6

Artificial Intelligence Advancements for Digitising Industry

Ovidiu Vermesan[1], and Reiner John[2]

[1]SINTEF AS, Norway
[2]AVL List GmbH, Austria

Abstract

In the digital transformation era, when flexibility and know-how in manufacturing complex products become a critical competitive advantage, artificial intelligence (AI) is one of the technologies driving the digital transformation of industry and industrial products. These products with high complexity based on multi-dimensional requirements need flexible and adaptive manufacturing lines and novel components, e.g., dedicated CPUs, GPUs, FPGAs, TPUs and neuromorphic architectures that support AI operations at the edge with reliable sensors and specialised AI capabilities.

The change towards AI-driven applications in industrial sectors enables new innovative industrial and manufacturing models. New process management approaches appear and become part of the core competence in the organizations and the network of manufacturing sites.

In this context, bringing AI from the cloud to the edge and promoting the silicon-born AI components by advancing Moore's law and accelerating edge processing adoption in different industries through reference implementations becomes a priority for digitising industry.

This article gives an overview of the ECSEL AI4DI project that aims to apply at the edge AI-based technologies, methods, algorithms, and integration with Industrial Internet of Things (IIoT) and robotics to enhance industrial processes based on repetitive tasks, focusing on replacing process identification and validation methods with intelligent technologies across

automotive, semiconductor, machinery, food and beverage, and transportation industries.

Keywords: Artificial intelligence, edge computing, industrial internet of things, deep learning, hardware, silicon technologies, silicon-born AI, computer vision.

6.1 AI at the Edge in Industrial Processes

With more increased computing power, intelligent sensors and IIoT devices can collect large volumes of data these devices generate, reason over that data, and turn it into knowledge. AI can process this data closer to where it is produced and getting distributed to the edge. Multi-parameter sensing IIoT devices, AI everywhere, and serverless computing drive this new intelligent edge era.

Defining "intelligence" in the AI context requires a careful approach since different choices lead the research in different directions. The current field of AI is a mixture of multiple research fields, each with its own goal, methods, practical applications, etc. They are all called "AI" mainly for historical rather than theoretical reasons. Many AI definitions are provided in the literature (published papers, articles, books, research studies) to reflect the activities of research fields that the definitions mirror [2], [3]. However, the definitions of AI systems are too vague and broad, requires further clarification.

As a starting point, AI in the context of the AI4DI project was defined as a machine's ability to collect information, perform logical analysis, acquire/produce knowledge, and adjust to an environment that varies over time or in a given context [1], [5]. These abilities include the collective attributes of a machine (e.g., computer, robot, intelligent IIoT device, etc.) capable of performing functions such as learning, decision making, or other intelligent functions and tasks that mimic human behaviours [5].

The manufacturing industry is in transition, driven by Industry 5.0 concepts that transform the entire manufacturing value chain through a technology-driven change in capabilities and expectations. It is not simply about substituting people with machines, but instead about how people, interconnected sensors, machines, IIoT devices, distributed ledger technologies (DLTs), digital platforms, and AI can work together more effectively, using fewer resources and minimising the carbon dioxide footprint. Technological advancement drives manufacturers to increase productivity, efficiency, growth, deliver quality products, satisfy customers, and achieve higher profitability and sustainability. The Industry 5.0 AI-based driven processes change the tasks execution and impact manufacturing at the individual operation level. These digital-driven capabilities advance manufacturing across industries, value chains and value networks. That means that to remain competitive, manufacturers must adopt new AI technologies and integrate them into the manufacturing processes.

AI becomes a critical element to advanced manufacturing, product life cycle management and enterprise asset management.

Despite its potential, AI has several drawbacks that prevent the full exploitation of AI-based technologies. A few of these drawbacks are listed below:

- **Insufficient reliability and robustness of AI systems -** despite the numerous relevant technological advances, AI systems are still associated with low reliability, reflected in their relatively low penetration and utilisation.
- **High complexity -** the complexity of AI tasks has increased steadily to address new paradigms for automating, conceptualising, designing, and implementing such AI-based systems that include sensors, hardware, software, models, and algorithms.
- **High costs -** The development, implementation and deployment of AI-based solutions require vast investments as AI-based systems are unusually complex. Their repair and maintenance require significant effort. The AI systems call for frequent upgrades to meet to the changing environment's needs and make machines more "intelligent" day by day. In severe breakdowns, the procedure to recover lost codes and re-instating the AI-based system might require enormous time and cost.
- **Energy consumption -** AI models consume a relatively extensive amount of energy, and these energy requirements are increasing as AI technology is deployed in different industrial applications. Using deep learning (DL) algorithms, the computational resources needed to produce performant AI models increase significantly every year. In this context, AI has a significant carbon footprint, and if industry trends proceed, it will soon become considerably more severe.
- **Training data shortage -** AI-based models require large amounts of data, and their performance relies heavily on the size of training data sets available. For most industrial sectors, it is not easy to create training datasets that are large enough and include information that allow different industrial stakeholders to use the same data sets for

benchmarking similar AI models. These data sets exhibit tremendous potential for optimising industrial processes in cases where traditional approaches, like stochastics, analytical or numerical models, can no longer be used.

- **Absence of improvement with experience -** technical drawbacks related to the lack of progress with experience of the AI systems is a challenge. AI-based systems store a large amount of data. The data can be accessed and used differently by human and machine intelligence. Machines today can still not alter their responses to changing environments without being re-trained or updated/upgraded.
- **Lack of original creativity -** while AI systems can help humans create, they do not match yet the power of thinking that the human brain has or the originality of a creative mind.

The AI4DI project addresses several drawbacks mentioned above and provides solutions that enable AI to optimise industrial processes, energy efficiency, and processing at the edge.

6.2 A pan-European AI Framework for Manufacturing and Process Technology

The purpose of AI4DI is to benefit from recent research in the fields of semiconductor, intelligent systems, process control, IIoT, connectivity and edge computing to extend the potential of state-of-the-art AI technology for industrial applications to address the current main challenges in the industry:

- **Flexibility**: Factories and processes need to adapt to dynamic demand and compensate for failures quickly. AI can support and automate processes, planning, decisions, and system optimisation during all design phases.
- **Complexity**: Design processes, supply chains, manufacturing sites and the final products become increasingly complex. Managing this complexity can no longer be handled by humans. AI can reduce this complexity by generating simplified representations and even automating control. Importantly, this also includes the complexity of the AI system itself, which means that humans must interpret and understand its decisions and actions.
- **Process locally and think holistic**: Strong requirements for low latency, high reliability, and data privacy in industrial applications preclude outsourcing AI to existing cloud services. To successfully adopt AI in

the industry, AI must be deployed at the edge to support distributed on-site data processing with state-of-the-art AI components, algorithms, techniques, and methods.

