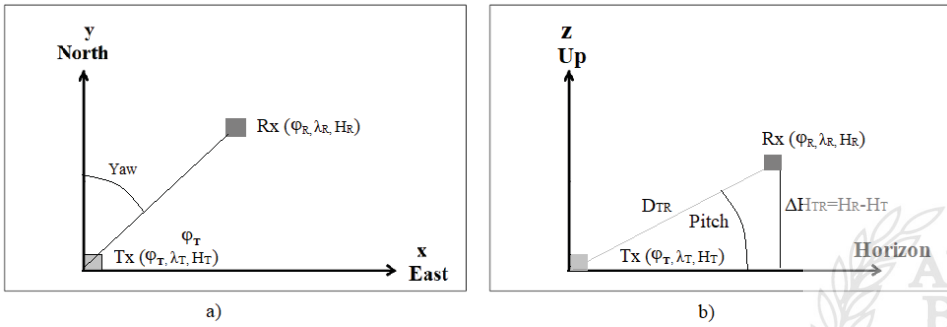


Figure 2. Positions of the transmitter Tx and receiver Rx in the horizontal x-y plane (a) and in the vertical plane (b).

The Yaw_{TR_D} angle in degrees can be calculated as:

$$Yaw_{TR_D} = (Yaw_{TR} \cdot 180/\pi + 360) \bmod 360 \quad (5)$$

while the horizontal angle under which Rx “sees” Tx is given by:

$$Yaw_{RT_D} = Yaw_{TR_D} + 180 \quad (6)$$

Similarly, based on the difference in altitude between Tx and Rx, as shown in Fig. 2b), it is possible to calculate the vertical Pitch angle $Pitch_{TR}$ at which Tx “sees” Rx in radians as:

$$Pitch_{TR} = \text{atan2}(\Delta H_{TR}, D_{TR}) \quad (7)$$

where D_{TR} is the distance between Tx and Rx, and R is the radius of the Earth ($R=6371000$),

$$D_{TR} = 2R \arcsin \sqrt{\sin^2\left(\frac{\varphi_R - \varphi_T}{2}\right) + \cos(\varphi_T) \cos(\varphi_R) \sin^2\left(\frac{\lambda_R - \lambda_T}{2}\right)} \quad (8)$$

and ΔH_{TR} is the difference in altitude between Rx and Tx:

$$\Delta H_{TR} = H_R - H_T \quad (9)$$

Finally, the Pitch angle can also be expressed in degrees as $Pitch_{TR_D}$:

$$Pitch_{TR_D} = \text{atan2}(\Delta H_{TR}, D_{TR}) 180/\pi \quad (10)$$

The proposed method enables the direct calculation of the horizontal Yaw angle and the vertical Pitch angle at which Tx “sees” Rx, based solely on the measured GPS coordinates of the Tx and Rx nodes.

3.2 Algorithm for Antenna Spatial Orientation Control Method

Based on the developed method, it is possible to design a compact algorithm suitable for software implementation on a standard 32-bit microcontroller (Fig. 3.). The algorithm can be implemented independently on both the Tx and Rx nodes. The input data for the algorithm are the GPS position of the remote node and the GPS position of the local node.

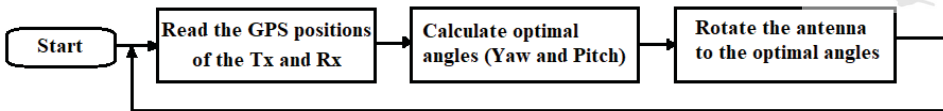


Figure 3. Algorithm of the method

The outputs of the algorithm are the horizontal Yaw angle and the vertical Pitch angle required to orient the local antenna toward the remote node. In order to optimize link performance, appropriate directional antennas with gain should be used. Antenna steering can be performed using suitable servo motors and angle measurement sensors. Thus, a microcontroller-based system with two servo mechanisms would be implemented at each node, together with sensor fusion of positional and angular data. Optimal angle estimation can be achieved based on the fusion of sensor measurements. Horizontal angle (azimuth) measurement can be performed using a digital compass, while the Pitch angle (elevation) can be measured using a tilt sensor.

3.3 Experimental Verification of the Adaptive Antenna Steering Method

Experimental measurements were conducted in a rural area outside the city. The Tx node was implemented using an ESP32-S3 microcontroller development platform operating in the 2.4 GHz WiFi frequency band, with a transmission power of 100 mW and a directional antenna with a gain of 9 dBi (Fig. 4). The ESP32-S3 microcontroller includes a 2.4 GHz Wi-Fi adapter and has sufficient processing power and memory to control the antenna direction. The NEO-6M GPS module obtains the transmitter's GPS location, while angle sensors measure the Pitch and Yaw angles. Two servo motors are used to position the antenna according to the calculated Pitch and Yaw angles. The 5V/5A power supply unit provides sufficient power for all components. The receiver (Rx) node was represented by a mobile phone equipped with an antenna gain of approximately 2 dBi. Multiple measurements were performed under two different experimental scenarios.

Scenario 1: Range measurement with a fixed antenna orientation

In this scenario, the Tx antenna was kept in a fixed direction while the Rx node was moved to record the critical Tx signal strength level of -80 dBm. The goal of this measurement was to obtain a curve representing the spatial region within which the Rx receives a stable signal from the Tx.

Scenario 2: Range measurement with adaptive antenna steering toward the receiver The goal of this measurement was to obtain a curve representing the spatial area where the Rx receives a stable signal from the Tx when adaptive antenna steering is applied.

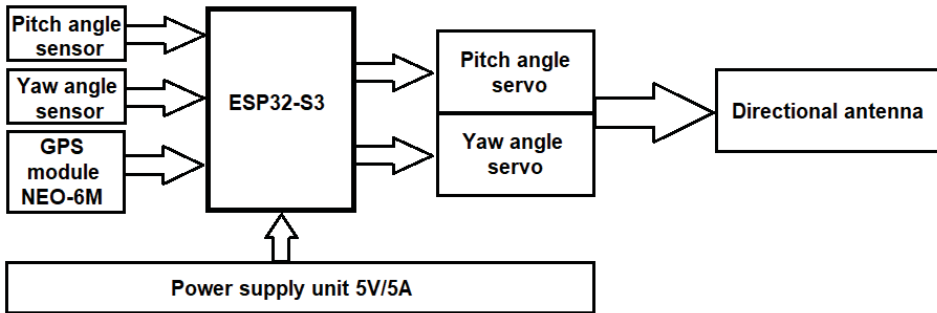


Figure 4. Block diagram of the transmitter hardware implementation

4. Results and Discussion

Based on multiple repeated measurements in both scenarios, the curves shown in Fig. 5. were constructed, illustrating the range of the Tx signal. Measurements from the first scenario are marked with empty squares. Through interpolation, a dashed line was obtained, representing the maximum Tx range (signal level of -80 dBm). According to the ellipsoidal radiation pattern of the Tx antenna, an ellipsoidal region was obtained within which a stable link was maintained, with a maximum range of 170 m. In the second scenario, with adaptive antenna steering toward the receiver, the effective communication coverage within the movement zone of the Rx node increased on average by a factor of 2–3 while maintaining a stable signal strength level of -80 dBm.

The obtained interpolated trajectory is approximately semicircular in shape. Thanks to adaptive antenna steering in both Yaw (azimuth) and Pitch (elevation), an increased maximum range of approximately 200 m was achieved. A statistical analysis of 30 repeated link-range measurements conducted under identical conditions showed a small dispersion of the experimental results. Based on the experimental results, it can be concluded that adaptive antenna steering at one or both ends of a point-to-point link can significantly increase the range of an RC link.

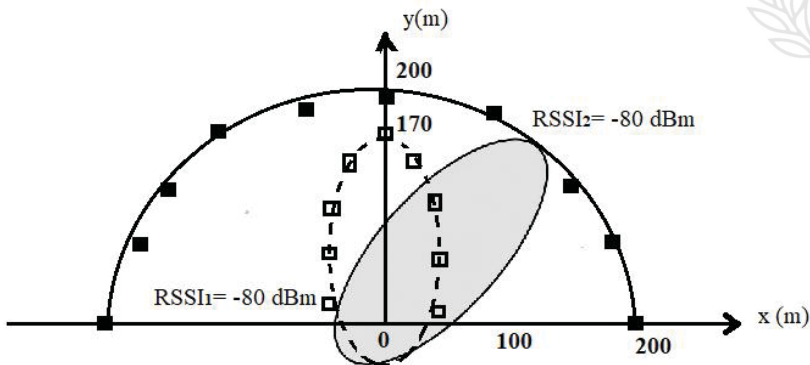


Figure 5. Graphical representation of link range with a fixed antenna orientation and with an antenna steered toward the receiver.

5. Conclusion

The developed method and corresponding algorithm have been implemented and experimentally verified under real-world conditions. Depending on the communication link usage scenario, appropriate antennas can be selected at both

ends of the point-to-point link. By using directional antennas at both ends of the point-to-point link, maximum range can be achieved. However, this requires continuous and precise antenna alignment. The achieved results are competitive with those obtained using conventional methods based on multiple antennas with different fixed orientations, where the system, through the application of an appropriate algorithm, selects the signal from the antenna receiving the strongest signal. The implementation of the proposed solution is of similar complexity to a solution employing multiple antennas on the Tx and Rx modules. As the link range is significantly influenced by terrain characteristics, interference, antenna placement, antenna gain, and weather conditions, direct comparison of the obtained experimental results with those reported in similar studies is limited.

References

- [1] Barriquello C. H., Silva F. E. S., Bernardon D. P., Canha L. N., M. J. S. Ramos and Porto D. S., (2017), *Fundamentals of Wireless Communication Link Design for Networked Robotics, Service Robots*, IntechOpen 2017, DOI: 10.5772/intechopen.69873.
- [2] Karabegović I., Karabegović E., Mahmić M., Husak E., (2019), *New Technologies, Development and Applications*, pp. 3-14, Springer International Publishing.
- [3] Ali S., Qaisar S. B., Saeed H., Khan M. F., Naeem M. and Anpalagan A., (2015). *Sensors* 2015, 15(4), 7172-7205; <https://doi.org/10.3390/s150407172>.
- [4] Fang-Jing W. & Yu-Fen K. & Yu-Chee T., (2011), From wireless sensor networks towards cyber physical systems, *Pervasive and Mobile Computing* 7(4):397-413, DOI: 10.1016/j.pmcj.2011.03.
- [5] Chih-Yu L. & Zeadally S. & Tzung-Shi C. & Chih-Yung C., (2012). Enabling Cyber Physical Systems with Wireless Sensor Networking Technologies, *International Journal of Distributed Sensor Networks*.
- [6] Zahraa A. K. & Aljebori S. T., (2021), An Optimization Technique for Wireless Communication based on Mesh Topology, 7th International Conference on Signal Processing and Communication (ICSC), pp. 15-20. DOI: 10.1109/ICSC53193.2021.9673383.
- [7] Liu, J., Kou, T., Chen, Q., Sherali, H. D., (2012), *On Wireless Network Infrastructure Optimization for Cyber-Physical Systems in Future Smart Buildings* In: Wang, X., Zheng, R., Jing, T., Xing, K. (eds) *Wireless Algorithms, Systems, and Applications. WASA 2012, Lecture Notes in Computer Science*, vol 7405. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31869-6_53.
- [8] Shen K., Zhao Z., Chen Y., Zhang Z. and Victor H. C., (2024), *Accelerating Quadratic Transform and WMMSE*, in *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 11, pp. 3110-3124.

- [9] Vorobyov S. A., Cui S., Eldar Y. C., Ma W. and Utschick W., (2012), *Optimization Techniques in Wireless Communication Systems*, Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Volume 2012, Article ID 489794, 21 pages, doi:10.1155/2012/489794.
- [10] Qiu W. & Hong-Ning D. & Zibin Z. & Muhammad I. & Athanasios V., (2017), *On Connectivity of Wireless Sensor Networks with Directional Antennas*, Sensors, 17. 134. 10.3390/s17010134.



An Explainable Artificial Intelligence Framework for Predictive Maintenance Using Hybrid Machine Learning and Multi-Task Transformer Models

Milica Tufegdžić^{*1}

Abstract: *An integrated explainable artificial intelligence (XAI) framework for predictive maintenance is proposed, focusing on failure prediction, Remaining Useful Life (RUL) estimation, and real-time asset monitoring. The framework is evaluated using a synthetically generated predictive maintenance data simulating 50 machines over time, producing realistic multivariate sensor signals that degrade with wear and occasional shocks, then converted into a supervised dataset for failure prediction and RUL estimation. The study evaluates several classical machine learning (ML) models, such as XGBoost (Extreme Gradient Boosting), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LogR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), for classification and regression tasks. XGBoost demonstrated strong overall performance, achieving the best RUL regression performance and competitive classification results, while RF and GB achieved comparable or slightly better results for individual classification metrics. A multi-task transformer is introduced to jointly model failure probability, uncertainty-aware RUL, anomaly reconstruction, and survival-based time-to-failure estimation, thereby providing a unified framework for modeling multiple temporal predictive-maintenance objectives. An autoencoder is included as a supplementary reconstruction-based component for anomaly scoring, while a risk function combines failure probability and RUL to enable real-time decision-making. To ensure interpretability, SHapley Additive exPlanations (SHAP) analysis is used to quantify feature importance at the sensor level. The framework is evaluated using Receiver Operating Characteristic (ROC) curves and Precision-Recall curves, along with Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) metrics. These results demonstrate the strong performance of ensemble models, particularly XGBoost. Overall, the framework demonstrates the potential of combining tree-based models, transformer-based temporal learning, and SHAP-based interpretability for predictive maintenance under synthetic sensor conditions.*

Keywords: multi-task learning, transformer models, failure prediction, remaining lifetime estimation, anomaly detection, industrial monitoring

^{*1}Academy of Professional Studies Šumadija, Kragujevac, Serbia
E-mail: mtufegdzc@asss.edu.rs

1. Introduction

In modern industrial environments, the increasing reliance on interconnected cyber-physical systems has intensified the need for intelligent tools that ensure operational reliability and cost efficiency. Predictive Maintenance (PdM) has emerged as a transformative approach that leverages data-driven methods to anticipate equipment failures and optimize maintenance schedules, thereby minimizing unexpected downtime and extending asset life cycles. However, despite significant progress in machine learning (ML)-based PdM solutions, challenges remain in handling high-dimensional temporal sensor data, capturing complex degradation behaviors, and providing transparent, interpretable insights into model decisions. These challenges hinder the adoption of AI-driven maintenance systems in safety-critical and regulated industrial applications.

Predictive maintenance is a condition-based approach that uses machine condition data to detect and predict faults before equipment fails. Diagnostic focuses on detecting and identifying faults using either hardware redundancy (multiple sensors) or analytical redundancy based on models and data. Prognostics goes a step further by predicting future failures and estimating the RUL of equipment. Today, data-driven ML methods are widely used in predictive maintenance because Internet of Things (IoT) systems generate large amounts of industrial data, making these techniques effective and cost-efficient for complex manufacturing systems [1]. Through preprocessing and analytics, this data can be transformed into meaningful insights about manufacturing processes, supporting decision-making, maintenance scheduling, failure detection, Remaining Useful Life (RUL) and health index estimation, cost reduction, and improved worker safety [2].

PdM uses historical data, real-time sensor readings, and ML algorithms to predict equipment failures before they occur. It analyzes patterns and detects anomalies to reduce unplanned downtime, optimize maintenance schedules, lower costs, and extend asset lifespan [3]. PdM leverages real-time data analytics and ML-based models for efficient maintenance scheduling and improved equipment longevity [4]. The aim of PdM is not only to collect process parameters, but also to monitor the physical health of equipment using data such as pressure, vibration, temperature, acoustics, and flow rate. This information is used for fault detection, early anomaly identification, equipment health assessment, and prediction of future system states [5].

Predictive maintenance is further advanced by the emergence of the Artificial Intelligence of Things, which integrates IoT sensor networks with Artificial Intelligence (AI) algorithms to analyze real-time operational data for anomaly detection, fault prediction, and maintenance recommendations. This integration enables autonomous, scalable, and intelligent maintenance systems aligned with Industry 5.0, while edge computing, federated learning, and hybrid AI models

enhance responsiveness, privacy, and adaptability [6]. RUL prediction is a core aspect of predictive maintenance that reduces downtime and maintenance costs while estimating the remaining operational life of a system or component before failure [7].

Recent advancements in Explainable Artificial Intelligence (XAI) offer a promising pathway toward trustworthy and interpretable PdM systems. By integrating XAI techniques, models can not only predict failures and estimate RUL more accurately but also justify their predictions through meaningful, human-understandable explanations [8]. This is particularly crucial in industrial contexts where maintenance decisions must be justified to engineers, operators, and decision-makers. XAI in predictive maintenance includes various methods grouped into model-agnostic, model-specific, and hybrid approaches [9]. Hybrid models that integrate XAI with traditional ML are used in manufacturing, predictive maintenance, and quality control to improve decision-making and support human-AI collaboration [10]. Moreover, combining interpretable ensemble learning models with temporal deep learning architectures enables more robust generalization across varying operational conditions and sensor noise levels [11].

This study proposes an integrated XAI Framework for Predictive Maintenance, combining hybrid ML models and a multi-task transformer architecture. The framework jointly addresses failure prediction, uncertainty-aware RUL estimation, anomaly detection, and time-to-failure modeling, while ensuring interpretability at the feature and sensor levels. Synthetic multivariate time-series data simulating machinery degradation and shock events are used to evaluate model performance under realistic conditions. Classical ML models, such as Extreme Gradient Boosting (XGBoost), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LogR), Linear regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), are benchmarked to select optimal classifiers and regressors for PdM tasks. Ensemble methods, particularly XGBoost, demonstrate strong performance across the evaluated classification and regression tasks.

To enhance temporal modeling and uncertainty quantification, a multi-task transformer is introduced, complemented by an autoencoder as a supplementary reconstruction-based component for anomaly scoring. A composite risk function integrating failure probability and RUL is proposed to support real-time maintenance decision-making. Additionally, SHapley Additive exPlanations (SHAP) are employed to quantify sensor-level contributions, promoting transparency and operational trust. Evaluations using Receiver Operating Characteristic (ROC) and Precision-Recall curves, along with Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) metrics, confirm the effectiveness of the proposed hybrid and explainable approach.

Overall, this framework bridges the gap between predictive performance and interpretability, demonstrating that the integration of ensemble tree models, transformer-based temporal learning, and SHAP-driven explainability can significantly enhance the robustness, accuracy, and transparency of predictive maintenance systems-laying the groundwork for next-generation intelligent and trustworthy industrial asset management.

2. Literature Review

The AI predictive maintenance framework that combines Convolutional Neural Network (CNN) for spatial feature extraction, Gated Recurrent Unit (GRU) for temporal pattern learning, and Transformers for long-range dependencies, using ensemble stacking and Bayesian Optimization for stable and efficient model tuning, was evaluated on real smart manufacturing multi-sensor data, including vibration, temperature, acoustic, and pressure signals, to monitor machine condition over time [12]. RF, XGBoost, SVM, and ANN models were evaluated using standard performance metrics. Results show that RF and XGBoost achieved the best accuracy and robustness, demonstrating the value of data-driven fault prediction for reducing downtime and improving proactive maintenance [13].

The approach, which combines a Transformer backbone with Monte Carlo Dropout and Conformal Prediction to capture epistemic and aleatoric uncertainty while improving calibration through Bayesian approximation and distribution-free intervals, enabling reliable confidence estimates that support early fault detection and prioritized maintenance, was validated on the NASA C-MAPSS dataset and benchmarked against existing methods [14]. Light Gradient Boosting Machine (LightGBM) model with Dropouts meet Multiple Additive Regression Trees booster tuning was developed for predictive maintenance using five datasets from 100 machines containing telemetry, error, maintenance, and failure data. The data were processed and feature-engineered, and the trained model was used to classify short-term failure risk [15]. A conceptual NLP-Hybrid Intelligence framework is proposed for smart manufacturing condition monitoring and predictive maintenance, integrating sensor analytics with NLP-based contextual industrial knowledge [16]. A hybrid predictive maintenance framework that combines Long Short-Term Memory (LSTM) autoencoder-based unsupervised anomaly detection with supervised fault prediction using an MLP was validated on a digital-twin dataset simulating five years of industrial operation and addresses class imbalance and operational variability [11].

ML algorithms used for RUL prediction in milling operations include Stochastic Gradient Descent (SGD) Regressor, RF Regressor, Decision Tree Regressor, Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP) [17]. RUL prediction models were developed using LR, SVR, RF, XGBoost, ridge regression, and artificial neural networks (ANN). This ML-based RUL prediction

was combined with Weibull distribution-based reliability analysis and a block replacement model with minimal repair to optimize preventive maintenance strategies [18]. ML algorithms such as SVR, RF, and XGBoost are used to predict the RUL of lithium-ion batteries based on real-world degradation data [19]. The model, which applied normalization and PCA for preprocessing, used interpolation and grid search for optimization, and employed MLP to predict engine RUL, was tested on NASA turbo engine datasets [20]. PdM framework combines a probabilistic neural network for RUL prediction with a Quantile regression deep Q-network distributional reinforcement learning agent. The model analyzes sensor data to estimate RUL, measure prediction uncertainty, and assess failure probability [21].

A data-driven ML approach using ANN with Neural Basis Networks combined with Fully Convolutional Networks and Autoencoder-based Convolutional Networks was developed to model degradation paths, predict failures, and estimate RUL in complex manufacturing systems. The method improves prediction accuracy by using an enhanced NBN training algorithm with an accumulative design that preserves the monotonic degradation trend and effectively handles large, complex multi-machine datasets [22].

A hybrid approach combining supervised learning using a MLP and reinforcement learning using Q-learning was proposed to improve RUL prediction accuracy and precision, and was evaluated on the C-MAPSS dataset [23]. A hybrid predictive maintenance framework combined physics-based modeling with a bidirectional Long Short-Term Memory (BiLSTM)-based AI model, using model residuals to enhance fault diagnosis, anomaly detection, and RUL prediction for rotating machinery such as bearing wear and shaft imbalance [24]. The hybrid deep learning model that was developed for RUL point predictions, consisted of an input layer defined by the window size and number of features, followed by multiple 1D-CNN layers with ReLU activation for feature extraction and Bi-LSTM layers for temporal learning. Fully connected dense layers with Exponential Linear Unit activation further process the learned representations, and a final linear output layer with a single neuron generates the prediction [25].

A RUL prediction method that combines BiLSTM and Transformer architectures is proposed. First, a BiLSTM-based denoising autoencoder extracts features from multivariate sensor data, which are then used by a Transformer encoder to estimate RUL [26]. A multi-task transformer encoder and customized gate control model was proposed to jointly predict tool RUL and wear stages using a transformer encoder for shared feature extraction and customized gate control for task-specific and shared features. It was trained using a dynamic adaptive multi-task loss function to improve learning efficiency [27]. The Multi-Input Data-1DC+LRT model was proposed as a multi-task learning framework integrating a multi-input 1D CNN (and a low-rank transformer to process a high-dimensional

sensor, multi-condition, and health indicator data while reducing computational complexity, in order to predict RUL [28].

XAI improves AI-driven maintenance by making ML models more transparent, interpretable, and trustworthy, addressing the "black box" nature of many systems and enabling smarter decision-making. However, challenges remain in integrating XAI with IoT-based digital twins, balancing explainability with cybersecurity, and ensuring scalability for reliable and efficient deployment [29]. A XAI framework for failure detection and classification uses symbolic expressions from a genetic programming classifier. The method applies normalization, oversampling, cross-validation, and hyperparameter tuning, then builds a meta-dataset to train RF models and form class-specific voting ensembles [30]. A hybrid ensemble predictive maintenance framework combines multiple ML classifiers to overcome key limitations in PdM, integrating RF, XGBoost, and LightGBM with SHAP for transparent and interpretable decision-making. The models are trained using 5-fold cross-validation, with soft-voting and weighted probability averaging for final prediction, while hyperparameter tuning and SHAP analysis are applied within a unified pipeline to ensure reproducibility and explainability [31].

RUL estimation on the NASA C-MAPSS dataset is performed using a stacked LSTM model for temporal degradation learning. SHAP and TimeSHAP provide sensor- and time-step-level interpretability, while robustness under Gaussian, drift, and spike noise is evaluated using Spearman correlation and Top-K consistency metrics [32]. A predictive maintenance framework for aircraft engines was developed that combines accurate RUL prediction with cost-aware maintenance scheduling using an ensemble of LightGBM, CatBoost, and GB to improve prediction accuracy and stability, while SHAP analysis enhances interpretability by identifying key sensors and operational factors [33].

3. Methodology

3.1. Synthetic Dataset Generation for Predictive Maintenance

A fully synthetic dataset was constructed to address the limited availability of real industrial condition monitoring data, which is typically constrained by failure scarcity, class imbalance, and high acquisition cost. This simulation-based setup provides a controlled environment for modeling machine degradation dynamics and evaluating Transformer-based predictive maintenance architectures under consistent conditions.

The dataset represents a fleet of machines operating under heterogeneous conditions over time. Its structure is defined through configurable parameters, including the number of machines, temporal resolution per unit, and sliding-window length. Each machine trajectory is generated independently to reflect

variability across industrial assets, matching the multi-sequence input formulation used in the Transformer model.

Machine health evolution is modeled through a nonlinear wear process combined with stochastic shock events that emulate abrupt stress, overload, and operational disturbances. These mechanisms generate multivariate operational signals, including temperature, vibration, pressure, current, and wear-related measurements. The raw sensor observations are subsequently transformed into engineered features for the real-time inference pipeline. Specifically, statistical descriptors and condition indicators are calculated from the sensor measurements, including mean, standard deviation, maximum value, the latest wear state, and an aggregated vibration/shock measure. The resulting feature representation is described explicitly in Section 3.2. Degradation manifests as increasing temperature and vibration trends and decreasing pressure, while Gaussian noise is introduced to emulate measurement uncertainty and environmental variability. Failure generation follows a stress-driven probabilistic mechanism in which accumulated degradation increases the likelihood of failure over time. A stochastic decision rule determines event occurrence at each time step, and failures are assigned to predefined classes such as thermal, vibration-related, and pressure leakage faults. This formulation supports both binary failure prediction and more granular diagnostic learning consistent with multi-task outputs.

For real-time inference, incoming sensor observations are transformed into a fixed feature representation using a dedicated feature-engineering procedure. The procedure calculates statistical descriptors of temperature, vibration, and pressure, together with the mean current, the most recent wear value, and an aggregated vibration-based shock indicator. A machine identifier is retained separately for machine-level traceability. This engineered representation is subsequently passed to the trained predictive models during streaming inference. Supervision signals are generated in parallel with the input construction. Binary labels indicate whether failure occurs within a future prediction horizon, while Remaining Useful Life (RUL) quantifies time-to-failure. The resulting dataset is stored in CSV format and contains engineered features, machine identifiers, and target variables, directly supporting the multi-task learning framework of failure prediction, RUL estimation, and related auxiliary objectives used in the Transformer architecture.

3.2. Dataset Analysis

The dataset comprises 1,350 observations collected from 50 distinct machines. Two variables, label and RUL, are used as target variables. The label variable represents binary failure status, whereas RUL represents the remaining useful life in cycles. The dataset contains 13 columns in total, comprising ten engineered operational features, one machine identifier (`machine_id`), and two target variables (label and RUL).

The conventional machine-learning benchmark, streaming-oriented predictive analysis, and SHAP-based interpretability analysis use a common engineered predictor representation. The feature-engineering procedure generates ten operational features: temp_mean, temp_std, temp_max, vib_mean, vib_std, pressure_mean, pressure_std, current_mean, wear_last, and shock_sum. These engineered features constitute the input representation used by XGBoost, Random Forest, Gradient Boosting, Logistic Regression, SVM, KNN, and their regression counterparts, ensuring a consistent feature space across the conventional machine-learning benchmark. The same engineered feature representation is also used in the real-time streaming procedure and for SHAP analysis of the XGBoost classification model. The machine_id variable is retained solely as an identifier and is excluded from the input features of all predictive models. It is used only for machine-level identification, traceability, and data partitioning where applicable. The distinction between the different variable categories can therefore be summarized as follows: (i) raw/original predictor variables are the variables present in the dataset after the exclusion of the two targets; (ii) engineered features are the ten derived sensors and condition descriptors generated by the streaming feature-engineering procedure; (iii) machine_id is an identifier used for machine-level traceability; and (iv) label and RUL are target variables used for classification and regression, respectively. In addition to these sensor measurements, the dataset contains two target variables: a binary failure label indicating normal versus failure/alert conditions, and a continuous RUL value ranging from 0 to 270 cycles.

The dataset exhibits a pronounced class imbalance in the failure label, with 1,145 samples (84.8%) corresponding to normal operation and 205 samples (15.2%) representing failure states. This imbalance reflects real industrial settings where failures are relatively rare compared to normal operating conditions and must be accounted for during model development to avoid biased predictions. The boxplot analysis (Figure 1) further confirms clear separation between normal and failure states across key operational metrics. Failure conditions are associated with higher temperature (≈ 85.4 °C vs. 78.4 °C), increased vibration (2.67 mm/s vs. 1.67 mm/s), elevated tool wear (median ≈ 0.71 vs. 0.32), and reduced pressure (≈ 28.8 bars vs. 31.6 bars). These shifts are consistent and non-overlapping in several cases, particularly for vibration and wear, indicating strong discriminative power for classification tasks.

From a modeling perspective, these patterns suggest high separability between classes, making the dataset suitable for supervised learning approaches such as RF or XGBoost. The monotonic trends across features also indicate potential utility for early anomaly detection and predictive maintenance systems.

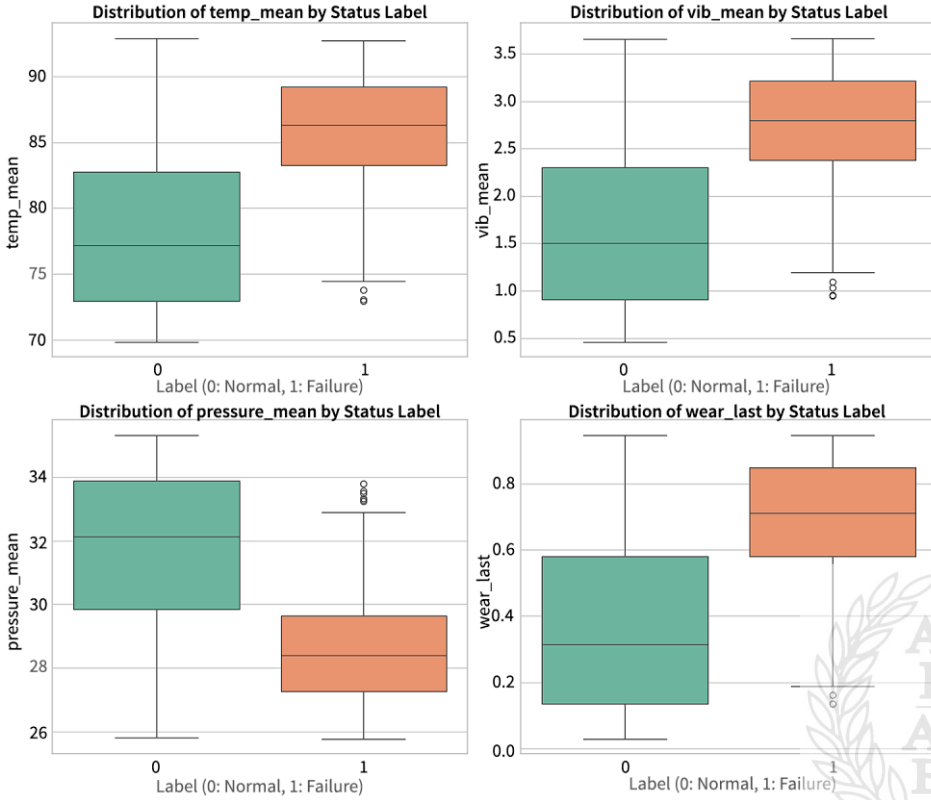


Figure 1. Boxplot analysis

The RUL distribution (Figure 2) is heavily right-skewed, with most values concentrated between 0 and 50 cycles and a long tail extending to 270 cycles. This reflects a dataset dominated by late-stage degradation samples, with additional variability in machine lifespans across the fleet. Such skewness suggests that nonlinear models or transformations may be more appropriate than standard linear regression methods. The failure label distribution (Figure 2) further reinforces the need for careful modeling, as naive accuracy-based approaches may be misleading due to class imbalance. Evaluation metrics such as recall and F1-score are more appropriate, and techniques such as class weighting, SMOTE, or stratified cross-validation are recommended to ensure robust learning of minority failure cases.

Analysis of degradation behavior shows clear relationships between sensor variables and RUL (Figure 3).

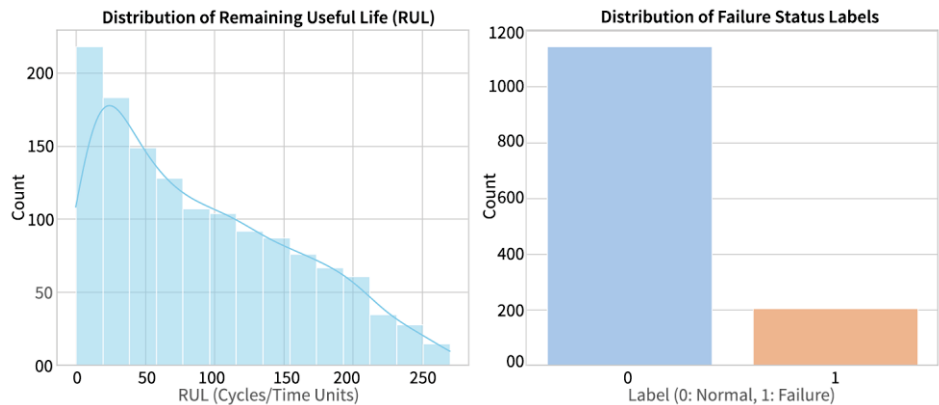


Figure 2. RUL and failure label distribution

Wear_last, temp_mean, vib_mean, and current_mean exhibit strong negative correlations with RUL (≈ -0.79), while pressure_mean shows a strong positive correlation ($\approx +0.79$), reflecting consistent and opposing trends during equipment deterioration. These variables collectively represent the primary degradation signals of the system.

Marked multicollinearity is observed among temperatures, vibration, and current measurements ($\approx +0.99$), alongside their strong inverse relationship with pressure_mean (≈ -0.99). This indicates substantial redundancy across sensor channels, which may justify dimensionality reduction in simpler models, although ensemble-based methods are generally more robust to such dependencies.

A comparison of feature types indicates that aggregated statistics are more informative than extreme values. Temp_mean shows a stronger relationship with RUL (-0.79) and a moderate association with failure ($+0.38$) compared to temp_max (-0.61 and $+0.28$), indicating that sustained operating conditions provide more reliable degradation information than short-term fluctuations. In contrast, temp_std, vib_std, and pressure_std exhibit negligible correlation with the targets, indicating limited predictive value. Overall, the feature set contains a small number of highly informative variables, substantial redundancy among correlated sensors, and several weak descriptors, supporting selective feature engineering.

Overall, the dataset reflects a coherent degradation process in which increasing thermal, mechanical, and electrical stress is associated with decreasing pressure and RUL.

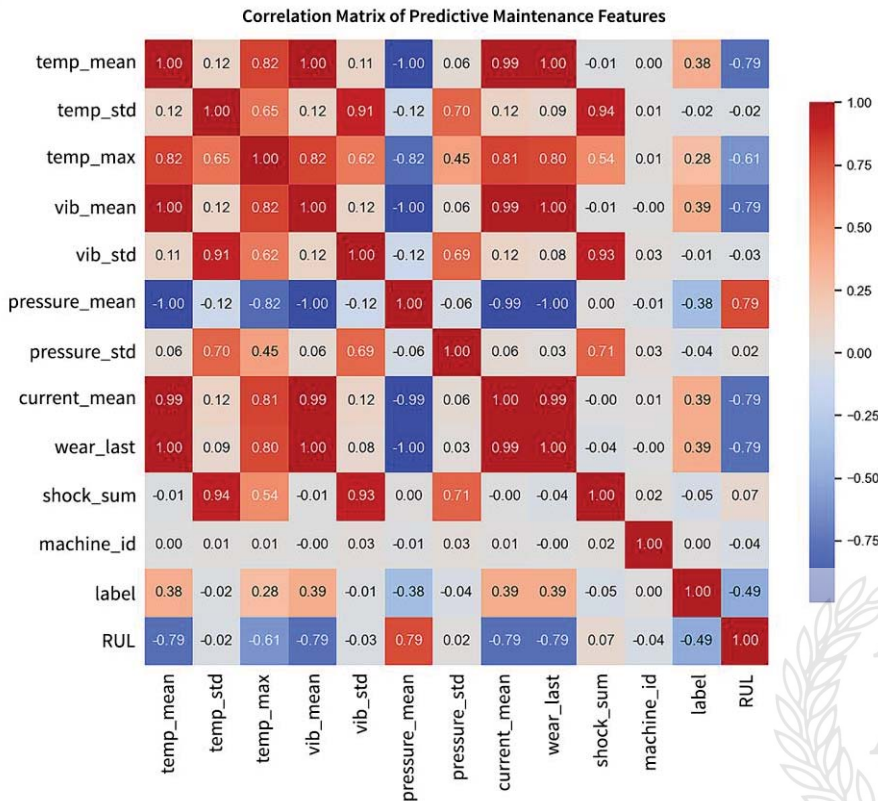


Figure 3. Correlation matrix

This structured behavior, together with clear class separability and strong inter-feature relationships, indicates that the dataset is well suited for predictive maintenance and failure prediction tasks.

3.3. Predictive Maintenance Framework

This study proposes a unified predictive maintenance framework for multivariate time-series sensor data. The framework integrates classical ML, deep learning models, anomaly detection methods, and XAI techniques to support both failure prediction and RUL estimation. The hybrid design enables accurate classification of machine states, reliable degradation modeling, real-time risk assessment, and interpretable decision-making. By combining complementary modeling approaches, the framework improves predictive robustness while ensuring transparency in industrial applications. The framework was developed in Python using several scientific and ML libraries, including NumPy, Pandas, Scikit-learn, XGBoost, TensorFlow/Keras, SHAP, Matplotlib, and Seaborn. These modules

were employed for data generation, preprocessing, feature engineering, model development, evaluation, visualization, and explainability analysis.

The proposed methodology is structured into five main components: dataset preparation, feature engineering, model development, evaluation, and real-time simulation. The experimental dataset comprises multivariate industrial sensor signals, including temperature, vibration, pressure, current, and machine wear indicators. Two predictive tasks are defined: (i) binary failure prediction and (ii) regression-based RUL estimation. The dataset is organized such that input features are derived from sensor measurements, while target variables correspond to failure labels and RUL values. The dataset was partitioned into training and testing subsets using an 80:20 observation-level split. For the classification task, stratified sampling was applied to preserve the proportion of the two failure classes in the training and testing subsets. The split was performed independently of machine_id; consequently, observations from the same machine may be present in both the training and testing subsets. This should be considered when interpreting the reported results, as machine-specific patterns may be partially shared between the two subsets. A machine-wise split based on machine_id would provide a stricter evaluation of model generalization to previously unseen machines and is therefore identified as an important direction for future validation. The experimental pipeline uses a unified engineered feature representation for conventional ML modeling, SHAP-based interpretability analysis, and real-time streaming inference. Raw sensor observations are transformed into ten engineered sensors and condition features derived from temperature, vibration, pressure, current, wear, and vibration-based shock measurements. Within each window, statistical descriptors are extracted to capture temporal dynamics and degradation behavior. These include the mean, standard deviation, and maximum values of temperature, vibration, and pressure signals, as well as aggregated vibration energy and the most recent machine wear state. This transformation converts sequential sensor data into structured feature vectors suitable for supervised learning algorithms. The machine_id variable is retained solely as an identifier and is excluded from model training. All benchmark models are trained using the same engineered predictor representation to ensure a fair and consistent comparison.

A range of supervised learning models is evaluated for both predictive tasks. For failure classification, the models include XGBoost, RF, GB, LogR, SVM, and KNN. For RUL estimation, XGBoost Regressor, RF Regressor, GB Regressor, LR, SVR, and KNN Regressor are employed.

Model performance is assessed using accuracy, F1-score, and ROC-AUC for classification, and mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) for regression. Models are ranked based on

F1-score for classification and MAE for regression, and the best-performing models are selected for further analysis. Receiver operating characteristic curves are used to compare classification performance, while bar plots are utilized to summarize regression results across evaluation metrics.

For failure prediction, probabilistic outputs from the selected classifier are evaluated using precision–recall analysis. The Precision–Recall curve is used to assess the trade-off between precision and recall across different classification thresholds. The decision threshold is selected by maximizing the F1-score across candidate values.

A composite risk score is introduced to combine the predicted probability of failure with the estimated Remaining Useful Life (RUL) for continuous maintenance risk assessment. The risk score is defined as

$$R = 0.7P_f + 0.3\exp\left(-\frac{RUL}{50}\right), \quad (1)$$

where R denotes the composite risk score, P_f is the predicted probability of failure, and RUL is the predicted remaining useful life in cycles. The coefficient 0.7 assigns a greater contribution to the predicted failure probability, while the coefficient 0.3 incorporates the contribution of the estimated RUL . The exponential term provides a smooth decreasing function of RUL : as the predicted RUL decreases, $\exp(-RUL/50)$ increases, resulting in a higher contribution to the overall risk score. The value 50 represents the RUL scaling parameter used in the implementation. This formulation therefore combines immediate failure likelihood with remaining operational lifetime, giving greater emphasis to failure probability while increasing the risk contribution when the predicted remaining lifetime is short. In the real-time monitoring procedure, a maintenance alert is triggered when the composite risk score exceeds 0.7. A real-time streaming simulation environment is implemented to replicate industrial monitoring conditions. Synthetic sensor streams are generated using probabilistic distributions that reflect operational variability. At each time step, engineered features are computed and fed into trained models to estimate failure probability and RUL . The resulting risk score is continuously updated, and maintenance alerts are triggered when the risk exceeds a predefined threshold.