- **Transparency**: The advance of digitisation will yield massive amounts of heterogeneous and possible unstructured data from many different sources. AI-based analysis and synthesis methods will be mandatory to analyse these large data sets and make them transparent for human operators and decision-makers.

Current AI tools are optimised for cloud services and therefore do not always fulfil the robust requirements of production environments (real-time capability, safety, reliability, guaranteed service quality, etc.). The AI4DI demonstrators provide and adopt AI methods and tools suited for moving the intelligence and analytics at the edge by addressing the following areas:

- **Hardware**: Develop AI-based microcontrollers, embedded systems and IIoT devices designed for industrial environments and edge processing. AI systems in the industry are implemented using a decentralised architecture with AI components distributed over a heterogeneous set of devices aligned with the software infrastructure.
- **Software and Libraries**: Adapt the existing software and libraries and develop new ones to address the critical requirements for AI in the industry regarding safety and reliability. As these requirements are not reflected in current AI tools, the project partners are extending the features of the existing tools, including appropriate workflows for testing and validation.
- **Algorithms**: Existing AI algorithms, in most cases, run on high-performance hardware. The AI4DI is advancing the optimisation of AI algorithms and models at the edge for running on IIoT devices and embedded systems with limited resources in terms of processing power, connectivity, and energy sources. The new optimisation methods and techniques for reducing the computational requirements are applied to neural networks (e.g., reducing the number of layers, adaptation to less precision) and quantify the impact of these optimisations on performance. New AI methods are investigated to improve the real-time operation capabilities and online learning.
- **Data**: Data and its quality play a critical role in AI industrial applications both for learning/training and testing the AI-based models and algorithms. In many manufacturing facilities, data is extremely sensitive and expensive to collect. Consequently, there is a strong need for an

explanation that can describe the needed information for various AI methods and techniques. Automated data anonymisation methods are required to allow data sharing for training without exposing confidential information.

AI4DI's mission is bringing AI from the cloud to the edge and making Europe a leader in silicon-born AI by advancing Moore's law and accelerating edge processing adoption in different industries through reference demonstrators [1]. The project focuses on five objectives to achieve its mission, as illustrated in Figure 6.1 and listed below.

1. Develop AI applications to be demonstrated under conditions as close as possible to real-life.
2. Formulate roadmaps, exploitation studies, business cases for AI technologies applications in industrial environments.
3. Provide a deployment plan showing how to develop and valorise AI technology in industrial sectors.
4. Establish an AI community in Europe, which is complementary to other initiatives.
5. Build and sustain dynamic AI technology ecosystems in Europe, ensuring ethical, responsible, and trusted AI for safety-critical real-time applications.

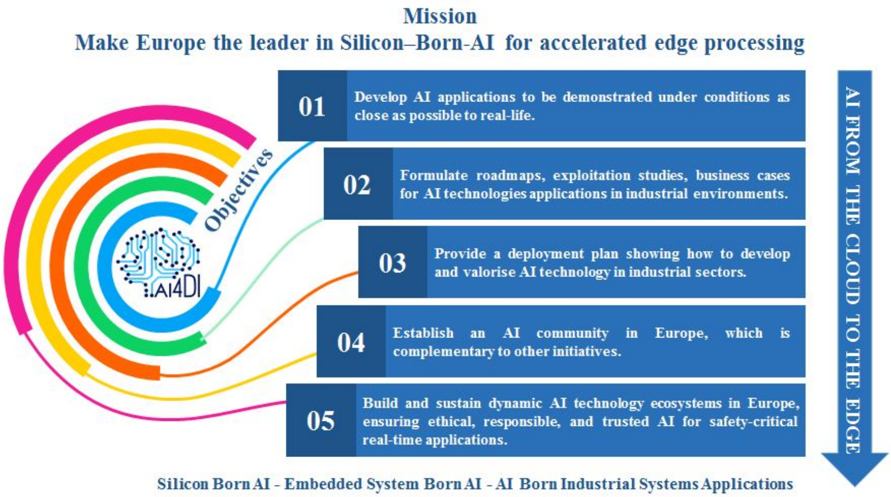

Figure 6.1 AI4DI Objectives.

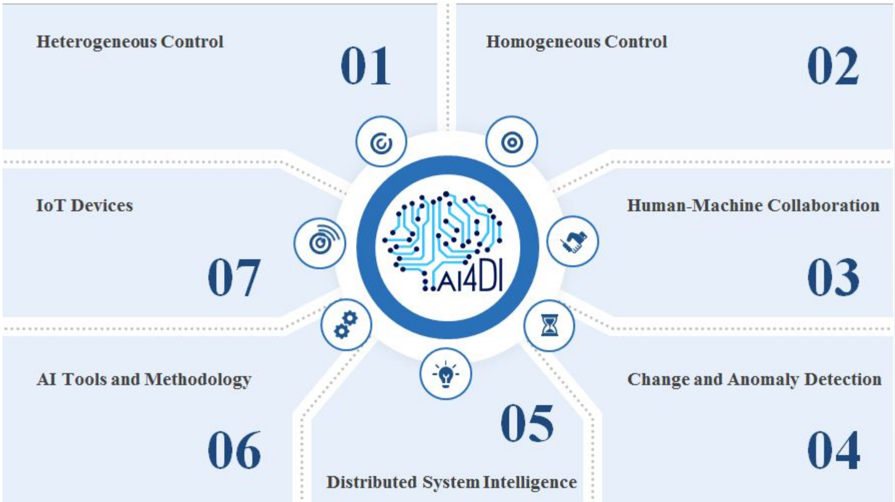

Figure 6.2 AI4DI Key Targets.

The AI4DI reproducible approach for implementing AI methods in the industrial sectors includes a structure that comprises seven key targets (KTs), as illustrated in Figure 6.2.

Each key target is a generic element representing a common characteristic of different processes, collaborations, methods etc., that is necessary for the successful cross-industry implementation of AI methods. A short description of each key target is provided in the following paragraphs.

Heterogeneous control - addresses the development and implementation of AI methods for heterogeneous control in manufacturing facilities. The main feature of heterogeneous production is assembling differentiable products from individual parts in a discrete manufacturing process with many different sequential steps. The exact sequence may vary from product to product and strongly depends on customer needs. Examples include the manufacturing of vehicles, aircraft, computers, and furniture. Various production steps make heterogeneous manufacturing highly complex and require human experts to identify and trace back issues. In this context, AI methods can collect knowledge, help to make transparent decisions based on current and saved data, and enable the optimisation of single production steps on the global level. One main task in the control of heterogeneous manufacturing processes is the scheduling of materials and production facilities. Production planning methods like material requirements planning rely heavily on correct data about the manufacturing facility, production history, predicted future required demands etc. AI methods have recently proven a considerable capability in generative modelling and can considerably contribute to more efficient production planning by providing more reliable external and internal factors models in the planning process.

AI can provide tools and methods for continuous quality monitoring at every process step and using the collected data to optimise the global control of the manufacturing process. Applications include automatic visual inspection with neural networks, pattern detection in measurement data, and proactive control based on predictive modelling. With distributed AI, devices already operating in the field can provide feedback directly available to the production process. A unique advantage of using AI to model and store data about the production process is that knowledge can be directly shared between the different plants. Introducing AI to control heterogeneous production processes become a fundamental prerequisite for managing increasingly complex products and growing product diversity (e.g., lot size one production in Industry 5.0). AI methods can model processes based on ontologies and knowledge databases. This information can be used to interpret data from the plant in an abstract, condensed, and semantically meaningful way, proactively recommending actions to the operator.

Homogeneous control - addresses the control of homogeneous production processes that manufactures products by combining raw materials from different supplies in a continuous and non-interruptible process. Typical examples of produced goods include pharmaceuticals, plastics, liquids, wine, and food. The various steps of homogeneous production processes are not isolated and usually irreversible. Control is therefore highly time-critical and directly related to the quality of the product. Very often, homogeneous processes are part of a larger heterogeneous manufacturing environment. The control of homogeneous production processes is based on continuous control loops often implemented as proportional–integral–derivative (PID) controllers. The parameters of these loops are directly related to the process output and the quality of the end product. Setting these parameters cannot be done locally only since changes at one stage of the system can affect all later stages due to the continuous nature of the process.

Moreover, specific sets of parameters can increase the overall robustness of the system at the global level even though local control loops might not be operating at the optimum level. The identification and dynamic application of these parameter sets require collecting and evaluating system data collected at all stages of the manufacturing process. Analysing this data, correlating them to favourable production parameters, and dynamically generating appropriate control parameters is a high-dimensional optimisation problem that is hard to track with traditional methods. Machine learning-based AI methods have achieved increased performance in control tasks but have not been applied in process control yet. Adopting these methods in industrial environments is highly promising and requires establishing standards to quantify important aspects such as reliability and real-time capability.

In large manufacturing facilities, homogeneous processes are typically observed in dedicated control rooms where experts have access to all sensors and actors in the system. Operating the control level requires a high level of expertise and concentration. Pre-processing and interpreting this data with domain-aware AI methods enable adding semantic information to raw measurements and identifying system states at an abstract level. Process control equipment in these facilities has specific safety and real-time requirements, and there is a strong need to move the required AI methods from the cloud to the edge.

Human-Machine collaboration - includes how AI methods can empower and enhance human-machine collaboration, including new ways of interaction, ethics, and value standards. The human-machine interactions in industrial environments can apply AI-based techniques locally at the edge close to industrial machinery by using latency communication and distributing intelligence between humans and machines on the industrial manufacturing floor. The collaboration approach is comparable to how an ants colony works, with each ant applying its separate intelligence towards the suitable solution of a common task. Examples of human-machine collaboration include spatiotemporal semantic awareness for enhanced worker safety, human-machine natural interaction for better design-to-manufacturing information transfer, correct instruction for acceptance of AI, and high-level understanding of continuous logistics flow on-premises and global level. The spatiotemporal semantic awareness for enhanced worker safety comprises cases, where machines are capable of a sophisticated semantic understanding of the environments and can move industrial robots to reduced risk towards the human worker, with less compromise on overall productivity. AI-based edge perception and activity recognition technologies applied at the level of the edge and IoT devices can be used to automatically understand the spatiotemporal relative position between workers and machines operating under his/her supervision, thus identifying potentially dangerous situations with ultra-low latency and much higher precision. The human-machine natural interaction for better design-to-manufacturing information transfer consists of on the edge node speech recognition AI-based solutions. Natural language processing provides a way to naturally interact with machines using voice-based commands even in very noisy environments. More advanced and friendly AI-based interfaces can improve human-machine collaboration on the manufacturing floor. Correct instruction for acceptance of AI includes advanced human-machine interfaces enabled by AI to deal with human knowledge. The correct instruction brings issues related to the ethical dimension, e.g., workers might be worried about the security and privacy aspects of interacting with an "aware" machine. Correct and transparent instructions on how to interact with AI-augmented devices encourage acceptance in the industry. A high-level understanding of continuous logistics flow on-premises, and global level relates to the capability to extract information at all manufacturing stages and understand complex logistic relationships to enhancing manufacturing facility-level productivity. AI performed at the edge provides a clear pathway in this direction, as too-abundant raw data can be locally characterised, inferred upon using AI techniques (ML, DL), summarised in a high information density format, and then immediately used to act accordingly. For industrial environments characterised by complex integrated value chains often spanning many geographically distant manufacturing facilities and industrial plants, utilising data analytics collected directly at the manufacturing level to smooth out logistics provides a significant productivity improvement.

Change and anomaly detection - addresses all methods and tools required to continuously monitor and analyse both heterogeneous and homogeneous production processes, both in real-time at runtime and offline after data collection. Tasks such as failure detection and quality control require in-depth domain knowledge and are usually limited to covering a predefined set of problem cases. Complex issues that depend on the specific production environment or a particular load profile are hard to detect and cannot be modelled in advance. More importantly, many changes in a production environment are not limited to a single stage in the production process but accumulate or spread over multiple phases. Detecting such changes becomes increasingly hard in complex production processes and can no longer be done by humans in dynamic reconfigurable factory environments of Industry 5.0. AI and ML are key factors for managing this complexity, and topics such as diagnostics and predictive maintenance, security and anomaly detection, distributed ledger are addressed. Diagnostics and predictive maintenance deals with the detection of failure or wear and tear on a single machine.