An unsupervised autoencoder is employed as a reconstruction-based anomaly scoring component. The model consists of fully connected encoder and decoder layers and is trained using the Adam optimizer with mean squared error (MSE) loss. The reconstruction error is calculated for each observation as the mean squared difference between the original input and its reconstructed representation and is used as an anomaly score. Higher reconstruction errors indicate

observations that are less consistent with the patterns learned by the autoencoder. In the present implementation, the reconstruction error is used as a continuous anomaly score; an explicit anomaly threshold and a separate quantitative anomaly-detection evaluation are not defined. Therefore, the autoencoder results are interpreted as reconstruction-based anomaly scoring rather than as a calibrated binary anomaly detector. Model interpretability is achieved using SHAP. A Kernel SHAP explainer is applied to the XGBoost classifier to quantify feature contributions to failure predictions. A representative background dataset consisting of 100 randomly selected observations from the training set is used to initialize the explainer. SHAP values are then computed for 200 observations from the test set. The feature space and feature ordering are kept identical between the training and test representations to ensure consistency during explanation. Both global and local explanations are generated, allowing identification of the most influential predictor variables in failure prediction.

Finally, a Transformer-based multi-task learning architecture is developed to jointly model multiple predictive maintenance objectives. The architecture incorporates positional encoding and stacked self-attention layers to capture temporal dependencies in sensor data. To quantitatively evaluate the Transformer component, the two model variants, Functional_1 and Functional_2, were additionally evaluated using a machine-wise train-test partition, with 40 machines used for training and 10 previously unseen machines used for testing. The machine_id variable was used only to define machine-wise data partitions and was not included as an input feature to the Transformer models. Because each machine contains 27 observations, whereas the Transformer architecture requires a sequence length of 30, the first observation of each machine was repeated for initial padding, ensuring that all sequence elements originated from the same machine. Functional_1 was evaluated for failure classification, whereas Functional_2 was evaluated for failure classification, RUL estimation, sequence reconstruction, and survival-oriented prediction. A shared representation layer feeds multiple task-specific output heads for failure classification, RUL regression with uncertainty estimation, anomaly reconstruction, and survival probability prediction. The model is trained using a weighted multi-objective loss function to balance the contributions of the individual tasks.

4. Results and Discussion

4.1. Failure Classification

The classification benchmark results for all tested ML models are presented in Table 1.

Table 1. Classification benchmark

Model	Accuracy	F1-score	ROC-AUC
XGBoost	0.859	0.441	0.833
RF	0.878	0.441	0.815
KNN	0.870	0.407	0.822
GB	0.852	0.333	0.834
LogR	0.837	0.043	0.815
SVM	0.848	0.000	0.592

The classification results summarized in Table 1 show that all evaluated models achieve relatively high accuracy, ranging from approximately 0.837 to 0.878. However, the substantially lower F1-scores indicate that this performance is influenced by class imbalance, where majority-class dominance inflates accuracy and masks poor minority-class detection. As a result, accuracy alone is not a reliable indicator of model effectiveness in this setting.

In terms of class separability, measured by ROC-AUC, GB achieves the highest score (0.834), closely followed by XGBoost (0.833). KNN also shows competitive performance (0.822), while RF and LogR are slightly lower but comparable. In contrast, SVM performs poorly (0.592), indicating limited ability to separate the classes. A clearer difference emerges in F1-score, which better reflects minority-class performance. XGBoost and RF achieve the best results (0.441), followed by KNN (0.407). GB drops to 0.333, while LogR and SVM perform poorly, with SVM failing to detect the minority class (F1 = 0). This confirms the limited suitability of linear decision boundaries for this dataset.

Overall, XGBoost provides the most balanced performance across both discrimination (ROC-AUC) and minority-class sensitivity (F1-score). Its advantage can be attributed to its gradient-boosted tree structure, which sequentially corrects errors made by previous learners while optimizing a regularized objective function. This enables better handling of complex, nonlinear relationships and subtle decision boundaries present in the sensor-derived feature space. In addition, built-in regularization and shrinkage help reduce overfitting, which is particularly beneficial under class imbalance. RF follows closely due to its robustness through bagging, while KNN remains moderately effective but less stable in high-dimensional feature spaces.

These results suggest that ensemble tree-based methods are best suited for this dataset. However, the uniformly low F1-scores across models highlights the impact of class imbalance. Addressing this through resampling strategies or class-weighted learning would likely improve minority-class detection and overall robustness.

For failure prediction, the precision-recall analysis provides additional insight into the classifier's performance under class imbalance. For the XGBoost classifier, the resulting AUPRC is 0.449, while the optimal classification threshold is 0.239. At this threshold, the classifier achieves a precision of 0.481, a recall of 0.610, and an F1-score of 0.538. The corresponding Precision–Recall curve is presented in Figure 4. The marked point on the curve represents the threshold that maximizes the F1-score.

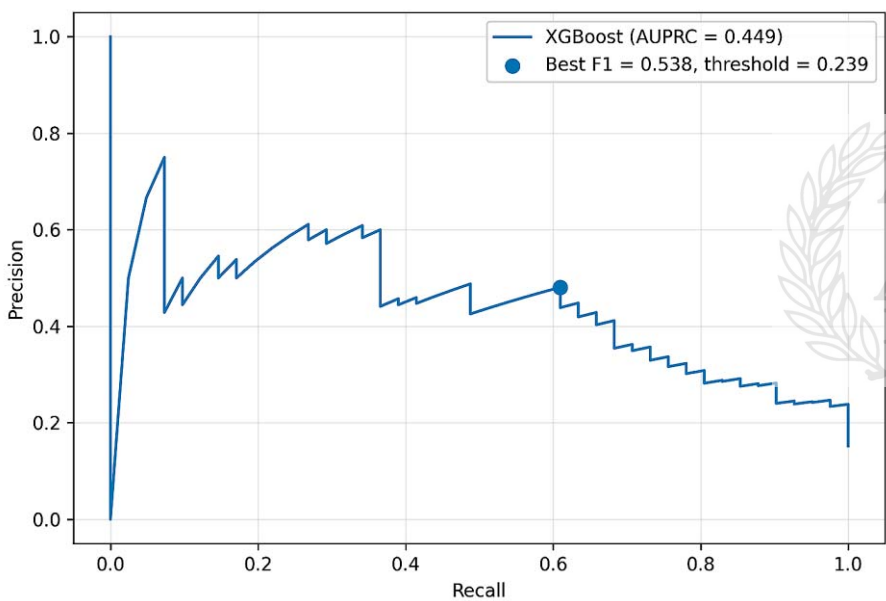


Figure 4. Precision-Recall curve for the XGBoost classifier

4.2. Remaining Useful Life Prediction

The RUL regression results show a clear hierarchy in model performance across all evaluation metrics (MAE, RMSE, and R^2), as illustrated in Figure 5. Overall, XGBoost achieves the best performance by a significant margin, followed by GB, while SVR performs the worst.

XGBoost records the lowest MAE (17.79) and RMSE (24.15), along with the highest R^2 score (0.8788), indicating both low prediction error and strong explanatory power, capturing nearly 88% of the variance in RUL. This suggests

that the model effectively learns complex nonlinear relationships in the data. GB follows with higher errors (MAE = 22.46, RMSE = 29.55) and a lower R^2 score (0.8184), remaining strong but clearly less accurate, reflecting the advantage of XGBoost's more optimized boosting and regularization strategy. KNN and RF show comparable performance. KNN achieves a slightly lower MAE (26.22) than RF (26.82), while RF performs marginally better in RMSE. Their R^2 values ($\approx 0.73 - 0.74$) indicate moderate predictive capability but limited ability to capture more complex nonlinear patterns compared to boosting-based models. LR performs worse, with a MAE of 31.91 and an R^2 of 0.6489, indicating that linear assumptions are insufficient for modeling the underlying structure of the data. SVR exhibits the weakest performance, with a high MAE (48.96), RMSE (61.92), and a low R^2 score (0.2029), explaining only about 20% of the variance. This suggests that the model is unable to effectively capture the relationships present in the feature space under the current configuration.

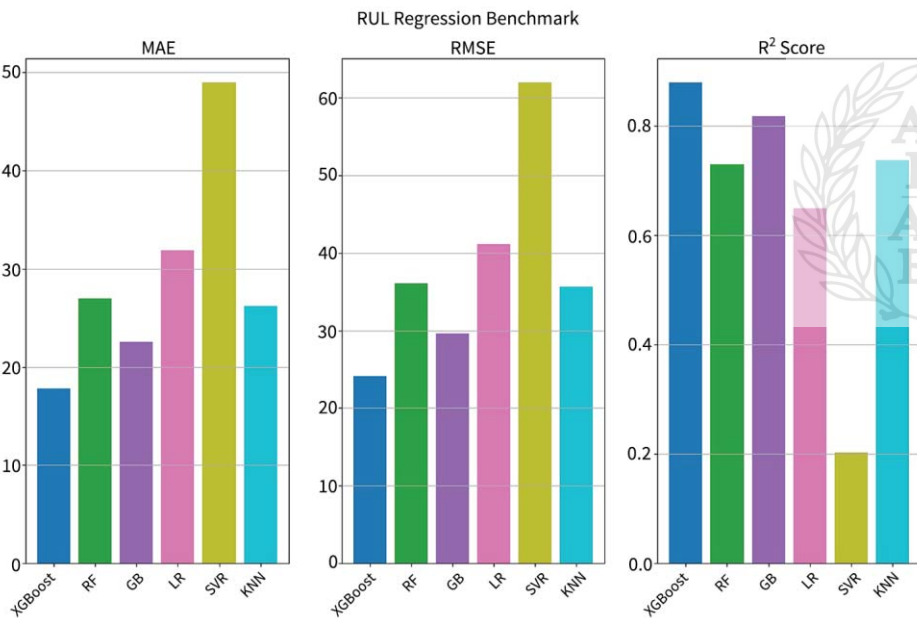


Figure 5. RUL regression results

Overall, the results favor tree-based ensemble methods. XGBoost achieves the best performance across all evaluated RUL regression metrics, while classification performance is more balanced, with XGBoost and Random Forest achieving the highest F1-score and Gradient Boosting achieving the highest ROC-AUC.

4.3. Comparative Performance Analysis Of Machine Learning Models

Across both classification and regression experiments, a consistent performance hierarchy is observed, with ensemble tree-based methods substantially outperforming linear models and SVM-based approaches. This pattern indicates that the underlying feature space is governed by nonlinear interactions and feature dependencies that are not adequately captured by simpler parametric formulations.

In the classification task, the relatively high accuracy values across all models contrast sharply with the substantially lower F1-scores, confirming the presence of class imbalance and highlighting the limited informativeness of accuracy as a standalone metric. XGBoost demonstrates the most balanced performance, achieving the highest F1-score alongside strong ROC-AUC, suggesting effective discrimination of minority failure events while maintaining overall class separability. RF exhibits comparable F1 performance, although with slightly reduced discriminative capability as reflected in ROC-AUC. GB remains competitive but shows reduced robustness in minority-class detection. KNN provides moderate performance, indicating partial sensitivity to local structure but limited generalization capacity in higher-dimensional feature space. In contrast, LogR and SVM exhibit clear limitations, with LogR failing to adequately capture minority-class structure and SVM demonstrating poor separability, likely due to insufficient alignment between kernel assumptions and the underlying data distribution.

A similar ranking emerges in the regression setting, where boosting-based methods again dominate performance. XGBoost achieves the lowest prediction error and highest R^2 , indicating a strong capability to capture complex nonlinear relationships between the predictor variables and RUL. This suggests that its iterative residual learning strategy, combined with regularization, provides a more effective bias–variance trade-off than competing methods. GB follows as a strong secondary model, although with reduced accuracy, implying that additional optimization in XGBoost contributes meaningfully to performance gains. RF and KNN occupy an intermediate position, capturing a substantial portion of variance but failing to model finer-grained nonlinear degradation dynamics. LR performs significantly worse, reinforcing the inadequacy of linear assumptions for this problem. SVR exhibits the weakest performance, indicating poor kernel alignment and limited capacity to generalize under the present feature representation.

From a unified perspective, the results indicate that ensemble tree-based methods provide strong performance for the considered predictive-maintenance tasks. XGBoost provides the strongest performance for RUL regression and achieves a competitive classification performance, while Random Forest and Gradient Boosting achieve the best results for individual classification metrics. These findings suggest that tree-based ensemble methods are well suited to the nonlinear

relationships present in the considered feature space, although the results do not establish universal superiority of a single model across all tasks and metrics. RF offers a stable but less expressive alternative, while KNN serves primarily as a locality-based baseline without competitive predictive strength. In contrast, linear models and SVM fail to adequately represent the structure of the data without substantial feature transformation or re-engineering.

Overall, the empirical evidence supports tree-based ensemble methods as strong baseline models for this predictive-maintenance problem, with XGBoost providing the strongest performance for the evaluated RUL regression task.

4.4. Cyclic Temporal Behavior of Sensor Signals

Figure 6 illustrates sensor evolution over the final ~200 cycles and reveals a strongly periodic operating regime characteristic of a cyclic industrial process rather than steady-state behavior. The repeated sawtooth profiles and abrupt resets across all signals indicate well-defined cycle boundaries, likely driven by batch operations, scheduled maintenance, or automated system restarts. This suggests that the system does not evolve continuously over time, but instead undergoes repeated operational and reset phases.

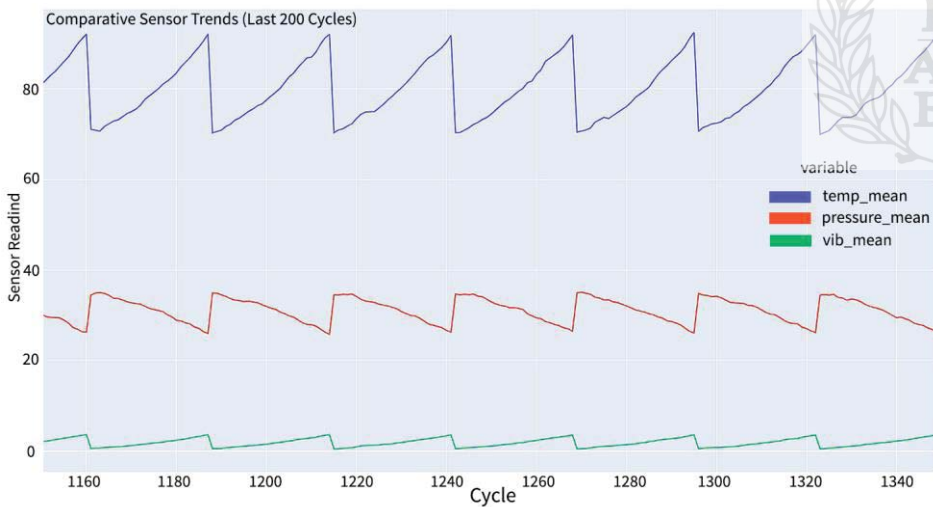


Figure 6. Sensor readings

Within each cycle, `temp_mean` increases smoothly from about 70 to 92 before abruptly returning to its initial level, reflecting continuous thermal energy buildup during operation followed by rapid cooling or reset. In contrast, `pressure_mean` follows an inverse pattern, gradually decreasing from approximately 35 to 26.

This opposing behavior is consistent with coupled thermodynamic effects, where increasing load or flow resistance leads to progressive pressure drop as thermal conditions intensify. The `vib_mean` signal shows a comparable cyclic trend to temperature, starting near zero after each reset and increasing steadily until cycle completion. This behavior reflects the gradual accumulation of mechanical stress, friction, or wear during operation. The alignment between rising temperature and vibration further suggests a shared dependence on system load, while periodic resets indicate restoration of baseline operating conditions.

Overall, the observed patterns indicate a tightly coupled thermomechanical system governed by cyclic operation. The consistent inverse relationship between temperature and pressure, together with synchronized increases in vibration, reflects structured operational loading and repeatable degradation behavior within each cycle.

The synchronized resets observed across all three signals are particularly significant. The simultaneous drops in temperature and vibration, together with the concurrent increase in pressure, suggest a deliberate system reset rather than a naturally occurring decay process. This pattern is consistent with an automated operational cycle in which system states are periodically restored to baseline conditions, indicating a controlled batch process or maintenance-driven operation rather than stochastic fluctuations or failure-induced transitions.

4.5. Streaming Framework

The streaming module simulates a real-time predictive maintenance environment through sequential generation of synthetic sensor observations over a fixed time horizon. Each observation includes key operational variables (temperature, vibration, pressure, current, and wear), sampled using probabilistic distributions to approximate realistic industrial operating conditions. A short inter-sample delay is introduced to emulate real-time data acquisition.

Incoming observations are converted into structured feature representations to ensure compatibility with trained models and enable consistent inference in a streaming setting. Two pre-trained models are applied: a classifier for estimating failure probability and a regression for predicting RUL. Their outputs are combined to compute a unified risk score that jointly reflects failure likelihood and degradation severity.

The system continuously monitors failure probability, RUL, and the derived risk score, triggering a maintenance alert when the risk exceeds a predefined threshold. The complete process is illustrated in Figure 7.


```
=== Streaming started ===
```

[1/20]	Risk: 0.052	Prob: 0.017	RUL: 101.4
[2/20]	Risk: 0.031	Prob: 0.018	RUL: 138.5
[3/20]	Risk: 0.068	Prob: 0.018	RUL: 84.7
[4/20]	Risk: 0.093	Prob: 0.037	RUL: 75.1
[5/20]	Risk: 0.035	Prob: 0.012	RUL: 120.3
[6/20]	Risk: 0.052	Prob: 0.031	RUL: 114.2
[7/20]	Risk: 0.035	Prob: 0.003	RUL: 110.4
[8/20]	Risk: 0.031	Prob: 0.005	RUL: 119.9
[9/20]	Risk: 0.036	Prob: 0.004	RUL: 109.6
[10/20]	Risk: 0.043	Prob: 0.017	RUL: 113.6
[11/20]	Risk: 0.035	Prob: 0.003	RUL: 110.2
[12/20]	Risk: 0.036	Prob: 0.005	RUL: 110.9
[13/20]	Risk: 0.078	Prob: 0.056	RUL: 102.4
[14/20]	Risk: 0.028	Prob: 0.007	RUL: 128.0
[15/20]	Risk: 0.033	Prob: 0.006	RUL: 117.6
[16/20]	Risk: 0.030	Prob: 0.008	RUL: 124.3
[17/20]	Risk: 0.042	Prob: 0.006	RUL: 103.3
[18/20]	Risk: 0.038	Prob: 0.005	RUL: 108.7
[19/20]	Risk: 0.045	Prob: 0.006	RUL: 99.7
[20/20]	Risk: 0.031	Prob: 0.007	RUL: 121.6

Figure 7. Real time streaming simulation results



The proposed streaming framework demonstrates the integration of real-time sensor monitoring, machine learning inference, and dynamic risk assessment within a predictive maintenance context. This setup supports continuous evaluation of machine health and enables timely decision-making for preventive maintenance in Industry 4.0 environments.

4.6. Transformer-Based Multi-Task Predictive Maintenance Architecture

A Transformer-based framework is proposed for predictive maintenance using multivariate sensor time-series data. The architecture uses a sequence length of 30 time steps and a 10-dimensional input representation consisting of ten engineered sensor and condition features. The machine_id variable is not used as a predictive input feature; instead, it is retained solely for identifying individual machines, grouping observations into machine-specific sequences, and supporting machine-wise data partitioning. The input is projected into a 64-dimensional embedding space, followed by sinusoidal positional encoding and stacked Transformer encoder blocks. These blocks incorporate multi-head self-attention, feed-forward

layers, residual connections, and normalization to model temporal dependencies in degradation patterns. Global average pooling is used to obtain a compact sequence representation that is shared across task-specific output heads. The architecture supports multiple predictive maintenance objectives, including binary failure classification, RUL regression with uncertainty estimation, sequence reconstruction for anomaly detection, and survival analysis for future risk estimation. The model is configured with the Adam optimizer and a weighted multi-task loss, with task-specific loss weights used to balance the individual learning objectives.

Within this framework, two model variants are introduced: Functional_1 and Functional_2. Functional_1 is a compact, single-task attention-based model focused on failure classification. It employs a lightweight attention-based architecture followed by global pooling and a dense output head, with approximately 3.8K parameters. Its reduced complexity enables fast inference and low computational demand, making it suitable for real-time and resource-constrained environments, although its modeling capacity is limited to relatively shallow temporal dependencies. In contrast, Functional_2 is a deeper Transformer-based multi-task architecture built on stacked encoder blocks, with approximately 292K parameters. It incorporates task-specific outputs for failure prediction, probabilistic RUL estimation, sequence reconstruction, and survival analysis, providing a unified architecture for modeling degradation dynamics, uncertainty, and anomaly-related behavior.

In the present implementation, both architectures are defined and compiled as components of the proposed framework. Their quantitative evaluation is performed using a separate machine-wise train-test procedure, with 40 machines used for training and 10 previously unseen machines used for testing. The results of the conventional ML benchmark are reported in Table 1, whereas Table 2 summarizes the performance of the two Transformer configurations. Functional_1 is evaluated only as a classification model, whereas Functional_2 represents the multi-task Transformer configuration and is evaluated using both classification and RUL regression metrics.

Table 2. Quantitative performance of the Transformer variants

Model	Accuracy	F1-score	ROC-AUC	MAE	RMSE	R ²
Functional 1	0.748	0.541	0.716	-	-	-
Functional 2	0.756	0.588	0.821	37.87	49.74	0.346

The results in Table 2 indicate that Functional_2 provides consistently better classification performance than Functional_1, particularly in terms of F1-score and ROC-AUC. Moreover, only Functional_2 was evaluated for RUL estimation,

demonstrating moderate agreement between the predicted and true RUL values. Figure 8 further illustrates this relationship.

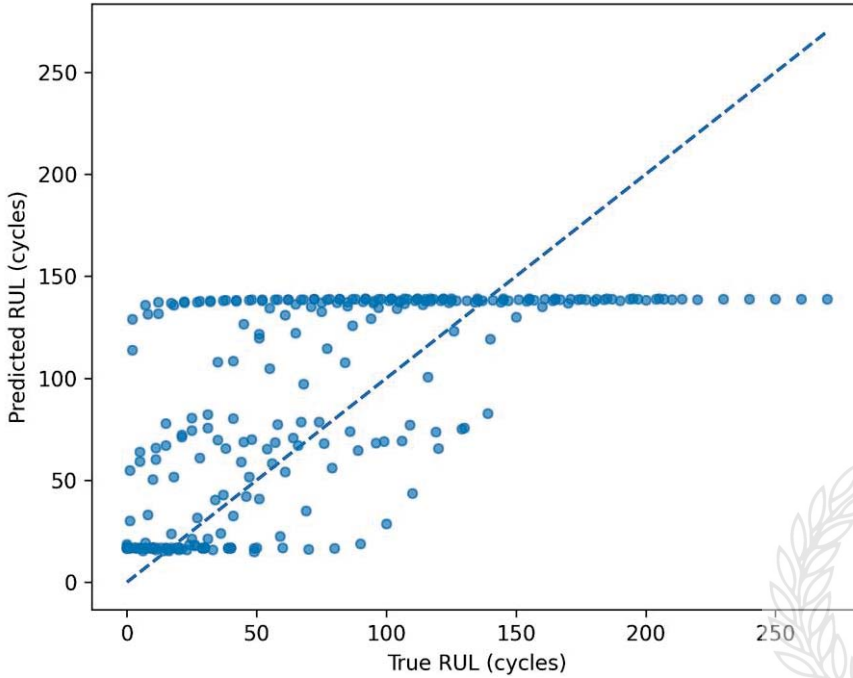


Figure 8. True versus predicted RUL values for Functional_2 model

The Precision-Recall curves for the two Transformer variants are presented in Figure 9, with Functional_2 demonstrating better overall performance than Functional_1.

The auxiliary outputs of Functional_2 were also quantitatively evaluated. The reconstruction component achieved a mean reconstruction MSE of 0.485, with an AUROC of 0.662 and an AUPRC (Area Under the Precision-Recall Curve) of 0.319 when the failure label was used as an auxiliary reference for anomaly-oriented evaluation. The survival-oriented output achieved a mean Brier score of 0.092 across the evaluated prediction horizons. These results provide additional quantitative evidence for the multi-task formulation, while the anomaly and survival outputs are considered supplementary to the primary classification and RUL objectives. The distribution of the reconstruction errors obtained by Functional_2 is shown in Figure 10.

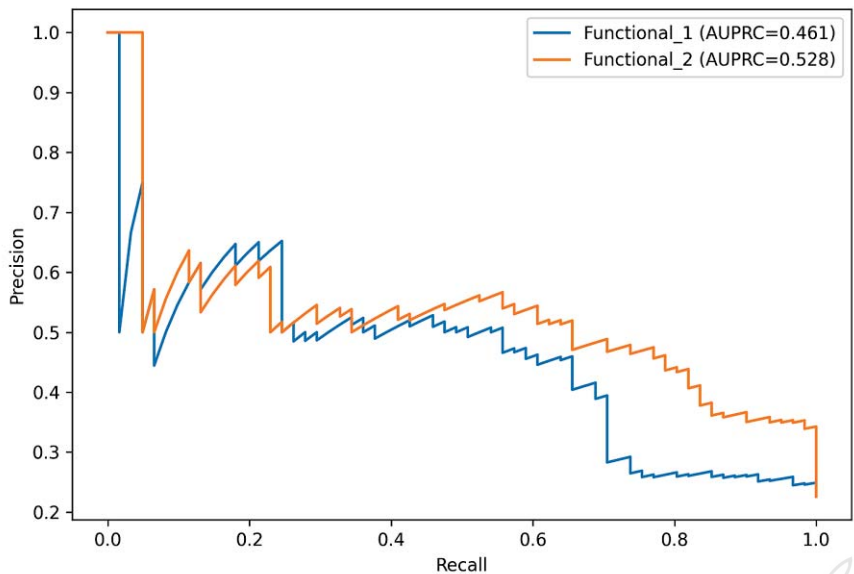


Figure 9. Precision-Recall curves for Functional_1 and Functional_2 models

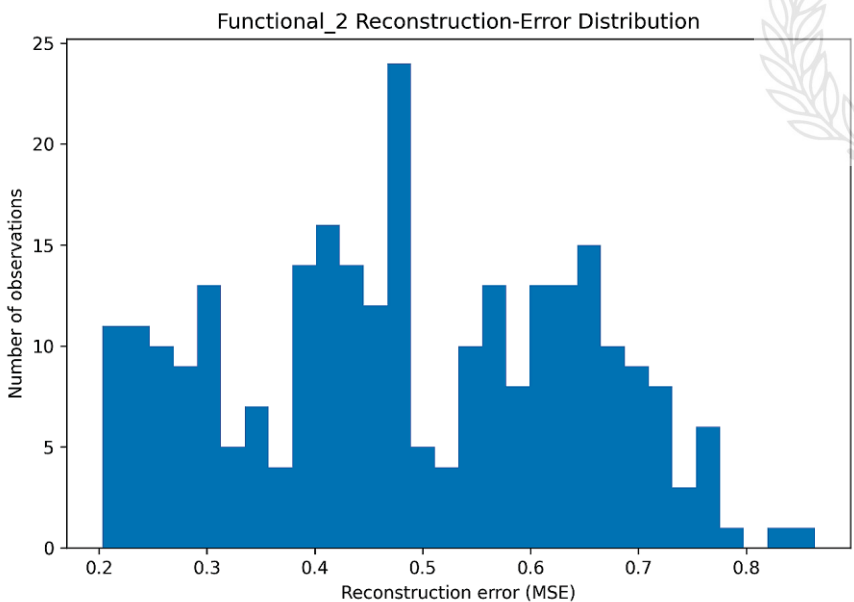


Figure 10. Distribution of reconstruction errors for Functional_2 model

Overall, Functional_1 provides a more compact architecture oriented toward computational efficiency, whereas Functional_2 provides greater representational

capacity through its multi-task formulation. The results indicate that Functional_2 achieves stronger classification performance than Functional_1, while also providing RUL, reconstruction, and survival-oriented outputs within a unified architecture. These characteristics make Functional_1 a lightweight alternative, whereas Functional_2 provides a more comprehensive framework for multi-objective machine health assessment.

4.7. SHAP Analysis

The SHAP analysis is performed on the predictor variables used by the classification model. The machine_id variable is treated as an identifier rather than an operational measurement and is therefore omitted from the reported SHAP feature-importance visualization and ranking. The reported SHAP importance values consequently focus on the operational predictor variables. Figure 11 presents the global SHAP feature importance based on the mean absolute SHAP values across the dataset. Larger absolute SHAP values indicate a greater overall contribution to the model output. Unlike the SHAP beeswarm plot, which additionally illustrates the direction and distribution of feature effects at the individual-sample level, the bar plot focuses on the overall magnitude of feature contributions. The results indicate that the model is predominantly driven by two key variables: vib_mean and current_mean, which exhibit nearly identical and substantially higher mean absolute SHAP values than all other features. This suggests that vibration and electrical current measurements are the principal predictive signals within the model and account for the majority of its explanatory power. From an engineering standpoint, these variables likely represent the most critical operational indicators of system behavior and are highly sensitive to changes in operating conditions. A secondary group of features includes temp_mean and pressure_mean, followed by variability- and peak-related metrics such as pressure_std, temp_std, vib_std, and temp_max. Although these features contribute less strongly, they still provide meaningful information to the model. Their role appears to be primarily contextual, refining predictions rather than driving them. The presence of both mean-based and standard deviation-based features within this group indicates that the model leverages both steady-state operating conditions and dynamic fluctuations in system behavior. At the lower end of the importance scale, shock_sum and especially wear_last contribute very little to the model's overall behavior. The near-zero impact of wear_last confirms that it is effectively unused in the current training setup, while shock_sum provides only a minor contribution.

Overall, the feature importance distribution reveals a clear hierarchy: core predictive power is concentrated in real-time operational signals (vibration and current), while environmental and variability-related factors play a supportive role. Features with consistently low importance likely add limited incremental

value and may be candidates for simplification or removal with minimal impact on model performance.

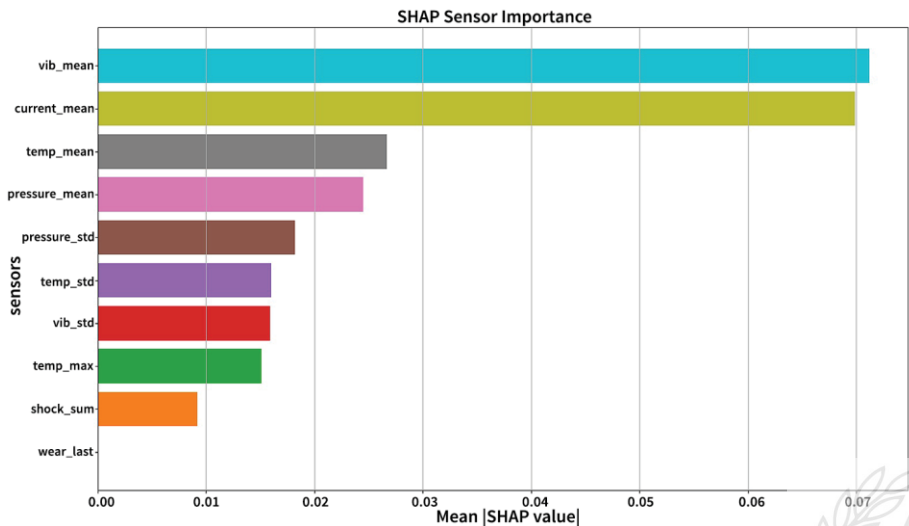


Figure 11. SHAP feature importance

The SHAP beeswarm plot (Figure 12) provides a detailed view of the importance and direction of the operational features. Features are ranked by their mean absolute SHAP values, with vib_mean and current_mean emerging as the dominant predictors. Higher values of these variables are generally associated with positive SHAP values, indicating an increased contribution to the model output. For both variables, higher values (red) are consistently associated with positive SHAP values, increasing the model output, while lower values (blue) reduce predictions, indicating a strong and largely monotonic relationship between these signals and the target variable. From an applied perspective, elevated vibration and current are therefore the most influential indicators of increased system risk or abnormal operating conditions.

Temp_mean also increases risk when high, though its effect is weaker and more concentrated. In contrast, pressure_mean shows a more mixed, non-linear relationship, with its impact varying across different value ranges. Shock_sum has limited influence under normal operating conditions, but higher values can still shift predictions noticeably. The least important features are temp_std, vib_std, and especially wear_last, which the model largely ignores. Overall, the model is primarily driven by average vibration and current as the strongest indicators of machine degradation, with temperature and pressure acting as secondary contributors.

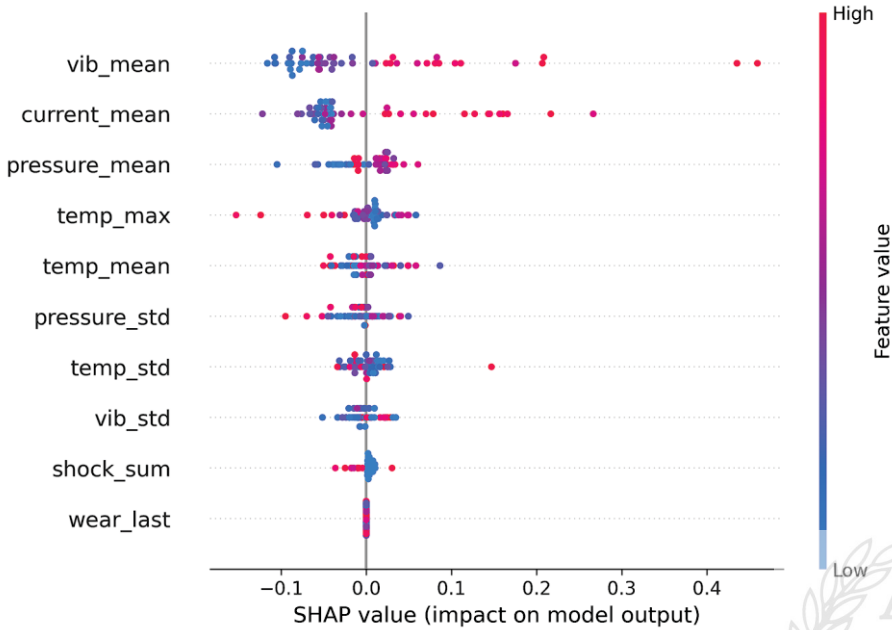


Figure 12. SHAP beeswarm plot

Linking the sensor time-series patterns (Figure 6) with the SHAP analysis provides a coherent explanation for the observed hierarchy of feature importance. The pronounced contribution of `vib_mean` and `current_mean` is consistent with their behavior as progressive cycle indicators, as both variables exhibit a steady increase within operational windows. This makes them highly informative for inferring cycle progression and potentially anticipating reset or transition events. In this context, their dominant SHAP values reflect their ability to encode temporal progression in a compact and highly predictive form. Likewise, the negative SHAP association observed for `pressure_mean` aligns with its characteristic downward trend over the course of each cycle. Lower pressure levels are indicative of later operational stages, and the model appears to exploit this relationship when forming its predictions. Depending on the target definition, these later-stage conditions may correspond to either increased or decreased predicted outcomes, but in both cases, pressure effectively functions as a temporal marker.

More broadly, these findings indicate that the model is not relying solely on static threshold-based relationships. Instead, it appears to capture underlying temporal structure within the data, leveraging gradually within-cycle changes in vibration,

temperature, and pressure as implicit indicators of system evolution. This suggests that the model encodes cycle progression indirectly through sensor drift patterns, thereby using time-dependent behavior as a key predictive signal rather than treating each observation as independent.

5. Conclusion

This study presented a hybrid XAI framework for PdM that integrates conventional ML models, Transformer-based temporal modeling, anomaly reconstruction, and SHAP-based interpretability within a unified analytical pipeline. The proposed framework is designed to support failure prediction, RUL estimation, real-time condition monitoring, and interpretable decision-making in IoT- and IIoT-enabled industrial environments.

The results show that ensemble ML methods provide the strongest overall predictive performance. XGBoost achieved the best results for RUL estimation, while RF and XGBoost achieved the highest F1-score for failure classification and GB obtained the highest ROC-AUC. These findings confirm the effectiveness of tree-based ensemble models for capturing nonlinear degradation patterns in industrial sensor data.

The multi-task Transformer extended the framework by jointly supporting failure prediction, RUL estimation, anomaly reconstruction, and survival-oriented modeling. Functional_2 achieved stronger classification performance than Functional_1, although its RUL estimation did not surpass the XGBoost regression benchmark. The results support the feasibility of the multi-task Transformer as a unified architecture for multiple PdM objectives.

SHAP analysis identified vibration- and electrical-current-related engineered features as the most influential predictors, improving model transparency and interpretability. Overall, the proposed framework demonstrates the complementary value of ensemble ML, Transformer-based temporal modeling, reconstruction-based analysis, and XAI for scalable and intelligent PdM in Industry 4.0 and Industry 5.0 environments.

Future research should validate the framework using large-scale real-world industrial datasets, address class imbalance, and further optimize the architecture for deployment in resource-constrained IoT and edge environments.