Usually, this task is carried out during maintenance or by a skilled worker operating the machine. Using time series analysis and prediction methods enables continuous monitoring based on either internal data recorded by the equipment or externally accessible state information that humans also use (e.g., vibration, sound, temperature sensing, image, video). While change detection can be computed on embedded devices directly at the machine, the actual training of the required models will, in most cases, rely on data analysis performed in the cloud. The results are transferred to edge devices that perform inferencing. Security and anomaly detection relates to the detection of intrusion and malicious system behaviour. In the ongoing digitising of production systems, more and more devices in the production process communicate via the network. The network connectivity is essential for many IIoT and AI-based tools and makes the system vulnerable to threats. Security is a significant issue in production systems and an important application for anomaly detection methods that detect deviations from a processing equipment's expected behaviour or output. For distributed systems used in future Industry 5.0 production environments, anomaly detection at the whole system level becomes particularly relevant since a single issue can propagate throughout the entire network. Distributed ledger offers the possibility of implementing a mechanism that guarantees the overall consistency of a distributed production system. This is particularly important for small lot sizes where the production process can differ considerably based on customer requirements. Implementing a distributed ledger enable guaranteed traceability of all process steps and is indispensable for safety- critical products. It also can increase the system's security since it allows for the direct detection of illegal operations. All mechanisms for change detection contribute to the awareness of the system and enable the detection of inconsistent, inefficient, and unsafe states. The detection performance heavily depends on the data available. Interfaces and exposed state data usually differ for every single machine, which makes the integration highly challenging. AI-based methods are a crucial contribution since they enable the processing of natural input data from IIoT devices that human experts also use to detect failure, wear out and other anomalies.

Distributed system intelligence - addresses the infrastructure on which the AI methods are implemented. Its focus is distributed intelligence systems, consisting of several subsystems taking each decision, with no need for central intelligence. As such, it constitutes one of the fundamental infrastructural blocks necessary for deploying the other key targets. The principal driver for the deployment of AI in complex industrial settings is collecting information, analysing it, and reacting immediately. Various pieces of machinery operate in concert towards the manufacturing of a single object or device, making it unfeasible to centralise intelligence in a unique central driver, as in the classical cloud-based AI model. Instead, like in an ant colony, a more effective strategy is distributing intelligence between decentralised actors located on the field, at the industrial machinery level. AI embedded in the edge devices and directly in the many IIoT sensors deployed inside the industrial equipment can progressively build hierarchical, advanced representations of the raw data collected. Edge and IIoT devices augmented with capabilities to perform AI-based functions for vision, perception, sensor fusion etc., can collect information, distil it in a high information density format, and then transfer it to a more performant on-premises edge processing capable of using this information for building a context awareness and acting accordingly to it. The edge processing unit might then be reporting to a centralised broker for data collection or a cloud-based AI service. However, the main functionality of the system is deployed locally, at the edge, enabling advanced low-latency local behaviour such as human-machine cooperation and change detection.

AI tools and methodology - addresses the development of tools and methods required for implementing AI in production systems. The focus is on establishing a toolchain that moves AI services from the cloud to the edge and applies them to safety-critical domains with real-time solid capability and reliability requirements. This toolchain includes the technical implementation and a principled methodology that allows system developers and integrators to apply AI in their specific domain. The fast progress of AI provides tools that helped speed up research and development by standardisation and state-of-the-art open-source methods to a broad community. Software tools like Google TensorFlow have set a standard that enables simple code sharing and quick reproduction of results. The development is fuelled by openly accessible project repositories such as GitHub, free databases with training data such as ImageNet, and dynamic training environments such as OpenAI Gym. Current AI tools are optimised for research or cloud services and therefore do not fulfil the robust requirements of production environments (real-time capability, safety, reliability, guaranteed service quality, etc.). Adopting AI methods and moving them from cloud to edge therefore addresses the areas such as hardware, software/libraries, data, and algorithms. The AI-based industrial equipment requires hardware that addresses the factory floor's needs (e.g., high-performance, energy efficiency, reliability, real-time capabilities etc.) and can be integrated into IIoT devices. AI systems in the industry are implemented in decentralised architecture and distributed over a heterogeneous set of devices, which define the requirements for the software infrastructure. Software and libraries are needed to address the safety and reliability requirements for AI-based applications in the industry. These requirements are not reflected in current AI tools, and deep analysis is required to identify and add missing features, including appropriate workflows for testing and validation. Data from manufacturing facilities is critical for the training/learning of AI-based models and algorithms.

In many cases, the industrial data is classified and costly to collect. New AI methods and techniques for data anonymisation are needed that allow data sharing for training without exposing confidential information. AI algorithms and run in many applications on high-performance hardware. To enable analytics and processing at the edge, optimisations of these algorithms are needed to run at the edge on resource-constrained IIoT devices without compromising the performance. This also applies to the investigation of AI methods capable of real-time operation and online learning. Introducing AI methods requires a structured methodology for selecting appropriate AI tools for a given task and integrating them following the requirements of the domain [4].

IoT Devices - considers the hardware/software aspects related to the practical implementation of the other key targets on IIoT devices deployed in industrial machinery. This includes components for sensing, actuating, connectivity, and end node IoT processing.

In the industrial scenario, the primary constraint about IIoT sensor/smart devices is their physical footprint. Sensor devices must be placed directly within parts of moving machinery, putting severe limitations on their size and capability to be wired. A second constraint is that these devices need long battery life and require "zero" maintenance, ideally outliving the piece of machinery mounted on with "zero" human intervention. Fulfilling both constraints require sophisticated and state-of-the-art IIoT device design. Traditional architectures for these end node devices leverage small-scale microcontrollers (e.g., ARM Cortex-M0 class) to minimise the compute power envelope, coupled with button batteries to ensure long life and minimal footprint. End nodes of this kind can collect sensory information, send it to a central server via wireless communication and go back to sleep.

Implementing distributed intelligence and deploying advanced AI-based functions directly on the IIoT devices requires much higher performance available on the device, which makes fitting within the constraints discussed above much more difficult. Therefore, devices designed within the project must use advanced high-performance, AI-dedicated/capable microcontrollers with much better power efficiency techniques. The wireless capabilities need to include energy-efficient communication and exploit an overall more advanced architecture, integrating multiple sensors with embedded in-plane analytics and dedicated hardware architectures for AI and inference.