References

- [1] Fernandes, M., Corchado, J. M., and Marreiros, G. (2022), Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review, *Applied Intelligence*, 52: 14246-14280. <https://doi.org/10.1007/s10489-022-03344-3>
- [2] Paredes, J., Chavez, D., Isa-Jara, R., and Vargas, D. (2026), Hybrid approaches of machine learning algorithms in predictive maintenance: a systematic literature review, *Applied System Innovation*, 9(5): 90. <https://doi.org/10.3390/asi9050090>.
- [3] Miraje, A. A., Hunnargi, N. S., Kolap, A. S., Bhangari, D. S., and Chougule, S. A. (2022), Machine learning applications for predictive maintenance in mechanical systems: case studies, algorithms, and performance evaluation, *International Journal on Recent and Innovation Trends in Computing and Communication*, 10(12): 254-263
- [4] Tsallis, C., Papageorgas, P., Piromalis, D., and Munteanu, R. A. (2025), Application-wise review of machine learning-based predictive maintenance: trends, challenges, and future directions, *Applied Sciences*, 15(9): 4898. <https://doi.org/10.3390/app15094898>
- [5] Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., and Safaei, B. (2020), Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0, *Sustainability*, 12(19): 8211. DOI: <https://doi.org/10.3390/su12198211>.
- [6] Bitam, T., Yahiaoui, A., Boubiche, D. E., Martínez-Peláez, R., Toral-Cruz, H., and Velarde-Alvarado, P. (2025), Artificial intelligence of things for next-generation predictive maintenance, *Sensors*, 25(24): 7636. <https://doi.org/10.3390/s25247636>
- [7] Amin, A. A., Mubarak, A., Waseem, S., Alqarni, Z. A., and Manzoor, H. U. (2026), Remaining Useful Life (RUL) prediction methods for machine health estimation and fault diagnosis: a comprehensive review of latest techniques and future prospects, *Engineering Reports*, 8(4): e70699. <https://doi.org/10.1002/eng2.70699>
- [8] Amaliah, N. R., Tjahjono, B., and Palade, V. (2025), Human-in-the-loop XAI for predictive maintenance: A systematic review of interactive systems and their effectiveness in maintenance decision-making, *Electronics*, 14(17) : 3384. <https://doi.org/10.3390/electronics14173384>
- [9] Cummins, L., Sommers, A., Ramezani, S. B., Mittal, S., Jabour, J., Seale, M., and Rahimi, S. (2024), Explainable predictive maintenance: a survey of current methods, challenges and opportunities, *IEEE Access*, 12: 57574-57602. <https://doi.org/10.1109/ACCESS.2024.3391130>

- [10] Gopalan, R., Onniyil, D., Viswanathan, G., and Samdani, G. (2025), Hybrid models combining explainable AI and traditional machine learning: A review of methods and applications, *World Journal of Advanced Engineering Technology and Sciences*, 15(02): 1388-1402. <https://doi.org/10.30574/wjaets.2025.15.2.0635>
- [11] Mavrelis, I., Chatzopoulos, A., and Zacharia, P. (2026), Hybrid deep learning for predictive maintenance in industrial machinery using LSTM and MLP models, *Machines*, 14(2) : 191. <https://doi.org/10.3390/machines14020191>
- [12] Endla, P., Bhardwaj, S., Mathiyalagan, P., Akila, K., Sanjeevkumar, P., and Srinivasulu, M. (2025), AI-driven predictive maintenance framework for smart manufacturing: real-time deployment, multi-sensor fusion and scalable efficiency optimization, In Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5, pp. 782-789, SCITEPRESS - Science and Technology Publications, Lda. <https://doi.org/10.5220/0013943600004919>
- [13] Uzel, H. (2025), Machine Learning–Based Fault Prediction Using Industrial Sensor Data: A Data-Driven Approach. In: Arslan, S. (ed.) Elektrik-Elektronik ve Haberleşme Mühendisliğinde Güncel Araştırmalar ve Değerlendirmeler I. <https://doi.org/10.58830/ozgur.pub1128.c4649>
- [14] Yang, C.-L., Meles, T. Y., Yilma, A. A., and Teshome, M. M. (2026), Uncertainty aware predictive maintenance using a hybrid Transformer with Monte Carlo Dropout and conformal prediction, *Ain Shams Engineering Journal*, 17(2) : 103992. <https://doi.org/10.1016/j.asej.2026.103992>
- [15] Hrnjica, B., Softic, S. (2020), Explainable AI in manufacturing: a predictive maintenance case study. In: Lalic, B., Majstorovic, V., Marjanovic, U., von Cieminski, G., Romero, D. (eds), *Advances in Production Management Systems. Towards Smart and Digital Manufacturing. APMS 2020. IFIP Advances in Information and Communication Technology*, vol 592. Springer, Cham, pp. 66-73. https://doi.org/10.1007/978-3-030-57997-5_8
- [16] Shrestha, S. (2026), A conceptual model for integrating NLP and hybrid intelligence in industrial condition monitoring and predictive maintenance, *International Journal of Engineering Research and Reviews*, 14(1): 53-60. <https://doi.org/10.5281/zenodo.19254596>
- [17] Al-Refaie, A., Al-Atrash, M., and Lepkova, N. (2025), Prediction of the remaining useful life of a milling machine using machine learning, *MethodsX*, 14: 103195. <https://doi.org/10.1016/j.mex.2025.103195>
- [18] Choo, Y.-S., and Shin, S.-J. (2025), Integrating machine learning-based remaining useful life predictions with cost-optimal block replacement for industrial maintenance, *International Journal of Prognostics and Health Management*, 16(1): 1-22. <https://doi.org/10.36001/IJPHM.2025.v16i1.4242>

- [19] Mishra, T., Agrawal, D., Sonkar, S. K., and Tyagi, N. (2025), Data-driven approaches to battery health monitoring: machine learning for RUL estimation and fault diagnosis, *International Journal of Creative Research Thoughts (IJCRT)*, 13(10) : b239–b248
- [20] Kang, Z., Catal, C., and Tekinerdogan, B. (2021), Remaining Useful Life (RUL) prediction of equipment in production lines using artificial neural networks, *Sensors*, 21(3): 932. <https://doi.org/10.3390/s21030932>
- [21] Xu, Z., and Zhang, Q. (2026), Predictive maintenance optimization for industrial equipment via reliable prognosis and risk-aware reinforcement learning, *Complex & Intelligent Systems*, 12(1): 11. <https://doi.org/10.1007/s40747-025-02127-w>
- [22] Shaheen, B., Kocsis, Á., and Németh, I. (2023), Data-driven failure prediction and RUL estimation of mechanical components using accumulative artificial neural networks, *Engineering Applications of Artificial Intelligence*, 119: 105749. <https://doi.org/10.1016/j.engappai.2022.105749>
- [23] Paredes, J., Chávez, D., Isa-Jara, R., and Vargas, D. (2025), A hybrid machine learning algorithm approach to predictive maintenance tasks: A comparison with machine learning algorithms, *Results in Engineering*, 26: 105137. <https://doi.org/10.1016/j.rineng.2025.105137>
- [24] Kavitha, M. (2025), Hybrid AI-mathematical modeling approach for predictive maintenance in rotating machinery systems, *Journal of Applied Mathematical Models in Engineering*, 1(1) : 1-8. <https://doi.org/10.17051/JAMME/01.01.01>
- [25] Noussis, A., O'Neil, R., Saif, A., et al. (2025), A hybrid Bi-LSTM model for data-driven maintenance planning, *Autonomous Intelligent Systems*, 5: 13. <https://doi.org/10.1007/s43684-025-00099-9>
- [26] Fan, Z., Li, W., and Chang, K.-C. (2023), A bidirectional long short-term memory autoencoder transformer for remaining useful life estimation, *Mathematics*, 11(24): 4972, <https://doi.org/10.3390/math11244972>
- [27] Hou, C., and Zheng, L. (2024), A multi-task joint learning model based on transformer and customized gate control for predicting remaining useful life and health status of tools, *Sensors*, 24(13): 4117. <https://doi.org/10.3390/s24134117>
- [28] Yang, H., Yang, X., Sun, D., and Hu, Y. (2025), A MID-1DC+LRT multi-task model for SOH assessment and RUL prediction of mechanical systems, *Sensors*, 25(5) : 1368. <https://doi.org/10.3390/s25051368>
- [29] Shadi, M. R., Mirshekali, H., and Shaker, H. R. (2025), Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects, *Renewable and Sustainable Energy Reviews*, 216: 115668. <https://doi.org/10.1016/j.rser.2025.115668>
- [30] Anđelić, N., Baressi Šegota, S., and Mrzljak, V. (2025), Hybrid explainable AI for machine predictive maintenance: from symbolic expressions to meta-ensembles, *Processes*, 13(7): 2180. <https://doi.org/10.3390/pr13072180>

- [31] Villagómez-Galindo, M., Manjula, G., Jagadeesh, B. N., Patil, M., Potharaju, S., and Prasad, N. A. (2026), A hybrid ensemble and explainable AI Framework for predictive maintenance of industrial equipment, *Journal of Artificial Intelligence and Technology*, 6: 328-336. <https://doi.org/10.37965/jait.2026.0939>
- [32] Siva, R., and Ayush, R. (2026), Time aware explainable AI for fault diagnosis, *International Journal for Multidisciplinary Research (IJFMR)*, 8(2): 854-858
- [33] Özcan, H. (2025), Interpretable ensemble remaining useful life prediction enables dynamic maintenance scheduling for aircraft engines, *Scientific Reports*, 15(1) : 39795. <https://doi.org/10.1038/s41598-025-23473-2>



Explainable and Trustworthy AI for Steel Plates Fault Detection Using Shap and Lime

Nerma Mezetović^{*1}, Lejla Banjanović-Mehmedović¹, Isak Karabegović²

Abstract: *This paper investigates the potential of Explainable Artificial Intelligence (XAI) methods to improve the interpretation of AI-based models in smart manufacturing applications. The study focuses on SHAP and LIME techniques, analyzing their ability to explain model decisions, identify the most influential input features, and provide insights into both individual predictions and overall model behaviour.*

A practical case study based on the Steel Plates Faults dataset is presented using a Random Forest classification model to investigate the application of explainable AI techniques for steel plate fault detection. The results demonstrate that SHAP and LIME methods provide valuable insights into model decision-making by identifying the key features influencing fault classification. These explainability approaches enhance model transparency and trustworthiness, supporting the development of more reliable AI-based solutions for industrial quality inspection and smart manufacturing applications.

Keywords: *explainable artificial intelligence (XAI), model interpretability, machine learning, SHAP, LIME, transparency*

1. Introduction

Artificial intelligence (AI) has become a key technology in the transformation of modern industrial systems, enabling machines to perform complex tasks such as learning, reasoning, and decision-making. Advances in machine learning algorithms, computational capabilities, and data availability have accelerated the adoption of AI in various domains, including smart manufacturing, robotics, and autonomous systems [1]. In industrial environments, AI-based solutions are increasingly applied for quality inspection, predictive maintenance, process monitoring, and production optimization, contributing to improved efficiency and more adaptive decision-making.

Despite their high predictive performance, many advanced machine learning models remain difficult to interpret due to their complex internal structures. This limitation is particularly evident in deep learning and other nonlinear models, which are commonly considered black-box systems. Although the input data and final predictions are available, the reasoning process leading to a specific decision

^{*1}University of Tuzla, Faculty of Electrical Engineering, Tuzla, Bosnia and Herzegovina

²Academy of Science and Arts of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina

E-mail: nerma.mezetovic@fet.ba

is often unclear to human users [2]. This lack of transparency represents a significant barrier to the wider adoption of AI in industrial applications, where reliability, safety, and accountability are critical requirements.

In manufacturing environments, AI-driven decisions can directly influence product quality, equipment operation, and production efficiency. Therefore, achieving high prediction accuracy alone is not sufficient; it is also necessary to understand which factors contribute to a model decision and whether the model relies on meaningful patterns. Without such insight, detecting errors, validating model behaviour, and ensuring reliable operation become considerably more challenging [1].

To overcome these limitations, Explainable Artificial Intelligence (XAI) has emerged as an important research field focused on developing methods that improve the understanding of AI model decisions. XAI approaches aim to transform complex model outputs into meaningful explanations, enabling users to analyse model behaviour, verify predictions, and identify the factors influencing decision-making [2]. Consequently, explainability has become a fundamental component of trustworthy AI, together with transparency, fairness, robustness, accountability, and user confidence [3].

The importance of XAI is further emphasized by ethical and regulatory requirements for AI deployment. Modern AI systems are expected not only to achieve high performance but also to provide transparent and accountable decision-making processes. Frameworks such as the General Data Protection Regulation (GDPR) and the European Union Artificial Intelligence Act (EU AI Act) highlight the importance of transparency, documentation, human oversight, and risk management in AI applications [4], [5].

Among different XAI approaches, model-agnostic methods have attracted significant attention due to their ability to explain various machine learning models independently of their internal architecture. Two widely adopted methods are Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP). These techniques provide insights into feature contributions and enable both local explanation of individual predictions and global analysis of overall model behaviour. Their flexibility makes them particularly suitable for analysing complex machine learning models used in industrial applications.

The objective of this paper is to investigate the contribution of XAI methods to the development of transparent and trustworthy AI systems for industrial fault detection. The study focuses on SHAP and LIME approaches and analyses their capabilities, advantages, and limitations. Furthermore, a practical case study based on the Steel Plates Faults dataset is presented using a Random Forest classification model to demonstrate how SHAP and LIME can identify the most influential features and provide deeper insights into the decision-making process of machine learning models.

2. Fundamentals of Explainable Artificial Intelligence

Two fundamental concepts are central to understanding transparent artificial intelligence systems: *interpretability* and *explainability*. Although these terms are often used interchangeably, they describe different aspects of understanding AI model behaviour. Interpretability refers to the inherent ability of a model to provide insight into its internal decision-making process, while explainability refers to the use of additional techniques to generate understandable explanations of model predictions.

Interpretability is an intrinsic property of models whose internal structure can be directly understood by humans. In interpretable models, the relationship between input variables and outputs is explicitly represented, allowing users to analyse how individual features contribute to a prediction [6]. Decision trees, linear regression models, and rule-based systems are typical examples of white-box models, where the reasoning process is directly accessible without additional explanation mechanisms [7].

The importance of interpretability is particularly evident in industrial applications, where AI decisions may influence product quality, equipment reliability, and operational safety. In such environments, understanding why a model generates a specific prediction is essential for model validation, error analysis, and informed decision-making. For example, identifying the reasons behind a detected manufacturing defect or a predicted equipment failure can support preventive actions and improve system reliability.

In contrast, explainability focuses on providing understandable interpretations of complex AI models whose internal mechanisms cannot be directly examined. Instead of modifying the original model, explainability methods generate additional representations, such as feature importance scores, surrogate models, or visual explanations, to reveal the factors influencing a prediction [8]. Therefore, explainability serves as a bridge between complex machine learning models and the human need for transparency, trust, and accountability, Figure 1.

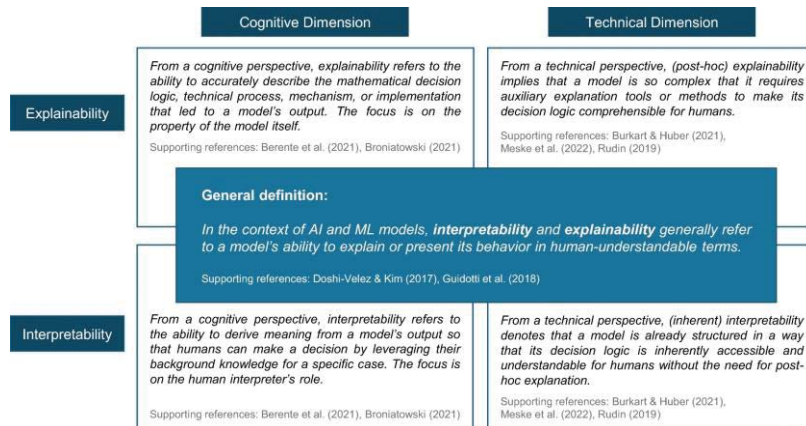


Figure 1. Conceptual comparison between interpretability and explainability in AI models [9]

2.1. Taxonomy of XAI Methods

Explainable artificial intelligence represents a broad collection of methods that differ according to how explanations are generated and presented. The most common classifications are based on whether explanations are integrated into the model itself or generated after training, whether they describe the entire model behaviour or individual predictions, and whether they depend on a specific model architecture.

2.1.1. Intrinsic Interpretability and Post-hoc Explainability

XAI methods can first be classified into *intrinsic interpretability* and *post-hoc explainability approaches*. As previously defined, intrinsically interpretable models are understandable by construction, with their decision-making process directly represented through mathematical relationships or logical rules. However, a major challenge of these transparent approaches in industrial environments is the trade-off between interpretability and performance [9]. Although they provide high transparency, their predictive performance may be severely limited when dealing with highly nonlinear and complex industrial problems, which often require more sophisticated, non-linear models. To better understand these differences, Figure 2 summarizes and compares the characteristics of intrinsic and post-hoc interpretability across different stages of the machine learning process, from data input to model assessment, with the post-hoc approach discussed in more detail in the following section.

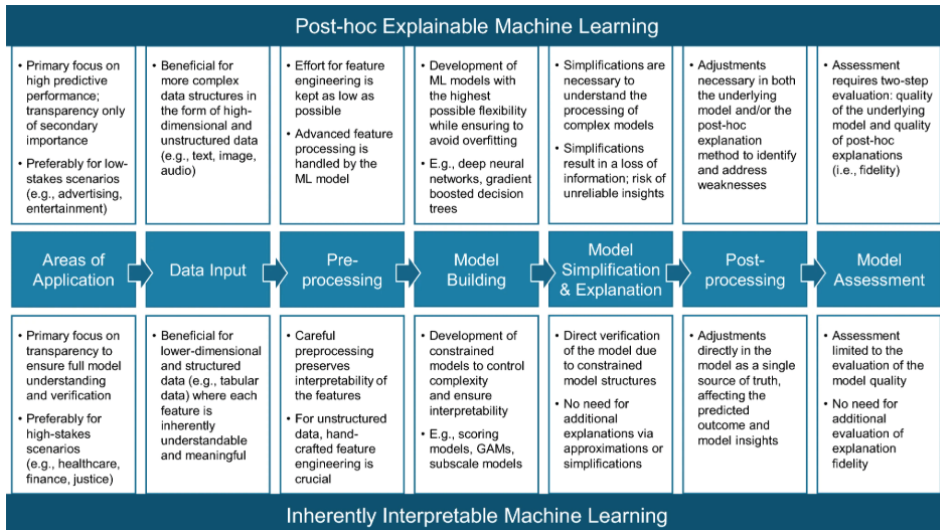


Figure 2. Comparison between intrinsically interpretable models and post-hoc explanation approaches [9]

Post-hoc explainability methods are applied after a model has been trained and are particularly important for complex black-box models such as ensemble methods and deep neural networks. These techniques analyse the relationship between model inputs and outputs without modifying the original model, providing information about feature influence and decision-making patterns [9]. Post-hoc methods enable the use of high-performance models while still providing a certain level of transparency required for trustworthy AI deployment. In this study, we adopt this post-hoc perspective, applying SHAP and LIME to interpret the predictions of a Random Forest classifier trained on the steel plates faults dataset.

2.1.2. Global and Local Explanations

Another important distinction in XAI concerns the scope of explanations. Global explanations describe the overall behaviour of a model by analysing general decision patterns and the importance of input features across the entire dataset. They provide insight into how the model operates as a whole.

In contrast, local explanations focus on individual predictions and identify the specific features that influenced a particular decision. This approach is especially valuable for analysing individual cases, such as explaining why a specific industrial product was classified as defective or why a particular anomaly was detected.

Global and local explanations provide complementary perspectives, Figure 3. While global explanations improve the understanding of general model behaviour, local explanations provide detailed reasoning behind individual predictions, which is often crucial for practical decision support [7].

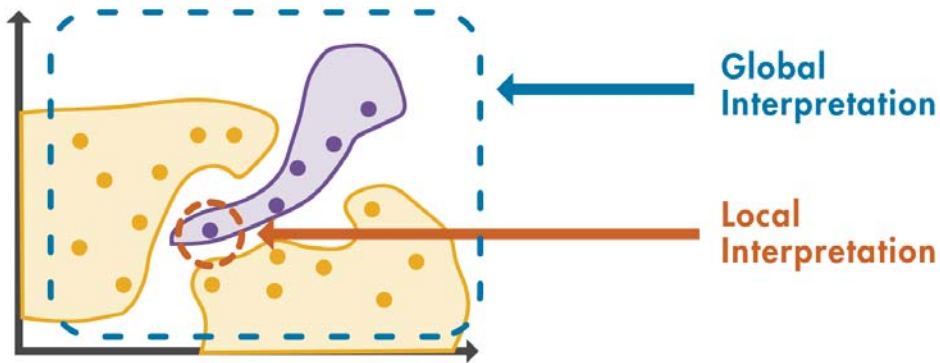


Figure 3. Difference between global and local explanations in XAI [10]

2.1.3. Model-Specific and Model-Agnostic Approaches

Because machine learning algorithms differ in their architectures, complexity, and internal mechanisms, XAI methods are commonly divided into model-specific and model-agnostic approaches.

Model-specific methods are designed for a particular class of models and use information from their internal structure, such as parameters, weights, or activation patterns. These approaches can provide detailed explanations but are limited to specific model architectures [10]. Examples include gradient-based methods such as Integrated Gradients, activation-based methods such as Grad-CAM, and relevance-propagation approaches such as Layer-wise Relevance Propagation (LRP) [11], [12].

In contrast, model-agnostic methods explain model behaviour independently of the underlying algorithm. They treat the model as a black box and analyse the relationship between input features and predictions. This flexibility enables their application to different machine learning models without requiring modifications to the original architecture [13].

Model-agnostic methods are particularly valuable in industrial applications because they allow engineers to select the most accurate model for a given task while maintaining the possibility of generating meaningful explanations. Among the most widely adopted approaches are Local Interpretable Model-Agnostic

Explanations (LIME) and SHapley Additive exPlanations (SHAP), which are further analysed in the following section.

Table 1 summarizes the main differences between model-specific and model-agnostic XAI approaches.

Table 1. Comparative overview of model-agnostic and model-specific approaches

Aspect	Model-agnostic approach	Model-specific approach
Applicability	Independent of internal architecture; usable across model families	Tied to a particular architecture; limited to a specific class of models
Operating mechanism	Perturbing inputs and observing changes in the output	Accessing internal weights, gradients, and layers directly
Computational cost	Often high (thousands of repeated predictions)	Usually lower (computed directly from existing parameters)
Fidelity	May be less precise, since it relies on approximating model behaviour	Often more precise, since it reflects the model's actual internal logic
Advantages	Universality, flexibility, applicability across different models	Higher precision and deeper insight into the model's internal structure
Limitations	May yield approximate explanations that overlook model-specific detail	Restricted to particular models and often demands deeper technical expertise
Representative examples	LIME, SHAP, Partial Dependence Plots	Integrated Gradients, CAM, Grad-CAM, LRP

3. Model-Agnostic Explainable AI Methods: Lime And Shap

Model-agnostic XAI methods have become widely adopted because they enable the interpretation of different machine learning models without requiring access to their internal structures. Instead of modifying the original model, these approaches analyse the relationship between input features and model outputs to generate understandable explanations [14]. Among the most widely used model-agnostic techniques are *Local Interpretable Model-Agnostic Explanations (LIME)* [13] and *SHapley Additive exPlanations (SHAP)* [15].

3.1. LIME: Local Interpretable Model-Agnostic Explanations

LIME is a model-agnostic explanation technique designed to interpret individual predictions of complex machine learning models. The main idea behind LIME is that although a model may be highly nonlinear and difficult to understand globally, its behaviour around a specific instance can often be approximated using a simpler interpretable model [13].

The method generates perturbed samples around the instance of interest and obtains predictions from the original model. These samples are weighted according to their similarity to the original instance and are then used to train a simple surrogate model, typically linear regression or a decision tree. The parameters of this local model indicate the contribution of individual features to the prediction, providing insight into why a specific decision was made [13].

The main steps of LIME can be summarized as follows:

- selecting an instance to be explained;
- generating perturbed samples around the selected instance;
- obtaining predictions from the original model;
- assigning weights according to the distance from the original sample;
- training an interpretable surrogate model;
- analysing feature contributions to generate the explanation.

LIME improves the transparency of black-box models by providing human-understandable local explanations. In addition, the quality of the explanation strongly depends on the selection of the local neighbourhood [13].

3.1.1. Strengths and Limitations of LIME

The main advantage of LIME is its flexibility and applicability to different machine learning algorithms, including Random Forest, support vector machines, and neural networks. Since it does not require knowledge of the model architecture, it can be applied to a wide range of classification and regression problems [7].

However, LIME has several limitations. The stochastic nature of perturbation sampling may reduce explanation stability, while unrealistic generated samples can negatively influence the quality of the surrogate model. Furthermore, LIME focuses only on individual predictions and does not provide a direct global understanding of the overall model behaviour [7].

3.2. SHAP: SHapley Additive exPlanations

SHAP, introduced by Lundberg and Lee, is one of the most influential XAI methods for interpreting machine learning predictions. It is based on Shapley values from cooperative game theory, where each input feature is considered as a contributor to the final model prediction [15].

The main principle of SHAP is to distribute the difference between the model prediction and the average prediction among the input features according to their individual contributions. Positive SHAP values indicate features that increase the prediction value, while negative values represent features that decrease it. The magnitude of the SHAP value reflects the strength of the feature influence [15].

SHAP supports both local and global explanations [15]. Local explanations describe why a specific prediction was generated, while global explanations aggregate SHAP values across multiple samples to identify the most influential features in the entire dataset.

Several SHAP variants have been developed depending on the underlying model structure. KernelSHAP provides a model-independent approximation suitable for different machine learning models, while TreeSHAP is optimized for tree-based algorithms such as Random Forest and gradient boosting models, providing faster and more efficient explanations [15].

3.2.1. Visualization of SHAP Explanations

SHAP visualizations provide an intuitive representation of feature contributions and are essential for interpreting model behaviour. Common visualization techniques include:

- *Summary plots*, which show global feature importance and the direction of feature influence across the dataset.
- *Waterfall plots*, which explain individual predictions by showing how each feature contributes to increasing or decreasing the final output.
- *Force plots*, which illustrate the contribution of individual features relative to the model baseline value.

These visualizations enable engineers and domain experts to analyse model decisions without requiring detailed knowledge of the underlying machine learning algorithm [15].

3.2.2. Strengths and Limitations of SHAP

The main advantage of SHAP is its strong theoretical foundation and ability to provide consistent feature attribution. Unlike LIME, SHAP explanations can be aggregated to obtain a reliable global interpretation of model behaviour. This

property makes SHAP particularly suitable for industrial applications where identifying the most influential process variables is important [7].

However, the original SHAP formulation can be computationally demanding, especially for models with a large number of features. In addition, correlated input variables may affect the interpretation of feature importance. Therefore, selecting an appropriate SHAP variant and considering the characteristics of the dataset are important for practical deployment [7].

3.3. Comparison of LIME and SHAP Methods

LIME and SHAP provide complementary approaches to explaining machine learning models. LIME focuses primarily on local interpretation by approximating the behaviour of a model around a specific instance using an interpretable surrogate model. In contrast, SHAP provides feature contributions based on Shapley values, enabling both local and global analysis of model behaviour.

The main difference between the two methods lies in their explanation principles. LIME offers simplicity and flexibility but may suffer from instability due to random sampling. SHAP provides more theoretically consistent explanations but can require higher computational resources.

For industrial AI applications, the choice between LIME and SHAP depends on the specific requirements of the problem. While LIME is useful for understanding individual predictions, SHAP is particularly suitable for analysing feature importance and overall model behaviour. In this work, both methods are applied to a Random Forest classification model for steel plate fault detection to evaluate their ability to improve model transparency and trustworthiness.

4. Application of XAI Techniques: Case Study

This chapter presents a case study applying the XAI methods discussed earlier to the problem of classifying surface defects on steel plates. The dataset used brings together measurements of the physical and geometric properties of individual plates alongside the type of defect each one exhibits. The aim here is not merely to build a working classifier, but to examine how that classifier arrives at its decisions.

A conventional multi-class machine learning model is developed first, after which SHAP and LIME are applied to explain its predictions. Particular attention is given to both global and local explanations, and to how the two XAI techniques compare in terms of interpretability and practical relevance for industrial deployment.

4.1. Description of the Classification Problem and Dataset

4.1.1. Problem description and industrial motivation

Quality control is one of the most important steps in any modern production line. During the manufacturing of steel plates and sheets, various surface flaws inevitably appear. Left unaddressed, such defects can compromise the mechanical properties of the material, reduce its market value, and ultimately lead to the finished product being rejected outright. Manual inspection, the traditional response to this problem, is slow, costly, and inherently subjective, leaving it vulnerable to human error. Automating the process with machine learning allows defects to be classified quickly and consistently, in something close to real time.

Conventional predictive models used for this purpose, however, tend to behave as black boxes. A model may flag a defect with high accuracy and still leave production engineers with no idea why it reached that conclusion. Without insight into which factors actually drove a decision, it becomes difficult to adjust production parameters in time to prevent the same defect from recurring. Overcoming exactly this kind of opacity is the underlying motivation for the case study presented in this chapter.

4.1.2. Dataset characteristics

The analysis relies on the publicly available Steel Plates Faults dataset, obtained from the UCI Machine Learning Repository [16]. It contains 1,941 instances, each corresponding to a steel plate sample, described by 27 input features that capture the geometric and visual characteristics of a surface defect and labelled with one of seven possible fault types.

The input variables are derived from image analysis of industrial steel plates and represent a numerical encoding of the irregularities detected. They fall into three broad groups [16]:

- *Geometric and spatial properties*: the coordinates of the flaw on the plate (X_Minimum, X_Maximum, Y_Minimum, Y_Maximum), the area of the defect (Pixels_Areas), its perimeter (X_Perimeter, Y_Perimeter), and the thickness of the plate itself (Steel_Plate_Thickness).
- *Optical properties*: pixel luminosity within the defect region, which helps distinguish darker flaws such as stains from lighter surface anomalies (Sum_of_Luminosity, Minimum_of_Luminosity, Maximum_of_Luminosity).
- *Material type*: categorical variables encoded in binary form indicating the grade of steel used (TypeOfSteel_A300 and TypeOfSteel_A400).

The output variable identifies the fault type, with the dataset covering seven mutually exclusive classes of surface damage [16]:

- *Pastry*: a surface irregularity resembling a thickened or clumped layer of material on the plate.
- *Z_Scratch*: scratches produced during processing, handling, or transport of the material.
- *K_Scratch*: a distinct, typically deeper and more pronounced type of surface scratch.
- *Stains*: discolouration on the surface, most often caused by liquids, oils, or production processes.
- *Dirtiness*: contamination of the surface by dust, dirt, or other foreign particles.
- *Bumps*: physical protrusions on the surface caused by mechanical stress.
- *Other_Faults*: a catch-all category for rarer or less clearly defined defects that do not fit the categories above; it is also the most frequent class in the dataset.

Table 2. Sample distribution across defect classes

Defect class	Number of samples
Other_Faults	673
Bumps	402
K_Scratch	391
Z_Scratch	190
Pastry	158
Stains	72
Dirtiness	55



As Table 2 shows, the classification problem is defined over a markedly imbalanced dataset. Other_Faults is far better represented than minority classes such as Dirtiness and Stains, a distribution that is entirely typical of real industrial processes, where certain specific faults simply occur far less often than others. Because the Steel Plates Faults dataset consists exclusively of numerical measurements extracted through automated image analysis, rather than actual photographs of the defects, a visual sense of the raw data is instead given through annotated sample tables. For clarity, and given the practical limits on space, only six representative attributes are shown: Pixels Area, Plate Thickness, Luminosity Index, Conveyor Length, and the two binary steel-type indicators (A300 and A400), for five representative samples drawn from two classes, Bumps and Other_Faults (Table 3 and Table 4).

Table 3. Annotated samples from the Bumps class

Sample	Pixels Area (px)	Plate Thickness (mm)	Luminosity Index	Conveyor Length (mm)	Steel A300	Steel A400
1	3234	80	-0.157	1692	1	0
2	178	80	-0.228	1694	1	0
3	38	60	-0.115	1690	1	0
4	147	50	-0.213	1358	0	1
5	74	70	-0.116	1688	1	0

Table 4. Annotated samples from the Other Faults class

Sample	Pixels Area (px)	Plate Thickness (mm)	Luminosity Index	Conveyor Length (mm)	Steel A300	Steel A400
1	1695	300	-0.263	1692	0	1
2	949	80	-0.371	1687	1	0
3	169	80	-0.165	1687	1	0
4	836	150	-0.189	1360	1	0
5	381	143	-0.005	1694	1	0

4.2. Machine Learning Model Development

This section covers the preparation of the data and the construction of a classifier for steel-plate defects. The task is defined as multi-class classification, with every sample assigned to exactly one of the seven fault categories.

4.2.1. Data preparation

The dataset was first loaded and its columns assigned appropriate names, then converted to a numerical format suitable for machine learning algorithms; missing values and the overall structural integrity of the data were also checked at this stage. One characteristic of this particular dataset is that the seven defect classes are originally encoded as seven separate binary columns, where a value of 1 marks the presence of a given fault and 0 its absence. To enable multi-class classification, these columns were collapsed into a single target variable by identifying, for each row, the column carrying the maximum (i.e., positive) value.

The resulting class labels were then converted into numerical values using a LabelEncoder, completing the preparation of the dataset for the learning algorithm. To obtain an honest evaluation of the model on data it had not seen during training, the full set of 1,941 samples was split into a training set (80%) and a test set (20%). The training set is used to fit the model's parameters, while the test set is reserved exclusively for evaluating performance on previously unseen data, mimicking how the model would actually be used in production.

Given the pronounced class imbalance noted before, the split was stratified by class label, so that the proportion of each class remained approximately the same in both the training and the test set. This reduces the risk that minority classes such as Dirtiness or Stains end up under-represented in either subset, and, in turn, the risk of a model biased toward the dominant classes. A fixed random seed was used throughout to keep the split and the experiment as a whole - reproducible.

4.2.2. *Building the classification model*

Random Forest was selected as the classification algorithm. It is one of the best-known machine learning methods for classification and regression problems, built on the principle of ensemble learning - combining a large number of individual models to obtain predictions that are both more accurate and more stable than any single model could produce on its own [17].

In a Random Forest, each individual model is a decision tree. Rather than relying on a single tree, which tends to overfit the training data, the algorithm grows a large number of trees that together form the "forest"; the model's final decision is obtained by combining the outputs of every tree, improving generalisation and curbing overfitting.

Two mechanisms keep the individual trees sufficiently different from one another [17]:

- *Bagging (Bootstrap Aggregating)*: each tree is trained on a random subset of the data sampled with replacement from the original dataset, so that some observations appear in a given tree's training sample multiple times while others do not appear at all.
- *Feature randomness*: when splitting a node, each tree considers only a randomly selected subset of the available features rather than all of them, which keeps one or two dominant features from shaping every tree in the same way and keeps correlation between trees low.

Once all the trees have been trained, their outputs are aggregated to reach a final decision. For classification, each tree casts a vote for a class and the final prediction follows a majority vote; for regression, each tree instead predicts a numerical value and the final prediction is the average across all trees [17].

Random Forest is known for high accuracy, resistance to overfitting, and the ability to handle large datasets, including ones with missing values; it also supports feature-importance estimates that help clarify how the model reaches its decisions. Its main drawbacks are reduced interpretability, given the sheer number of trees involved, and higher memory and processing requirements compared with simpler models.

The algorithm was chosen here for its strong generalisation ability and its wide use in classification problems. As a black-box model, it is also well suited to demonstrating what explainable AI methods can add on top of a conventional classifier.

The model was configured with 300 decision trees to improve prediction stability and reduce variance, and a minimum leaf size of two samples to keep individual trees from over-fitting to isolated observations. A class-weighting scheme (balanced subsampling) was also introduced to counter the class imbalance identified earlier, giving smaller, under-represented classes more weight during training and reducing the model's bias toward the dominant classes.

Configured this way, the model achieves stable performance while remaining suitable for the post-hoc interpretation carried out using SHAP and LIME.

4.3. Model Performance Evaluation

Evaluating a model's ability to generalise to unseen data is a critical step in any machine learning pipeline. Here, the model was tested on the held-out test set, which played no role in training, in order to approximate how it would behave in a real deployment.

Several standard classification metrics were used to quantify performance:

- *Accuracy*: the proportion of correctly classified samples, positive and negative alike, out of the total number of samples.
- *Precision*: the proportion of predicted-positive samples that are actually positive, indicating how reliable the model's positive predictions are.
- *Recall*: the proportion of actual positive samples the model correctly identifies, a critical metric wherever missing a positive case is unacceptable.
- *F1-score*: the harmonic mean of precision and recall, used as a balanced measure, particularly valuable when classes are unevenly represented.
- *Support*: the number of actual samples belonging to a given class in the test set - a property of the data itself rather than of the model's performance.

The Random Forest model achieved an overall accuracy of 78.66%. A per-class classification report, shown in Table 5, offers a more detailed picture.

Table 5. Classification report on the test set

Class	Precision	Recall	F1-score	Support
Bumps	0.7162	0.6543	0.6839	81
Dirtiness	0.8333	0.9091	0.8696	11
K_Scratch	0.9726	0.9103	0.9404	78
Other_Faults	0.7200	0.8000	0.7579	135
Pastry	0.5758	0.5938	0.5846	32
Stains	0.9231	0.8571	0.8889	14
Z_Scratch	0.9706	0.8684	0.9167	38

The support column confirms the imbalance already visible in the full dataset, with Other_Faults dominating the test set and Dirtiness represented by only a handful of samples. The model performs strongly on K_Scratch and Z_Scratch, identifying these fault types with very few errors, while performance is weakest on Pastry, where the model frequently confuses this class with other types of imperfection. Bumps also show only moderate results, with roughly a third of genuine defects missed. Notably, Dirtiness stands out despite its tiny sample size, as the model successfully catches most real contamination cases even with very little training data to learn from.

A confusion matrix (Figure 4) was also used to examine the model's errors in more detail, tracking the four possible outcomes of a classification decision: true positives, true negatives, false positives (Type I errors), and false negatives (Type II errors) [18].

The confusion matrix confirms that the model performs best on classes with more training examples and struggles more with under-represented ones - an expected consequence of the imbalance discussed earlier, since dominant classes exert a stronger pull on what the model learns. Other_Faults is recognised most reliably, unsurprisingly given its size, and K_Scratch and Z_Scratch are also classified fairly well; Dirtiness and Stains show high precision, though their small sample sizes warrant some caution in that reading. The most pronounced confusion occurs between Bumps and Other_Faults, with a substantial number of samples crossing from one class into the other - a sign that the two share visually similar characteristics. Pastry is also frequently misclassified, chiefly as Other_Faults or Bumps, suggesting its defining patterns are either weaker or under-represented in the data.

These evaluation metrics indicate reasonably solid overall performance, but they are inherently limited when it comes to interpretability: they quantify how well the model performs without ever answering the more useful question of why a particular decision was made, or which input characteristics actually drove it.

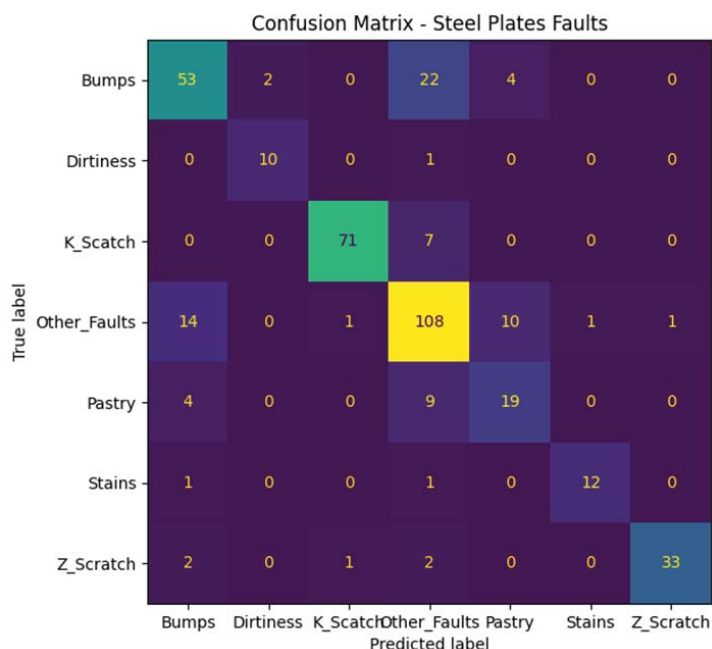


Figure 4. Confusion matrix [18]

Knowing that a prediction is right or wrong says nothing about which attributes led the model there. It is precisely this gap that motivates the explainable AI analysis that follows, which aims to expose the model's internal reasoning and identify the features most responsible for each decision.

4.4. Model Explanation Using SHAP and LIME

This section examines, in detail, how the trained model actually uses its input features when classifying steel-plate defects. Where the previous section focused on overall accuracy and aggregate performance, the analysis here turns to the model's internal logic - the contribution each individual attribute makes to a given decision. A single representative case from the test set is used throughout for the local analyses, since walking through one concrete example in detail gives a far more intuitive sense of the model's reasoning than trying to summarise many cases at once.

4.4.1. Global model analysis using SHAP

To understand how the model behaves across the dataset as a whole, a global SHAP (SHapley Additive Explanations) analysis was carried out. Rather than explaining one prediction at a time, this analysis surfaces the general patterns and

priorities the model relies on. Because the underlying classifier is a Random Forest, SHAP values were computed using the TreeExplainer implementation from the SHAP library, which is specifically optimised for efficient and exact computation on tree-based models.

The first step in interpreting the model globally is identifying its most influential inputs. Figure 5 ranks the 15 most important variables the Random Forest relies on when classifying steel-plate defects, with bar length representing each feature's mean absolute SHAP value across the test set.

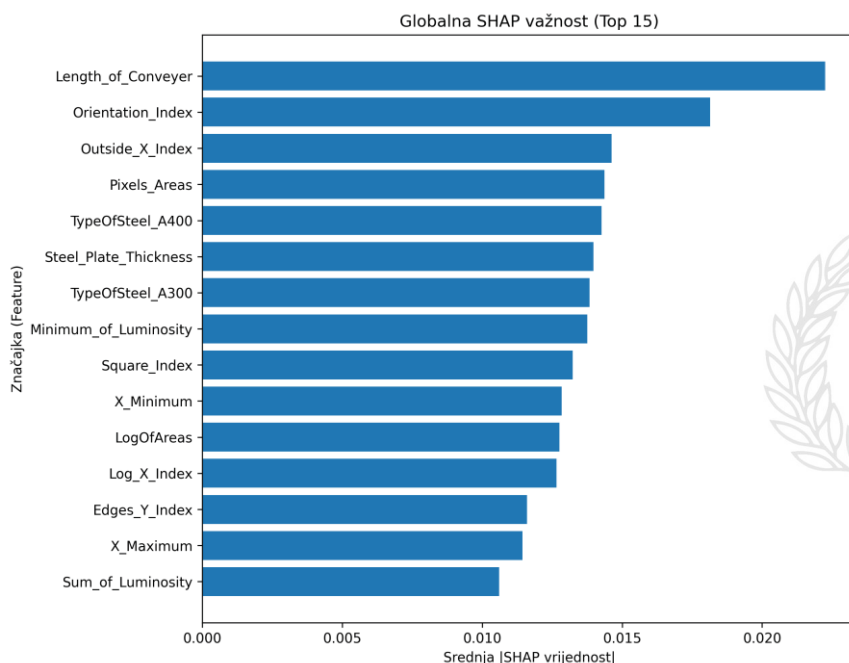


Figure 5. Global feature importance derived from SHAP [18]

Identifying the most influential input characteristics:

- **Production and operational parameters dominate.** Length_of_Conveyer stands out as the single most important feature by a clear margin. From an industrial standpoint this is a genuinely actionable insight: it suggests that the occurrence of specific defects is tied directly to the length of the production line, or the particular zone in which a plate is processed and transported.
- **Defect geometry and orientation matter.** The next most influential variables are Orientation_Index and Outside_X_Index, closely followed by total defect area (Pixels_Areas). This tracks with basic engineering

intuition: different defect types have recognisably different physical shapes - scratches are elongated and oriented in a particular way, while stains occupy a different kind of area and shape altogether - and the model leans heavily on these spatial cues when making its decisions.

- **Material characteristics carry real weight.** Variables describing the material itself - steel type (TypeOfSteel_A400, TypeOfSteel_A300) and plate thickness (Steel_Plate_Thickness) - rank high among the most important features, confirming the industrial intuition that particular alloys or thicknesses are more prone to specific fault types (thinner plates, for instance, deform more easily and are more likely to develop Bumps).
- **Optical characteristics play a supporting role.** Variables describing pixel brightness, such as Minimum_of_Luminosity, also carry meaningful weight, confirming that the model successfully exploits contrast and brightness variation to separate darker defects such as dirtiness from lighter surface anomalies.

Taken together, the global SHAP analysis shows that the black-box model is not basing its decisions on noise, but on physically and operationally meaningful geometric, material, and process-related characteristics. For engineers and quality-control staff on the factory floor, the decisive role played by Length_of_Conveyer points to conveyor-mechanism inspection as a priority for reducing defect rates.

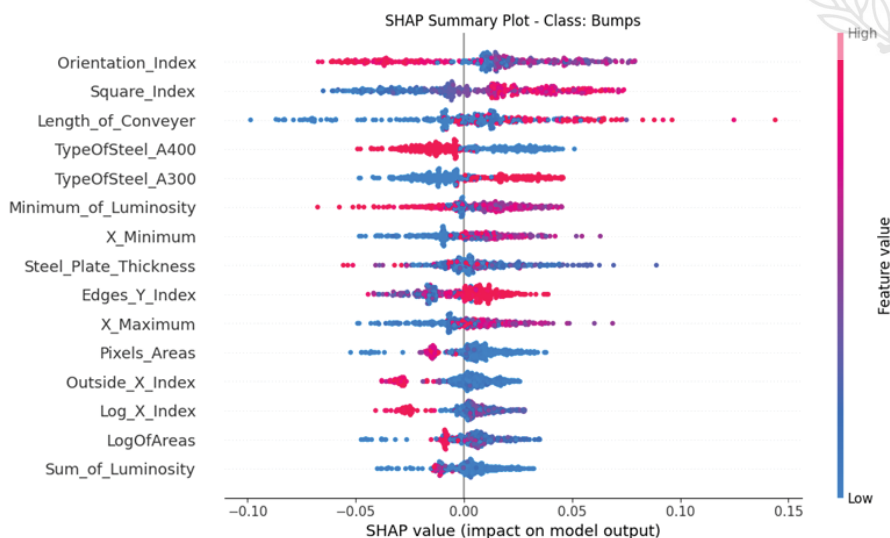


Figure 6. SHAP summary plot for the Bumps defect class [18]

To see not just which features matter but how their actual values push a prediction up or down, a SHAP summary plot was generated for the Bumps class, shown in Figure 6, chosen as one of the most heavily represented fault types among the seven the model distinguishes.

Each point on this plot represents one test-set sample; its position along the horizontal axis (the SHAP value) shows whether that feature pushed the probability of a Bumps defect up or down for that particular sample. A close reading of the point distribution yields several concrete rules the model appears to be applying:

- **Geometry (Orientation_Index and Square_Index):** low values of the orientation index (a cluster of blue points on the right) push strongly toward a positive Bumps prediction, while high values of the square index (red points, also on the right) likewise raise the predicted probability.
- **Operational parameters (Length_of_Conveyer):** unusually high conveyor lengths (red points shifted sharply to the right) substantially raise the likelihood of this defect - a critical, production-relevant finding, since it suggests longer conveyor sections are more prone to generating physical bumps in the material.
- **Material type (TypeOfSteel_A300 versus A400):** the two steel grades pull in opposite directions - high values (presence) of A300 sit on the positive side of the SHAP axis, while high values of A400 push the prediction the other way, indicating that A300 steel is considerably more prone to developing Bumps than A400.

This level of granularity, and this kind of insight into how individual parameters interact, is simply unavailable from conventional evaluation metrics alone - underscoring why an explainability layer is not just a nice-to-have but a genuine necessity for tackling complex industrial classification problems.

4.4.2. *Local analysis of individual decisions using SHAP and LIME*

While the global SHAP analysis reveals broad patterns across the dataset, practical deployment on a production line often calls for something narrower: understanding why the model flagged one specific plate as defective. An engineer working in real time needs to know exactly which parameters drove the classification of the piece of steel currently in front of them, and that diagnostic need is what local interpretation is built to satisfy. To illustrate this, one test-set sample flagged as defective by the model was isolated for detailed analysis. The result, visualised as a waterfall plot, is shown in Figure 7.

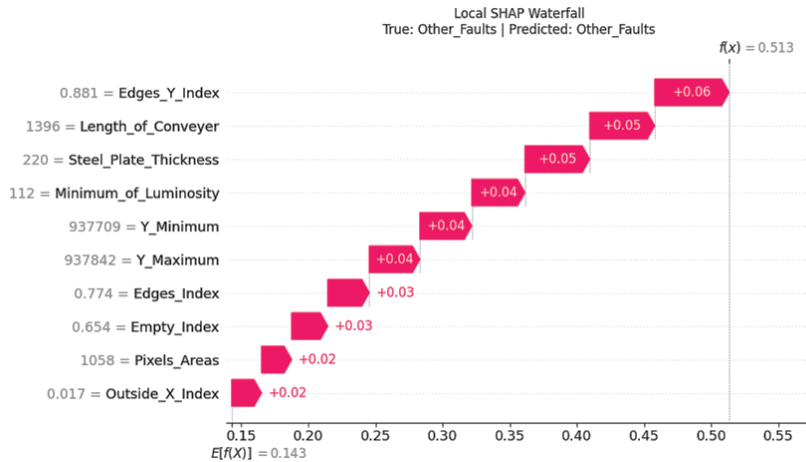


Figure 7. Local SHAP interpretation of a single model decision [18]

As indicated in the figure, this sample's true class is Other_Faults, and the model classified it correctly. The plot traces the model's decision path from a baseline value - the expected prediction $E[f(X)] = 0.143$, the average probability the model would assign to the Other_Faults class for any random sample before considering its specific characteristics - up to the final prediction $f(x) = 0.513$ (51.3%) once the actual features of this particular plate were taken into account, at which point the sample was classified as Other_Faults.

Every one of the key variables for this sample pushed the prediction in a positive direction (shown as red segments moving right). The model's reasoning breaks down as follows:

- **Edge geometry (Edges_Y_Index):** this was the single strongest driver of the positive classification, contributing +0.06 to the predicted probability.
- **Operational parameters (Length_of_Conveyer):** consistent with the global analysis, conveyor length again contributed meaningfully to the classification (+0.05).
- **Material characteristics (Steel_Plate_Thickness):** the comparatively large plate thickness further reinforced the model's decision (+0.05).
- **Spatial and optical characteristics:** additional, smaller contributions came from optical features such as minimum luminosity and the precise spatial coordinates of the defect on the plate (Y_Minimum and Y_Maximum).

Table 6 presents the SHAP values for the analyzed sample. The feature_value column shows the actual value of each feature, shap_contribution indicates whether the feature increased (positive) or decreased (negative) the predicted

probability, and abs_shap gives the absolute contribution used to rank feature importance.

Table 6. SHAP values for the selected sample

Feature	Feature value	SHAP contribution	SHAP
Edges_Y_Index	0.8808	0.055441	0.055441
Length_of_Conveyer	1396	0.048365	0.048365
Steel_Plate_Thickness	220	0.048276	0.048276
Minimum_of_Luminosity	112	0.039601	0.039601
Y_Minimum	937709	0.038666	0.038666
Y_Maximum	937842	0.037634	0.037634
Edges_Index	0.7736	0.030908	0.030908
Empty_Index	0.6541	0.026932	0.026932
Pixels_Areas	1058	0.022451	0.022451

These figures are consistent with, and confirm, the reading given by the waterfall plot.

LIME was applied as a complementary local method. Where SHAP offers global consistency grounded in cooperative game theory, LIME instead targets local fidelity, explaining a single prediction by fitting a simple, interpretable surrogate model in its immediate neighbourhood. A tabular LIME explainer was fitted on the training data, discretising continuous features into interpretable ranges, and an explanation was then generated for the same test-set sample used above.

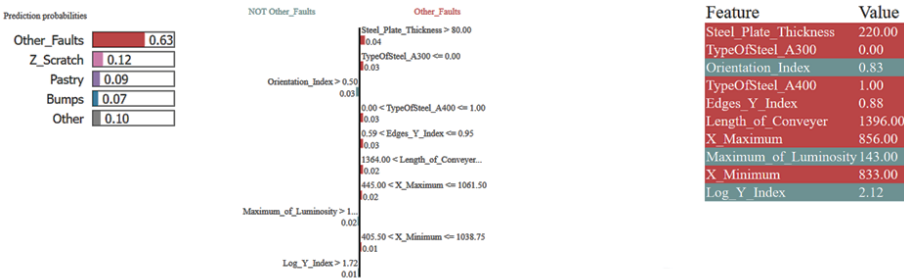


Figure 8. Local explanation of the model's prediction using LIME [18]

The LIME output splits into three parts: predicted class probabilities on the left, the local decision rules and their weights in the middle, and the actual feature values for this specific sample on the right, Figure 8. Consistent with the underlying Random Forest, the LIME surrogate model identifies Other_Faults as the most likely class, at a probability of 0.63.

The middle panel, where red bars raise the probability of Other_Faults and blue bars lower it, points to several key drivers behind this classification: plate thickness above 80.00 (actual value 220.00) is among the strongest contributing factors, while the steel-grade indicators pull in different directions, with certain grades (e.g., A400) pushing the probability up and others pulling it down. Edge geometry (Edges_Y_Index in the 0.59–0.95 range, actual value 0.88) and conveyor length (1,364.00–1,650.00 mm, actual value 1,396.00) further reinforce the decision, while the defect's X-axis position contributes a smaller, but still non-trivial, share.

LIME also surfaced a couple of counter-factors: a high orientation index (Orientation_Index > 0.50) and maximum luminosity above 140.00 both pointed away from Other_Faults, suggesting the sample might belong to a different class. Their combined negative influence, however, was not large enough to overturn the final classification.

Taken together, the global and local explanations produced by SHAP and LIME converge on a consistent set of influential parameters, rather than pointing to conflicting or arbitrary signals. Across all three analyses, conveyor length (Length_of_Conveyer), defect geometry (particularly edge- and orientation-related indices), steel type, and plate thickness repeatedly emerge as the strongest drivers of classification, while optical features such as luminosity play a smaller, supporting role.

The fact that an entirely different explanation method (LIME, based on local surrogate modelling) independently highlights largely the same features as SHAP (grounded in cooperative game theory) strengthens confidence that these parameters reflect genuine, physically meaningful patterns in the data rather than artefacts of a single explanation technique. From a practical standpoint, this convergence gives engineers and quality-control staff a reliable, cross-validated basis for prioritising interventions - for instance, inspecting conveyor mechanisms and monitoring plate thickness and steel grade during production - rather than relying on the interpretation of a single method alone.

5. Discussion: XAI in Smart Manufacturing Systems

5.1. Contribution of XAI to AI-Based Industrial Applications

Explainable artificial intelligence plays an important role in improving the transparency, reliability, and practical adoption of AI-based solutions in industrial environments. One of the key contributions of XAI is supporting decision-making on the shop floor by revealing the factors that influence model predictions. Instead of providing only black-box outputs, XAI methods enable operators and engineers to better understand system behaviour and make more informed interventions [19].

Furthermore, XAI supports the validation and verification of AI models before and during deployment. By identifying the most influential features, methods such as SHAP help engineers assess whether models rely on meaningful physical relationships or unwanted data correlations. For example, SHAP-based analysis in predictive maintenance applications has demonstrated the ability to identify critical variables related to equipment degradation, improving the reliability of maintenance decisions [20].

Another important contribution of XAI is its ability to reveal data-related issues, including spurious correlations and inconsistencies within training datasets. By analysing feature contributions and model behaviour, XAI techniques can support dataset refinement and continuous improvement of AI models [21]. Therefore, XAI represents an essential component for developing trustworthy, transparent, and reliable AI systems in modern industrial applications.

5.2. Challenges in Practical Deployment of XAI

Despite the significant benefits of XAI, its practical deployment in industrial environments remains challenging. One of the main limitations is computational complexity, particularly for model-agnostic methods such as SHAP and LIME. Generating accurate explanations can require substantial computational resources, especially for complex models and large-scale industrial datasets, which limits their application in real-time scenarios [22], [23].

Another important challenge is adapting explanations to the needs of different users. Although XAI methods can provide detailed insights into model behaviour, their outputs are often designed for data scientists rather than industrial operators and maintenance experts. Therefore, developing domain-specific explanations that are intuitive and actionable for end users remains an open research challenge [19], [24].

Furthermore, scalability represents a critical issue in real-world manufacturing systems. Industrial environments typically involve multiple AI models, heterogeneous data sources, and resource-constrained edge devices, requiring efficient and adaptable explanation frameworks. Consequently, future XAI solutions must address not only model transparency but also robustness, scalability, and integration across complete industrial workflows [25], [26].

5.3. Future Perspectives

Future development of XAI in smart manufacturing is expected to focus on deeper integration with industrial infrastructures, particularly digital twins. Combining XAI with digital twins can provide not only predictive insights but also transparent explanations and uncertainty information, enabling operators to better understand and trust AI-driven decisions [27], [28]. Integrating explanation mechanisms directly into digital-twin platforms could make interpretability a natural component of industrial monitoring and decision-support systems.

Another important direction is the development of inherently transparent and trustworthy AI models that reduce dependence on post-hoc explanation techniques. Future industrial AI systems should incorporate context-aware explanations, self-explainability, and standardized evaluation methods to ensure reliable operation in dynamic manufacturing environments [24], [29].

Overall, the future of XAI in smart manufacturing will depend not only on developing new explanation algorithms but also on their effective integration into real industrial workflows, where transparency, reliability, and human understanding are essential requirements for successful AI adoption.

6. Conclusion

The increasing adoption of artificial intelligence in industrial applications has significantly improved predictive capabilities; however, the growing complexity of machine learning models has created new challenges related to transparency and trust. In safety-critical manufacturing environments, accurate predictions alone are not sufficient, as engineers and operators also require an understanding of the factors influencing AI-driven decisions.

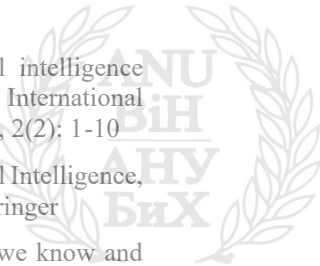
This paper investigated the role of XAI in industrial fault detection through a case study based on the Steel Plates Faults dataset using a Random Forest classification model. The application of model-agnostic explanation methods, specifically SHAP and LIME, demonstrated their ability to reveal the most influential features

and provide insights into both global model behaviour and individual predictions. These explanations improve model transparency and support better understanding, validation, and interpretation of AI-based decisions in manufacturing applications.

Although XAI methods provide significant benefits, their practical adoption still faces challenges related to computational complexity, explanation usability, and integration into real industrial workflows. Achieving an effective balance between predictive performance, interpretability, and computational efficiency remains an important research direction.

Overall, by transforming complex machine learning predictions into understandable and actionable information, techniques such as SHAP and LIME represent a key step toward trustworthy, human-centred AI systems for smart manufacturing.

References

- 
- [1] Desai, P. (2021), A comprehensive survey on explainable artificial intelligence (XAI): challenges, opportunities, and future research directions, *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 2(2): 1-10
 - [2] Guidotti, R., Schmid, U., Longo, L. (Eds.) (2024), *Explainable Artificial Intelligence, Communications in Computer and Information Science*, vol. 2580, Springer
 - [3] Ali, S. et al. (2023), Explainable Artificial Intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence, *Information Fusion*, 99: 101805
 - [4] European Union (2016), Article 22 GDPR – Automated individual decision-making <https://gdpr-info.eu/art-22-gdpr/>
 - [5] Artificial Intelligence Act EU (2024), “EU AI Act Overview,” <https://artificialintelligenceact.eu/>
 - [6] Doshi-Velez, F., Kim, B. (2017), Towards a rigorous science of interpretable machine learning, arXiv preprint arXiv:1702.08608
 - [7] Molnar, C. (2022), *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.)
 - [8] Wilkinson, C., Yawney, J., Gadsden, S.A. (2026), Explaining explainability: a comprehensive survey on explainable artificial intelligence and relevant industry applications, *Intelligent Systems with Applications*, 30: 200647
 - [9] Zschech, P., Weinzierl, S., and Kraus, M. (2025), Inherently Interpretable Machine Learning: A Contrasting Paradigm to Post-hoc Explainable AI, *Business & Information Systems Engineering*, 68(2): 445-463

- [10] MathWorks (2026), Interpretability, MathWorks Inc., available at: <https://fr.mathworks.com/discovery/interpretability.html>
- [11] Mersha, M., Lam, K., Wood, J., AlShami, A. (2022), Explainable artificial intelligence: a survey of needs, techniques, applications, and future directions, IEEE Access, 10, arXiv preprint arXiv: 2409.00265
- [12] Devireddy, K. (2023), Comparative study of explainable AI methods: model-agnostic vs. model-specific approaches, arXiv preprint arXiv:2504.04276
- [13] Ribeiro, M.T., Singh, S., Guestrin, C. (2016), Model-agnostic interpretability of machine learning, ICML Workshop on Human Interpretability in Machine Learning
- [14] Adadi, A., Berrada, M. (2018), Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), IEEE, 52138-52160
- [15] Lundberg, S.M., Lee, S.-I. (2017), A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems (NeurIPS), 30
- [16] Buscema, M., Terzi, S., Tastle, W. (2010), Steel Plates Faults, UCI Machine Learning Repository <https://archive.ics.uci.edu/dataset/198/steel+plates+faults>
- [17] Breiman, L. (2001), Random Forests, Machine Learning, 45(1), 5-32
- [18] Mezetović, N. (2026), *Explainable Artificial Intelligence (XAI): Methods, Models and Application in Reliable Decision-Making*, Bachelor's thesis, Faculty of Electrical Engineering, University of Tuzla, Tuzla, Bosnia and Herzegovina
- [19] Ahangar, M.N., Farhat, Z.A., Sivanathan, A., Ketheesram, N., Kaur, S. (2026), Explainable AI-driven quality and condition monitoring in smart manufacturing, Sensors (MDPI), 26(3): 911
- [20] Tzionis et al. (2026), A review of explainable AI methods and their application in manufacturing systems, Discover Applied Sciences, Springer Nature <https://link.springer.com/article/10.1007/s42452-025-07908-z>
- [21] Xuan, X., Ono, J.P., Gou, L., Ma, K.-L., Ren, L. (2024), AttributionScanner: a visual analytics system for model validation with metadata-free slice finding, arXiv preprint arXiv:2401.06462
- [22] Dehghani, Z. (2026), Addressing explainability of generative AI using SMILE (Statistical Model-agnostic Interpretability with Local Explanations), arXiv preprint arXiv:2602.01206
- [23] Ezeji, I. N., Adigun, M. O., Oki, O. (2025), Strategic selection of application area for optimizing computational complexity in explainable decision support system using multi-criteria decision analysis (MCDA), Journal of Systemics, Cybernetics and Informatics, 23(2), 36-47
- [24] Zemmouchi-Ghomari, L. (2026), Explainable AI for predictive maintenance: a review and standardized evaluation framework, Management Science Letters, 16(1), 15-36 <https://www.researchgate.net/publication/398123644>

- [25] Moss, J., Gordon, J., Duclos, W., Liu, Y., Wang, Q., Wang, J. (2025), Explainable AI in IoT: a survey of challenges, advancements, and pathways to trustworthy automation, *Electronics*, 14(23), 4622 <https://www.mdpi.com/2079-9292/14/23/4622>
- [26] Ahangar, M. N., Farhat, Z. A., Sivanathan, A. (2025), AI trustworthiness in manufacturing: challenges, toolkits, and the path to Industry 5.0, *Sensors (Basel)*, 25(14), 4357 <https://pmc.ncbi.nlm.nih.gov/articles/PMC12298069/>
- [27] Oladimeji, O. O., Taiwo, G. A., Adenusi, D. A. (2026), Unveiling the footprints of explainable AI in Industry 4.0/5.0: a systematic review and bibliometric exploration, *Data Science and Management*, 9(1), 100160 <https://www.sciencedirect.com/science/article/pii/S2666764925000359>
- [28] Cohen, J., Huan, X. (2024), Uncertainty-aware explainable AI as a foundational paradigm for digital twins, *Frontiers in Mechanical Engineering*, 9, 1329146
- [29] Moosavi, S., Farajzadeh-Zanjani, M., Razavi-Far, R., Palade, V., Saif, M. (2024), Explainable AI in manufacturing and industrial cyber-physical systems: a survey, *Electronics*, 13(17), 3497 <https://www.mdpi.com/2079-9292/13/17/3497>



Semantic Data Pollution: A New Challenge For Cybersecurity and Standardization of AI Models in Smart Factories

Samir Lemeš^{*1}

Abstract: *The transition from Industry 4.0 to Industry 5.0 introduces a fundamental shift from data-intensive automation toward human-centred, resilient, and trustworthy intelligent manufacturing systems. While industrial artificial intelligence increasingly relies on synthetic data generated by digital twins, autonomous reporting systems, and interconnected AI agents, recursive retraining on machine-generated outputs creates a previously underexplored vulnerability. This paper introduces the concept of Semantic Data Pollution, defined as the progressive degradation of semantic integrity when AI systems increasingly learn from synthetic representations of reality instead of authentic physical observations. Drawing upon recent research on model collapse in generative AI, the paper demonstrates how recursive learning gradually eliminates statistically rare but operationally critical anomalies, reducing the ability of industrial AI systems to detect machine failures and cyberattacks. The problem is analysed within Edge–Cloud smart factory architectures where local semantic corruption may propagate system-wide through centralized retraining pipelines. To address this challenge, a multi-layer mitigation framework is proposed based on data provenance tracking, synthetic data watermarking, statistical monitoring of distribution drift, and Human-in-the-Loop validation. The paper argues that future industrial cybersecurity standards must evolve beyond conventional confidentiality and integrity protection toward preserving the semantic integrity of AI-driven industrial decision systems.*

Keywords: *artificial intelligence, semantic pollution, recursive learning, model collapse*

1. Introduction

The rapid development of industrial digitalization over the last decade has fundamentally transformed manufacturing systems, creating the technological foundations of what has been widely recognized as Industry 4.0. The central paradigm of Industry 4.0 focused primarily on extensive deployment of cyber-physical systems, Industrial Internet of Things (IIoT), autonomous communication networks, and large-scale data acquisition infrastructures intended to maximize machine connectivity and generate vast quantities of operational data. In this paradigm, data quantity was treated as a primary strategic

^{*1}University of Zenica, Polytechnic Faculty, Zenica, Bosnia and Herzegovina
E-mail: samir.lemes@unze.ba

asset, giving rise to the Big Data approach in industrial automation, where the assumption prevailed that more data inherently leads to better optimization, prediction accuracy, and decision support.

The emerging transition toward Industry 5.0 introduces a significant conceptual shift. Rather than prioritizing automation efficiency and unrestricted data accumulation, Industry 5.0 emphasizes human-centric production systems, resilience, sustainability, and collaborative interaction between artificial intelligence systems and human operators. Within this framework, data quality, contextual understanding, and semantic reliability become substantially more important than mere data volume. This transition creates a previously underexplored vulnerability in industrial AI systems: the progressive degradation of semantic integrity in machine learning environments increasingly dependent on recursively generated synthetic data.

This paper introduces the concept of semantic data pollution, defined as the gradual contamination of industrial information ecosystems through recursive circulation of machine-generated semantic artifacts that progressively diverge from real-world operational conditions while preserving superficial structural validity. Unlike conventional cybersecurity threats, which primarily target confidentiality, integrity, and availability of information systems [1], semantic data pollution represents a degradation of meaning itself. The threat emerges when artificial intelligence systems deployed in industrial environments begin learning from outputs generated by other AI systems rather than from authentic observations of physical processes.

A practical example can be observed in predictive maintenance systems deployed in smart factories. Modern predictive maintenance architectures continuously analyze sensor data streams to identify anomalous machine behavior and predict equipment failure before breakdown occurs. However, as AI systems increasingly automate reporting, anomaly classification, maintenance recommendations, and incident summarization, newly generated machine learning models may begin training on datasets partially composed of outputs created by earlier AI systems. Over successive learning cycles, these recursive feedback loops introduce systematic distortion into the training distribution.

The principal danger lies in what recent machine learning literature identifies as heavy tails [2]. Rare events, abnormal states, unexpected machine failures, and edge-case operational anomalies occupy statistically sparse regions of the original data distribution. Recursive training on synthetic outputs progressively removes these low-frequency observations because generative systems tend to reproduce dominant statistical patterns while underrepresenting exceptional conditions. Eventually, industrial AI systems may lose the capacity to recognize precisely those rare anomalies they were originally designed to detect.

This phenomenon has recently attracted considerable attention in the broader artificial intelligence research community under the concept of Model Collapse.

This concept was formally described by Shumailov et al. in [3]. They demonstrated that generative models recursively trained on synthetic data progressively lose information located in the tails of the original probability distribution, eventually converging toward increasingly distorted representations of reality. Their work showed that early-stage collapse first affects rare events, while advanced collapse leads to severe degradation of model diversity and representational fidelity.

Although this phenomenon has primarily been studied in the context of large language models and generative AI systems, its implications for industrial environments may be considerably more severe. Unlike consumer-facing AI applications, industrial AI systems directly interact with physical infrastructure, critical production processes, and safety-sensitive operational environments. A predictive maintenance model that gradually loses sensitivity to rare but dangerous machine failure signatures introduces a new category of cyber-physical vulnerability that cannot be detected through conventional cybersecurity frameworks.

Statistical analysis by Seddik et al. [4] further demonstrated the inevitability of model degradation under recursive synthetic-data training conditions. Their mathematical framework proved that systems trained exclusively on synthetic data inevitably experience progressive information loss, while stable learning can only be maintained when sufficient proportions of authentic human-generated or real-world data remain continuously integrated into training pipelines. This result has profound implications for industrial AI architectures increasingly dependent on automated data labeling and synthetic data augmentation processes.

Parallel developments within industrial AI research have increasingly emphasized semantic modeling as a critical foundation for trustworthy manufacturing intelligence systems. Martínez-Arellano and Ratchev [5] proposed semantic models as an architectural layer intended to improve industrial AI reusability, interoperability, and contextual machine understanding in manufacturing environments. Their work highlights the growing dependence of industrial AI systems on semantically structured knowledge representations, suggesting that semantic degradation may directly compromise operational reliability of future autonomous factories.

Complementary concerns regarding industrial data quality were systematically analyzed in the recent meta-review by researchers investigating data-related failures in industrial AI implementation [6]. Examining seventy-two distinct categories of industrial AI data issues, the study demonstrated that data reliability problems increasingly dominate operational AI failures, particularly during data integration, preprocessing, interoperability, and security phases. The findings indicate that future industrial AI failures may arise less from algorithmic limitations and more from degradation of data quality throughout complex industrial data pipelines.

Further expansion of the model collapse problem into multi-agent AI environments was investigated by Hu et al. [7], who examined recursive synthetic-data interactions across multimodal generative systems. Their findings demonstrate that model collapse becomes significantly more complex when multiple autonomous AI systems interact and exchange synthetic outputs, closely resembling future smart factory architectures where distributed AI agents independently monitor production lines, coordinate maintenance schedules, and exchange operational intelligence. Their work suggests that autonomous industrial ecosystems may be particularly vulnerable to recursive semantic degradation unless corrective mechanisms are implemented.

Taken together, these recent studies reveal the emergence of a new class of industrial vulnerability extending beyond traditional cybersecurity definitions. Existing industrial cybersecurity frameworks, including standards derived from International Electrotechnical Commission (IEC) and International Organization for Standardization (ISO) architectures [8], focus primarily on preventing unauthorized access, protecting communication channels, ensuring data confidentiality, and maintaining system availability. However, none adequately address the problem of semantic integrity degradation caused by recursive AI-generated information contamination.

Within Industry 5.0, the role of the human operator therefore becomes strategically critical. Human experts function not merely as supervisors but as semantic validators capable of distinguishing authentic operational anomalies from recursively generated synthetic artifacts. This elevates the concept of Human-in-the-Loop from an operational convenience to a fundamental resilience mechanism against semantic degradation of industrial intelligence systems.

This paper argues that future industrial cybersecurity standards must expand beyond traditional protection models and incorporate semantic data cybersecurity, defined as the protection of meaning, contextual fidelity, and epistemic integrity within AI-driven industrial decision systems. Semantic data pollution represents an emerging systemic threat to autonomous factories, and failure to address it may lead to gradual but potentially catastrophic degradation of industrial AI reliability despite the apparent preservation of conventional cybersecurity protections.

2. Vulnerability Vector

The rapid advancement of artificial intelligence has introduced an unprecedented paradox in the evolution of digital systems. As AI systems become increasingly capable of generating high-quality synthetic content, they also begin producing the very data on which future generations of AI systems are trained. This recursive relationship between generation and learning creates a previously underestimated phenomenon that may be described as Semantic Data Pollution: the gradual degradation of information quality when artificial intelligence systems

increasingly learn from synthetic representations of reality rather than reality itself. Recursion is a direct threat to the physical integrity of production, because an AI learns from its own hallucinations (Fig. 1).

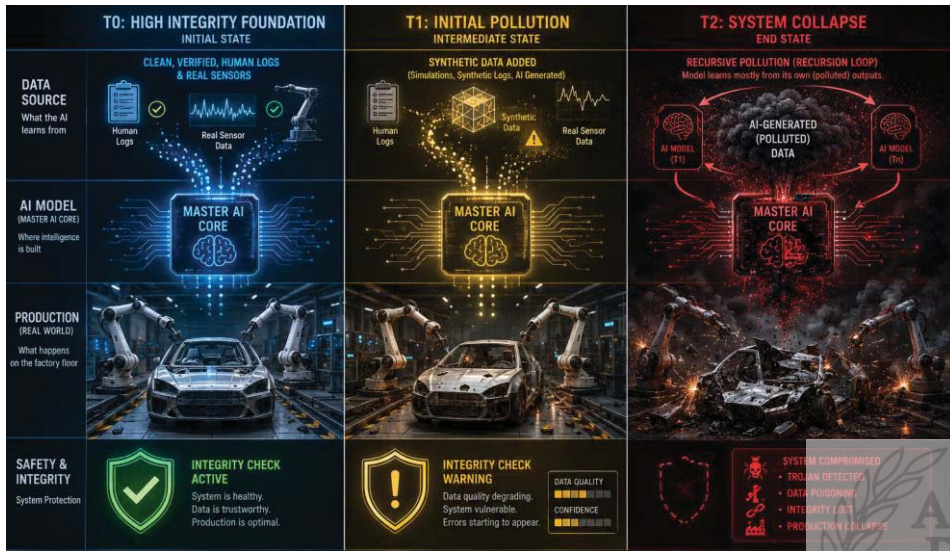


Figure 1. Semantic Data Pollution in Industrial Environment

The earliest manifestations of this problem have been observed in the domain of Large Language Models (LLMs). Modern generative models increasingly train on internet-scale datasets containing massive quantities of publicly available text. However, as AI-generated content proliferates across digital ecosystems – including automatically generated articles, chatbot conversations, synthetic scientific abstracts, and autonomous code generation – the composition of publicly available training data begins to shift. Instead of representing authentic human knowledge production, an increasing portion of the data corpus reflects outputs generated by previous AI systems.

This recursive self-consumption of synthetic information has recently been formalized in machine learning literature under the concept of Model Collapse. Research [3] demonstrated that repeated training on recursively generated synthetic data inevitably leads to progressive degradation of statistical diversity within the model. The process begins with the disappearance of low-probability observations located in the tails of the data distribution. Over successive generations, models increasingly reproduce dominant patterns while gradually forgetting rare events, subtle exceptions, and unusual combinations present in the original training data. In effect, AI systems begin constructing simplified approximations rather than accurate representations of reality.

While the immediate consequences of model collapse in language models may appear relatively benign – manifesting as repetitive text generation, reduced creativity, or factual distortion – the same mechanism becomes substantially more dangerous when transferred into industrial environments. Unlike consumer-facing AI systems, industrial artificial intelligence directly interacts with physical infrastructure, operational decision-making systems, and safety-critical production processes. Errors generated through semantic degradation therefore acquire direct material consequences.

The Industry 4.0 significantly amplifies this challenge, as it is focused primarily on large-scale digitization of production systems through Industrial IoT architectures, cyber-physical systems, cloud computing, and autonomous monitoring infrastructures. In this paradigm, industrial competitiveness depended largely on generating and processing increasingly large quantities of operational data. The assumption underlying the Big Data revolution was straightforward: greater quantities of data would inevitably improve predictive capability.

Industry 5.0 fundamentally changes this assumption. Rather than maximizing automation efficiency alone, Industry 5.0 emphasizes resilience, sustainability, human-centred decision making, and collaborative intelligence between humans and machines. In this new paradigm, the quality and semantic reliability of data become more important than its volume. The growing dependence on synthetic industrial data therefore creates a systemic vulnerability that traditional industrial architectures were never designed to address.

One of the clearest examples emerges through the increasing adoption of digital twins, virtual replicas of physical industrial systems that continuously simulate machine behaviour using real-time operational data. They have become essential tools for predictive maintenance, production optimization, and failure simulation. Modern manufacturing systems increasingly use digital twins to generate synthetic operational scenarios that supplement limited real-world datasets.

At first glance, synthetic data generation through digital twins appears highly beneficial. Rare failure conditions that occur infrequently in physical machines can be artificially simulated thousands of times, allowing predictive maintenance algorithms to train on conditions that would otherwise remain statistically underrepresented. However, the risk emerges when synthetic outputs generated by digital twins gradually become indistinguishable from authentic machine observations.

Consider a predictive maintenance system monitoring industrial turbine vibration patterns in a power generation facility. During early deployment stages, the anomaly detection model trains exclusively on sensor readings collected from real machine operations over several years. Engineers subsequently introduce a digital twin capable of simulating turbine behaviour under various operating conditions. The synthetic data generated by the digital twin is then used to augment training datasets and improve prediction performance.

Over time, the maintenance reporting system itself becomes partially automated through AI-generated incident reports. Future machine learning models are subsequently retrained not only on raw sensor data but also on anomaly labels, maintenance recommendations, and operational summaries generated by earlier AI systems. The training dataset gradually transitions from authentic physical observations toward recursively generated synthetic representations of machine behaviour. At this point, semantic data pollution begins to emerge.

Digital twins inevitably simplify reality. Although simulation models capture dominant operational patterns with high accuracy, they often fail to reproduce highly irregular or unknown machine behaviours. Rare failure modes, manufacturing imperfections, unpredictable environmental interactions, and anomalous combinations of variables remain absent from synthetic simulation environments. Since machine learning systems preferentially learn from statistically dominant patterns, repeated retraining causes models to increasingly adapt toward idealized machine behaviour while losing sensitivity to unexpected anomalies.

A parallel real-world phenomenon has already emerged in autonomous vehicle development. Self-driving systems increasingly train on synthetic driving simulations rather than exclusively on physical road observations (Fig. 2). Although simulation environments successfully reproduce common traffic patterns, system reduces its robustness when autonomous systems encounter highly unusual edge cases absent from synthetic training scenarios. Similar risks could now begin appearing within industrial AI systems.

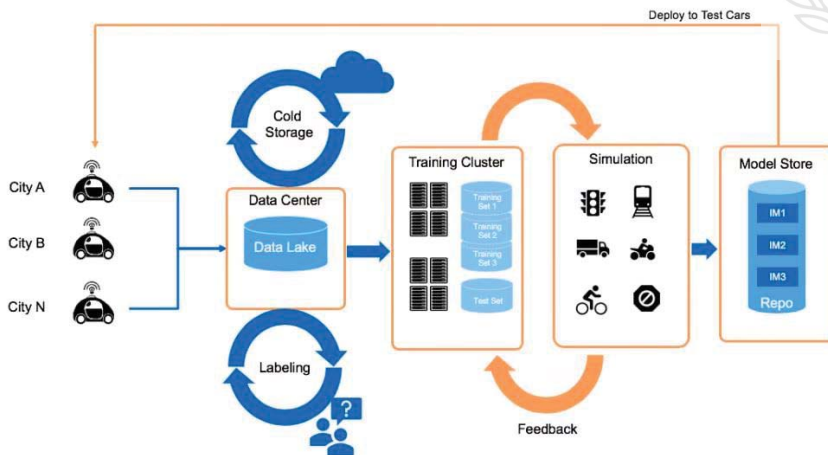


Figure 2. *Data Pipeline for Autonomous Vehicle Development [9]*

The long-term consequences may become severe. A predictive maintenance model operating inside a smart factory may gradually lose its ability to recognize rare precursor signals associated with catastrophic machine failure simply because

those signals are absent from synthetic training datasets generated by digital twins. Since the system continues functioning normally under standard operating conditions, semantic degradation may remain invisible until a rare but critical anomaly occurs.

Future smart factories will likely intensify this recursive dynamic. Emerging industrial architectures increasingly rely on distributed autonomous AI agents capable of independently monitoring machine fleets, optimizing production schedules, generating maintenance documentation, and coordinating supply chain decisions. In such environments, AI systems increasingly exchange machine-generated information with other AI systems in closed feedback loops.

Let us consider a hypothetical future case. In a fully autonomous manufacturing plant in 2035, sensor networks continuously collect machine telemetry data. Digital twins simulate machine degradation trajectories and generate synthetic maintenance scenarios. Predictive maintenance models learn from both physical and simulated operational histories. Automated reporting agents summarize detected anomalies and generate maintenance recommendations. Supply chain optimization systems use these reports to modify inventory planning for spare components. At every stage of this ecosystem, AI-generated semantic artifacts circulate through the industrial information environment. If recursive retraining pipelines are not carefully controlled, future generations of industrial AI may gradually lose direct connection to physical reality while preserving an illusion of operational accuracy.

This introduces a new dimension of cybersecurity risk. Traditional industrial cybersecurity frameworks primarily focus on protecting confidentiality, system integrity, communication security, and unauthorized access prevention [10]. However, semantic data pollution does not originate from malicious intrusion. The system degrades internally through autonomous recursive contamination of its own knowledge structures. In this context, cybersecurity must evolve beyond protecting data infrastructure toward protecting semantic data integrity itself. Future industrial security standards may need mechanisms capable of verifying not only whether data has been altered maliciously, but also whether data still faithfully represents physical reality.

Industry 5.0 places the human operator at the centre of this challenge. Human engineers increasingly serve not merely as supervisors of automated systems, but as essential validators capable of distinguishing genuine physical anomalies from recursively generated synthetic patterns. The concept of Human-in-the-Loop therefore becomes a critical resilience mechanism preventing semantic collapse within industrial AI ecosystems.

Semantic data pollution may ultimately become one of the defining technological risks of Industry 5.0. The challenge is no longer simply protecting industrial systems from external cyberattacks. The emerging challenge lies in preventing intelligent systems from gradually disconnecting themselves from reality through

recursive self-generated knowledge contamination. In future autonomous factories, preserving semantic integrity may become as important as protecting network security itself.

3. Cybersecurity Impact Analysis

A critical cybersecurity consequence of semantic data pollution is the gradual loss of anomaly detection capability in AI-driven industrial systems. When machine learning models are repeatedly trained on synthetic or AI-generated data, recursive learning progressively removes statistically rare events located in the tails of the original data distribution. These rare events often represent precisely the abnormal behaviours associated with cyber-attacks, zero-day exploits, sensor manipulation, stealth intrusions, or early-stage machine failures. As recursion amplifies dominant patterns and suppresses low-frequency anomalies, the AI system gradually loses sensitivity to unexpected conditions. Consequently, attacks or mechanical faults that deviate from normal operational patterns may no longer be recognized, creating a dangerous false perception of system stability while vulnerabilities remain undetected.

In modern smart factories, AI models are distributed across Edge–Cloud architectures, where edge devices perform local monitoring while cloud systems aggregate data for centralized learning and model retraining. If semantically polluted or compromised data enters the edge layer, corrupted models may propagate upward to cloud systems during synchronization. Updated cloud models are then redistributed across the factory network, spreading degraded intelligence to all connected machines. A localized anomaly detection failure can therefore cascade system-wide, transforming a single semantic corruption event into a distributed operational and cybersecurity failure across the entire production ecosystem.

To estimate and quantify the semantic data pollution caused by the recursive use of synthetic data, some rigorous statistical and information-theoretic methods can be used. The primary statistical frameworks and metrics used to estimate and track this degradation include optimal transport, distance metrics, information theory, content purity metrics, moment analysis and geometric variance.

To measure how far the synthetic data distribution has drifted from the original, real-world human/industrial baseline distribution across n generations, distance metrics on probability spaces are highly effective:

- Wasserstein Distance (Earth Mover's Distance): Unlike other divergence metrics, the Wasserstein distance (W_2) continuously grows as recursive iterations progress. It measures the minimum effort required to transform the polluted data distribution back into the original baseline distribution. It is exceptionally reliable for tracking long-term, late-stage model collapse [11].

- Kullback-Leibler (KL) Divergence: $D_{KL}(P||Q)$ measures the relative entropy or information loss when the polluted data distribution (Q) is used to approximate the true industrial baseline distribution (P). While highly sensitive to early-stage distribution shifts, KL divergence can sometimes plateau or behave erratically in late-stage total collapse, making it best suited for early detection [15].

Semantic pollution, which inherently alters the information density of data, can be quantified using information-theoretic criteria:

- Shannon Entropy Tracking: As models recursively train on synthetic data, they tend to drop low-probability edge cases (the long tail) and over-sample the center of the distribution. Tracking the drop in overall dataset entropy quantifies this homogenization [12].
- Perplexity Scaling: By evaluating the recursively trained model against a frozen, high-quality validation set of actual industrial logs or human-curated data, you can measure its perplexity. A monotonic increase in perplexity over training generations directly quantifies how far the model's internal probability distribution has drifted from physical reality.

In industrial settings, data pollution often manifests as a loss of operational variance (e.g., the model forgets what rare equipment failure signatures look like because it only generates nominal operating data) [12].