6.3 AI Technologies

The success of industry-grade embedded AI on edge devices directly depends on the availability of dedicated central processing units (CPUs), graphics processing units (GPUs), and hardware accelerators architectures for electronics components and systems that are powerful enough to fulfil the full potential of state-of-the-art AI methods.

By extending Moore's law, as illustrated in Figure 6.3, silicon-born AI takes full use of advances in semiconductor technology and even improve

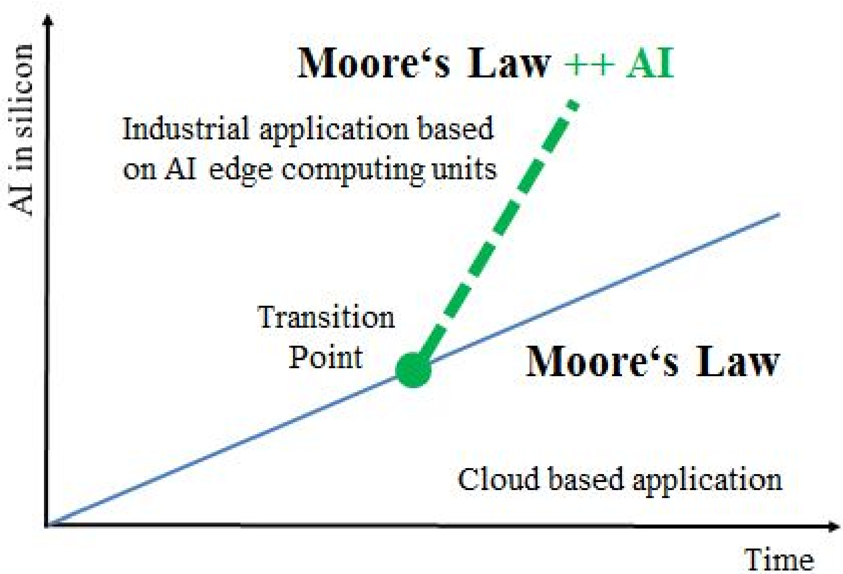

Figure 6.3 Silicon-born AI effect on Moore's Law beyond the current silicon technology developments.

the performances by leveraging additional scaling effects of “More Moore” and “More than Moore”. The adoption of AI technologies in the industry can be substantially accelerated by making the required compute power available at the edge and thereby enable completely new AI applications that are not available for industrial applications yet.

In the short term, AI is implemented on readily available AI capable CPUs that include essential support for accelerating AI models (e.g., extension for vector operations). This is realised by model compression techniques that can reduce parameters in neural networks by factors of up to 50.

Dedicated AI methods can be implemented purely in software to enable industry-grade AI for data processing at the edge. Industry-grade AI must comply with the robust requirements of industrial applications and can therefore not be implemented by applying only the existing AI algorithms to industrial tasks.

AI techniques and methods are optimised to execute on AI-based hardware architectures designed for industry and featuring enhanced connectivity capabilities for fast communication in distributed networks of embedded IIoT devices.

However, growing amounts of data and new AI methods require larger models and more AI performance support. In the mid-term, “More Moore” will enable the design of AI-enhanced processing units that handle more data and larger models at lower response times.

Dedicated silicon-born AI hardware components, design languages, application generators, design automation tools, and respective standardisation can address AI features directly in the chip design to leverage performance speedup through the advancement of Moore’s law.

“More than Moore” technologies allow heterogeneous functional processing units on a single chip in the long term. This makes novel neuromorphic processing units (NPUs) tailored to the execution of large- scale neural models in real-time with maximum power efficiency available for industrial applications. Neuromorphic computing can also support future brain-inspired AI technologies.

More dense system integration enabled by “More Moore” technologies increases the AI algorithms performance, and more heterogeneous integration enabled by “More than Moore” technologies increases the AI functionalities. Both enlarge the potential for industrial application of AI even further.

The silicon-born AI maximises the benefits of Moore’s law and revives it beyond the current semiconductor/silicon technologies while enabling native AI computing and the native embedding of AI algorithms directly in silicon.

Through this approach, progress in semiconductor technology automatically translates to better performance for AI applications running on embedded edge computing. "More than Moore" technologies additionally address the integration of AI-specific computational units and sensor/actuator devices on a single chip and thereby also accelerate the speedup of silicon-born AI for industrial applications.

Implementing a roadmap that builds on silicon-born AI supports and accelerates the adoption of AI by European's industry to address its most urgent priorities in digitisation, such as mastering complexity, increasing flexibility, maximising efficiency by moving the intelligence to the edge and providing new distributed reference architectures [6] that are aligned with the industrial requirements. The demand for high AI performance is fuelled by technological (e.g., intelligent sensors/IIoT devices generating more useful data) and industrial factors (e.g., moving from linear to network processes).

The new industry-grade AI methods require to be tailored to the European industry's specific needs, and the development of AI-based hardware keep up with the growing demand for AI through the advancement of Moore's law.

6.4 AI Application Areas

Various industrial sectors are currently experiencing the most radical changes since assembly lines and the rise of mass production. Product complexity is continually increasing while customers simultaneously demand individually configured and manufactured products.

Many industrial AI systems are built around a centralised paradigm where machine learning solutions are delivered as a part of cloud-based APIs and software packages deployed on remote servers of AI providers. The future requires a paradigm shift by moving toward decentralised and distributed AI that can run and train at the edge on local intelligent devices in industrial applications or make decisions in decentralised networks like blockchain. The transition to decentralised and distributed AI is enabled by new technologies that allow for crowd-training of ML algorithms, device-centred AI that runs and trains ML models on mobile IIoT devices, and AI in decentralised autonomous organisations on heterogeneous networks.

Intelligence on an edge device allows it to process information locally and respond quickly to situations instead of communicating with a central cloud or server. For instance, an autonomous AI system must respond in real-time to what's happening on the production line. Decisions are time-sensitive, and latency is critical for many mission-critical industrial processes.

The AI4DI provides AI-based technologies at the edge for digitising the industry by reducing costs, save time, optimising/improving processes/products/services, increasing quality by enhancing industrial processes, and built and sustain a dynamic AI technology ecosystem in Europe.