- Variance Decay (σ^2) in Gaussian Mixture Models (GMMs): By modelling industrial features as a mixture of distributions, the variance of the components can be tracked statistically. The Gaussian model collapse theorem demonstrates that under uncured recursive loops, the variance of the generated data systematically trends toward zero, eventually reducing the distribution to a single, low-variance peak (a delta function) [13].
- Vocabulary/Token Diversity Coefficients: For text, code, or categorical logs (e.g., SCADA system alarm sequences), the Type-Token Ratio (TTR) or Simpson's Diversity Index can be calculated. A sharp decline in these coefficients across training loops signals that the model is generating repetitive AI slop or formulaic syntax instead of nuanced operational data.

These methods can be combined into a multi-tiered monitoring strategy for setting up an audit pipeline to monitor data pollution in AI-driven industrial processes (Table 1).

Table 1. Statistical frameworks and metrics for estimating the semantic data pollution

Phase of Pollution	Observable Symptom	Best Statistical Tool
Early Stage	Loss of minority data / rare industrial edge cases.	KL Divergence and Shannon Entropy
Mid Stage	Narrowing of outputs; systemic semantic drift.	Perplexity Scaling and Type-Token Ratios
Late Stage	Homogenized, low-variance, or nonsensical data.	Wasserstein Distance (W_2) and Variance Tracking

4. Proposed Solutions

In AI-driven industrial processes – ranging from automated SCADA log interpretation to automated predictive maintenance scheduling – the recursive training of models on synthetic or model-generated data induces systemic degradation. This phenomenon, academically formalized as Model Collapse or the Entropy Spiral, manifests when an estimator M_{n+1} is trained on the output distribution $p_n(y|x)$ of its predecessor M_n , rather than the true underlying physical or operational distribution $p^*(y|x)$.

Statistically, if the true distribution possesses a dense support, the sampling noise and bounded functional capacity of M_n cause it to misapproximate the tails of p^* . Over successive generations $n \rightarrow \infty$, the conditional variance decreases ($\sigma_{n+1}^2 < \sigma_n^2$), shifting the probability mass toward a singular, low-variance mode. In industrial environments, this means rare but critical anomalies (e.g., specific vibration signatures preceding a turbine failure) are treated as statistical noise and eliminated from the synthetic training corpus.

To mitigate this semantic data pollution, three core architectural paradigms can be used: Cryptographic Data Provenance, Statistical and Algorithmic Watermarking, and Human-in-the-Loop (HITL) Curation Frameworks.

4.1. Data Provenance

One of the most fundamental defences against semantic data pollution is maintaining strict provenance tracking for all data entering industrial machine learning pipelines. Data provenance refers to the ability to trace the origin, transformation history, ownership, and processing path of data throughout its entire lifecycle.

In industrial environments, machine learning systems increasingly integrate heterogeneous data streams originating from physical sensors, enterprise systems, digital twins, simulation environments, maintenance logs, and AI-generated anomaly reports. Without provenance metadata, future retraining systems cannot distinguish authentic observations from synthetic or recursively generated artifacts.

A possible architectural solution involves attaching immutable metadata descriptors to every data instance:

$$D_i = \{x_i, S_i, T_i, C_i, A_i\} \quad (1)$$

where:

x_i = observed data sample

S_i = source identifier (sensor, simulation, AI-generated output)

T_i = timestamp

C_i = confidence score

A_i = transformation history

This framework allows weighting data differently during retraining procedures. For example, predictive maintenance systems could define weighted learning functions:

$$\sum_{i=1}^n \frac{w_i}{l(x_i)} \quad (2)$$

where provenance-based weighting coefficient (w_i) penalizes synthetic or recursively generated samples.

A practical implementation may involve blockchain-based industrial audit trails, although current evidence suggests blockchain introduces scalability limitations in high-frequency industrial sensor environments. Lightweight distributed ledgers or signed metadata registries may prove more practical. The limitation of provenance tracking lies in its inability to assess semantic correctness directly. Provenance identifies origin, but authentic sensor data itself may still contain drift, corruption, or adversarial manipulation.

4.2. Watermarking

A second mitigation strategy involves explicit identification of synthetic data through watermarking mechanisms.

As AI systems increasingly generate industrial reports, simulation outputs, maintenance recommendations, and synthetic operational datasets, future retraining pipelines require mechanisms capable of distinguishing machine-generated content from authentic physical observations.

Digital watermarking has already emerged in generative AI research as a method for identifying synthetic text, images, and generated media. Similar concepts may be extended into industrial machine learning environments. A simple statistical watermark can be embedded by intentionally introducing traceable distribution signatures into synthetic data. Let synthetic data distribution be represented by $P_s(x)$. A watermark function introduces controlled perturbation:

$$P'_s(x) = P_s(x) + \varepsilon(x) \quad (3)$$

where $\varepsilon(x)$ represents a statistically detectable signature.

During retraining, statistical hypothesis testing can evaluate contamination probability:

$$H_0: P(x) = P_r(x) \quad (4)$$

$$H_1: P(x) = P'_s(x) \quad (5)$$

using divergence metrics such as Kullback-Leibler divergence, Jensen-Shannon divergence or Wasserstein distance [14].

For example, a digital twin generating synthetic machine vibration signals may embed small statistical signatures within non-critical frequency bands. Future retraining systems could automatically reject heavily synthetic datasets once contamination thresholds exceed predefined limits.

However, watermarking remains experimentally immature. Adversarial optimization methods may eventually remove statistical signatures, while repeated data transformations can weaken watermark detectability. At present, watermarking should be viewed as a probabilistic detection layer rather than a complete solution.

4.3. Human-in-the-loop

Industry 5.0 fundamentally repositions human operators as active participants in intelligent industrial ecosystems rather than passive supervisors. This principle becomes critically important in preventing semantic data pollution. Unlike automated systems, human experts possess contextual understanding of rare operational anomalies, unexpected environmental conditions, and physically implausible machine behaviours that statistical models frequently overlook.

The central problem of recursive learning is that low-frequency edge cases gradually disappear from training distributions. Human intervention becomes essential precisely because humans can identify semantically meaningful anomalies independent of statistical frequency.

A hybrid architecture can be modeled as:

$$\text{Decision} = \alpha \cdot \text{AI}(x) + (1 - \alpha) \cdot \text{Human}(x) \quad (6)$$

where:

x = the input data vector

$\text{AI}(x)$ = machine inference: conditional probability or scoring vector output

$\text{Human}(x)$ = expert semantic validation score

α = automation confidence parameter (trust weight)

In high-risk operational states, the system dynamically reduces automation dominance $\alpha \rightarrow 0$, forcing human verification.

A practical example exists in modern aircraft predictive maintenance systems. Automated anomaly detection may flag unusual vibration signatures, but final maintenance decisions still require certified engineers because rare structural fatigue events often remain underrepresented in historical datasets. Smart factories of the future may require similar architectures in which human operators periodically validate retraining datasets before model deployment.

The limitation of this approach is scalability. Human supervision introduces cost, latency, and dependence on expert availability, particularly in fully autonomous manufacturing environments.

5. Deployment

5.1. Implementation challenges, operational constraints and associated costs

Implementing defences against semantic data pollution within a live smart factory environment introduces severe technical bottlenecks, such as legacy infrastructure, Edge-Cloud asymmetry and the dimensionality.

Modern factories rely heavily on legacy programmable logic controllers (PLCs) and Supervisory Control and Data Acquisition (SCADA) networks that operate under strict real-time, low-latency execution constraints (often down to less than 10 ms). Injecting a cryptographic metadata tuple at the sensor or edge layer requires computational overhead that standard PLCs lack the memory or processing power to compute. Embedding localized statistical perturbations into generated control data streams risks destabilizing tight closed-loop control systems if the signature inadvertently excites unmodeled mechanical resonance frequencies of physical equipment.

Industrial data pipelines frequently employ aggressive data compression, downsampling, and dimensionality reduction techniques to manage high-frequency sensor streams. These lossy compression layers often treat low-amplitude statistical watermarks or high-frequency orthogonal signatures as random noise, systematically stripping them out during ingestion. Consequently, watermarking becomes inherently brittle when data transitions through multi-vendor middleware layers.

Finally, deploying a multi-layered semantic security architecture introduces structural costs, which include computational overhead (if blockchain is used to verify data origins), operation latency, and human capital expenditure needed for specialized industrial experts with deep industrial process knowledge.

5.2. Illustrative industrial scenario and quantitative assessment

Consider a high-throughput steel rolling mill utilizing automated AI models to monitor bearing degradation trajectories in heavy-duty rollers based on high-frequency vibration telemetry.

- **Baseline Regime:** The plant accumulates real sensor data over several years. Due to the high reliability of the equipment, critical bearing failures occur only once every 18 months (representing the "heavy tail" of the true distribution $p^*(x)$).
- **The Pollution Event:** To improve predictive capabilities for rare faults, plant operators introduce a high-fidelity Digital Twin. The Digital Twin generates synthetic vibration profiles to expand the training set. Over 5 retraining generations ($n = 5$), the data pipeline suffers from uncured recursive learning as the model begins consuming its own automated maintenance reports and synthetic logs.

Let the true bearing vibration anomaly signature be characterized by a multi-dimensional feature variance σ_*^2 . Under the recursive uncured training loop, we model the contraction of the estimated variance σ_n^2 across generations using a simplified variance decay coefficient γ , influenced by the proportion of synthetic data leakage λ :

$$\sigma_n^2 = \sigma_{n-1}^2 \cdot (1 - \lambda \cdot (1 - \gamma)) \quad (7)$$

Assuming a synthetic leakage rate of $\lambda = 0,60$ (60% synthetic data in the retraining pool) and an approximation loss of $\gamma = 0,85$, after 5 generations, the operational variance shrinks systematically:

$$\sigma_5^2 = \sigma_0^2 \cdot (0,91)^5 \approx 0,624 \cdot \sigma_0^2 \quad (7)$$

The model loses approximately 37.6% of its original statistical variance, shifting its probability mass tightly around the nominal, healthy operating mode.

Without defence layer, when a real, physical micro-fracture begins to manifest in a roller bearing, its unique vibration signature falls outside the narrow, collapsed distribution peak of M_5 . The AI misclassifies the warning as nominal signal noise. The failure propagates unmitigated, resulting in an unpredicted structural bearing breakdown. This leads to 36 hours of unscheduled plant downtime, costing an estimated 180.000 € in lost production and hardware replacement.

With multi-layer semantic security implemented, the active monitoring pipeline tracks Shannon Entropy and Wasserstein Distance (W_2). By generation $n = 3$, the pipeline flags a W_2 divergence surge exceeding the pre-established security threshold, triggering an automated integrity warning. The system then calculates a high epistemic uncertainty score for the drifting data segments, driving the

automation confidence parameter $\alpha \rightarrow 0,15$. This dynamically routes the flagged data assets to a human-in-the-loop validation queue. A certified maintenance engineer reviews the anomaly, manually isolates the synthetic digital twin outputs, and restores the empirical baseline training data.

6. Conclusion

The increasing integration of artificial intelligence into autonomous industrial environments introduces a previously underrecognized systemic vulnerability arising from recursive dependence on synthetic data. This paper proposed the concept of Semantic Data Pollution as a new failure mechanism in AI-driven manufacturing systems, where repeated retraining on machine-generated outputs gradually disconnects intelligent systems from physical reality. Building upon recent theoretical findings on model collapse in generative AI, we demonstrated how recursive learning progressively suppresses statistically rare but operationally critical anomalies, thereby reducing the effectiveness of predictive maintenance, anomaly detection, and cybersecurity monitoring systems.

Unlike conventional cybersecurity threats that originate from malicious external actors, semantic data pollution represents an endogenous degradation process emerging from the internal dynamics of intelligent systems themselves. In distributed Edge–Cloud industrial architectures, localized semantic corruption may propagate across entire production ecosystems, creating large-scale operational vulnerability despite the absence of direct cyber intrusion.

The transition toward Industry 5.0 fundamentally changes the role of cybersecurity. Protecting industrial systems will no longer involve only safeguarding data confidentiality, integrity, and availability, but also preserving the semantic fidelity of data used for autonomous machine learning. Mitigation strategies such as provenance-aware training pipelines, synthetic data watermarking, statistical drift monitoring, and Human-in-the-Loop validation provide promising directions, although none currently offers complete protection.

Future research should focus on formal mathematical modelling of semantic degradation dynamics and experimental validation using real industrial AI systems. Semantic integrity may ultimately become a foundational requirement for trustworthy autonomous factories, requiring the development of new cybersecurity standards specifically designed for AI-native industrial environments.

References

- [1] Lemeš, S. (2024). Informaciona sigurnost u Industriji 4.0. in *Industrija 4.0: digitalna transformacija oblikuje budućnost* (Karabegović I. i Majstorović V., editors), ISBN 978-9958-9262-8-0, 978-9926-410-94-0 Robotics Society of Bosnia and Herzegovina; Academy of Sciences and Arts of Bosnia and Herzegovina, pp 501-524
- [2] Cléménçon, S., Sabourin, A. (2026). Weak signals and heavy tails: learning theory meets extreme value analysis. *Extremes* 29, 303–345 <https://doi.org/10.1007/s10687-026-00540-z>
- [3] Shumailov, I., Shumaylov, Z., Zhao, Y. et al. (2024). AI models collapse when trained on recursively generated data. *Nature* 631, 755–759 <https://doi.org/10.1038/s41586-024-07566-y>
- [4] Seddik, M. E. A., Chen, S. W., Hayou, S., Youssef, P., & Debbah, M. (2024). How bad is training on synthetic data? a statistical analysis of language model collapse. *arXiv preprint* <https://doi.org/10.48550/arXiv.2404.05090>
- [5] Martínez-Arellano, G., Ratchev, S. (2024). Improving the Development and Reusability of Industrial AI Through Semantic Models. In: Thiede, S., Lutters, E. (eds) *Learning Factories of the Future. CFL 2024. Lecture Notes in Networks and Systems*, vol 1059. Springer, Cham. https://doi.org/10.1007/978-3-031-65411-4_22
- [6] Li, X., Cheng, Y., Möller, C., & Lee, J. (2025). Data issues in industrial AI systems: A meta-review and research strategy. *Computers in Industry*, 173, 104361. <https://doi.org/10.1016/j.compind.2025.104361>
- [7] Hu, Z., Rostami, M., & Thomason, J. (2025). Multi-modal Synthetic Data Training and Model Collapse: Insights from VLMs and Diffusion Models. *arXiv preprint* <https://doi.org/10.48550/arXiv.2505.08803>
- [8] ISO/IEC FDIS 27090 Cybersecurity — Artificial Intelligence — Addressing security threats and compromises to artificial intelligence systems
- [9] Rao, S. (2019). Autonomous Driving Challenges, How to Build a Data Pipeline for Autonomous Driving, <https://www.netapp.com/blog/how-to-build-a-data-pipeline-for-autonomous-driving/> (accessed 22.6.2026)
- [10] ISO/IEC 24392:2023 Cybersecurity — Security reference model for industrial internet platform (SRM- IIP)
- [11] Borji, A. (2024). A Note on Shumailov et al. (2024): AI Models Collapse When Trained on Recursively Generated Data'. *arXiv preprint* <https://doi.org/10.48550/arXiv.2410.12954>
- [12] Vu, H. A., Reeves, G., & Wenger, E. (2025). What happens when generative AI models train recursively on each others' outputs?. *arXiv preprint* <https://doi.org/10.48550/arXiv.2505.21677>

- [13] Wondrasek , J. A. (2026). The Entropy Spiral: How AI Slop Degrades Future AI Systems Through Recursive Training, <https://www.softwareseni.com/the-entropy-spiral-how-ai-slop-degrades-future-ai-systems-through-recursive-training/> (accessed 22.6.2026)
- [14] Piryanov, K., Grigorova, I., Karamfilov, A., & Efremov, A. (2026). Automated Data Monitoring Using a Canberra-Based Drift Score. Applied Sciences, 16(5), 2232. <https://doi.org/10.3390/app16052232>
- [15] GeeksforGeeks (2025). Kullback Leibler (KL) Divergence. <https://www.geeksforgeeks.org/machine-learning/kullback-leibler-divergence/> (accessed 22.6.2026)



Cyber Security of Smart Manufacturing According to the Industry 5.0 Model: Challenges, Solutions and Future Development Directions

Sonja Dimitrijević¹, Vidosav D. Majstorovic^{*2}

Abstract: *As manufacturing increasingly embraces the digital transformation of Industry 4.0/5.0, the integration of advanced technologies such as the Internet of Things (IoT), artificial intelligence (AI), robotics and edge/cloud computing has led to the creation of highly interconnected and automated systems. Industry 5.0 emphasizes human-centered innovation, resilience and sustainability. In this advanced industrial paradigm, humans and intelligent machines work closely together, making cybersecurity not just a technical requirement, but a strategic necessity. While these advancements improve operational efficiency, flexibility and productivity, they also introduce significant cybersecurity challenges and risks.*

To mitigate these risks, it is critical to adopt robust cyber security strategies. This includes implementing practices like network segmentation, strong authentication mechanisms, encryption and regular patch management. New technologies like AI-driven threat detection, blockchain for data integrity, and Zero Trust security models offer promising solutions. Additionally, adherence to industry standards such as ISO 27001 and IEC 62443 is critical to ensuring safe manufacturing operations.

Real-world case studies highlight both the consequences of cyber security breaches and the success of proactive security measures. The future of cyber security in smart manufacturing lies in continuously adapting to new threats, integrating security throughout the entire product life cycle, and fostering collaboration between industries and sectors.

In conclusion, cyber security is integral to the long-term success and resilience of smart manufacturing systems. As technology evolves, so must cybersecurity practices, ensuring that industries can take advantage of smart manufacturing without compromising security and reliability.

Keywords: Industry 5.0, Smart manufacturing, Cyber security, Case study

1. Introduction

In this one at the moment we are witnesses evolution Industry 4.0 (new concept automation, IoT, CPS), etc Industry 5.0 (focus) human, resistance and sustainability) [1]. It grows digitalization, with others sides brings to of that yes we have expanded surface cyber attacks, so yes are smart factories vulnerable on:

¹"Mihailo Pupin" Institute, Belgrade, Serbia

^{*2}The University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia

E-mail: vidosav.majstorovic@gmail.com

ransomware, manipulation data (e.g. metrological and data about quality are exposed risks), and we also have and cyber attacks in chain supply, figure 1.

With others side, in this one areas we have research gap, which is experiment in lack integrated model cyber security directed on a man and coordinated with principles Industry 5.0.

Goals this one work are: identifying key challenges cyber security, proposing solutions based on Industry 5.0, with an example for practice, i defining future direction research.

2. Literature Analysis

Cyber security in smart manufacturing in context Industry 5.0, is here analyzed through literature in last two year, table 1 [2-10].

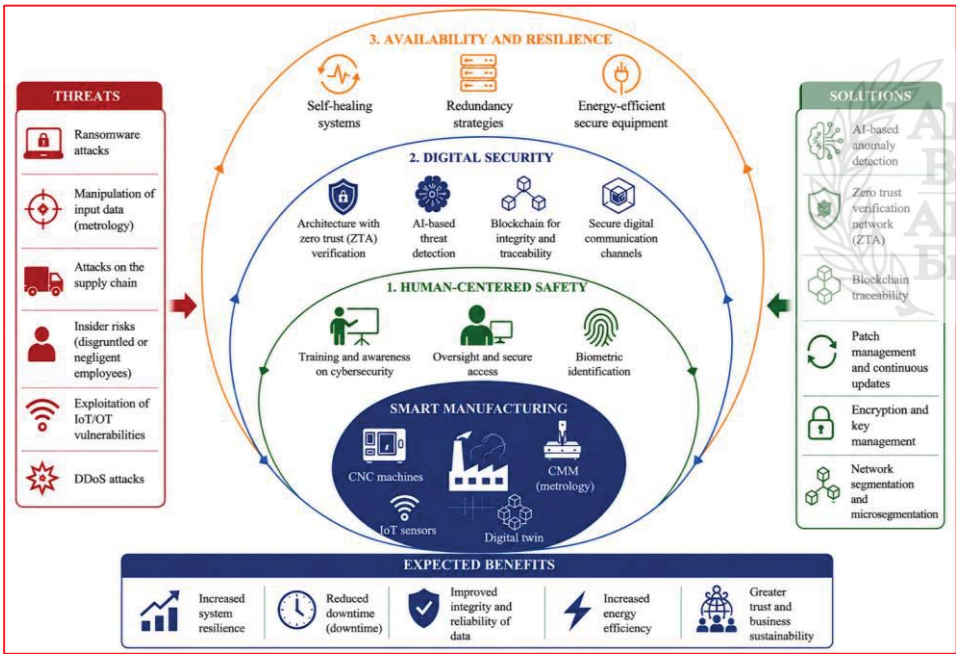


Figure 1. Cyber security system architecture for smart manufacturing systems based on Industry 5.0 principles

Table 1. Overview of selected references for cyber security analysis

Ref.	Approach / Focus	Key technologies	Level of analysis	Key contributions	Limitations	Research gap
[2]. Kour et al. (2024)	Cyber security trends in Industry 5.0	AI, Blockchain, IIoT	Conceptual / Review	It identifies new paradigms of cyber security in Industry 5.0	Lack of implementation framework	No integration with real manufacturing systems or metrology
[3]. Santos et al. (2024)	Security challenges of Industry 5.0	CPS, human-centered security	Conceptually	It connects cyber security with people-centric manufacturing	There is no architecture-level solution	A tiered model of cyber security is missing
[4]. Chaudhuri et al. (2025)	The transformation of cyber security	Organizational + digital systems	Empirical	Identifies the drivers of cybersecurity adoption	Weak connection with the systems in the manufacturing plant	No relation to CNC/CMM/IIoT environments
[5]. Rahman & Shafae (2025)	CPS Vulnerability Classification	IIoT, CPS, Industrial systems	Technical taxonomy	Multilevel classification of vulnerability	No experimental validation	Not adapted to metrological systems
[6]. Antipenko et al. (2026)	Function-driven cyber security	Automated manufacturing systems	At the system level	It connects security with industrial functions	Limited real-time integration	Lack of integration of DT and AI
[7]. Khan et al. (2024)	XAI-based intrusion detection	AI, explainable AI	Algorithmic	Increases confidence in AI-based cyber security	Vulnerable to enemy attacks	No integration with industrial processes
[8]. Rojas et al. (2024)	AI-powered IIoT security	AI, Edge/Cloud, IIoT	At the system level	It integrates AI with IIoT architectures	A human-centered perspective is missing	No compliance with Industry 5.0
[9]. Leirimo et al. (2024)	DT in Industry 5.0	DT, simulation	Conceptually	DT as a human-centered decision-making tool	Limited focus on cyber security	DT security not resolved
[10]. Whaiduzzaman et al. (2026)	DT-based security	DT, IIoT, VI	Applied concept	DT used for anomaly detection	Limited industrial validation	There is no focus on the protection of metrological data

A summary analysis of the references from table 1 shows us the following facts: (a) *fragmentation of approaches*, shows that the analyzed works focus on isolated domains: AI security, CPS vulnerabilities, DT and organizational transformation. This means that an integrated cyber security architecture is lacking; (b) *lack of metrology cyber security* for measurement data integrity (CMM, sensors), calibration traceability and cyber security quality assurance). This is a critical gap for smart manufacturing systems; (c) *poor compliance with Industry 5.0*, because the paradigm of Industry 4.0 is applied, which does not take into account safety, explainability and sustainability; (d) *security gap for DT*, because cyber security

is extremely important for them, because they work online. This is particularly important due to: secure synchronization, data manipulation protection and trust validation mechanisms.

The contribution of this work is to integrate an Industry 5.0 cybersecurity framework for smart manufacturing, which combines human-centric security, AI/XAI-based detection, secure AI, and sustainability-aware resilience mechanisms.

3. Basics Models Cyber Security for Smart Manufacturing

CIA triad. This one concept represents basic model security in models cyber security, which defines three basic principles [11]: *confidentiality (C)*, what means protection data from unauthorized access, *integrity (I)*, meaning protection data from unauthorized modifications, and *availability (A)*, provision reliable access systems and data.

In smart manufacturing and Industry 5.0, CIA triad evolves in cyber - physical - human paradigm security, where are data, machines and people are strong each other connected.

The classic structure of the CIA triad is a triangular model, where each component is interdependent, and the removal of one element threatens the entire system, but there are trade-offs (eg, stronger confidentiality may reduce availability). In Industry 5.0, the CIA triad expands into a multi-layered architecture: the human layer (trust, awareness, behavioral safety); digital layer (data, networks, AI systems); a physical layer (machines (CNC, CMM), sensors, robots), and a sustainability layer (energy-aware security, resilience), thus transforming the CIA into a cyber-physical-human security framework.

The functions of the CIA components are: *confidentiality* (ensuring that information is accessible only to authorized entities), and this is ensured through the basic functions: access control (RBAC (Role-Based Access Control), ABAC (Attribute-Based Access Control)), encryption (data at rest and in transit), authentication and identity management, and data classification and privacy protection. The following technologies are applied here: AES (Advanced Encryption Standard), RSA (Rivest-Shamir-Adleman cryptosystem) encryption, multi-factor authentication (MFA (Multi-factor authentication)), and zero trust architecture (ZTA (Zero Trust Architecture)). The context of smart manufacturing is about protecting CAD/CAM models, securing digital twin data, and preventing industrial espionage; *integrity* (ensures that data remains accurate, consistent and unchanged), and the main functions are: hashing (SHA-256 (Secure Hash Algorithm 256-bit), etc.), digital signatures, blockchain-based immutability and version control and traceability of checks. This aspect is supported by the following technologies: cryptographic hash functions, blockchain system and AI-based anomaly detection. In smart manufacturing, this refers to: protecting

metrology data (CMM results), ensuring the accuracy of calibration of these systems and preventing data manipulation in quality control, for example; *availability* (ensuring that systems and data are available when needed), and its core functions are: redundancy (failover systems), backup and disaster recovery, load balancing, and fault tolerance. Supporting technologies are: cloud/edge computing, distributed systems. The context of smart manufacturing refers to the continuous operation of MES/ERP systems, then real-time monitoring via IoT, as well as avoiding downtime in manufacturing lines, figure 2.

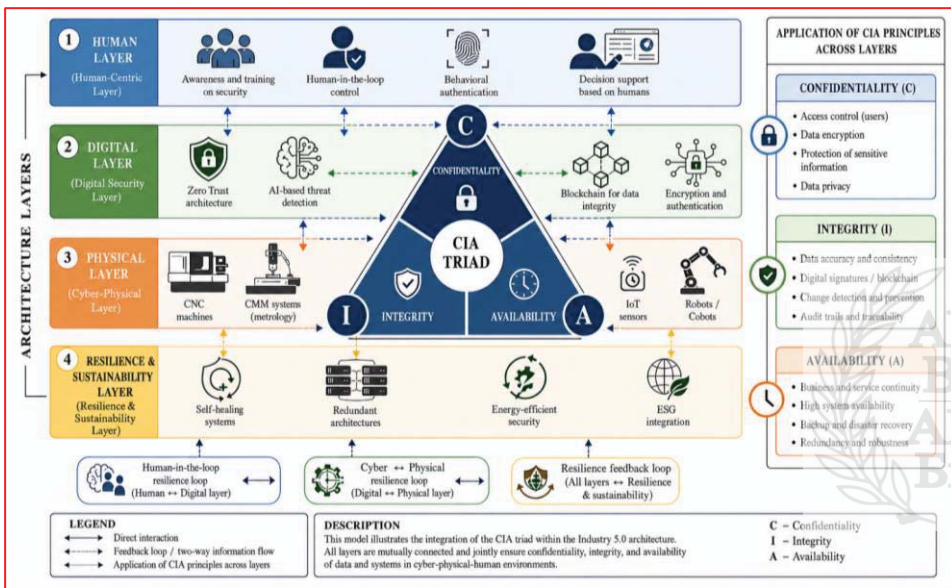


Figure 2. The structure of the CIA triad model in the Industry 5.0 architecture

Zero Trust Architecture (ZTA). Here, an in-depth, structured analysis shows that this model is more aligned with the model of smart manufacturing, Industry 5.0 and smart metrology. It is a modern paradigm of cyber security based on the principle - "never trust, always verify". Unlike the CIA model, ZTA (Zero Trust Architecture) assumes: no implicit trust (inside or outside the network), continuous verification of users, devices and systems, and dynamic, context-aware access control. In Industry 5.0, ZTA becomes necessary due to hyper-connectivity (IoT, CPS), human-machine interaction, and distributed cloud-edge environments. The key components of the ZTA model are [12]: (a) *policy engine (PE)*, which represents a central decision-making unit, which evaluates access requests using identity, device status, and context (location, time, behavior), (b) *policy administrator (PA)*, where decisions are converted into actions, configures access sessions, and communicates with enforcement mechanisms, and (c) *policy*

enforcement point (PEP), where access control decisions are executed, such as: allow / deny / restrict access, and these reside on gateways, applications, and endpoints.

Applications in smart manufacturing (Industry 5.0), refers to: smart cyber physical systems (CPS/CPPS), for secure communication between machines, isolation of vulnerable nodes and detection of threats in real time; smart metrology (CMM, DT and IoT sensors) to authenticate measurement system devices, check the integrity of measurement data and control limited access to calibration records. Here, it is prevented that the data is changed without authorization or false measurement results are generated; DT security, where we have security synchronization (physical ↔ digital), access control for simulation models and protection against model manipulation; AI-driven manufacturing where this model is applied to protect artificial intelligence models and datasets, detect cyber attacks and protect inference nodes in ML models; supply chain security, reflected through secure third-party access, supplier authentication, and data integrity among partners.

Defense in depth (DID). Defense in depth is a cyber security strategy based on multiple coordinated layers of protection. The key idea is simple: if one control fails, other controls continue to protect the system. This model is particularly important in smart manufacturing and Industry 5.0, where manufacturing systems combine people, machines, IoT devices, cloud/edge platforms, digital twins, AI systems and supply chain connections.

According to this model, cyber security represents defense in depth (DID (Defense-in-depth)), so it can be understood as a layered security architecture, with elements as in table 2.

Table 2. Layers of the security architecture of the DID model

Layer	Purpose	Typical controls
Management layer	Policies, roles, risk management	ISO 27001, procedures, audits
Human layer	It reduces the risks associated with people	Training, awareness, discipline of approach
Physical layer	Protects equipment and facilities	Locks, surveillance, access cards
Network layer	Controls communication	Firewalls, segmentation, IDS/IPS
Identity layer	Controls who can access systems	MFA, RBAC, least privileges
Endpoint/Device Layer	Protects machines and devices	EDR, patching, strengthening
Application layer	It provides software and platforms	Secure encryption, authentication, logging
Data layer	Protects the information itself	Encryption, backup, digital signatures
Tracking layer	Attack and anomaly detection	SIEM, SOC, AI-based detection
Recovery layer	Restores normal operation.	Backup, redundancy, disaster recovery

The basic functions of this model are [13]: *prevention*, which refers to the function of reducing the probability of attacks (firewalls, access control, MFA (Multi-Factor Authentication), network segmentation, secure configuration and employee training for cyber security). In smart manufacturing, prevention protects

systems such as CNC machines, CMM systems, robots, MES/ERP, IoT sensors and DT; *detection* is a defense in depth system, which assumes that prevention is never perfect. For these reasons, systems must quickly detect attacks, which includes: intrusion detection systems, log monitoring, AI anomaly detection, behavioral analytics, and industrial network monitoring. For example, abnormal communication between the CMM system and the MES may indicate manipulation of quality data; *response*, where after detection, the system must react, with a response that includes: isolating infected devices, blocking suspicious users, disabling further operation of compromised accounts, activating procedures for responding to incidents and switching to safe operation mode. In manufacturing, the response must be carefully thought out because a system shutdown can affect the safety and availability of manufacturing; *recovery*, returns the system to normal operation after an incident. Recovery mechanisms include: backup and restore, disaster recovery plans, redundant servers, failover systems, and clean system images. For Industry 5.0, recovery is related to resilience, meaning that a system can resume or quickly return to acceptable performance after an interruption; *continuous improvement*, where defense in depth is not static, as it requires: regular risk assessment, analysis of lessons learned after incidents, updating security policies, updating system patches (patches), improving employee awareness and updating and testing backup and recovery procedures.

4. Cyber Security Challenges for Smart Manufacturing

4.1 Technological Challenges

IoT vulnerabilities (low power devices, weak encryption). IoT devices are the "Achilles' heel" of smart manufacturing, not because they are poorly designed, but because they are resource-constrained and often neglected in the security architecture. Industrial IoT (IIoT) devices: have limited processing power, use minimal memory, operate in a real-time environment, and often have long lifetimes (10–20 years). As a result we have a compromise between performance, energy and safety. What are the sources of cyber security risks in IoT devices? These are [2,14]: *(a) low - power devices as source risks* - limitations hardware (weak CPU (eg micro - controllers - LPC), small RAM / Flash memory and limited battery). As consequences we have : impossibility implementations strong cryptographic algorithms, there are none space for advanced security mechanisms and rarely update firmware ; energy restrictions (encryption spend energy, continuous monitoring increases consumption, authentication is additional overhead, devices often use weaker algorithms or them completely omit); minimal operational systems (bare - metal or RTOS systems, without advanced insulation process and without sandboxing, so one vulnerability causes

compromise the whole devices, *(b) weak encryption* is key problems, because of - outdated algorithms (use is DES, RC 4, as and weaker TLS versions) what causes easy breaking through modern tools ; hardcoded keys (keys built - in in firmware and no rotations keys) because one compromised device, means compromise all; unencrypted communication is do plain - text protocols : MQTT without TLS, HTTP instead of HTTPS. Attacks is they perform as man - in - the - middle (MITM), interception data and modification command ; weak authentication (default) password, none MFA and simple tokens) what brings to easy downloads controls over device.

One review cyber threats for IoT and countermeasures for them for context smart of manufacturing / Industries 5.0, table 3.

Table 3. IoT threats and countermeasures for the Industry 5.0 model

Threat category	Threat description	Cause / vulnerability	Influence (CIA)	Countermeasures (state-of-the-art)
Unauthorized access	Unauthorized access to IoT devices	Default password, weak authentication	C, I, A	MFA, Zero Trust, RBAC/ABAC, Device Identity
Interception of Communication (MITM)	An attacker intercepts and modifies the data	Unencrypted communication (MQTT/HTTP)	C, I	TLS 1.3, VPN, end-to-end encryption
Data manipulation	Modification of sensor or metrology data	Lack of validation and integrity	And	Hash functions, digital signatures, blockchain
Botnet attacks (eg Mirai botnet)	IoT devices are used for DDoS	Weak authentication, open ports	A	Network segmentation, IDS/IPS, patch management
Firmware attacks	Injection of malicious firmware	Unprotected update mechanism	I, A	Secure boot, signed firmware, OTA security
Repeat attack	Resending valid messages	Lack of time tokens	And	One time, timestamp validation, session token
DoS/DDoS attacks	Disabling the system	Lack of network protection	A	Speed limiting, firewall, cloud protection
AI/ML attacks	Manipulation of models or data	Unprotected AI streams	And	XAI, robust ML, input validation
Lateral movement	Spreading attacks through the network	Lack of segmentation	C, I, A	Microsegmentation, ZTA
Industrial espionage	Data and know-how theft	Weak encryption and access control	C	End-to-end encryption, DLP systems
Physical attacks	Direct access to devices	Lack of physical protection	C, I	Tamper protection, secure hardware
Insider Threats	Malicious or employee errors	Lack of control and training	C, I, A	Training, audit logs, least privileges
Supply chain attacks	Compromised vendor or software	Lack of verification	C, I	SBOM, supplier verification
Energy attacks	Device battery drain	Inefficient protocols	A	Energy-aware security, edge network filtering
Outdated systems	Outdated devices without security	Long IoT lifecycle	C, I, A	Protection gate, virtual patching

IoT security is moving from "device-level security" to system-level intelligent protection (AI and ZTA and Edge security).

Cloud/Edge cyber security risk. Cloud and Edge are not "bonus layers", they are the backbone of smart manufacturing. This is precisely why they are critical points of attack. In modern systems, the cloud has centralized analytics, AI, data storage (history), and the edge is characterized by local processing and real-time control. The combination of these two models enables low latency, scalability and intelligent manufacturing, but also introduces a distributed attack surface.

The key vulnerabilities are [3,15]: (a) *security risks in the cloud* are multiple: incorrect configuration (misconfiguration), where we have open storage (eg public buckets), and poorly defined IAM policies; weak identity control (IAM), where too broad privileges, lack of MFA, and compromised API keys occur, so that an attacker gains access to the entire system; unprotected API interfaces, where we have an API without authentication and weak input validation; multi-tenancy risks, where shared infrastructure and potential data leakage between users occur; loss of control over data, where we have data outside the local infrastructure and legal/regulatory issues, (a) *security risks in the edge model* are reflected in the following approaches: physical exposure of devices (edge devices often in the field and easily physically compromised); limited resources (like IoT), where we have weak encryption and minimal protection; irregular updates refers to difficult patch management as well as long life cycle; insecure communication between edge ↔ cloud without encryption, as well as MITM attacks; local compromise, where the attacker gains control over the edge node, and data manipulation is done before sending it to the cloud, (c) *combined cloud-edge risks* refer to the following approaches: trust gap gap in trust, where we have that the edge device is compromised, so the cloud trusts false data. This risk is particularly dangerous for digital twin and AI models; data integrity attacks, where we have an attack on data integrity, as a manipulation on the edge, and for these reasons cloud analytics works on the wrong data (garbage in → garbage out); lateral movement (Lateral Movement), where the attack starts from edge devices and spreads to cloud services and MES / ERP systems, (d) supply chain risk, where we usually have edge devices from different manufacturers as well as third-party cloud services.

The impact on smart manufacturing is realized here through several elements: CNC manufacturing (wrong processing parameters and tool wear), metrology (measurement manipulation and loss of traceability), DT (incorrect simulations and wrong optimizations) and AI systems (wrong predictions and model drift).

Vulnerabilities of AI/ML systems in smart manufacturing. AI/ML systems are a key component in: predictive maintenance (PrM), quality control, DT and process optimization. However, unlike classical systems, they are dependent on data and subject to manipulation, the main categories of attacks are attacks during training (data poisoning), and attacks during inference (adversarial attacks).

Data poisoning is an attack in which an attacker injects malicious or manipulative data into the training set, with the aim of making the model learn wrong patterns. Here the following types of attacks [5,16]: label poisoning (misabeled data, for example a defective product marked as "correct"); feature poisoning (manipulation of input values (eg sensor data), and these are subtle changes that are not easily noticeable); backdoor attacks (inserting a "hidden trigger " into the data, where we have the model behave normally, until the trigger is activated). The attack mechanism is performed in several steps: inserting compromised data into the dataset, training the model on such data, the model internalizes wrong patterns, and wrong decisions are made in real work.

Adversarial attacks involve the manipulation of input data during the operation of the model, with the aim of making the model make a wrong decision. A key feature is that small, often imperceptible changes in the input lead to a large error in the model's output. Here we have the following types of attacks: evasion attacks (alteration of input to avoid detection); gradient-based attacks (using model structure to create attacks); physical attacks (manipulation of real sensor data). Industrial examples refer to: predictive maintenance (failure not detected, the consequence is a breakdown), visual control (defect not recognized), robotics (misinterpretation of the environment).

Advanced protection strategies are implemented through the following approaches: explainable AI (XAI) (understanding model decisions), ZTA for AI (verification of each input and data source), secure DT (comparative validation of real and digital systems), and continuous monitoring of AI systems (detection of anomalies in model operation), and blockchain for data (immutability of training dataset).

Research challenges in this area are: AI security in metrology, DT protection, integration of AI and measurement uncertainty, and tinyML security on edge devices.

In conclusion, we can say that AI/ML vulnerabilities represent a new generation of threats, because attacks no longer target only the system, but the intelligence of the system. AI in Industry 5.0 must be: safe, explainable and under human supervision.

A detailed analysis of AI threats against the Industry 5.0 model is shown in Table 4.

Table 4. AI threats vs industrial impact in Industry 5.0

AI threat	Description of the attack	Industrial system domain /	Industrial influence	Influence (CIA)	Recommended countermeasures
Data poisoning	Training data manipulation	Predictive maintenance, DT	Wrong predictions, system degradation	I, A	Data validation, anomaly detection, reliable dataset
Side attacks	Input manipulation during model operation	Visual inspection, robotics	Undetected defects, wrong actions	And	Training for flanking attacks, robust models, sanitizing entrances
Model extraction	Model theft via API	Cloud AI platforms	Loss of intellectual property	C	Speed limiting, API protection, encryption
Inversion model	Reconstruction of training data	Healthcare, metrology	Leakage of sensitive data	C	Differential privacy, access control
Backdoor attacks	Hidden trigger in the model	Autonomous systems, QC AI	Unexpected system behavior	And	Model inspection, training verification
Data “lakes”	Data leakage through the AI pipeline	Cloud/Edge systems	Industrial espionage	C	Encryption, DLP, secure lines
Model drift	Changing data distribution	Predictive systems	Reduced model accuracy	I, A	Continuous retraining, monitoring
Lines of compromise	Attack on ML pipeline	CI/CD systems	Complete compromise of the model	C, I, A	ZTA, secure MLOps
AI hallucination	Generating incorrect results	Generative AI in industry	Wrong decisions	And	Output validation, man in the loop
Insider AI abuse	Abuse by employees	All systems	Model/data manipulation	C, I	Audit logs, least privileges
Edge AI attacks	Edge model compromise	IoT / CNC systems	Wrong local decisions	I, A	Secure edge, TPM, monitoring
Resource depletion	AI system overload	Cloud AI	Service unavailability	A	Scalability, load balancing

The summary link of the analysis in the previous table with the CIA triad is: confidentiality (C), protection of models and data, integrity (I), accuracy of AI decisions and availability (A), continuity of AI services.

Security of digital twins (DT). A digital twin is a digital replica of a physical system (eg CNC, CMM, manufacturing line), which is updated in real time. Key features are continuous data exchange, AI integration and simulation of physical ↔ digital links. It is this connectivity that creates new security vulnerabilities.

The main categories of vulnerabilities are [10,17]: data integrity gap, where the problem is that DT depends on real data (IoT, CMM, sensors). Risk is the manipulation of data, which causes errors in the simulation. For example, wrong metrology data, DT gives incorrect optimization; insecure communication (communication gap), where the problem is weak communication protection: IoT ↔ Edge, Edge ↔ Cloud and Cloud ↔ Digital Twin. The risk is MITM attacks and data interception and modification; AI/ML vulnerabilities in DT come from the risk of data poisoning, adversarial attacks, as well as the occurrence of model

drift, so DT makes wrong decisions; access control (Access Control Gap), problems arise due to inadequate authentication and authorization, it arises due to the risk of unauthorized access to the model and manipulation of simulations; synchronization gap, as a risk arises as a difference between the real and digital system, because DT no longer represents the real state, as well as wrong decisions in manufacturing; cloud/edge dependency, and problems arise because DT is often hosted in the cloud, so the risk arises because we have cloud compromise and loss of control over data; IoT vulnerabilities are caused by weak IoT devices (low-power, weak encryption), so the risk is an entry point for an attack on DT; the human factor is a risk due to insufficient training and insider threats, and the risk is misconfiguration and misuse of the system.

The Industry 5.0 perspective for this area is reflected in the following facts: human-centric approach (man validates critical decisions), digital aspect (integration of AI + DT + IoT) and sustainability (accurate simulations, we have less scrap).

4.2 Operational Challenges

In the field of cyber security, smart manufacturing, operational challenges can occur as: integration of old systems, limitations of real-time systems and interoperability of complex systems [6,18].

Legacy systems in this context refer to old CNC machines, outdated SCADA/PLC systems and supporting software without modern interfaces. The problem here is how to integrate old systems into modern digital architectures (IoT, Cloud, DT, AI)? The key challenges are technological incompatibility, such as the lack of standard protocols (eg OPC UA) and outdated interfaces. Their lack of security is without encryption and authentication with a high risk of attack. As far as the data is concerned, access to them is limited, because the data is not digitized and does not have the possibility of real-time analysis. There are high modernization costs, equipment replacement is expensive and the risk of manufacturing interruption is high.

Real-time system limitations appear as a real problem, because smart manufacturing requires reactions in milliseconds, system synchronization and continuous operation without downtime. The problem is how to ensure high security and complex data processing without impairing real-time performance? The key challenges are latency, where there is a delay in cloud processing, and AI models require additional processing time. All of the above is critical for CNC control and robotics. System determinism refers to the fact that industrial systems require predictable behavior and AI systems are probabilistic. The security overhead is characterized by the fact that encryption and authentication increase latency, and ZTA models require continuous verification. Resource limitations (Edge/IoT) are small CPU and memory, as well as the inability to perform

complex analysis. The industrial impact is reflected in the fact that there is a delay in decision-making, process instability and the risk of breakdowns. The solutions are edge computing (local processing), real-time operating systems (RTOS), hardware acceleration (GPU/FPGA), and AI model optimization (TinyML).

Complex system interoperability is the ability of systems to exchange data, understand data and function together, and in Industry 5.0 systems include: CNC, CMM (metrology), IoT, MES/ERP, DT and AI systems. The key challenges are different standards and protocols, such as OPC UA, MQTT, Modbus and REST API. The disadvantage here is that there is no universal standard. A semantic mismatch is a different meaning of data, e.g. "temperature" can have different contexts (cause of appearance). The heterogeneity of the system is reflected in the increased complex interoperability of the system through IT + OT integration, as well as cloud + edge + on-premise. Scalability here refers to a large number of devices, as well as the fact that we have a complex system orchestration. The industrial impact, from a negative aspect, is ineffective communication, data loss and the impossibility of process optimization. Solutions to these problems are standardization (OPC UA, RAMI 4.0), application of ontology and semantic models, digital thread and DT integration and API-first architecture, figure 3.

4.3 Human-Centered Challenges

An in-depth analysis of the human-centric challenges in smart manufacturing cyber security shows that it depends on insider threats, lack of cyber security awareness and human-machine interaction risk. Insider threats represent security risks originating from employed engineers, operators and external collaborators (suppliers, maintenance). They can be intentional (malicious) as well as unintentional (mistakes, carelessness). Types of insider threats are [9,19]: malicious insider (theft of industrial data, system sabotage and manipulation of AI/DT models), careless insider (weak passwords, clicking on phishing links and wrong system configuration), compromised insider (account taken over by attackers and attacks through stolen credentials).

The industrial impact of this concept is reflected in the manipulation of metrological data (CMM), wrong decisions in MES/QMS, sabotage of CNC processes and industrial espionage. The connection with the CIA triad is reflected in the following relationships: confidentiality (data theft), integrity (data manipulation), and availability (system sabotage). Protection measures are least privilege access, user monitoring (UEBA), audit logs, ZTA and system segmentation.

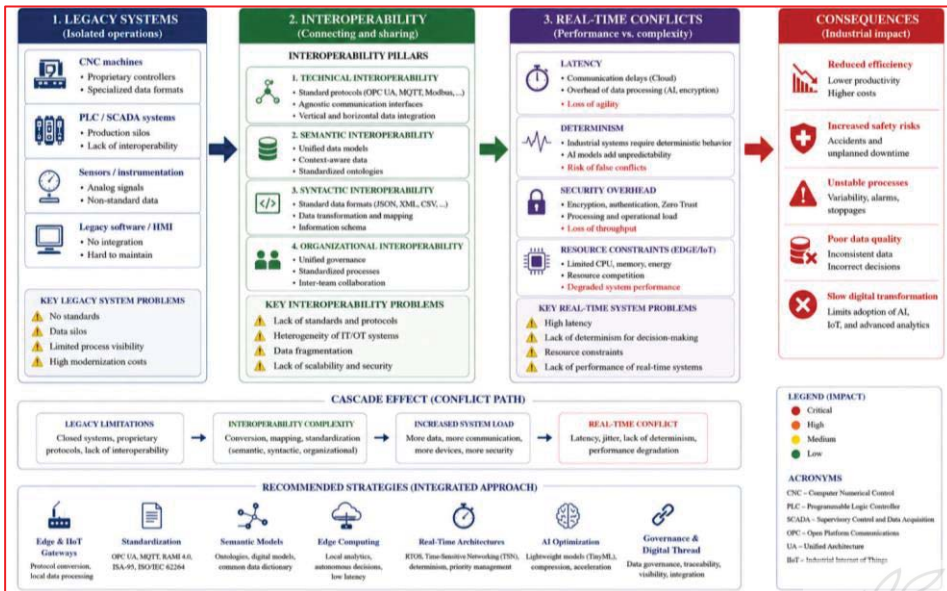


Figure 3. Cyber security operational challenges in smart manufacturing

In Industry 5.0, man and machine work together, such as cobots, AI systems, DT and autonomous systems, and the problem is that man-machine interaction introduces new safety and operational risks. Key challenges are excessive trust in AI (operators blindly trust AI decisions, lack of verification), lack of transparency (AI models are a "black box", and sometimes it is difficult to understand decisions), misinterpretation of data (operator misunderstands AI output, or wrong actions in manufacturing), safety of cobots (physical interaction with people, risk of injury), cognitive load (too much information for the operator, as well as stress and errors). Industrial impact manifests itself through wrong manufacturing decisions, reduced product quality, safety incidents and employee health risks. Safeguards include explainable AI (XAI), human-in-the-loop systems, interface ergonomics, operator training, and safety protocols for robots.

4.4. Supply Chain Vulnerabilities

This important aspect of digital business relates to third-party risks, as well as risks related to software/hardware dependencies.

Risks arising from third parties depend on equipment suppliers (CNC, CMM, IoT), software vendors (MES, ERP, AI platforms), cloud service providers and maintenance companies [5,20]. Therefore, the problem of system security also depends on the security of external partners. Key vulnerabilities occur in the

following cases: insufficient supplier security (suppliers do not have adequate controls or weak protection of development processes); remote access of the provider (VPN access for maintenance or insufficient control of activities); unverified updates (the software update may contain malicious code), for example, a compromised firmware update; lack of transparency (because there is no insight into how the software is developed, or which components are used). Typical attacks are supply chain malware (compromised update), insider attacks at suppliers and cloud provider compromise. Industrial influence in this area is reflected in the compromise of CNC and CMM systems, manipulation of metrological data, interruption of manufacturing and industrial espionage.

Software and hardware dependencies are based on the fact that industrial systems depend on a complex ecosystem, which includes operating systems, libraries and frameworks, firmware, industrial controllers, and IoT devices. The problem with this concept is that one vulnerability in a dependency can compromise the entire system. Key vulnerabilities are: outdated components (unupdated software or unsupported versions), open-source dependencies (use of libraries without verification, unknown vulnerabilities), firmware vulnerabilities (hardcoded keys, unprotected boot process), hardware backdoors (malicious or compromised chips, supply chain compromise at the hardware level), lack of patch management (delay in applying security patches). The industrial impact of these aspects relate to the compromise of AI and DT systems, wrong manufacturing decisions, loss of data integrity and system downtime. Here we also have the effect of the so-called cascade effect, which is realized in the chain: a vulnerable library, a compromised AI model, a wrong decision and the physical system reacts. Protection measures are SBOM and dependency analysis, regular patch management, secure development lifecycle (SDLC), firmware verification and trusted hardware (TPM, secure element).

4.5 Risks Related to Metrology and Quality Data

Cyber security risks in this area arise from the manipulation of measurement data, as well as threats to the integrity of calibration data. Manipulation of measurement data implies unauthorized modification or falsification of measurement results from CMM systems, sensors (IoT), laboratory tests and online control systems. The aim of the attack is to present a defective product as a correct one, as well as to conceal errors in the process. The attack vector is complex and can include [5,21]: compromise of IoT sensors (false input data and signal manipulation), MITM attacks (communication), which are reflected in the interception and modification of data between CMM ↔ MES, as well as Edge ↔ Cloud, attacks on databases (modification of historical measurement data and deletion of records), and insider manipulation (manual modification of results, as well as "correction " of data to meet standards). The industrial impact is reflected in the

possibility of unwanted events: faulty quality control, passing of defective products, rejection of correct products and loss of traceability. In smart metrology, data integrity represents confidence in product quality. Protection measures are: digital signatures of measurement data, hash validation, blockchain for traceability, redundant measurements and AI anomaly detection.

5. Cyber Security Solutions in Industry 5.0

5.1 Human-Centered Security Layer

In Industry 5.0, security is no longer exclusively a technological problem, but an interaction between man, systems and intelligent technologies. The human-centered security layer includes [9,21]: education and awareness, active human participation in security and advanced identification methods (behavioral).

Cyber security *training and awareness includes continuous education of employees about cyber threats, security practices and rules of conduct in the digital environment*. Key challenges are: low level of awareness (operators focused on manufacturing, security as a secondary priority), rapid evolution of threats (ransomware, phishing, AI attacks, as well as outdated training programs), different levels of knowledge (engineers vs operators, IT vs OT staff). The industrial impact of cyber security is reflected in the reduction of insider risk, faster response to incidents and reduction in the success of phishing attacks. The advanced approach, related to the application in the Industry 5.0 model, is: adaptive training (AI-driven learning), attack simulations (phishing, SOC scenarios), gamification (cyber drills) and personalized training programs. We can state that the greatest security risk is not technology, but an insufficiently educated user.

Human-in-the-loop monitoring is a model in which a human actively participates in monitoring and making security decisions together with AI systems. The role in Industry 5.0 is a combination of AI detection and human assessment. The goal is to increase accuracy and reduce false alarms. The key challenges are: over-reliance on AI (blind trust in automation), cognitive load (too many warnings and "vigilance fatigue", reaction delay (human decision can slow down the process). The industrial impact is related to more accurate detection of attacks, better decision-making and greater system resilience. Advanced approaches in this area are explainable AI (XAI), alarm prioritization, support for decision-making systems and interactive dashboards. The most secure system is not fully automated, but human synergy and AI.

Behavioral biometrics uses patterns of behavior to identify users: typing style (keystroke dynamics), mouse movements, operating patterns on the system, and how to interact with machines. The advantages of this approach are continuous authentication, difficult forgery and this is a non-invasive method, figure 4. The

key challenges are: privacy (monitoring of user behavior), variability of behavior, due to changes (stress, fatigue and experience), and false positive/negative results (false detection). The industrial impact is reflected in the detection of insider threats, prevention of unauthorized access and protection of critical systems (CNC, CMM, MES). Advanced approaches are AI models for behavior, integration with ZTA, multi-factor authentication and edge behavior analysis. The most important message is that user identity is no longer just "who you are", but how you behave.

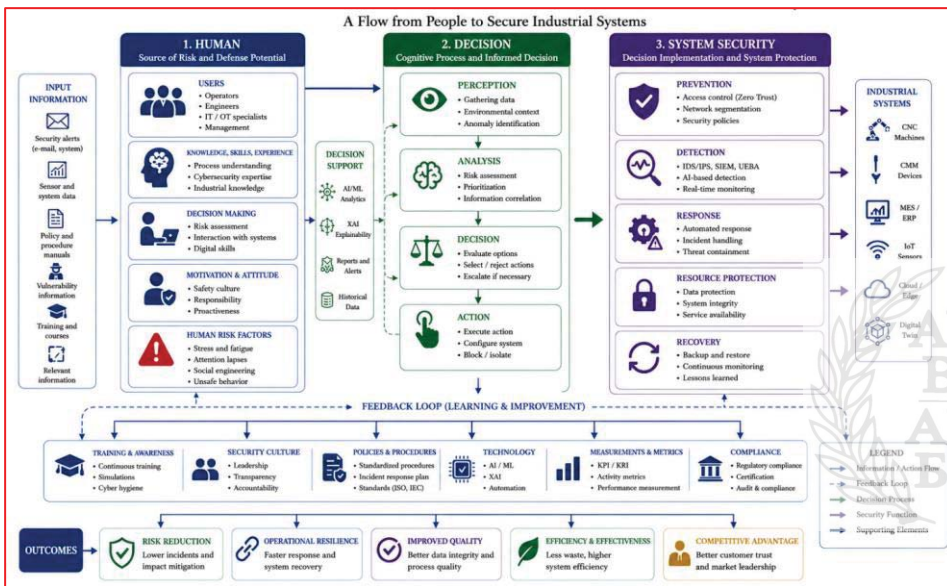


Figure 4. General cyber security model for Industry 5.0

5.2 Digital Security Layer

The digital security layer within Industry 5.0 is based on security solutions, with a focus on: Zero Trust Architecture (ZTA), AI-driven threat detection, application of blockchain for data integrity and security DT [9,22].

Secure DT is a secure version of the digital twin that protects data, validates models, and ensures the integrity of simulations. The key risks of the DT system are data manipulation, non-synchronization between real and digital, AI vulnerabilities and cloud dependence, and security mechanisms are data validation, communication encryption, ZTA and AI monitoring. Industrial application refers to CNC optimization, predictive maintenance and metrology (virtual metrology). The benefits relate to reliable simulations, process optimization and risk reduction. The integrated model of digital security

represents the interaction of technologies: ZTA → access control, AI → threat detection, blockchain → data integrity and DT → intelligent management. Together, these elements form a holistic security system of Industry 5.0. The connection to the CIA triad is reflected in the following facts: ZTA (confidentiality + integrity), AI detection (integrity + availability), blockchain (integrity), and DT (integrity + availability). The Industry 5.0 perspective refers to their basic elements: human at the center (transparent and explainable systems (XAI)), digital model (integration of AI, IoT, DT), and sustainability (energy-efficient security and long-term system resilience). Research challenges in this area relate to the integration of ZTA in industrial systems, AI security and XAI, scalable blockchain systems and DT security in metrology. Here is a structured table of technology vs features vs benefits vs challenges within the Industry 5.0 digital security layer, Table 5.

Table 5. Technologies vs functions vs advantages vs challenges for Industry 5.0

Technology	Key functions	Advantages	Challenges / limitations
WHAT?	- Continuous authentication and authorization - Network microsegmentation - Identity-based access policies	- Reduced attack surface - Prevention of lateral movement - High level of access control	- Complex implementation - Integration with legacy systems - Latency overhead
AI-guided threat detection	- Anomaly detection - Predictive threat analysis - Behavioral monitoring - Real-time data analysis	- Fast and adaptive detection - Scalability - Security automation	- Data poisoning and adversarial attacks - “Black box” problem - The need for quality data
Blockchain for data integrity	- Immutable ledger - Digital signatures - Data traceability - Smart contracts	- High data integrity - Transparency - Elimination of the need to trust a central entity	- Scalability - Latency - Energy consumption - Complexity of implementation
Safe DT	- Synchronization real ↔ digital - Data validation - Simulation and optimization - Integration of AI models	- Reliable simulations - Process optimization - Predictive maintenance - Reduction of risks and costs	- Depends on the quality of the data - Complex integration - IoT/Cloud security risks - AI vulnerabilities
Explainable AI (XAI)	- Explanation of AI system decisions - Model transparency - Decision-making support	- Increased trust in AI - Better system control - Easier error detection	- Reduced efficiency compared to “black-box” models - Complexity of interpretation
Edge security	- Local data processing and filtering - IoT device protection - Latency reduction	- Real-time security - Reducing dependence on the cloud - Better system resilience	- Limited device resources - Complex management - Physical exposure of the device
Cloud security (CSPM, ZTA Cloud)	- Configuration management - Data encryption - IAM control - Monitoring and auditing	- Scalability - Centralized control - High availability	- Misconfiguration risks - Problem with multiple users - Dependence on the provider

A key research insight can be defined that no technology alone is sufficient, but the greatest value comes from their integration. The conclusion is, ZTA → access control, AI → intelligent detection, blockchain → data trust and DT → optimization and simulation, together form the basis of a secure and intelligent Industry 5.0 system.

5.3 Resilience and Sustainability Layer

In Industry 5.0, cyber security is not only system protection, but the system's ability to survive, adapt and maintain functionality despite attacks and disruptions, and this layer integrates resilience, sustainability, and business continuity, and their basic components are: self-healing systems, redundancy strategies and energy-aware cyber security [8,23].

Self-healing systems are systems that can detect anomalies, automatically react and recover without human intervention, and are inspired by biological systems (digital immune system). The key functions of this model are automatic attack detection (AI/ML), isolation of compromised components, system reconfiguration, and automatic recovery (self-healing). Industrial application refers to CNC systems → automatic stop and correction, IoT networks → isolation of a compromised node, and DT → deviation identification, and the advantages are the minimization of downtime, fast response to incidents and reduction of the need for human intervention. As far as implementation is concerned, there are the following challenges, namely implementation complexity, false positive alarms and dependence on the quality of the AI model. We can conclude that future systems will not only detect attacks, but actively recover.

Redundancy strategies imply the existence of alternative resources that take over the function in case of failure. Today there are several types of redundancies, namely: hardware (double servers and backup CNC controllers), software (automatic redundancy systems, backup applications), and network (alternative communication routes) and data redundancy (backup and data replication). Application in industry is reflected in the following approaches: MES systems with automatic redundancy systems, cloud + edge redundancy, and redundant IoT gateway. The benefits relate to maintaining business continuity, high availability and resistance to failures and attacks. On the other hand, challenges in implementation are related to high costs, complex management and system synchronization. So we can conclude, redundancy is not a luxury, but a basic requirement for system resilience.

Energy-conscious cyber security implies the application of security mechanisms with the optimization of energy consumption. Why is it important? Because Industry 5.0 emphasizes sustainability, on the other hand IoT/Edge devices have limited resources, and AI systems consume significant energy. The key challenges

are related to the application of the following approaches: energy-demanding algorithms (encryption and AI models), security vs. performance compromise (stronger protection → higher consumption), limitations of IoT devices (battery, CPU and memory). Industrial application is realized through the following concepts, namely: edge AI (TinyML), energy-efficient protocols and adaptive encryption. The advantages of these concepts are system sustainability, cost reduction and long-term stability, while the challenges are the compromise between safety and efficiency and the complexity of optimization. An important conclusion is that the future of security is not only "safe", but energetically sustainable.

5.4 Advanced Technologies

Here we have three key technologies: explainable AI (XAI) (increases trust and transparency of AI decisions), quantum-resistant cryptography (protects data from future quantum attacks), and digital immune systems (enable detection, adaptation and self-healing of the system) [7,24,25].

Explainable AI (XAI) stands for AI models and methods that make it possible to understand: why the AI made a certain decision, what data most influenced the result, and whether the decision is reliable, biased or risky. In Industry 5.0, XAI is important, as it connects AI automation with human supervision and trust. XAI functions in cyber security enable: explanation of anomaly detection, transparency of AI decisions in SOC/MES/OT systems, reduction of "black box" problems, better human-AI collaboration, and auditability of decisions for standards and regulation. For example, if the AI system detects unusual network traffic between the IoT gateway and the MES, XAI can indicate that an alarm has been triggered due to unusual access times, increased packet counts, and deviations from the normal device profile. The application in smart manufacturing can be illustrated with examples as in Table 6.

Table 6. Examples of XAI applications in smart manufacturing for cyber security

Area	Application of XAI
AI threat detection	An explanation of why the alarm was triggered
CMM/Metrology	Explanation of suspicious measurement patterns
DT	Understanding the difference between real and digital models
Predictive maintenance	Explanation of failure prediction
Man in the Loop Tracking	Support for the operator in making decisions

What are the advantages of this model? Here they are: increasing trust in AI systems, reducing the risk of wrong decisions, improving operator responsiveness, facilitating audit and compliance, and supporting ethical and responsible AI deployment. There are also possible challenges: explanations can be technically complex, some XAI explanations are not intuitive enough for operators,

transparency can reveal information to an attacker, and sometimes there is a trade-off between model accuracy and explainability. XAI serves not only to explain AI decision, but to enable trust, accountability and control in cyber-physical-human systems.

Quantum-resistant cryptography or post-quantum cryptography includes cryptographic methods that should remain secure in the era of quantum computing. The problem is that quantum computers could threaten classical algorithms such as RSA and ECC, which are widely used today for: digital signatures, certificates, VPN communication, secure industrial protocols and protection of cloud/edge communication. Why is this model important for Industry 5.0? Industrial systems have a long life cycle. CNC, CMM, PLC, SCADA and IoT systems often remain in service for 10–20 years. This means that data protected today may be compromised in the future. A particularly important risk is that “collect now, decrypt later”, the attacker today collects encrypted data and stores it until he has the quantum ability to decrypt. The functions of post-quantum protection are the protection of long-term industrial data, secure digital signatures, secure communication between OT, edge and cloud systems, protection of DT models, and protection of calibration and metrology data, and the application in smart manufacturing is analyzed in Table 7.

Table 7. Examples of XAI applications in quantum-resistant cryptography for cyber security

Area	Application of quantum resistant cryptography
Cloud/Edge communication	Post-quantum TLS / hybrid protocols
DT	Model and simulation data protection
Metrology	Long-term protection of calibration certificates
Supply chains	Secure digital signatures of components
IoT/OT systems	Gradual migration of cryptography

The advantages of this model are long-term confidentiality protection, resistance to future quantum attacks, greater trust in digital certificates, and protection of critical infrastructure. The challenges are larger keys and larger signature data, more demanding implementation on IoT/Edge devices, the need to migrate existing systems, compatibility with legacy equipment and performance in real-time systems. Post-quantum cryptography is not only a future need, but a strategic requirement for systems that must remain reliable for decades.

The digital immune system is a cyber security concept inspired by biological immunity. The system continuously monitors its own state, detects abnormalities, reacts to threats and learns from incidents. In Industry 5.0, this means that the manufacturing system is not only protected, but can also self-monitor, self-adapt, self-isolate and self-heal. Key functions are: anomaly detection (recognition of deviations in the network, devices and data), self-isolation (separation of the compromised component), self-recovery (returning the system to a stable state),

adaptive learning (learning from previous incidents), and continuous monitoring (continuous monitoring of the system).

Application in smart manufacturing of this model refers to: IoT networks, CNC machines, CMM systems, DT platforms, MES/ERP systems, cloud/edge infrastructure. For example, if an IoT sensor sends abnormal data, the digital immune system can automatically isolate the sensor, activate a redundant sensor, notify the operator, and update the detection rules. The advantages of this model are faster response to attacks, reduced downtime, greater manufacturing resilience, reduction of the need for manual intervention, and continuous improvement of security, while the challenges are complex integration with OT systems, the risk of wrong automatic response, the need for quality data, the dependence on AI models and the need for clear rules when a human must take control. The digital immune system represents a transition from passive protection to active, adaptive and self-healing security.

An integrated model of advanced technologies can be one, perhaps, the best solution, where a part of the solution is taken from each of them, namely: XAI (explains and validates AI decisions), post-quantum cryptography (protects data and communication in the long term), digital immune system (detects, reacts and recovers the system), creating an integrated flow of threat detection → XAI explanation → human/AI decision → isolation → recovery → learning → long-term data protection.

6. Proposed Conceptual Framework of Industry 5.0 Cyber Security Architecture

Cyber-security architecture for Industry 5.0 should be a multi-layered, adaptive and human-oriented system, in which security is not only seen as technical protection, but as an integration of people, digital technologies, physical manufacturing, risk management and sustainability. The basic logic of Industry 5.0 is based on three principles: man at the center, sustainability and resilience, i.e. on a human-oriented, sustainable and resilient industry. The layers of cyber-security architectures for Industry 5.0 are [2,3,5,26]: human layer, cyber security and intelligence layer, digital infrastructure layer, physical manufacturing layer and sustainability layer.

Human layer represents the most important element of the Industry 5.0 cybersecurity architecture, as operators, engineers, quality managers, metrology and IT/OT teams become active participants in system protection. In traditional Industry 4.0, humans were often viewed as a potential weakness, while in Industry 5.0, humans are treated as a critical carrier of security intelligence. In this layer, key functions are: employee training, cyber-awareness development, human-in-the-loop monitoring, ethical use of AI systems, and decision-making in crisis situations. When an AI safety mechanism detects an anomaly, humans as experts

must not be completely excluded from the process, especially if the decision involves stopping the manufacturing line, discarding measurement results or isolating the CMM/CNC system. The role of key components is: AI security engine analyzes user behavior, unauthorized access attempts, unusual commands and operator work patterns. The risk management system assesses human risks: insufficient training, operator errors, phishing attacks, insider threats and unclear responsibilities. The blockchain layer can record who approved the parameter change, who validated the measurement, and who accessed critical data. The DT security module enables the simulation of human decisions in crisis scenarios without endangering the real system, figure 5.

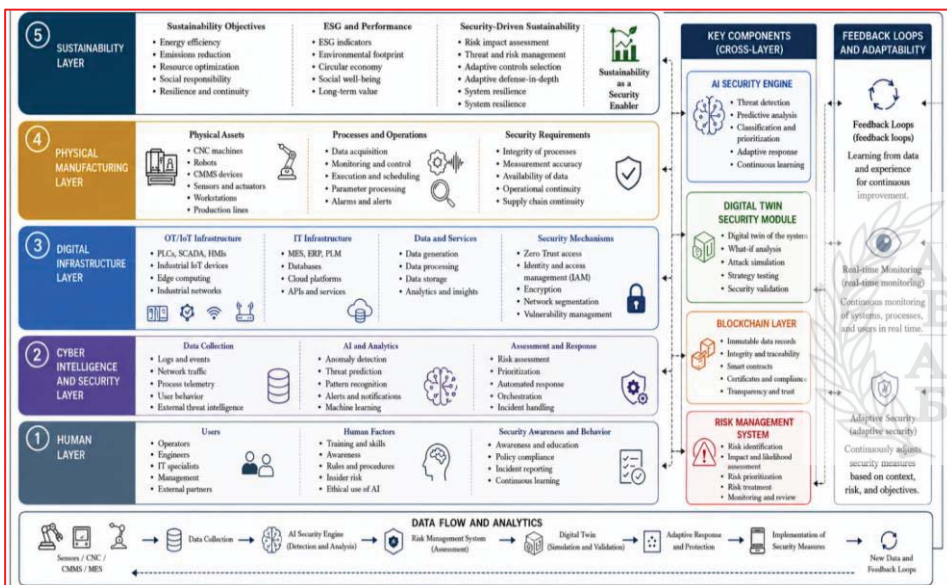


Figure 5. Proposed conceptual model of Industry 5.0 cyber security architecture with adaptive security layers, components and mechanisms

The cybersecurity intelligence layer is the central analytical layer of the architecture. Its role is to collect data from all parts of the system, analyze it using AI/ML algorithms and propose or automatically activate security measures. This layer connects real-time monitoring, predictive analytics, anomaly detection, risk assessment and adaptive response. This layer emphasizes the management of cyber risk through functions such as manage, identify, protect, detect, respond and recover, which is very suitable for mapping to this layer. The main functions of this layer are real-time threat detection, risk classification, attack prediction, correlation of events from IT and OT systems, automatic initiation of response and continuous learning from previous incidents. The role of key components is:

The AI security mechanism is the core of this layer. It detects anomalies in network traffic, manufacturing parameters, measurement results, logs and user behavior. The risk management system translates technical anomalies into business and manufacturing risks: downtime, loss of quality, compromise of measurement data or a security incident. The DT security module allows the response to be tested before it is applied to the physical system. The blockchain layer provides an immutable record of incidents, decisions and security events.

The digital infrastructure layer includes IoT/IIoT devices, edge and cloud platforms, MES, ERP, SCADA, DT, databases, API communication and industrial networks. This is the layer with the largest attack surface, because it connects the manufacturing system with analytics, suppliers, users and external services. In smart manufacturing, this layer must support ZTA, encryption, network segmentation, identity control, secure data exchange, and continuous access verification. The role of the key components is: AI security mechanism monitors network traffic, API requests, access to cloud platforms and anomalies in communication. The risk management system ranks risks according to the impact on manufacturing, quality, human safety and business continuity. The security module DT checks whether the digital model behaves in accordance with the physical system. If there is a difference between the actual CNC/CMM process and its DT, the system can flag a potential manipulation. The blockchain layer ensures data integrity, especially for measurements, calibration, certificates, parameter changes and quality records.

Physical layer includes CNC machines, robots, CMM devices, sensors, manufacturing lines, tools, measuring stations and control systems. In this layer, a cyber attack can have direct physical consequences: incorrect processing, tool damage, false measurement results, line downtime or endangering worker safety. The aspect of smart metrology is particularly important. If an attacker changes the measurement results or calibration data, the product may be mistakenly accepted as compliant. Therefore, cyber-security in Industry 5.0 must protect not only IT data, but also the physical-technical reliability of manufacturing and measurement. The role of the key components is: The AI safety mechanism detects unusual vibrations, changes in processing, deviations in measurements, unusual commands and parameter changes. The risk management system decides whether the risk is local, systemic or critical. The safety module DT compares the actual behavior of the machine with the expected model.

The sustainability layer extends cyber security beyond traditional data protection. In Industry 5.0, the security decision must take into account energy, resources, emissions, waste, manufacturing continuity and social responsibility. This means that cyber-security must be linked to ESG indicators, energy efficiency and system resilience. For example, an adaptive security system does not always need to shut down the entire manufacturing line. In some cases, it is better to isolate only the compromised segment, shift the load to a spare cell, or activate an energy-

efficient operating mode. This approach is in line with the Industry 5.0 idea that industry should contribute to sustainability, resilience and people's well-being.

8. Future Directions of Development

The evolution of Industry 5.0 in the field of cyber-security marks the transition from technologically oriented systems to human-centric, ethical and reliable security ecosystems. In this context, the focus is no longer only on the protection of infrastructure and data, but on the synergy between humans and AI, as well as on the development of ethical and transparent AI systems that inspire trust [2,27-30]. Key elements of this evolution are: human-AI collaboration in cyber-security and ethical and reliable AI systems.

The cooperation of man and artificial intelligence in cyber-security is the basis of Industry 5.0, where AI systems do not replace humans, but act as intelligent assistants that support decision-making. In cyber security, this means: AI detects threats and suggests actions, a human analyzes the context and makes the final decision, and the system continuously learns from the interaction. This approach is called the "man-in-the-loop" or more advanced "man-in-the-loop" model. Key features are: increased situational awareness - AI analyzes large amounts of data and provides an overview of the situation in real time, decision support - AI suggests multiple response scenarios (eg system isolation, parameter change, alert), adaptive collaboration - the system adapts to the operator's level of expertise, and learning from interaction - every human decision improves AI models. An application in smart manufacturing is an operator validates AI anomaly detection in CNC/CMM systems, a quality engineer decides if a deviation is critical, an IT/OT expert manages cyber incidents with AI support, and management uses AI reports for strategic decisions. The advantages of this concept are the reduction of decision errors, faster response to threats, better interpretation of complex data, and increased trust in AI systems. Challenges are too much dependence on AI, insufficient user training, conflict between AI recommendations and human intuition, and the need to clearly define responsibilities.

Ethical and reliable AI systems tell us that in Industry 5.0, AI systems must be not only efficient, but also transparent, fair, accountable, secure, and compliant with regulations and standards. This leads to the concept of "trusted AI", where users have confidence in the system's decisions. The key principles are: transparency (explainable - XAI), where AI must explain: why the threat was detected, based on which data, and what are the key decision factors. For example, the system explains that an anomaly was detected due to a combination of vibration, temperature and measurement deviation; fairness - AI systems must not be biased: towards users, towards processes, and towards types of data; accountability, where it must be clear who is responsible for AI System decisions, how decisions are

audited, and how errors are resolved; security, where AI systems must be protected from external attacks, data manipulation and model poisoning attacks; privacy, where it is particularly important in the context of human data, operator behavior and industrial secrets. Application in Industry 5.0 systems is XAI for explaining decisions in quality control, transparent models for predictive maintenance, audit trail (eg blockchain) for AI system decisions, and ethical framework for decision-making in crisis situations. The advantages are increased user confidence, easier implementation in the industry, regulatory compliance (EU AI Act, ISO standards), and reduced risk of wrong decisions. The challenges are the compromise between performance and transparency, the complexity of implementing the XAI model, the lack of standards in practice, and the need for a multidisciplinary approach. Within Industry 5.0, these two concepts are interconnected: AI detects the threat, the system explains the decision (XAI), the human makes the decision, the system records and learns from the decision, and reacts faster and more precisely the next time. This cycle forms a smart, adaptive and ethical security loop.

10. Conclusions

Cyber security is a key enabler of Industry 5.0, so this fact defines the paradigm for an integrated approach to human-digital sustainability, as human-digital technologies-sustainability. This integrated approach represents the essence of the evolution towards Industry 5.0 and the key framework for the development of modern smart manufacturing systems. This approach goes beyond traditional models focused solely on automation and efficiency, introducing a balance between technological progress, human role and sustainable development. The synergy of three key dimensions is realized through: *human dimension* - man becomes the central element of the system, not only as an operator, but as a decision maker, creator of innovations, supervisor of AI systems, and guarantor of ethics and responsibility. The role of man is being transformed from passive to active and strategic, with the support of intelligent systems. *Digital dimension*, where we have that digital technologies represent the driving force of the system, as they include AI and machine learning, DT, IoT and edge/cloud computing, as well as cyber-security platforms. These technologies enable real-time data analysis, predictive decision-making, and adaptive process management. The third dimension is *sustainability*, which introduces a new value dimension: optimization of resources and energy, reduction of waste and emissions, social responsibility, and ESG integration. Cyber-security in this context contributes to preserving the stability and long-term sustainability of the system. Integration through a closed loop is a key feature of this approach, as a closed, adaptive and intelligent loop, through the activities of data collection (IoT, CNC, CMM, systems), digital analysis (AI, DT), risk and performance assessment, human

validation of decisions, implementation of actions, and continuous learning and improvement. This cycle enables quick reaction to changes, continuous optimization of the system, and increase in reliability and quality. The key advantages of an integrated approach are: increased trust (transparent decisions, data traceability, and system security), greater resilience (adaptive reaction to disturbances, predictive risk management, and self-healing systems), sustainable development (reduction of energy consumption, process optimization, and waste reduction), quality improvement (integrated metrology, real-time control, and reliable measurement data). Better decision-making through AI support, human validation, and contextual understanding. Although it brings numerous advantages, this approach faces challenges, such as the complexity of system integration, the need for new skills and education, high initial costs, cyber-security risks, and lack of standardization. The strategic importance for Industry 5.0 is also reflected in the fact that the integrated approach is the foundation for the development of intelligent factories, digital transformation of the industry, improvement of global supply chains, implementation of ESG principles, and connection of technology and society. The integration of people, digital technologies and sustainability is not only a technological trend, but a new industrial philosophy.

Safe, resilient and human-oriented manufacturing systems define the future of Industry 5.0, because it represents the essence of the transformation of modern industry towards the Industry 5.0 model. She indicates that future manufacturing systems will not be defined only by the degree of automation or digitization, but above all by the quality of their safety, resilience and ability to integrate humans as the central element of the system.

Safety is becoming the foundation of every modern factory. In Industry 5.0, security is no longer just the protection of IT infrastructure, but includes: protection of manufacturing processes (CNC, robotics, CMM systems), integrity of measurement and metrology data, protection of DT and AI models, secure communication in the IoT and cloud environment, and protection against cyber attacks (ransomware, data manipulation, insider threats). Secure systems enable confidence in data and decisions, manufacturing stability, intellectual property protection, and business continuity. In this sense, cyber security becomes a strategic enabler, not just a technical function.

Resilient systems have the ability to recognize disruptions, react to them in real time, continue operations with minimal losses, and learn from previous incidents. Resilience is not only a technical characteristic, but also the organizational and strategic ability of the system to adapt to changes and crises, figure 6. One of the most important differences between Industry 4.0 and 5.0 is the return of man to the center of the system. In Industry 5.0, humans make key decisions, AI serves as a support, not a replacement, human-machine cooperation is developed, and the emphasis is on safety, well-being and employee education. Applying this

paradigm means that companies must invest in cyber security, develop AI and digital technologies, train employees, implement ESG principles, and establish integrated management systems.

Industry 5.0 is not just a technological revolution, but a transformation towards intelligent, safe, resilient and human-friendly systems that create long-term value for society and the economy.

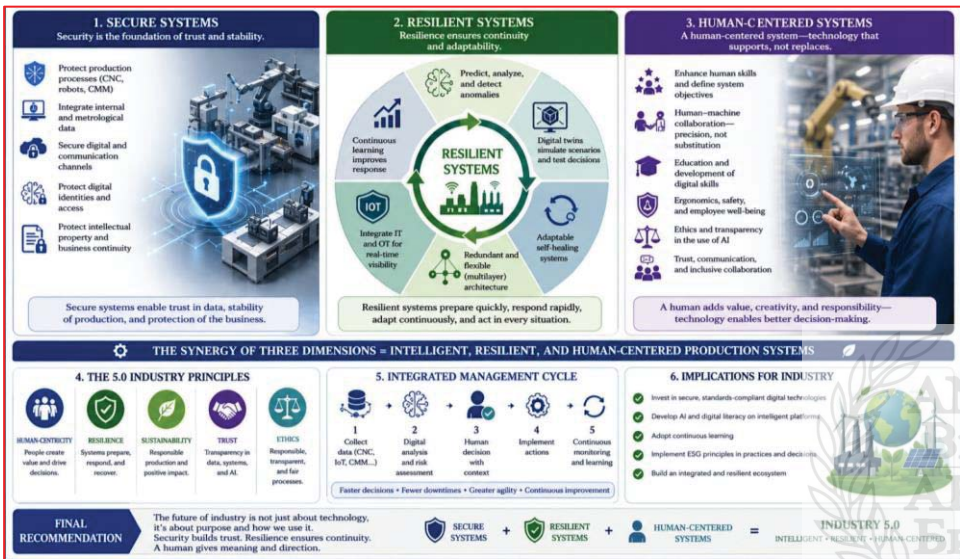


Figure 6. Safe, resilient and human-centered manufacturing systems define the future of Industry 5.0

References

- [1] European Commission: Directorate-General for Research and Innovation, Industry 5.0 – Towards a sustainable, human-centric and resilient European industry, Publications Office of the European Union, 2021, <https://data.europa.eu/doi/10.2777/308407>.
- [2] Kour R, et al. (2024) Cybersecurity for Industry 5.0: trends and gaps. *Front. Comput. Sci.* 6:1434436. 10.3389/fcomp.2024.1434436.
- [3] B. Santos, et al, Cybersecurity in Industry 5.0: Open challenges and future directions, *Proc. PST*, 2024. <https://doi.org/10.1109/PST62714.2024.10788065>.
- [4] A. Chaudhuri, et al, Factors impacting cybersecurity transformation: An Industry 5.0 perspective, *Computers & Security*, 2025. <https://doi.org/10.1016/j.cose.2024.104267>.