The project develops IIoT technologies, AI-based hardware, software, models, and algorithms to enhance processes based on repetitive tasks, focusing on replacing process identification and validation methods with intelligent technologies across automotive, semiconductor, machinery, food/beverage, and transportation industries.

The following sub-sections provide an overview of the topics covered in the five industrial sectors. The different use cases are presented are presented at different levels of detail in [7].

6.4.1 Automotive

Digitisation is an essential prerequisite for tracing the production process along the entire supply chain and enabling future innovations in the automotive manufacturing industry. Growing data and new non-linear manufacturing paradigms yield massive data sets that humans can no longer interpret. AI, therefore, becomes an essential tool for processing this data. It will accelerate the automotive industry also have a significant impact on automotive companies' finance and control. Maximum data transparency is essential for AI-enabled analysis, optimisation of automotive production processes and supply chains. When the required data is available in real-time, the potential of AI methods such as DL, ML, expert systems and distributed autonomous agents is enormous. So far, the planning and operation of automobile production processes still require human planning and could be made more responsive and automated with AI.

Improving the responsiveness and automation using AI-based technologies covers the complex logistic processes across deep supplier networks with the potential of optimising the complete supply chain, including prediction of future system states or even autonomous control.

AI4DI addresses the AI-based technologies and applications for optimising logistic processes to reduce transport costs and the environmental impact.

The AI-based technologies and applications in the automotive industry cover two main areas, AI-supported automotive manufacturing and logistics and real-time predictive maintenance. A list of the demonstrators implemented under each application areas is presented below:

Inbound logistic process optimisation

- *Inbound logistics process optimisation* - addresses the systemic analysis and decision-making for responding to critical supply chains. Different data streams are injected into the AI core to react to disruptions as quickly as possible with suitable measures.
- *Assembly process optimisation* - based on computer vision systems and deep learning methods, ensures correct installation, and enables an ergonomic evaluation of the workers' activities.
- *Autonomous reconfigurable battery system* - aims to combine various retired batteries with very heterogeneous performance characteristics within one battery system. For this purpose, it is essential to accurately determine the state parameters, like the state of health and state of charge (SoC) of each single battery cell during the operation.
- *Virtual AI training platform for robot learning* - uses reinforcement learning (RL) to address the challenge of bringing autonomy in industrial robotic manipulation. Advanced simulations are used to virtually train the policy network by providing multitudes of realistic synthetic data.
- *Bluetooth low energy (BLE) localisation in asset tracking* - focused on indoor asset tracking based on Bluetooth wireless technology. The functionality of low-cost and BLE based components are enhanced by AI technology designed to analyse the non-deterministic signal received from tracking tags.
- *Autonomous mobile robotic agent* – addresses a multi-purpose robotic platform for indoor use intended for autonomous transportation of the factory's material, goods, or tools. AI algorithms deliver autonomous and cooperative behaviour even in complex environments, and AI trajectory planning manage distributed intelligent traffic control ensuring fast and reliable delivery within the factory.

Real-time predictive maintenance

- *Predictive health-monitoring system for machines on the level of a digital twin* – addresses a combination of AI methods and mathematical damage models connected to the operation of an e-motor unit for real-time failure prediction and diagnostics. To evaluate and monitor the asset's current health status, the system processes operation data in real-time at the edge to detect anomalies and conclude upcoming failure occurrences or required maintenance actions.

6.4.2 Semiconductor

The AI technologies open various opportunities for semiconductor manufacturing by using AI-based systems in different co-existing models in the datacentres and on-premises at the edge and embedded in the semiconductor manufacturing equipment. These AI-based systems optimise and improve the efficiency of processes for different semiconductor technology nodes and support the acceleration of the design and manufacturing of multiple hardware architectures (e.g., CPU, GPU, NN accelerators, FPGA, dedicated ASICs, etc.), addressing a large set of heterogeneous applications.

AI-based technologies support semiconductor manufacturing facilities optimise and improve efficiency during the research and chip-design phase. The AI methods are used for eliminating defects and out-of-tolerance process steps that can decrease/avoid time-consuming iterations, accelerate yield ramp-up, and lower the costs required to maintain yield. The AI-based techniques are used to automate the time-consuming physical layout design and verification processes.

The AI-based technologies and applications in semiconductor manufacturing industry, address the following areas AI-based failure modes and effects analysis (FMEA) generator, AI-based 3D inspection for quality assurance, fault package detection, automatic interpretation of scanning electron microscope (SEM) images from semiconductor devices, silicon package fault detection and digitised support for product definition. A list of the demonstrators implemented under each application areas is presented below:

AI-based FMEA generator

AI-based FMEA assistant – development of an FMEA assistant tool to support the engineers to analyse the existing information efficiently. FMEA assistant is created by using existing data from the manufacturing process like structured or semi-structured FMEA, Failure Analysis (FA), 8D documents, and other domain-specific unstructured texts like production tools manuals, handbooks, and process descriptions patents and similar.

AI-based 3D inspection for quality assurance

- *Neural network for predicting critical 3D dimensions in MEMS inertial sensors* – addresses the use of ML to predict product parameters of inertial sensors, which are determined by the 3-dimensional shape and dimensions of the MEMS device. Data is collected from several process sources, including product measurements in various process steps and processing machine conditions.

- *Machine vision system developed in the wafer inspection production line* - processes the microscopic images of semiconductor wafers and detect surface defects, providing the results in a readable form, either in a table with coordinates and size of each defect or in the form of a heatmap of defect location on a wafer.

Fault package detection

- *Wafer fault classification* - provides a device-integrated solution for the wafer classification problem. The device gets pictures from a camera and perform real-time data analysis, giving the category to which, the wafer default belongs and binary faulty/non-faulty information.

Automatic interpretation of SEM images from semiconductor devices

- *Automatic inspection of SEM cross-section images for technology verification* – addresses a fully automated measurement toolchain. Research focuses on computer vision tasks and additionally on methods for automated analysis techniques of semiconductor front-end technologies.

Silicon package fault detection

- *Anomaly detection on wire bond process trace data* - covers the supply chain's relevant functionalities: developing the AI-based model, deployment, and visualisation. The work addresses the limitations regarding the availability, scalability (number of eq. and recipes) and degree of integration into the production data landscape.
- *Optical outgoing inspection* - provides an optical inspection solution working on the same or similar hardware and software environment providing anomaly detection with a pre-trained neural network (NN) for detecting deviations, image labelling for supervised learning and deployment of the AI-based model for image analysis and prediction.