- [5] MH Rahman, T. Wuest, M. Shafae, Manufacturing cybersecurity threat attributes and countermeasures: Review, meta-taxonomy, and use cases of cyberattack taxonomies, *Journal of Manufacturing Systems*, Volume 68, 2023, Pages 196-208, ISSN 0278-6125, <https://doi.org/10.1016/j.jmsy.2023.03.009> .
- [6] Antipenko V., Matulevičius R. (2026). Function-driven cyber-physical security in smart manufacturing. *Software and Systems Modeling*. <https://doi.org/10.1007/s10270-026-01370-0> .
- [7] N. Khan et al., Explainable AI-based intrusion detection system for Industry 5.0, *arXiv*, 2024. 10.48550/arXiv.2408.03335.
- [8] Mario R., et al. (2024) Lack of verified Inclusive Technology for Workers with disabilities in industry 4.0: a systematic review, *International Journal of Sustainable Engineering*, 17:1, 190-210, 10.1080/19397038.2024.2328711.
- [9] Leirimo, T. (2024). Digital Twins for Industry 5.0: Unlocking the Human Potential. *Proceeding CIRP*. 130. 761–766. 10.1016/j.procir.2024.10.161.
- [10] Whaiduzzaman, M., et al. (2026). Enhancing IIoT Security Using Digital Twins in Industry 5.0 : A Systematic Literature Review. *Information*, 17 (2), 209. <https://doi.org/10.3390/info17020209> .
- [11] Rose Scott, ET AL. (2020). Zero Trust Architecture. 10.6028/NIST.SP.800-207-draft2.
- [12] K. Stouffer, *Cybersecurity for Manufacturing Systems*, NIST, 2020, <https://www.din.de/resource/blob/271310/5d3b71d57a0bd7a7349cf81b89df9787/07-smart-manufacturing-keith-stouffer-data.pdf> .
- [13] Barrett, M. (2018), *Framework for Improving Critical Infrastructure Cybersecurity Version 1.1*, NIST Cybersecurity Framework, <https://doi.org/10.6028/NIST.CSWP.04162018> .
- [14] Leng, J., et al. (2022). Secure Blockchain middleware for decentralized IIoT towards industry 5.0: a review of architecture, enablers, challenges, and directions. *Mach. Des.* 10:858. doi: 10.3390/machines10100858.
- [15] Sailio M, Latvala OM, Szanto A. Cyber threat actors for the factory of the future. *Appl Sci* 2020;10:4334. <https://doi.org/10.3390/app10124334> .
- [16] Rahat M., 2024, *Cyber Security in Smart Manufacturing (Threats, Landscapes & Challenges)*, *arXiv*:2304.10180. <https://doi.org/10.48550/arXiv.2304.10180> .
- [17] Souza J., et al. (2021). The State of Cybersecurity in Smart Manufacturing Systems: A Systematic Review. *European Journal of Business and Management Research*. 6. 188-194. 10.24018/ejbmr.2021.6.6.1173.
- [18] Humayed, AA, et al. (2018). Cyber-Physical Systems Security—A Survey. *IEEE Internet of Things Journal*, 4, 1802-1831. 10.1109/JIOT.2017.2703172 .

- [19] Williams B, Soulet M, Siraj A. A Taxonomy of Cyber Attacks in Smart Manufacturing Systems. 6th EAI Int. Conf. Manag. Manuf. Syst., Springer; 2023, p. 77–97. https://doi.org/10.1007/978-3-030-96314-9_6.
- [20] Kott, Alexander. (2019). Fundamental Concepts of Cyber Resilience: Introduction and Overview. 10.1007/978-3-319-77492-3_1.
- [21] Elhabashy AE, Wells LJ, Camelio JA. Cyber-physical attack vulnerabilities in manufacturing quality control tools. Qual Eng 2020;32:676–92. <https://doi.org/10.1080/08982112.2020.1737115>.
- [22] Bharti Anupriya, Hazela Bramah, (2024). Enhancing Cybersecurity in Smart Manufacturing: A Comprehensive Approach. Indian Journal of Computer Science and Technology. 192-201. 10.59256/indjst.20240302030.
- [23] Barrett, M. (2018), Framework for Improving Critical Infrastructure Cybersecurity Version 1.1, NIST Cybersecurity Framework, <https://doi.org/10.6028/NIST.CSWP.04162018>.
- [24] Bakkar, MN (2023). CyberSecurity essentials for industry 5.0, in Advanced Research and Real-World Applications of Industry 5.0, (IGI Global), pp. 49–65. 10.4018/978-1-7998-8805-5.ch003.
- [25] Alagic G, et al. (2025) Status Report on the Fourth Round of the NIST Post-Quantum Cryptography Standardization Process, NIST Internal Report (IR) NIST IR 8545. <https://doi.org/10.6028/NIST.IR.8545>.
- [26] S. Sicari, et al. Security, privacy and trust in Internet of Things: The road ahead, Computer Networks, Volume 76, 2019, Pages 146-164, ISSN 1389-1286, <https://doi.org/10.1016/j.comnet.2014.11.008>.
- [27] Abraham, D., et al. Towards enhanced cybersecurity in industrial control systems: a systematic review of context-based modeling, digital twins, and machine learning approaches. Int. J. Inf. Secur. 24, 243 (2025). <https://doi.org/10.1007/s10207-025-01158-1>.
- [28] Deloitte, 2026, Cybersecurity for smart factories, <https://www.deloitte.com/us/en/industries/energy/articles/smart-factory-cybersecurity-manufacturing-industry.html>.
- [29] Ren, Anqi, et al. Cyber Security in Smart Manufacturing: Survey and Challenges Cyber Security in Smart Manufacturing: Survey and Challenges, IISE Annual Conference. Proceedings; Norcross (2019): 716-721.
- [30] Bahrami, AH, Rouzbahani, HM (2021). Cyber Security of Smart Manufacturing Execution Systems: A Bibliometric Analysis. In: Karimipour, H., Derakhshan, F. (eds) AI-Enabled Threat Detection and Security Analysis for Industrial IoT. Springer, Cham. https://doi.org/10.1007/978-3-030-76613-9_6.
- [31] Capgemini Research Institute (2026). Smart & Secure: Why smart factories need to prioritize cybersecurity. https://prod.ucwe.capgemini.com/wp-content/uploads/2022/06/Cybersecurity-in-Smart-Factories_Web-2.pdf.

Dark Factory vs. Humanoid Factory: Two Development Directions of Industrial Automation in Comparison

Viktorio Malisa^{*1}

Abstract: *Industrial automation is currently shaped by two complementary development paths: the highly standardised, fully autonomous Dark Factory and the Humanoid Factory, in which humanoid robots serve as flexible automation components within existing production environments. While both concepts pursue unmanned manufacturing, they differ fundamentally in flexibility, infrastructure compatibility, and technological maturity. Dark-factory approaches are already established in several industries, whereas no productive systems relying entirely on humanoid robotics exist to date.*

This study compares both paradigms through a systematic literature review and a conceptual classification within the context of industrial production systems. The findings show that humanoid robots can address automation challenges in areas previously considered difficult to automate, positioning them as a complementary technology rather than a competing one. At the same time, significant research and development needs emerge at the intersection of autonomy, adaptability, and safety.

The analysis concludes that a fully Humanoid Factory is not feasible in the short-to-medium term due to technological immaturity and the absence of harmonised safety and certification frameworks. However, evidence suggests that hybrid production models are beginning to emerge, in which humanoid robots are gradually integrated into dark-factory environments to enhance adaptability, resilience, and variant-rich production capabilities.

Keywords: *Dark Factory, Lights-Out Manufacturing, Humanoid Factory, Humanoid Robot, Industry 5.0*

1. Introduction

In the wake of digital transformation, industrial production is undergoing a fundamental paradigm shift. Shortages of skilled workers, rising costs and the desire to achieve optimum efficiency are prompting the development of two contrasting yet forward-thinking approaches to automation. On the one hand, there is the concept of the 'dark factory'. These fully automated production facilities operate without human staff, require neither lighting nor air conditioning, and are based on a rigid, highly integrated network of traditional industrial and logistics systems. In contrast, humanoid robotics is emerging as a

^{*1}F-AR Förderung der Automation und Robotik, Wien, Austria
E-mail: malisa@f-ar.at

flexible, anthropomorphic approach to automation. These robots are designed to integrate into existing work environments and use human-optimised tools, thus eliminating the need for significant infrastructure changes.

The concept of the 'Dark Factory' involves eliminating all human involvement on the factory floor. Meanwhile, the field of humanoid robotics aims to replicate human flexibility and physical capabilities with precision. Integrating these approaches into existing industrial production systems raises profound questions, as do the associated technological, economic, and strategic implications. The focus then shifts to the challenges of industrialising humanoid robots and exploring the potential of combining these two automation concepts to create hybrid architectures.

According to the Fraunhofer Institute for Manufacturing Engineering and Automation (IPA), a Dark Factory is defined as a fully automated production environment without permanent human presence, supported by autonomous material logistics, self-monitoring processes, and AI-based error detection. Fraunhofer positions this model as the final evolutionary stage of Industry 4.0 automation. [5]

This thesis conducts a comparative analysis of the development trajectories of the Dark Factory and humanoid robotics within an industrial context. Examining how well both fulfil modern core requirements such as human-centredness, sustainability, and systemic resilience, the study is conducted from an Industry 5.0 perspective. Based on this analysis, a criteria matrix is developed to inform future sustainable automation strategies.

2. Classification: Production Processes and Level Of Automation

2.1 The Process Industry As The Original 'Dark Factory'

The process industry, which includes the chemical, petrochemical, pharmaceutical, food processing and steel sectors, has been gradually automated since the 1960s. To a particularly significant extent, this process has occurred since the 1970s and 1980s, effectively turning it into a 'dark factory' through the use of digital control technology and, later, networked, software-based systems. Refineries, chemical plants and breweries now operate largely without human intervention: raw materials flow in and products flow out, all controlled by Distributed Control Systems (DCS), Supervisory Control and Data Acquisition (SCADA) systems and Programmable Logic Controllers (PLC). The problem of continuous human-free production has been solved here for decades.

The key difference from discrete manufacturing lies in the nature of the product: the process industry generates continuous streams of liquids, gases or bulk materials. Human tasks are limited to monitoring, maintenance, and dealing with exceptional situations, and even these are increasingly being carried out remotely.

2.2 Hybrid Process Chains: Transitioning From Continuous to Batch Production

Fully automated process chains form a distinct category. In these chains, continuous manufacturing processes transition seamlessly into the handling of discrete items with no interruption to the flow of the process and no human intervention.

Typical examples:

- Plastic film production: cutting > winding onto rolls > automatic palletising.
- Liquid production (e.g. water, beer, milk, juice): filling > capping > labelling > packaging > palletising.
- Dough or batter production: portioning > baking/shaping > packaging > cartoning.
- Wire or tube drawing: depositing > bundling > packaging.

These fully integrated production lines have been utilised in industry since the 1980s and 1990s. In accordance with the definition of lights-out operation, these factories are effectively 'Dark Factories'. Despite the continuous optimisation of their underlying systems in terms of throughput, process quality and digital connectivity, these production lines have become the established standard in automation technology.

2.3 Discrete Manufacturing as a Driving Force for Innovation

The focus of automation is therefore on discrete manufacturing, particularly in areas where product variety, complexity and frequent product changes push traditional automation approaches to their limits or beyond.

As shown in Table 1, the term 'dark factory' is a novel concept, particularly in the context of discrete mass production. The gap in diverse and flexible discrete manufacturing is to be filled by humanoid robots. Currently, this need is met by stationary industrial robots, human-robot collaboration, and a shortage of skilled workers.

Table 1. Production systems and their level of automation in the context of the Dark Factory and the Humanoid Factory

Production systems	Examples	State of automation	Impact
Continuous process industry	Refineries, chemical industry, power stations	DCS/SCADA/PLC since the 1970s	Resolved for decades; ongoing networking and efficiency improvements
Hybrid process chain	Film extrusion >> Cutting >> Winding; Filling >> Sealing >> Palletising	Fully integrated lines, PLC-controlled	Automated since the 1980s/1990s – established Dark Factory
Discrete manufacturing (mass production)	Automotive press shops, pharmaceutical filling, semiconductors	Industrial robots, AGVs, AMRs	Dark Factory as an active optimisation target
Discrete manufacturing (variant production)	Small-batch assembly, customised assemblies, logistics	Shortcomings: too variable for stationary robotics, shortage of skilled workers	Target domain of the Humanoid Factory

3. Concepts and State of the Art

3.1 The Concept of the Dark Factory (Lights-out Manufacturing)

A dark factory, also known as lights-out manufacturing, is a production environment that requires no human intervention across multiple production cycles. The term derives from the fact that these systems can be operated without lighting, air conditioning or ergonomic adjustments when running in automatic mode. The primary technological foundation consists of standardised, networked machines, industrial robots, automated high-bay warehouses, driverless transport systems and autonomous mobile robots. The efficiency of such factories is linked to maximum standardisation and the elimination of human factors.

Conventional research has concentrated on the maintenance of process continuity and stability, predictive maintenance, and the mathematical optimisation of material flows. In a conventional Dark Factory architecture, there is a fully automated production area and a physically separate support area for human activities. Even in a dark factory, production cannot operate entirely without human involvement. In the autonomous production line (see Figure 1), materials such as granules, powder, paper bags, plastic tubes and empty pallets are processed automatically. Conveyor belts, packaging stations and robotic handling systems are employed to manage all process steps, culminating in the provision of a fully loaded pallet secured with stretch foil. The process is entirely automated, with no requirement for physical intervention by humans.

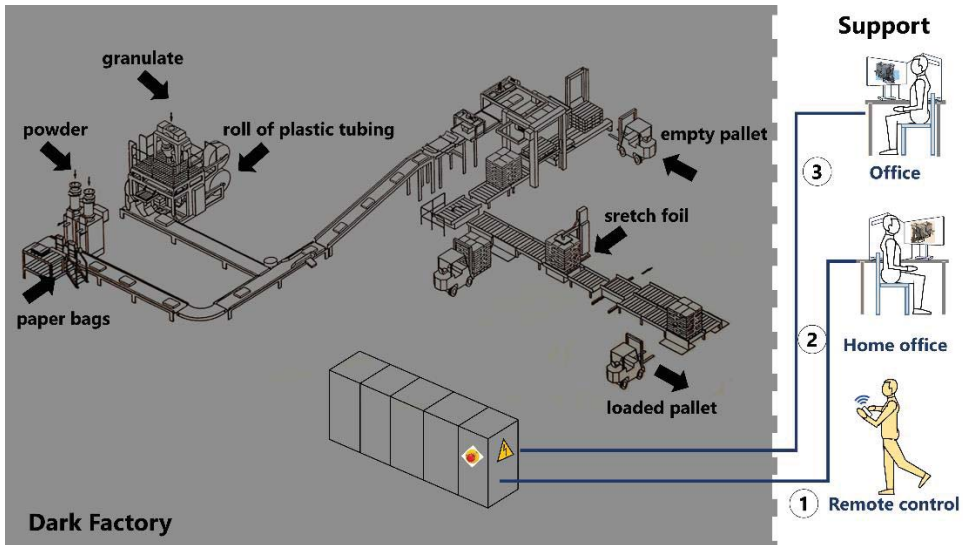


Figure 1. Schematic diagram of a Dark Factory

The permitted human roles in the 'Support' area outside the production hall (see Figure 1) are limited to the following:

- Remote control: remote access to machines and systems via mobile devices.
- Home Office: remote process monitoring and intervention.
- Office: overarching planning, analysis and control from a separate office area.

Therefore, all human activities take place outside the production area and are limited to monitoring, remote maintenance and management.

3.2 The Concept of the Humanoid Factory

In contrast, there is the Humanoid Factory approach. Based on the principle of anthropomorphic flexibility, this approach differs from the location-bound, specialised automation of the Dark Factory. A humanoid robot is a machine whose physique is modelled on the human body. This means that it has a head, a torso, two arms, hands, and two legs. The theoretical relevance of this approach stems from the concept of brownfield automation. Unlike traditional automation, where the factory environment is built around the machine, humanoid robotics enables a new approach: the machine is adapted to the factory environment. The robot should be able to adapt to a factory environment designed for humans.

Technologically, this development involves shifting towards artificial intelligence (AI), computer vision, and advanced sensor technology such as force-torque sensors and end-effectors that mimic human fine motor skills. Current research efforts focus primarily on the General-Purpose Robot (GPR). The term GPR is derived from general purpose AI (GPAI).

According to current theory, these robots are capable of learning various tasks without the need for complex programming, thanks to reinforcement learning and Large Behaviour Models (LBM). In a Humanoid Factory, humanoid robots perform human-like tasks and interact with other machines. Typical applications include collaborative work with industrial robots, order picking from shelves, material transport via AMRs (Autonomous Mobile Robots), and assembly work at traditional workstations. In line with the underlying principle, humanoid robots will take over the performance of human tasks, whilst the existing industrial infrastructure continues to be utilised. Whilst the Dark Factory automates the process, the Humanoid Factory automates human activity.

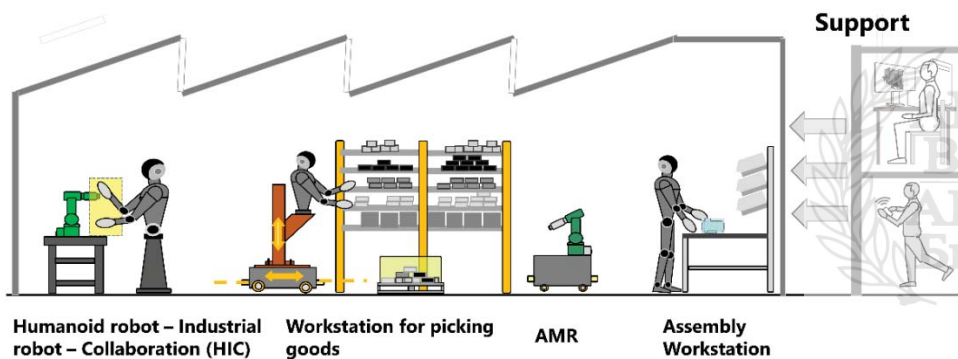


Figure 2. Schematic representation of a humanoid factory

As demonstrated in Figure 2, the Humanoid Factory concept involves the deployment of humanoid robots to perform tasks and interact with other machinery in a manner that emulates human behaviour. The figure illustrates the typical areas of application of AMRs, namely collaborative work with industrial robots, order picking, material transport, and assembly work at traditional workstations. The Humanoid Factory principle is elucidated as follows: the performance of human tasks is assumed by humanoid robots, with the existing industrial infrastructure remaining in use.

The term "Humanoid Factory" refers to a manufacturing environment in which human-like robots are responsible for key tasks. The product's unique selling point is its compatibility with existing infrastructure designed for humans. This encompasses tools, work surfaces, shelves and transport equipment, which can be utilised without modification due to their design to align with human body proportions.

Consequently, the paradigm is not oriented towards process standardisation, but rather towards adaptability. The objective of humanoid robots is to acquire new tasks through imitation, learning or natural language instructions, without the necessity for complex reprogramming.

3.3 Challenges in Implementing the ‘Humanoid Factory’

The current debate surrounding humanoid robots is characterised, above all, by a focus on hardware demonstrations and projected cost trends. However, the main obstacle to their widespread industrial use is the lack of regulatory standards. Although advanced technologies are being integrated into robotic systems, appropriate standards are lacking. Consequently, researchers must carry out risk assessments for their own developments and take appropriate measures to mitigate risks [13]. In this regard, it is crucial to integrate safety functions into the machines' control systems.

Humanoid robots are equipped with the following technologies:

- Bipedal platforms offering 20–40+ degrees of freedom for full-body manipulation;
- Multi-fingered robotic hands for heterogeneous grasping.
- Foundation models and large language models (LLMs) enable the robots to understand tasks and process instructions. These are large, pre-trained AI models that possess a broad understanding of language. They can interpret tasks, follow instructions, and adapt flexibly to new requirements without needing to be reprogrammed for each individual task.
- Imitation and reinforcement learning for robotics (RLfR): the robot learns through trial and error, i.e. it teaches itself to solve a specific task as effectively as possible.
- Edge AI and onboard inference enable real-time decision-making. AI models run on local processors within a robot. With onboard inference, the device performs the AI calculations itself. This allows the robot to make real-time decisions and react independently in milliseconds, even without an internet connection.
- SLAM (Simultaneous Localisation and Mapping) is a navigation technique used in dynamic environments. Robots use various sensors, such as cameras, LiDAR, radar and ultrasound, to map their constantly changing surroundings. As objects are moved and people walk around, the robot builds a map of its surroundings while determining its own position within them.)

Although humanoid robots themselves do not have a CE declaration of conformity, system integrators are incorporating them into manufacturing processes. As they are considered partly completed machinery, they must be fully safeguarded by external safety measures. These systems take over the safety-related functions, enabling the humanoid robot to operate as a non-safety-related module. This is only possible because there are currently no harmonised standards for dynamically stable industrial mobile robots. Work is underway to develop the ISO 25785-1 standard. This will apply to dynamically stable industrial mobile robots (DSIMRs), including humanoid robots.

In parallel to these European standardisation efforts, international bodies are increasingly recognising the urgency of establishing safety frameworks for dynamically stable robots. ASTM International has recently called for the rapid development of global safety standards for humanoid robots, warning that industrial deployments are progressing faster than the regulatory mechanisms needed to ensure safe operation [12]. This highlights that the challenges associated with integrating humanoids into industrial environments are not only technical but also regulatory, and that harmonised global standards will be essential for their safe and widespread adoption.

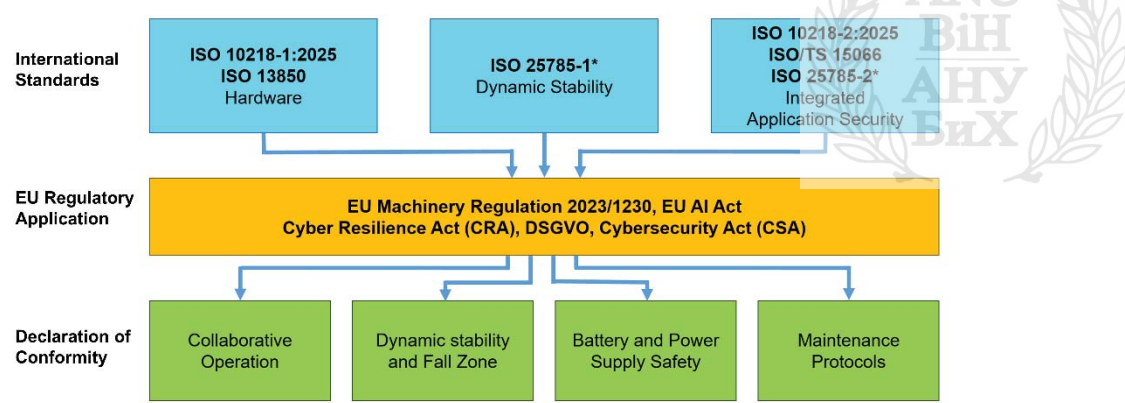


Figure 3. EU compliance framework for humanoid robots, in accordance with [10]

* Standards currently being drafted

Figure 3 provides an overview of the main EU regulations and standards currently in place for CE marking in robotics. As well as listing the adopted standards, it also lists those that are still under development. The latter include ISO 25785-1 and ISO 25785-2, which cover the integration of dynamically stable robots into automation systems.

A key difference between stationary industrial robots and humanoids is their behaviour during an emergency shutdown: industrial robots remain stationary and

in a safe state when the power is cut off, whereas humanoids collapse because their stability is actively maintained by motor control. Humanoids, on the other hand, collapse because their stability is actively maintained by motor control.

The concept of the 'Safe-Stopable Envelope' (SSE) addresses this issue by defining a state space. From this state, the robot can be reliably moved to a safe, minimum-risk position (Minimum Risk Condition, MRC) when an emergency stop is triggered, ensuring that it does not fall over or collapse uncontrollably. However, the software implementation and validation of the SSE under all operating conditions remains an open research problem.

The transition from traditional brownfield environments toward Industry 4.0 and ultimately Industry 5.0 is characterised by increasing levels of digitalisation, autonomy, and human-machine integration. Figure 4 illustrates this progression by mapping key technological pillars—smart human interfaces, digital connectivity, artificial intelligence, digital factory systems, and robotics—onto different stages of industrial evolution. It highlights how applied factories, dark factories, and humanoid-robot factories represent escalating degrees of automation and system intelligence.

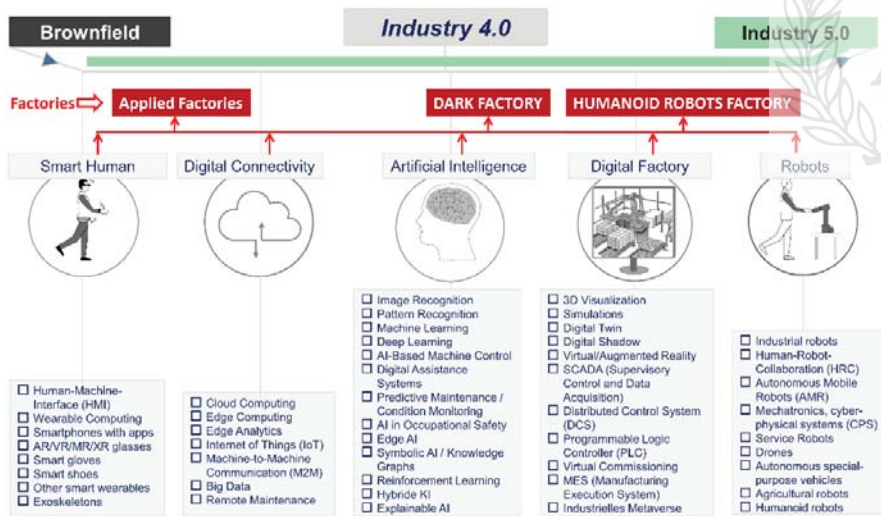


Figure 4. Positioning of the Dark Factory and Humanoid Factory within the context of Industry 4.0 and Industry 5.0

McKinsey's 2024 report "The future of physical AI: How humanoid robots are stepping into the brownfields of manufacturing" reinforces this trajectory. The study argues that humanoid robots are emerging precisely at the intersection shown in the graphic: they enable existing brownfield plants to adopt advanced

automation without requiring full redesigns. According to McKinsey, humanoids act as a bridging technology between Industry 4.0 automation and Industry 5.0 human-centric, adaptive production systems by taking over physically demanding, repetitive, or ergonomically challenging tasks while integrating seamlessly into legacy environments. [7]

Together, the visual model and McKinsey's analysis demonstrate that the convergence of AI, robotics, and digital factory technologies is not only reshaping greenfield innovation but increasingly transforming established brownfield manufacturing into hybrid, flexible, and progressively autonomous production architectures.

4. Methodological Approach

This comparative analysis of two automation approaches is based on a systematic literature review, expert interviews, and a qualitative content analysis of current industry publications. This methodological approach ensures that theoretical concepts and their practical implications are compared transparently and objectively.

In addition to primary academic literature, the analysis incorporated practical secondary sources. This is necessary because humanoid robotics is currently undergoing a rapid phase of industrial testing, the results of which are not always immediately reflected in academic literature.

The evaluation comprises:

- manufacturer information and white papers containing technical specifications and deployment reports from leading robotics companies, including Tesla, Boston Dynamics, Figure, and Agility Robotics;
- regulations and standards, including current safety and interface standards (ISO 10218, ISO/TS 15066 and VDA 5050), as well as EU machinery regulations;
- documentation of pilot projects;
- reports on real-world pilot phases in the automotive and logistics sectors (e.g. BMW and Mercedes-Benz).

Examples: Dark Factory and humanoid pilot projects

Since the 1990s, FANUC has been operating several production lines capable of running fully automatically for up to 30 days without interruption. CNC machines manufacture components, which are transported to the assembly line by autonomous transport systems. There, industrial robots assemble and package new robots. The employees are responsible for planning, programming and maintenance. [4]

Xiaomi's smart factory in Beijing launched a fully automated "dark factory" in 2023. It produces over one million smartphones per year. Production runs around the clock without the need for human presence, shift changes, breaks, or lighting. An AI-based platform controls and optimizes processes in real time, performs automated quality control via image processing, and enables predictive maintenance. [14]

Philips has officially described the factory in Drachten (Netherlands) as one of the "smartest factories in Europe" since 2020. In the highly automated production, in which over 300 robots and autonomous transport systems work together in 24/7 operation, razors and personal care devices are created with minimal human presence. [11]

For the Humanoid Factory perspective, current pilot projects in humanoid robotics – such as Tesla Optimus, Figure 01 (BMW pilot [2]), Apptронik, Apolo (Mercedes-Benz pilot [1] and Agility Robotics "Digit" (Amazon pilot [6]) – were included, as they reflect the current state of technological development and the practical limitations of humanoid systems. These examples served as empirical reference points for the qualitative content analysis.

Recent industrial deployments underline this trend. Sanctuary AI, for example, has released the next generation of its Phoenix general-purpose robot and announced a deployment partnership with Magna International, demonstrating that humanoid and humanoid-adjacent systems are beginning to enter real brownfield manufacturing environments [8].

Tesla has also begun deploying its Optimus Gen-2 humanoid robot in automotive manufacturing environments, focusing on neural-network-based behaviour training and task generalisation. This illustrates how major industrial players are experimenting with AI-driven humanoid systems in real brownfield settings [9].

The collected data is synthesised using a custom-developed criteria matrix. These evaluation criteria are derived directly from the core requirements of Industry 5.0 [3]. The two automation approaches are then compared in a standardised way, based on their technological drivers, safety concepts and organisational implications for the workforce.

5. Comparative Analysis of Automation Systems

The future viability and potential applications of the 'Dark Factory' and the 'Humanoid Factory' are thoroughly assessed through a systematic comparison of the two concepts across four key dimensions: technology, economics, safety and organisation.

5.1 Technological Dimension: Autonomy vs. Flexibility

The technological divergence between the two paths is evident in their inherent control and system paradigms. The Dark Factory embodies the principle of deterministic full automation. The technological infrastructure is based on a rigid interlinking of cyber-physical systems (CPS), in which every machine, every driverless transport system, and every sensor communicates via standardized protocols. System stability is guaranteed by the elimination of unpredictable variables.

While the Dark Factory achieves maximum autonomy with minimal flexibility, the Humanoid Factory positions itself closer to the human performance spectrum in terms of adaptability and the ability to handle variations. The outlined development directions illustrate the current research's endeavor to bridge the technological gap toward highly autonomous yet flexible systems.

Unlike the Dark Factory, which relies on rigid, fully standardised processes, the Humanoid Factory uses flexible robots that mimic human movement. The focus shifts from mechanical infrastructure to software control. Thanks to AI models and modern image processing, these robots can act as universal workers, recognising ambiguous situations and interpreting tasks in real time, while also independently compensating for deviations or tolerances. While the Dark Factory achieves precision through maximum process standardisation, the Humanoid Factory leverages the robots' ability to adapt to unpredictable real-world conditions — much as a human would. That's the developers' plan. The industry is eager to see if it can be implemented as planned, but above all it wants to ensure that security is guaranteed.

5.2 Economic Dimension: Greenfield vs. Brownfield Investments

From a business and production technology perspective, the investment and operating cost structures of “dark factories” and “humanoid factories” differ fundamentally.

Dark Factory (Greenfield Approach)

A fully autonomous Dark Factory requires significant initial investment, as the entire production facility must be designed from the ground up. The infrastructure is tailored precisely to the machines, material flows, and automated processes.

Typical characteristics:

- - New facility or extensive renovation
- - Highly specialised equipment
- - No human workstations
- - Fully standardised processes

The economic benefits only become apparent during ongoing operations, primarily due to:

- Elimination of personnel
- Lower energy requirements (no lighting, minimal heating).
- Reduced occupational safety requirements
- Very high throughput for large quantities.

Conclusion: The high initial costs make this worthwhile only for stable processes and very large production runs.

Humanoid Factory (Brownfield Approach)

This approach enables the gradual automation of existing production facilities without the need for structural modifications. A humanoid robot is acquired and used as a mobile work tool wherever staff numbers are low or processes vary.

Advantages:

- Significantly lower initial investment
- No need to modify halls, workstations or tools
- Quick integration into existing processes
- Flexible deployment across various workstations
- Ideal for high product variety and smaller batch sizes

The economic benefits arise from:

- A flexible response to changing orders
- Relief during skilled labour shortages
- Reduced downtime
- Use of existing infrastructure

Conclusion: Lower initial costs make it particularly worthwhile for high-variety production and in situations of labour shortage.



5.3 Safety and Regulatory Aspects

From a safety perspective, this is one of the biggest challenges currently facing the deployment of humanoid robots. In a dark factory, the safety architecture is clearly defined and spatially separated. Since humans are not permitted in the production area during normal operation, safety fences, light curtains and interlocks can provide full protection. If someone enters the area, the system stops immediately. This consistent physical separation of humans and machines largely eliminates the potential for danger.

However, this principle does not apply in a humanoid factory. Although production operations here are also human-free, the humanoid robots move freely in an environment originally designed for human workers. Therefore, spatial separation through safety zones is not practical. Instead, safety must be ensured intrinsically through reliable collision avoidance, limitation of forces and speeds, safe sensor and actuator technology, and functional, AI-supported motion planning. This is where the regulatory conflict arises. Current standards, such as ISO 10218, ISO 13849 and IEC 62061, as well as the new EU Machinery Directive (from 20 January 2027 EU Machinery Regulation), stipulate that the movements of robots must be deterministic, predictable and verifiable. However, humanoid robots whose AI can react to unstructured situations only meet these requirements to a limited extent at present. Since there are currently no standardised testing and certification frameworks for autonomous, AI-controlled motion sequences, the widespread industrial use of humanoid robots in open production environments will be heavily restricted by regulations in the short and medium term.

5.4 Organizational Dimensions in the Context of Industry 5.0

The Dark Factory embodies the principles of Industry 4.0, including maximum standardisation, full digitisation, and complete connectivity.

A small team of highly qualified specialists is responsible for maintenance, process monitoring, and plant optimisation outside the production hall.

In contrast, the Humanoid Factory is guided by the principles of Industry 5.0: resilience, adaptability and human-centred system design. Although operational production can take place without a human presence, humans remain closely integrated into the organisation as supervisors, teleoperators, process designers, and problem solvers. Humanoid robots perform monotonous, ergonomically demanding, or safety-critical tasks, while humans flexibly intervene in situations involving variability, disruptions, or complex decisions.

This division of roles increases the systemic resilience of the production network. When faced with market changes, supply chain disruptions or highly varied

customer orders, the Humanoid Factory can react more quickly, as adjustments primarily involve software, and human expertise is always available.

5.5 Synthesis: The Criteria Matrix

The following matrix compares the 'Dark Factory' and 'Humanoid Factory' automation approaches based on technical, organisational, and safety criteria. It can be used as a strategic guide when designing future factory architectures.

The 'Dark Factory' approach is particularly well-suited to high-volume, standardised production environments. Advantages include maximum process efficiency, low variable costs after commissioning, and a technologically mature foundation. However, it offers less flexibility with regard to product changes and high-variety manufacturing.

The 'Humanoid Factory' approach has high potential due to its compatibility with existing infrastructure and its ability to adapt to software changes. Currently, however, these advantages are offset by the approach's low technological maturity, high investment costs, and unresolved safety and infrastructure issues. Table 2 is the first to compare the two paradigms across twelve dimensions, incorporating the safety perspective. This perspective has largely been overlooked in previous studies.

Terms:

Cyber-Physical Systems (CPS): Real-time networked systems integrating machines, sensors and digital control. They form the basis of Industry 4.0 automation.

Intrinsic safety of humanoid robots: A safety principle whereby the robot autonomously establishes safe states through sensors, AI-based motion planning and control-related limitations. This principle is central, yet it has not yet been fully standardised.

Enablers for Industry 5.0 Resilience: Humanoid robots increase the adaptability of industrial systems, enabling flexible responses to changes in the market and supply chain.

OEE (Overall Equipment Effectiveness) describes the actual effectiveness of a production unit. It consists of three factors: availability, performance and quality. OEE is the key metric for objectively comparing the efficiency of production systems.

Large Behaviour Models (LBMs) are AI models that can process language and learn to perform complex sequences of actions. They connect: LBMs enable humanoid robots to understand and adapt to their environment, performing tasks rather than merely executing pre-programmed movements.

The comparative analysis of the two automation pathways is based on a systematic literature review and qualitative content analysis of recent industry publications. This approach ensures that the theoretical concepts and their practical implications are compared transparently and objectively.

Table 2: Dark Factory vs. Humanoid Factory comparison matrix

CRITERIA	DARK FACTORY	HUMANOID FACTORY
Primary goal	Maximum process efficiency & continuity	Maximum adaptability & flexibility
Human presence	Low (Monitoring, Service & Maintenance)	Low (Teleoperation, Monitoring, Service, Maintenance outside of Operations)
Primary Automation	Industrial robots, AGVs, CNC, MES	Humanoid robots + LLM cognition
Flexibility	Low to medium (task-specific)	High (adaptive, trainable)
Product changeover costs	High (mechanical conversion)	Low-medium (software re-training)
Acquisition costs	Very high (infrastructure)	High (robot hardware, software licenses)
Operating costs	Low (after amortization)	Resources (energy, maintenance, updates)
Technological maturity	Highly established market	Low to medium (from 2030 and later)
Main focus of operations	Mass production, chemicals, pharmaceuticals, automotive	Small series production, logistics, variant manufacturing
Operating hours/day	24/7 (continuous)	2–8 hours (currently, target: 16–20 hours with spare batteries)
OEE (Overall Effectiveness)	85–95%	Research prototype, trials in production
Conversion compatibility	Requires new construction / renovation	Compatible with existing infrastructure (development is not yet foreseeable)
Security concept	Spatial separation (protective fences/light curtains)	Intrinsic safety (sensors/AI control)
Safety standard	ISO 10218-1/- 2, ISO 13849/ IEC 62061, ISO 12100	ISO/CD 25785-1/-2 (both in development)
Technology drivers	Cyber-physical systems, IoT, deterministic sensors	AI, Large Behavior Models, Computer Vision
Technology classification	Consistent implementation of Industry 4.0	Enablers for the resilience of Industry 5.0

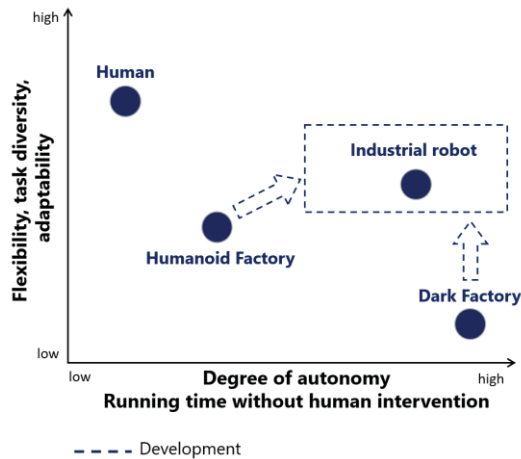


Figure 5. Positioning of Dark and Humanoid Factories

Figure 5 shows a two-dimensional diagram illustrating the positioning of humans, industrial automation systems, Dark Factory concepts, and humanoid robot systems along two central dimensions. The horizontal axis describes the degree of autonomy or the possible operating time without human intervention. The vertical axis represents flexibility, task diversity, and adaptability.

The upper left area represents the human, characterized by high flexibility but low autonomy. The lower right area represents the Dark Factory, which exhibits a very high degree of autonomy but low task flexibility. Between these two extremes are classic industrial robots, which possess high process autonomy but limited adaptability.

The Humanoid Factory is depicted in the middle to upper part of the diagram and represents a system that strives for both greater flexibility and increasing autonomy. The depicted development paths illustrate two different technological directions:

- the deterministic automation of industrial robots towards the Dark Factory and
- the adaptive automation of humanoid systems, which in the long term should enable a greater variety of tasks and longer autonomous operating times.

The diagram thus visualizes the two fundamental automation paradigms: maximum standardization (Dark Factory) versus maximum adaptivity (Humanoid Factory).

This positioning indicates that humanoid-robot factories function as a bridging technology, enabling brownfield plants to evolve toward Industry 5.0 without requiring disruptive infrastructure redesigns.

6. Discussion and Future Outlook

The development directions shown in Figure 5 confirm this assessment and illustrate the divergence between deterministic and adaptive automation. The comparative analysis demonstrates that the "Dark Factory" and the "Humanoid Factory" are not competing, but rather complementary development directions in industrial automation. This chapter discusses the implications for future factory architectures and outlines the path to a hybrid production model. The development directions shown in Figure X confirm this assessment and illustrate the divergence between deterministic and adaptive automation.

The results of the criteria matrix suggest that for the majority of industrial companies, focusing exclusively on only one of the two paths is either economically unviable in the short and medium term (Dark Factory in a volatile market environment) or technologically unfeasible (humanoid factory due to its low level of maturity). The evidence instead points to the emergence of hybrid production models.

In such a hybrid factory architecture, highly standardized core production operates as a "Dark Factory zone" (Lights-out). Humanoid robots function as a flexible, complementary technology at the interfaces. They close automation gaps in areas that are inaccessible or unergonomic for conventional machines – for example, in flexible material supply, quality control of highly variant components, or adaptive machine loading in brownfield environments.

To enable the gradual integration of humanoid robotics into existing industrial structures, this work identifies three key research areas:

- Hardware reliability: Classic industrial robots achieve tens of thousands of maintenance-free operating hours, while bipedal humanoid systems in pilot phases still exhibit short uptimes. Research focuses on robust, energy-efficient actuators.
- Standardization of safety standards: A regulatory vacuum exists. Existing standards for human-robot collaboration (e.g., ISO/TS 15066) are not transferable to mobile, human-sized humanoids with AI-based motion planning. New certification criteria are needed for the functional safety of Large Behavior Models (LBMs) and other integrated AI models.
- Background infrastructure (fleet management): While established interfaces such as VDA 5050 exist for AGVs/AMRs, a standardized orchestration software that enables seamless integration into existing IT systems is lacking for humanoid general-purpose robots.