Digitised support for product definition

- *Digitising product definition* – addresses the assessment of product definition via automated application simulations as planned via ML and formulate requirements human- and machine-readable to boost automation in the design and development phases.

6.4.3 Industrial Machinery

AI technologies are becoming a necessary part of manufacturing and automation across engineering, operations, and maintenance in the machinery and industrial equipment industry. AI applications start to be used more in high-end machinery and gradually migrate toward simpler machinery, such as palletisers and packaging machines.

In the machinery and industrial equipment industry AI is deeply embedded in the controller, the engineering tool, or the devices controlling the manufacturing line. The AI embedded solutions used for taking decisions and replacing the programmable logic controller (PLC), require techniques that are near 100% transparent to be accepted in a conservative industry such as industrial automation. The growth of industrial personal computers (IPCs) in manufacturing transformed what AI-based PLCs are capable of. In the machinery and industrial equipment industry, AI can be integrated into engineering and programming tools with embedded natural language processing (NLP) autocorrect features or automatically suggesting code and changing programming controllers. Using low-cost, robust and energy-efficient high-performance AI chips, AI becomes a necessary part of the controllers in the automated production lines.

New approaches to computer vision and ML open a variety of possibilities in industrial automation to optimise processes and improve the safety of human operators in the industrial environment. The areas of improvement cover areas from detecting defects of the goods and the erroneous behaviour of machinery to the detection and classification of all the objects present and acting in the working area.

The supply chain in machinery and industrial equipment develops to integrate the support of DL in the industrial environment, allowing the dynamic adaptation of the behaviour of the machinery with a re-training of the AI support on cloud level and the deployment at the edge of many precise high-efficient devices for the local processing and analytics. The interaction with the machinery and the data retrieved from different types of sensors and IIoT devices produce a consistent amount of data processed by the AI-based services to improve machine learning and the continuous re-training of the AI-based embedded modules.

The AI-based technologies and applications in the industrial machinery industry comprise two main areas, wood machinery with innovative human-machine interface (HMI) and smart robots. A list of the demonstrators implemented under each application areas is presented below:

Wood machinery with innovative HMI interface

- *Wood machinery with the perception of the surrounding environment* – addresses the use of specific sensors (e.g., ultrasonic sensor grid sensors) to detect the presence of obstacles near a woodworking machine, slowing down or stopping the machine's cabinet in case of detection. AI-based techniques are used for the refinement of the sensing and detecting capabilities to guarantee a higher level of reliability of the detection.

Smart robot

- *Smart robot* - addresses how to enable robots to "see", "feel", and interface with humans and the environment around them using a universal multi-modal cognitive sensing platform providing synthetic real-life like data generation for AI-training, intuitive human-machine interaction, and usage of Robot Operating System (ROS) for adaptability of different industrial robots, sensors, and other equipment.

6.4.4 Food and Beverage

The implementation of IIoT and robotics solutions in the food and beverage industry sector has supported overcome critical issues related to production and execution by eliminating the possible human errors while reducing the redundancy in work performed by manual labour. AI fuels innovation in the production and packaging of food and beverage to reach expectations regarding the quality of the products delivered to the consumers and their related impact on the cost. To attain the potential trade-off between quality and price, industry stakeholders are actively leveraging the potential of AI across various applications, such as product design, quality control, maintenance, and consumer engagement, among others.

The integration of AI technology increases the efficiency improvements in the food and beverage industry, with significant reductions in downtime, repair costs, and additional labour requirements and cost. Companies in the food and beverages production and manufacturing industry leverage the benefits of AI through the use of NNs, ML techniques, advanced analytical tools, combined with image recognition and computer vision technologies for optimising the manufacturing processes.

The food and beverage processing lines include continuous monitoring IIoT technologies used in the predictive maintenance process that collects real-time data from multiple and varied IIoT sources placed on motors/equipment, combines them and uses ML techniques to anticipate equipment failure before it happens. Predictive maintenance of production machinery is for instance based on sound or vibration analysis computed directly by IIoT devices and vision-based quality control of the product at the edge for production process optimisation. Parts of the collected data is sent to a cloud service that continuously optimises a detection model. Control applications directly influence the production process and are therefore especially critical considering their real-time capability and reliability.

The AI-based technologies and applications in the food and beverage industry comprise two main areas, beverage production - Champagne and food production - soya beans. A list of the demonstrators implemented under each application areas is presented below:

Beverage production - Champagne

- *Environmental monitoring system* – addressing the implementation of an industrial monitoring system that enables an accurate analysis of the production process in the vineyards and caves. The data obtained by this monitoring infrastructure enable accurate decision-making accordingly to an external environmental condition that can impact the production step under analysis.
- *Autonomous environment-aware* - addresses the implementation of an AI-based method of capturing images and data using an autonomous robot to support cameras and sensors. The data analysis from the vineyards allows precise decision-making regarding the yield, vine diseases and missing vines.
- *Quality control system* – addresses the setting up an image acquisition system in the Champagne presses facilities. The data analysis from the presses allows the neural network training based on the quality classification of the grapes.

Food production - Soya beans

- *Production process optimisation* – addresses soybeans production process optimisation using IIoT-based sensors for visual analysis, temperature, humidity, and moisture throughout the preparation phase and correlates real-time data in those parameters using AI-based models.
- *Predictive maintenance* – addresses a solution for implementing an intelligent monitoring system that separates the equipment's normal condition from abnormal conditions. IIoT-based sensors are installed to measure different parameters such as vibration, current, sound, temperature etc. Data from the IIoT devices are sent to AI-based models that correlate with normal and abnormal conditions and implement a predictive maintenance solution.

6.4.5 Transportation

Mobility-as-a-Service (MaaS) based on vehicle sharing changes people's transport habits and introduces new mobility modes. Future automated vehicles enable 24/7 driving and serving people with fewer vehicles on roads. MaaS steps taken without automation aim to improve the availability of public transportation when needed.

The main challenge with public transport is the mean of transportation available when needed. Buses, trams, etc., are bind to the schedules and cannot offer ad-hoc on-demand solutions.

Therefore, on-demand taxis and mass transportation have great potential to change people's mobility in cities and rural areas where the ageing population is suffering from available transport services when needed. In addition, fewer vehicles and buses mean fewer pollutions thus, leading to better transport sustainability.

The application of AI in the transportation industry is accelerating the next generation of Intelligent Transportation Systems (ITS). Intelligent edge computing technology supported by high-speed connectivity is used to process AI decision-making at the vehicle and edge level without connecting to a server in the cloud.

AI technologies in traffic management enhance the efficiency of the mobility systems it integrates with and play a significant role in developing and deploying new and innovative environmentally friendly solutions to operate vehicles for travel and transportation.

The AI-based technologies and applications in the transportation industry cover one central area, MaaS development of AI-based fleet management for supporting multi-modal transport. The demonstrator implemented under this application area is presented below:

MaaS, development of the AI-based fleet management for supporting multi-modal transport

- *MaaS - AI-based fleet optimisation tool* – addresses an AI fleet management of MaaS solution, in which two automated last-mile vehicles are controlled according to the transport demands of the users. The data processing is done in the vehicles and the infrastructure by reducing links to the cloud and increasing decentralisation. Novel computation platforms are utilised for accelerated processing. A neural network (NN) based analysis program for predicting travel times is implemented. The machine learning-based, improved data pipeline leads to improvements in terms of waiting times for passengers.

6.5 AI Technology Roadmap for Digitising Industry

The transition towards a more integrated technology converging combining AI with other cutting-edge technologies like IIoT, edge computing, and connectivity is essential to reframe the challenges that the European electronic components and systems community need to tackle in the future. AI4DI stakeholders drive activities for community consensus and take the lead on the compilation of multi-annual research roadmaps addressing human-centric AI technologies aligned with roadmaps of other related technologies. The roadmap guides how the European electronic components and systems community can obtain a competitive advantage in designing, developing and deploying silicon-born AI, AI-based embedded systems in industrial sectors.

Building the AI roadmap results from exchanging ideas and concepts at the European level and aligning the work with activities at the global level. The approach for interaction with related industrial sectors is primarily focused on analysing existing challenges and gaps of the related technology areas and specific workshops exploring the intersection of AI technologies with key stakeholder groups.

The perception of the AI technologies by European citizens and the industrial sectors that it affects play an essential role in the broader adoption debate of AI technologies. Industrial AI solutions may lack direct consumer scrutiny, but they are under the evaluation of the industry stakeholders that strongly and robustly influence the sector regulations and standards.

The AI4DI work on the AI road-mapping activities provides an excellent framework for the industrial stakeholders to prioritise resources and align the vision of the electronic components and systems community to focus on essential breakthroughs to reach the next level of AI technology evolution for digitising industry.

The shift of AI methods from cloud to edge is the primary approach of AI4DI for digitising industries and marks the starting point of a comprehensive transition regarding the control of industrial processes and functionality of devices. The AI4DI roadmap lists the significant milestones of this transition driven by AI methods operating on the edge.

The AI technology developments influence the evolution of IIoT devices, silicon-born AI, embedded systems-born AI, AI methods, models, algorithms, and integration in the manufacturing processes. The integration of complex AI-based systems is highly linked and dependent on all these elements. The increase in computing power and industrial user experience with single AI methods supports integrating interconnected machines and automated data analysis and optimisation of processes. This development enables the transition from linear manufacturing processes to AI-controlled value chains as a network of many flexible interconnected machines as part of a distributed production line. More powerful AI methods implemented on IIoT devices at the edge increase these devices' functionality aiming at self-learning and self-optimisation features. The ongoing parallel development of AI methods for cloud applications plays a significant role in developing efficient processes and functional devices at the edge along with the roadmap. Implementing industry-grade embedded AI on edge devices directly depends on the availability of dedicated silicon-born AI architectures for electronics components and systems that are powerful enough to fulfil the full potential of state-of-the-art AI methods.

6.6 Conclusion

Intelligence on the edge devices in industrial environments allows it to process information locally and respond fast to situations instead of communicating with a central cloud or server.

AI raises new ethical and legal questions related to liability or potentially biased decision-making in industrial environments. AI4DI actively supports the activities for progressing ethical guidelines on AI development in industrial environments by guiding the industrial stakeholders on the new challenges brought by AI and the interpretation of the liabilities in the light of technological AI developments to ensure legal clarity for industrial consumers and producers.

The article gives an overview of the ECSEL AI4DI project that develops AI-based solutions to bring intelligence processing from the cloud to the edge by providing intelligent technologies across industrial sectors such as automotive, semiconductor, machinery, food and beverage, and transportation.

The project aims to provide AI-based technologies at the edge for digitising the industry by reducing costs, saving time, and increasing quality by enhancing industrial processes. The project's advancements enable optimising/improving industrial processes, products, services, and support building and sustaining a dynamic AI technology ecosystem in Europe.

Acknowledgements

This work is conducted under the framework of the ECSEL AI4DI "Artificial Intelligence for Digitising Industry" project. The project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826060. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Austria, Czech Republic, Italy, Latvia, Belgium, Lithuania, France, Greece, Finland, Norway. The authors want to express their great appreciation to Dr Cristina De Luca for her valuable and constructive ideas, suggestions and inputs during the preparation and writing of this chapter. She was the ECSEL AI4DI project coordinator, working for Infineon Technologies AG until March 2022, and her dedication and hard work supported the project to obtain the best results.

References

[1] ECSEL AI4DI project. Artificial Intelligence for Digitising Industry. Available online at: https://ai4di.eu/

[2] The Artificial General Intelligence Sentinel Initiative (AGISI), "A working list: Definitions of Artificial Intelligence and Human Intelligence". Online at: http://agisi.org/doc/AGISI_DefinitionsIntelligence.pdf

[3] Wang, P. (2019). "On Defining Artificial Intelligence" Journal of Artificial General Intelligence, Vol.10, No.2, 2019, pp.1-37. https://doi.org/10.2478/jagi-2019-0002

[4] AI4DI Deliverable D1.12 - Requirements and specifications for smarter food and beverage production based on AI-technologies. 2019

[5] AI4DI Deliverable D1.2 - AI for digitising industry road-mapping requirements. 2020

[6] AI4DI Deliverable D2.1 - Report on hybrid reference system level architecture design for the digitising industry. 2020

[7] Vermesan, O., John, R., De Luca, C., and Coppola, M. (2021). Artificial Intelligence for Digitising Industry - Applications. Gistrup: River Publishers. Available online at: https://www.riverpublishers.com/pdf/ebook/RPE9788770226639.pdf