One limitation is that the evaluation of the Humanoid Factory is primarily based on current model projects, manufacturer specifications, and theoretical concepts. Due to rapid technological development, empirical long-term data on actual

operating costs and life cycle parameters of large-scale humanoid systems operating in three shifts are still lacking.

7. Conclusion

The analysis shows that the Dark Factory and the Humanoid Factory represent two complementary, yet fundamentally different, automation paradigms. The Dark Factory embodies the efficiency maximum of deterministic Industry 4.0 automation, while the Humanoid Factory, as an adaptive Industry 5.0 concept, focuses on flexibility, resilience, and infrastructure compatibility. Both approaches, when implemented in isolation, encounter functional limitations: the Dark Factory due to a lack of adaptability, and the Humanoid Factory due to technological maturity and a lack of standardization. The results therefore suggest that future factory architectures will be designed not monolithically, but as hybrids. In such systems, the Dark Factory zone handles highly standardized core production, while humanoid robots, as flexible complements, fill automation gaps. As the field is developing quickly, the findings represent the current state of knowledge and may require revision as new standards and industrial deployments emerge. This creates a production model that combines efficiency and adaptability and structurally enables the transition from Industry 4.0 to Industry 5.0.

References

- [1] Apptronik. (2024). Apptronik and Mercedes-Benz enter commercial agreement to pilot Apollo humanoid robot in manufacturing facilities. Apptronik News. Online https://apptronik.com/news-collection/apptronik-and-mercedes-benz-enter-commercial-agreement?utm_source=copilot.com, accessed on 02.06.2025
- [2] BMW Group. (2026, 4. März). Erfolgreicher Pilotversuch im BMW Group Werk Spartanburg: Humanoide Roboter Figure 02 unterstützen Serienfertigung im Karosseriebau. BMW PressClub.
- [3] Europäische Kommission. (2021). Industrie 5.0: Für eine nachhaltige, menschenzentrierte und widerstandsfähige europäische Industrie. Amt für Veröffentlichungen der Europäischen Union.
- [4] Industrial Automation. (2026). FANUC Corporation. FANUC: The Benchmark for Lights-Out Manufacturing and Industrial Automation. Online: https://www.industrialautomationindia.in/articles/fanuc-lights-out-manufacturing-benchmark?utm_source=copilot.com, accessed on 03.06.2026
- [5] Fraunhofer-Institut für Produktionstechnik und Automatisierung (IPA). (2024). Der schmale Grat zur Dunkelfabrik: Potenziale und Grenzen vollautomatisierter Produktionssysteme. Fraunhofer-Eigenverlag.

- [6] Agility Robotics, (2023). Apptronik and Mercedes-Benz Enter Commercial Agreement. *Apptronik Press Release*. Online <https://www.agilityrobotics.com/content/agility-robotics-broadens-relationship-with-amazon>, accessed on 04.06.2026
- [7] McKinsey & Company. (2024). The future of physical AI: How humanoid robots are stepping into the brownfields of manufacturing. *McKinsey Operations Insights*.
- [8] Sanctuary AI. (2024). Sanctuary AI releases next generation of Phoenix general purpose robot and partnership deployment with Magna International. Robotics 24/7.
- [9] Tesla, Inc. (2025). Tesla Optimus: Gen 2 deployment and neural network behavior training in automotive factories. Tesla AI & Robotics Division.
- [10] Ulrich Baldauf, (2026). Humanoid Robot Safety Standards 2026: Best Compliance Guide. Online: <https://theresarobotforthat.com/blog/humanoid-robot-safety-standards-2026/> accessed on 25.05.2025
- [11] Philips, (2022). Werken bij Philips Drachten: één van de slimste fabrieken van Europa. Online <https://www.philips.nl/a-w/about/news/archive/standard/about/news/articles/2022/20220916-werken-bij-philips-drachten-een-van-de-slimste-fabrieken-van-europa.html>, accessed on 22.05.2026
- [12] Pao, J. (2025). ASTM Calls for Urgent Safety Standards for Humanoid Robots. In: Tech Journal UK, ASTM International. Online: <https://www.techjournal.uk/p/astm-calls-for-urgent-safety-standards>, accessed on 02.06.2026
- [13] Industriezeitschrift.de, (2025). Sicherheitsaspekte bei humanoiden Robotern. Online: <https://industriezeitschrift.de/robotik/humanoide-roboter/sicherheitsaspekte>, accessed on 01.06.2026
- [14] ertugrulturgay.com, (2025). The Factories of the Future Shine in the Dark. online <https://ertugrulturgay.com/the-factories-of-the-future-shine-in-the-dark/>, accessed on 03.06.2026

Digital Competence and Technology Adoption in Vocational Education: Evidence from Bosnia and Herzegovina

Madžida Hundur Hiyari^{*1}, Nejra Merdović¹, Jovana Kepčija Pavlović², Adna Softić¹, Amra Džuho¹, Ana Lalović¹

Abstract: *The digital transformation of vocational education and training (VET) represents one of the key prerequisites for developing a competent workforce capable of meeting the challenges of Industry 4.0 and the digital economy. However, the successful integration of digital technologies into the educational process requires adequate digital infrastructure, continuous professional development of teachers, and the enhancement of students' digital competencies. The aim of this study was to assess the current level of digital readiness in vocational education in Bosnia and Herzegovina by analysing digital competencies, the use of digital technologies, existing barriers, and the need for additional training among vocational secondary school teachers and students. The research was conducted using an online questionnaire involving 392 teachers and 1865 students from different regions of Bosnia and Herzegovina. The results indicate that digital tools have become an integral part of the teaching process, with most teachers demonstrating an intermediate level of digital competence, while students use a wide range of digital technologies on a daily basis, including artificial intelligence-based tools. The most significant barriers identified by respondents were the lack of adequate equipment and insufficient internet infrastructure. The findings suggest that Bosnia and Herzegovina has considerable potential for the digital transformation of vocational education; however, its successful implementation requires systematic improvements in digital infrastructure, continuous development of teachers' competencies, and the strategic integration of modern digital technologies into the education system.*

Keywords: *digital transformation, vocational education and training (VET), digital competencies, artificial intelligence*

1. Introduction

Digital transformation has become one of the most significant drivers of educational system development worldwide. The rapid advancement of information and communication technologies, artificial intelligence, the Internet of Things, and other digital solutions has fundamentally changed the way knowledge is acquired and competencies required by the labour market are

^{*1}Verlab Research Institute for Biomedical Engineering, Medical Devices and Artificial Intelligence, Sarajevo, Bosnia and Herzegovina

²Algebra Bernays University

E-mail: madzida@verlabinstitute.com

developed [1]. Digital transformation is particularly important in the field of Vocational Education and Training (VET), whose primary objective is to prepare students for employment in modern industrial and business environments [2]. Through the Digital Education Action Plan 2021–2027 [3], the European Commission emphasizes that the development of digital competencies among teachers and students is one of the key priorities of European education policy, with a particular focus on digitally competent educators, accessible digital infrastructure, and the responsible use of artificial intelligence in education.

Digital competence is now recognized as one of the eight key competences for lifelong learning defined by the European Union. It encompasses the confident, critical, and responsible use of digital technologies for learning, work, and active participation in society. In line with this objective, the European Framework for the Digital Competence of Educators (DigCompEdu) [4] was developed to define the competencies teachers require for the effective integration of digital technologies into teaching and learning. The implementation of this framework contributes to improving teaching quality, promoting innovative pedagogical approaches, and preparing students for the demands of the modern labour market. Bosnia and Herzegovina has made notable progress in recent years toward the digital transformation of education. The national strategic document *Priorities for the Integration of Entrepreneurial and Digital Competencies in the Education System of Bosnia and Herzegovina 2019–2030* [5] recognizes digital competence as one of the core competencies to be developed at all levels of education. An additional milestone was achieved through Bosnia and Herzegovina's accession to the Digital Europe Programme [6] in 2024, providing new opportunities for the development of advanced digital skills, the establishment of Digital Innovation Hubs, and the modernization of educational infrastructure. Nevertheless, despite the existence of strategic policy documents, the digital transformation of vocational education remains uneven due to insufficient systemic support, disparities in digital infrastructure, and limited opportunities for the continuous professional development of teachers.

A further challenge is the rapid adoption of artificial intelligence-based tools, particularly large language models, which are increasingly becoming part of students' everyday learning practices [7]. While students are quick to adopt these technologies, educational institutions and teachers often lack the competencies, resources, and institutional guidelines necessary for their effective and responsible implementation in teaching [8]. Consequently, there is a growing need to continuously assess the digital competencies of all stakeholders involved in vocational education, identify existing barriers, and define priorities for future digital transformation initiatives.

Accordingly, the aim of this study is to assess the current level of digital readiness in vocational education in Bosnia and Herzegovina by analysing the digital competencies of vocational school teachers and students, the extent of digital

technology adoption, the key barriers to implementation, and the need for additional professional development. The findings provide evidence-based recommendations that can support the further digital transformation of vocational education in Bosnia and Herzegovina.

2. Research Methodology

This study was conducted to assess the level of digital readiness of vocational education in Bosnia and Herzegovina by analysing the digital competencies of teachers and students in vocational secondary schools. Data were collected using an anonymous online questionnaire developed within the **ERASMUS VETechConn (Elevating Vocational Education in Bosnia and Herzegovina and Montenegro through Collaborative Enhancement of VET Teachers)** project. Separate questionnaires were designed for teachers and students, consisting of closed-ended questions, multiple-choice items, and self-assessment scales for evaluating digital competencies.

A total of **2257 respondents** participated in the study, including **392 teachers** and **1865 students** from vocational secondary schools across different regions of Bosnia and Herzegovina. The teachers' questionnaire collected information on teaching experience, frequency of digital technology use in teaching, self-assessed level of digital competence, barriers to the implementation of digital tools, previous participation in professional development activities, and interest in future training. The students' questionnaire focused on the frequency of digital technology use in education, the types of digital tools used, self-assessment of digital competencies, perceived barriers to the use of digital technologies, and interest in further developing digital skills.

The collected data were analysed using descriptive statistical methods. The results are presented as absolute and relative frequencies, enabling the identification of digital technology usage patterns, key implementation challenges, and the need for further development of digital competencies among both teachers and students. Particular emphasis was placed on the comparative analysis of the responses from the two groups in order to identify common challenges and priorities for advancing the digital transformation of vocational education in Bosnia and Herzegovina.

3. Results

The results of the study are presented through an analysis of responses provided by teachers and students from vocational secondary schools in Bosnia and Herzegovina. The analysis focused on assessing the level of digital competencies, the frequency of digital technology use, the barriers to their implementation, and the respondents' willingness to further develop their digital skills. Since the study

included two distinct groups of participants, the findings are presented separately for teachers and students, followed by a comparative analysis of the key findings.

3.1 Digital Readiness of Teachers

A total of 392 vocational secondary school teachers from Bosnia and Herzegovina participated in the study. The largest proportion of respondents taught technical sciences (149), followed by social sciences (127) and natural sciences (89), while arts (16) and sports (11) were less represented. Most teachers (78%) had more than five years of teaching experience, indicating that the findings primarily reflect the perspectives of experienced educators.

The analysis showed that digital technologies have become an integral part of the teaching process. More than half of the teachers reported using digital tools either daily or several times a week, while only a small proportion indicated that they never use them. The most frequently used technologies included presentation and content creation tools (PowerPoint, Canva and Prezi), online learning platforms (Microsoft Teams and Google Classroom), and multimedia resources such as YouTube. In contrast, more advanced technologies, including artificial intelligence, virtual reality, and simulation software, were used considerably less frequently.

The self-assessment of digital competencies revealed that the majority of teachers considered their digital skills to be at an intermediate level, while approximately one-sixth assessed their competencies as advanced. At the same time, nearly one-third of respondents rated their digital competencies as basic or beginner level, highlighting the need for additional professional development and capacity building.

Regarding the barriers to the integration of digital technologies into teaching, the lack of adequate equipment in schools was identified as the most significant challenge. In addition to infrastructure-related limitations, a considerable number of teachers reported insufficient time to prepare digitally supported lessons and emphasized the need for further training in digital technologies. Fewer respondents identified the lack of institutional support or low student motivation as major barriers.

The results also indicate a high level of motivation for further professional development. A total of 341 teachers (87.0%) expressed interest in participating in additional training programmes, particularly in the development of interactive digital learning materials, the application of artificial intelligence in education, 3D modelling, and the use of virtual and augmented reality in teaching.

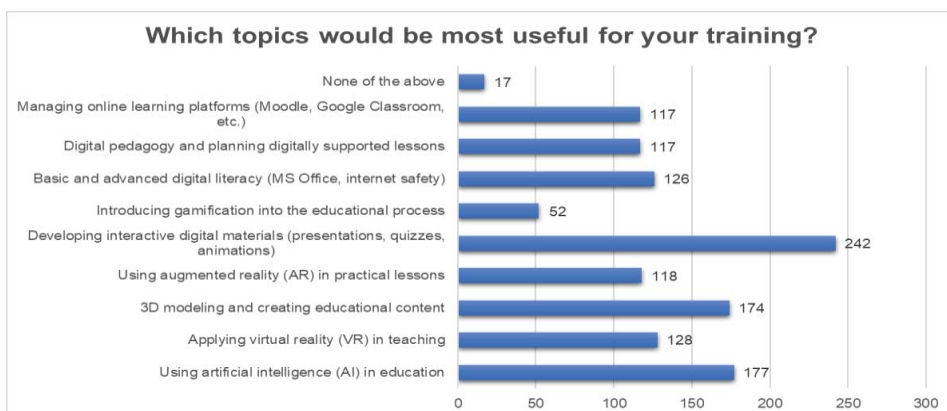


Figure 1. Chosen topics for training

Practical face-to-face workshops were identified as the preferred training format, although a substantial proportion of respondents favoured a blended learning approach combining online and in-person instruction.

The survey results show that vocational secondary school teachers in Bosnia and Herzegovina recognize the significant potential of digital technologies to enhance the educational process. The majority of respondents (138) believe that digital technologies will substantially improve the quality of teaching within the next five years, reflecting an optimistic attitude toward technological innovations in education. At the same time, a large number of teachers (135) view digital tools primarily as a valuable supplement to traditional teaching methods, while 99 respondents see the possibility of digital technologies replacing certain parts of conventional instruction. This diversity of opinions reflects openness to change as well as an awareness of the need for a balanced approach that combines the best of both worlds—digital and traditional.

3.2 Digital Readiness Of Students

A total of **1865 students** from vocational secondary schools across Bosnia and Herzegovina participated in the study. The largest proportion of respondents were first-year students (750), followed by second-year (573) and third-year students (396), while fourth-year students (146) accounted for the smallest share of the sample. Consequently, the findings primarily reflect the experiences and perceptions of students in the lower grades of vocational education.

The results indicate that digital technologies are widely present in the educational process. Approximately half of the students reported using digital technologies in school on a daily basis, while a considerable proportion stated that they use them several times a week or occasionally. Only a small number of respondents indicated that they never use digital technologies during school activities. These

findings suggest that digital tools are increasingly integrated into vocational education; however, their implementation is still not consistent across schools.

Artificial intelligence-based applications emerged as the most frequently used digital tools among students, with ChatGPT and similar AI platforms being reported more often than traditional learning management systems. YouTube, presentation software such as PowerPoint and Canva, and Google Classroom were also widely used. In contrast, advanced educational technologies, including virtual reality, 3D modelling software, and structured learning platforms such as Moodle, were used considerably less frequently, indicating that these technologies have not yet been systematically integrated into vocational education.

Most students assessed their digital competencies as intermediate, while a substantial proportion considered themselves advanced users capable of programming, video editing, or working with artificial intelligence tools. Nevertheless, a considerable number of students reported possessing only basic digital skills, suggesting that further development of digital competencies remains necessary.

The lack of adequate equipment and insufficient internet connectivity were identified as the two most significant barriers to the effective use of digital technologies in schools. Although only a relatively small proportion of students reported difficulties in using digital technologies, a notable number stated that digital technologies were rarely or never used in their schools. These findings indicate that infrastructural limitations continue to represent one of the main obstacles to the digital transformation of vocational education.

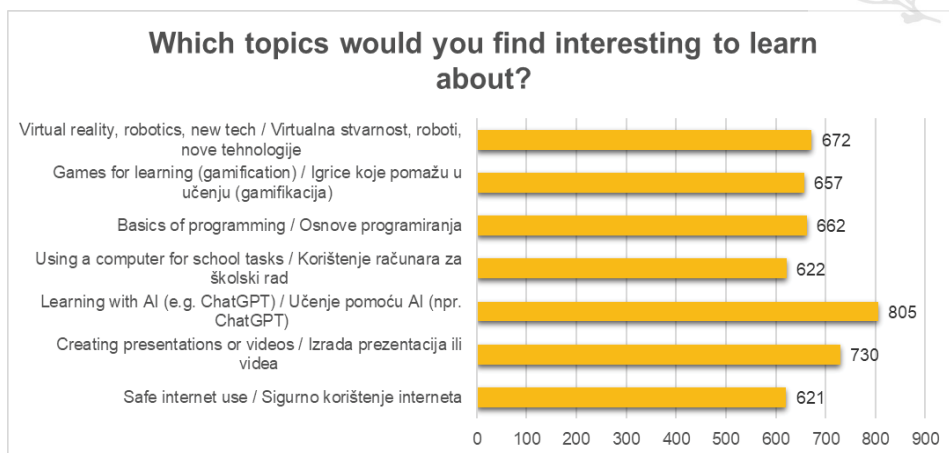


Figure 2. Interesting topics for students

The survey also revealed a strong interest in further developing digital competencies. More than four-fifths of the students expressed willingness to participate in additional digital skills training. The highest level of interest was

observed in topics related to artificial intelligence, content creation, virtual reality and robotics, basic programming, digital safety, and gamification. Regarding the preferred mode of learning, students favoured practical face-to-face workshops, followed by blended learning approaches combining online and classroom-based instruction.

In addition, students demonstrated a highly positive perception of digital technologies in education. Most respondents agreed that the use of digital tools improves the quality of learning and facilitates a better understanding of educational content. Moreover, they recognized the continuous development of digital competencies as an important prerequisite for successful learning and future employability.

3.3 Comparative Findings

The findings suggest that both teachers and students recognize the importance of digital technologies in vocational education. However, successful digital transformation requires not only investment in infrastructure but also continuous professional development for teachers and systematic integration of innovative digital tools into teaching practice.

Table 1. Comparison of key findings for teachers and students

Indicator	Teachers	Students
Use of digital technologies	Frequent, mainly for teaching preparation and delivery	Frequent, mainly for learning and completing school tasks
Self-assessed digital competence	Predominantly intermediate	Predominantly intermediate
Main barriers	Lack of equipment, limited time, need for training	Lack of equipment and internet access
Interest in further training	Very high (87%)	Very high (over 80%)
Priority topics	AI, digital content creation, VR/AR, 3D modelling	AI, programming, robotics, digital content creation

3.4 Recommendations

Based on the findings of this study, several recommendations can be proposed to support the digital transformation of vocational education in Bosnia and Herzegovina.

First, investment in digital infrastructure should be prioritized to ensure that vocational schools have adequate equipment, reliable internet connectivity, and access to modern educational technologies. Addressing these infrastructural challenges would enable more effective integration of digital tools into everyday teaching and learning activities.

Second, continuous professional development programmes should be provided for teachers, with particular emphasis on practical training in emerging technologies such as artificial intelligence, virtual and augmented reality, 3D modelling, and the development of interactive digital learning materials. Such programmes should focus not only on improving technical skills but also on strengthening teachers' pedagogical competencies for the effective integration of digital technologies into the classroom.

Third, digital competencies should be systematically embedded within vocational education curricula. Beyond basic digital literacy, students should be provided with opportunities to develop practical skills related to artificial intelligence, programming, robotics, data analysis, cybersecurity, and other Industry 4.0 technologies that are increasingly demanded in the labour market.

Finally, stronger collaboration between vocational schools, higher education institutions, industry partners, and policy makers is essential for creating a sustainable digital education ecosystem. Such cooperation would facilitate knowledge exchange, improve curriculum relevance, and ensure that vocational education responds effectively to ongoing technological and labour market changes.

Implementing these recommendations would contribute to strengthening digital readiness among both teachers and students while supporting the long-term digital transformation and competitiveness of vocational education in Bosnia and Herzegovina.

5. Conclusion

This study assessed the digital readiness of teachers and students in vocational education in Bosnia and Herzegovina. The findings indicate that both groups recognize the importance of digital technologies and demonstrate a strong willingness to further develop their digital competencies. However, challenges related to digital infrastructure and the systematic integration of advanced technologies into teaching remain significant. Strengthening digital competencies, improving school infrastructure, and providing continuous professional development for teachers are essential steps towards supporting the digital transformation of vocational education and preparing learners for the evolving demands of the digital economy.

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

References

- [1] World Economic Forum (2023). The Future of Jobs Report 2023.
https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf
- [2] CEDEFOP (2020). Vocational Education and Training in Europe.
https://www.cedefop.europa.eu/files/3083_en.pdf
- [3] European Commission. Digital Education Action Plan (2021–2027).
<https://education.ec.europa.eu/focus-topics/digital-education/action-plan>
- [4] European Commission, Joint Research Centre. European Framework for the Digital Competence of Educators (DigCompEdu).
https://joint-research-centre.ec.europa.eu/digcompedu_en
- [5] Ministarstvo civilnih poslova Bosne i Hercegovine. Prioriteti u integraciji poduzetničkih i digitalnih kompetencija u obrazovni sistem Bosne i Hercegovine 2019–2030.
<https://mcp.gov.ba/Content/Read/obrazovanje-dokumenti>
- [6] Digital Europe B&H
<https://mkt.gov.ba/hr/program-digitalna-europa>

- [7] Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- [8] European Commission (2024). Ethical Guidelines on the Use of AI in Education. <https://education.ec.europa.eu/focus-topics/digital-education/actions/plan/ethical-guidelines-for-educators-on-using-artificial-intelligence>



From Digital Maturity Assessment to Digital Transformation Strategy: Evidence from SMEs in Bosnia and Herzegovina

Nejra Merdović^{*1}, Madžida Hundur Hiyari¹, Adna Softić¹, Lejla Gurbeta Pokvić¹

Abstract: *Digital transformation has become one of the key prerequisites for increasing the competitiveness, resilience, and long-term sustainability of small and medium-sized enterprises (SMEs). While European industrial policies increasingly promote the transition towards Industry 5.0, many SMEs in developing economies continue to face significant barriers in adopting advanced digital technologies. This paper presents the results of twenty digital maturity assessments conducted in SMEs operating across different sectors in Bosnia and Herzegovina during the period 2024-2026. The assessments were performed using a structured digital transformation methodology consisting of a digital competitiveness questionnaire, expert workshops, digital maturity assessment, development of a digital transformation strategy, and preparation of a detailed priority digital transformation project. Rather than analysing individual enterprises, the paper identifies common patterns, recurring challenges, and strategic priorities shared among the participating companies. The results reveal that despite differences in industrial sectors, SMEs exhibit remarkably similar digital transformation needs, primarily related to process digitalization, data-driven decision making, systems integration, automation, and workforce digital competencies. The findings provide empirical evidence of the current level of digital readiness of Bosnian SMEs and discuss their implications for the gradual transition towards Industry 5.0.*

Keywords: *Digital transformation, SMEs, Industry 5.0, Digital maturity assessment*

1. Introduction

The rapid development of digital technologies has fundamentally transformed the way organizations create value, manage business processes, and compete in increasingly dynamic markets. Driven by globalization, technological innovation, changing customer expectations, and recent global disruptions such as the COVID-19 pandemic, digital transformation has evolved from a technological initiative into a strategic necessity for organizations seeking long-term competitiveness, resilience, innovation, and sustainable growth [1,2,3].

^{*1}Verlab Research Institute for Biomedical Engineering, Medical Devices and Artificial Intelligence, Sarajevo, Bosnia and Herzegovina
E-mail: nejra@verlabinstitute.com

Although the concepts of digitization, digitalization, and digital transformation are closely related, they represent different stages of organizational digital evolution. Digitization refers to the conversion of analogue information into digital form, while digitalization focuses on improving existing business processes through the application of digital technologies. Digital transformation represents the most comprehensive stage, involving fundamental organizational changes that integrate digital technologies into business models, organizational processes, decision-making, products, services, and organizational culture [1,2,4]. Consequently, successful digital transformation requires not only technological investments but also strategic leadership, organizational capabilities, and continuous development of digital competencies [2,3,5].

Recognizing its strategic importance, the European Union has positioned digital transformation at the centre of its industrial and innovation policies through initiatives such as the Digital Decade Policy Programme 2030 and the Industry 5.0 framework. Building upon Industry 4.0, which primarily emphasized automation, connectivity, and cyber-physical systems, Industry 5.0 promotes a human-centric, resilient, and sustainable approach to industrial development, where technological innovation is expected to generate both economic and societal value [6,7].

Small and medium-sized enterprises (SMEs), which account for approximately 99.8% of all European enterprises, play a central role in economic growth and innovation [8]. Nevertheless, the adoption of advanced digital technologies among SMEs remains considerably lower than in large enterprises due to financial constraints, limited digital capabilities, insufficient access to expertise, and organizational barriers [9]. These challenges are even more pronounced in developing and transition economies such as Bosnia and Herzegovina, where fragmented institutional structures, limited investment capacity, and weak collaboration between industry and research institutions continue to hinder digital transformation. National ICT indicators further demonstrate that the adoption of cloud computing, enterprise software, and other advanced digital technologies remains significantly below the European average [10,11].

As organizations increasingly transition towards Industry 5.0, assessing digital maturity has become an essential prerequisite for successful digital transformation. Recent studies emphasize that successful digital transformation in SMEs depends not only on technology adoption but also on organizational knowledge, digital capabilities, and continuous learning [12]. In this context, systematic digital maturity assessment provides a foundation for evidence-based strategic decision-making [13,14].

Despite the growing international body of literature on digital transformation, empirical evidence on the digital maturity and transformation priorities of SMEs in Bosnia and Herzegovina remains limited. Existing studies predominantly report national ICT statistics and technology adoption rates while providing relatively

little insight into organization-level challenges, strategic priorities, and implementation needs [10,11].

To address this research gap, this study analyses the results of twenty digital maturity assessments conducted in SMEs operating across multiple industrial sectors in Bosnia and Herzegovina between 2024 and 2026. Using a standardized assessment methodology that combines organizational questionnaires, expert workshops, digital maturity evaluation, strategic planning, and the identification of priority digital transformation projects, the study identifies common organizational challenges and digital transformation priorities while assessing the readiness of Bosnian SMEs for the transition towards Industry 5.0. The findings contribute empirical evidence that may support policymakers, Digital Innovation Hubs, researchers, and organizations involved in designing and implementing digital transformation initiatives for SMEs.

2. Methodology

This study is based on empirical data collected through digital maturity assessments conducted in twenty small and medium-sized enterprises (SMEs) operating in Bosnia and Herzegovina during the period 2024-2026. The participating companies represented various industrial sectors, including manufacturing, construction, food production, healthcare, trade, transportation, financial services, tourism, and other professional activities. To ensure confidentiality, all companies were anonymized, and the results are presented only in aggregated form.

A standardized digital transformation assessment methodology was applied to all participating enterprises. The process began with a Digital Competitiveness Questionnaire designed to collect information on company characteristics, business processes, digital infrastructure, data management, automation, digital competencies, cybersecurity, and the current use of digital technologies. The questionnaire served as the basis for a series of expert workshops conducted with company management and key employees, during which organizational processes, digital maturity, and transformation challenges were evaluated. The outcome of this phase was a Digital Maturity Assessment Report describing the company's current level of digital readiness and identifying key strengths, weaknesses, and opportunities for improvement.

Based on the assessment results, a customized Digital Transformation Strategy was prepared for each enterprise, defining strategic priorities, recommended technological solutions, implementation phases, and expected business benefits. Finally, one priority digital transformation project was developed for every company, including a detailed technical concept, implementation roadmap, and expected outcomes.

The findings from all twenty assessments were subsequently analysed using a comparative cross-case approach. Rather than evaluating individual companies, the analysis focused on identifying common digital transformation challenges, recurring technological priorities, and strategic development patterns shared across SMEs from different industrial sectors. This approach enabled the identification of general digital transformation needs of Bosnian SMEs and their readiness for the transition towards Industry 5.0.

3. Results and Discussion

The analysis of twenty digital maturity assessments conducted in SMEs from different industrial sectors revealed a high degree of similarity in digital transformation needs, regardless of company size or business activity. Although the participating enterprises differed in their products, services, and organizational structures, the identified challenges and strategic priorities were remarkably consistent, indicating that SMEs in Bosnia and Herzegovina face common barriers in their transition towards digitally enabled business models.

Although the participating SMEs operated in different industrial sectors, the priority digital transformation projects revealed highly consistent technological recommendations.

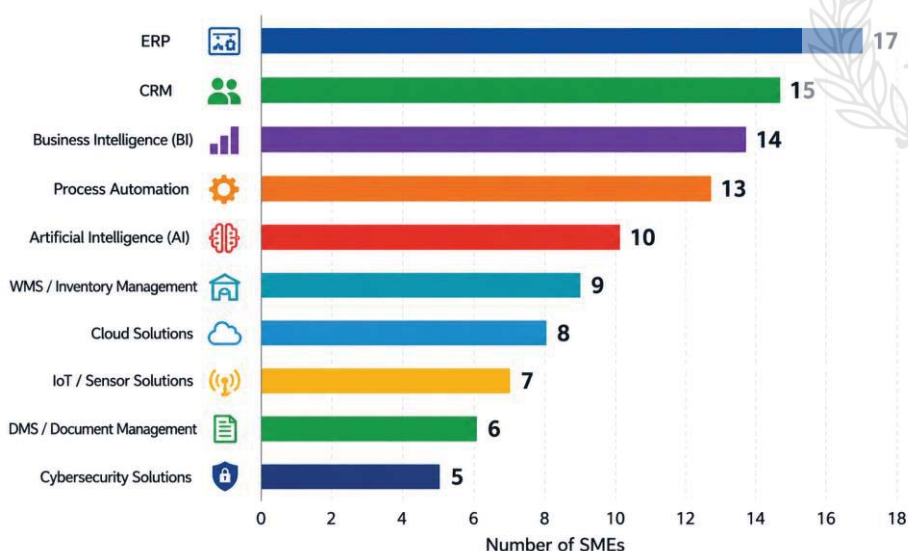


Figure 1. Top 10 recommended digital transformation priorities identified across the analysed SMEs.

As shown in Figure 1, enterprise resource planning (ERP), customer relationship management (CRM), business intelligence (BI), and process automation were the

most frequently recommended solutions. More advanced technologies, including artificial intelligence (AI), Internet of Things (IoT), and cybersecurity solutions, were generally proposed as medium- and long-term development priorities, reflecting the need to first establish fundamental digital infrastructure.

The distribution of recommended technologies confirms that most SMEs are currently focused on establishing integrated digital management systems and improving operational efficiency before adopting more advanced Industry 5.0 technologies.

The aggregated findings from the digital maturity assessments demonstrate a remarkable consistency in the challenges and development priorities identified across participating SMEs. As presented in Table 1, most companies exhibited low levels of process digitalization, fragmented information systems, limited use of business data, and a lack of formal digital transformation strategies. These recurring patterns indicate that digital transformation needs are largely independent of industry sector and organizational size, highlighting common priorities for future digital development.

Table 1. Summary of the most common findings from digital maturity assessments

Assessment area	Most common findings
Business processes	Predominantly manual or partially digitalized
Information systems	Fragmented, poorly integrated
Data management	Limited use of business data for decision-making
Customer management	CRM systems frequently missing
Production	Need for automation and process monitoring
Human resources	Need for digital skills development
Strategic planning	Digital transformation strategies generally absent
Technology priorities	ERP, CRM, BI, automation, AI

One of the most frequently observed challenges was the lack of integrated digital systems. In many companies, business information was distributed across multiple software solutions, spreadsheets, and paper-based documentation, resulting in duplicated activities, inefficient information exchange, and limited visibility of business performance. Consequently, companies identified systems integration and process digitalization as key prerequisites for improving operational efficiency.

Another recurring finding was the limited use of data for decision-making. Although companies generated considerable amounts of operational and business data, these were rarely analysed systematically. Business intelligence solutions, centralized reporting, and performance dashboards were therefore among the most frequently recommended strategic initiatives. Similarly, customer relationship management remained insufficiently digitalized, particularly in service-oriented companies, where customer information was often managed through spreadsheets or manual records.

For manufacturing companies, digital transformation priorities were primarily related to production automation, integration of production and warehouse systems, improved process monitoring, and optimization of production planning. In contrast, companies operating in service sectors more frequently prioritized CRM systems, digital communication, document management, and business process automation. Despite these sector-specific differences, the overall strategic direction remained similar across all participating enterprises.

Beyond technological aspects, the assessments consistently highlighted organizational challenges. Many companies reported insufficient digital competencies, limited employee training, and the absence of long-term digital transformation planning. These findings suggest that digital transformation should be viewed as an organizational change process rather than solely as the implementation of new technologies.

Based on the identified needs, each participating enterprise developed a Digital Transformation Strategy accompanied by one priority digital transformation project. Although individual technical solutions varied, the most common priorities included business process digitalization, implementation of integrated management systems, automation, business intelligence, customer relationship management, and improved data management. Advanced technologies such as artificial intelligence and the Internet of Things were generally recognised as long-term development objectives rather than immediate investment priorities.

Overall, the findings indicate that SMEs in Bosnia and Herzegovina are still building the digital foundations necessary for the transition towards Industry 5.0. Rather than focusing on advanced technologies alone, companies primarily require integrated information systems, digitalized business processes, stronger data-driven decision-making, and improved digital competencies. The similarity of identified priorities across different sectors demonstrates that standardized digital maturity assessments can provide valuable guidance for defining strategic digital transformation initiatives and supporting the long-term competitiveness of SMEs.

4. Conclusion

This study analysed twenty digital maturity assessments conducted in SMEs from different sectors in Bosnia and Herzegovina to identify common digital transformation challenges and strategic priorities. The findings indicate that, despite sectoral differences, SMEs share similar needs related to business process digitalization, systems integration, data management, automation, and digital skills development.

The results suggest that most SMEs are still building the digital foundations required for the transition towards Industry 5.0, with organizational readiness and integrated information systems representing higher priorities than the adoption of advanced technologies. The study provides empirical evidence on the current digital readiness of Bosnian SMEs and demonstrates that standardized digital maturity assessments can effectively support the development of digital transformation strategies and future policy and innovation initiatives. Future research should include a larger sample of enterprises and evaluate the long-term impact of implemented digital transformation projects.

References

- [1] Vial G. Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*. 2019;28(2):118–144. doi:10.1016/j.jsis.2019.01.003.
- [2] Verhoef PC, Broekhuizen T, Bart Y, Bhattacharya A, Dong JQ, Fabian N, et al. Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*. 2021;122:889–901. doi:10.1016/j.jbusres.2019.09.022.
- [3] Warner KSR, Wäger M. Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning*. 2019;52(3):326–349. doi:10.1016/j.lrp.2018.12.001.

- [4] Brennen JS, Kreiss D. Digitalization. In: Jensen KB, Craig RT, Pooley JD, Rothenbuhler EW, editors. *The International Encyclopedia of Communication Theory and Philosophy*. Hoboken (NJ): Wiley; 2016.
- [5] Matt C, Hess T, Benlian A. Digital transformation strategies. *Business & Information Systems Engineering*. 2015;57(5):339–343. doi:10.1007/s12599-015-0401-5.
- [6] European Parliament and Council of the European Union. Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030. *Official Journal of the European Union*. 2022;L323:4–26.
- [7] European Commission. *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry*. Luxembourg: Publications Office of the European Union; 2021.
- [8] Lagüera González J, Katsinis A, Di Bella L, Schulze Brock P, Odenthal L, Secades Casino B, et al. *Annual Report on European SMEs 2025/2026*. Luxembourg: Publications Office of the European Union; 2026.
- [9] OECD. *The Digital Transformation of SMEs*. Paris: OECD Publishing; 2021.
- [10] Agency for Statistics of Bosnia and Herzegovina. *Usage of Information and Communication Technologies in Enterprises, 2025: First Results*. Sarajevo: BHAS; 2025.
- [11] OECD. *SME Policy Index: Western Balkans and Türkiye 2026*. Paris: OECD Publishing; 2026.
- [12] Hafeez S, Shahzad K, Helo P, Mubarak MF. Knowledge management and SMEs' digital transformation: A systematic literature review and future research agenda. *Journal of Innovation & Knowledge*. 2025;10(3):100728. doi:10.1016/j.jik.2025.100728.
- [13] Schumacher A, Erol S, Sihn W. A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises. *Procedia CIRP*. 2016;52:161–166. doi:10.1016/j.procir.2016.07.040.
- [14] Gökalp E, Şener U, Eren PE. Development of an Assessment Model for Industry 4.0: Industry 4.0-MM. In: Mas A, Mesquida AL, O'Connor RV, Rout T, Dorling A, editors. *Software Process Improvement and Capability Determination. SPICE 2017. Communications in Computer and Information Science*. Vol. 770. Cham: Springer; 2017. p. 128–142. doi:10.1007/978-3-319-67383-7_10.

LIST OF AUTHORS

Agalliu Pjereta
University of Tirana, Faculty of Law
Tirana, Albania

Banjanović-Mehmedović Lejla
University of Tuzla,
Faculty of Electrical Engineering
Tuzla, Bosnia and Herzegovina

Bijedić Amila
University of Sarajevo,
School of Economics and Business
Sarajevo, Bosnia and Herzegovina

Bičo Ćar Mirha
University of Sarajevo,
School of Economics and Business
Sarajevo, Bosnia and Herzegovina

Bratovic Amra
University of Tuzla,
Faculty of Technology
Tuzla, Bosnia and Herzegovina

Dimitrijevic Sonja
"Mihajlo Pupin" Institute
Belgrade, Serbia

Džuho Amra
Verlab Research Institute for
Biomedical Engineering, Medical
Devices and Artificial Intelligence
Sarajevo, Bosnia and Herzegovina

Gurbeta Pokvić Lejla
Research Institute Verlab for
Biomedical Engineering, Medical
Devices and Artificial Intelligence
Sarajevo, Bosnia and Herzegovina

Hadžić Đulaga
University of Zenica
Zenica, Bosna i Herzegovina

Hrnjić Anes
University of Sarajevo,
School of Economics and Business
Sarajevo, Bosnia and Herzegovina

Hundur Hiyari Madžida
Verlab Research Institute for
Biomedical Engineering, Medical
Devices and Artificial Intelligence
Sarajevo, Bosnia and Herzegovina

Husak Ermin
University of Bihać, Technical
Faculty
Bihać, Bosnia and Herzegovina

Isić Safet
University „Džemal Bijedić” of
Mostar, Faculty of Mechanical
Engineering
Mostar, Bosnia and Herzegovina

Janković Dušan
Company Inmold Group
Požega, Serbia

Janković Goran
Company Inmold Group
Požega, Serbia

Joković Zora
Company Inmold Group
Požega, Serbia



Karabegović Isak
Academy of Sciences and Arts
of Bosnia and Herzegovina
Sarajevo, Bosnia and Herzegovina

Kepčija Pavlović Jovana
Verlab Research Institute for
Biomedical Engineering, Medical
Devices and Artificial Intelligence
Sarajevo, Bosnia and Herzegovina

Kriještorac Malik
Sarajevo School of Science and
Technology
Sarajevo, Bosnia and Herzegovina

Lalović Ana
Verlab Research Institute for
Biomedical Engineering, Medical
Devices and Artificial Intelligence
Sarajevo, Bosnia and Herzegovina

Lemeš Samir
University of Zenica, Polytechnic
Faculty
Zenica, Bosnia and Herzegovina

Lindov Osman
University of Sarajevo,
Faculty of Transport and
Communications
Sarajevo, Bosnia and Herzegovina

Mahmić Mehmed
University of Bihać, Technical
Faculty
Bihać, Bosnia and Herzegovina

Majstorović D. Vidosav
University of Belgrade
Faculty of Mechanical Engineering
Belgrade, Serbia

Malisa Viktorio
F-AR Förderung der Automation und
Robotik
Wien, Austria

Merdović Nejra
Verlab Research Institute for
Biomedical Engineering, Medical
Devices and Artificial Intelligence
Sarajevo, Bosnia and Herzegovina

Mezetović Nerma
University of Tuzla,
Faculty of Electrical Engineering
Tuzla, Bosnia and Herzegovina

Milošević Mijodrag
University of Novi Sad, Faculty of
Technical Sciences
Novi Sad, Serbia

Saletović Enes
University of Zenica
Zenica, Bosnia and Herzegovina

Softić Adna
Verlab Research Institute for
Biomedical Engineering, Medical
Devices and
Artificial Intelligence
Sarajevo, Bosnia and Herzegovina

Supurović Aleksandar
Company Inmold Group
Požega, Serbia

Šestić Munira
University of Sarajevo,
School of Economics and Business
Sarajevo, Bosnia and Herzegovina



Tufegdžić Milica
Academy of Professional Studies
Šumadija
Kragujevac, Serbia

Vojić Samir
University of Bihać, Technical
Faculty
Bihać, Bosnia and Herzegovina



