

Semantic Models and Machine Learning Approach in CPS :A Survey*

Hafidi Mohamed Madani^{1,2,*,†}, Meriem Djezzar^{1,3}, Hemam Mounir^{1,2,*,†},
Ahmed Seghir ZIANOU^{1,2} and Moufida MAIMOUR^{4,*,†}

¹University of , Khenchela, Algeria

²ICOSI Laboratory, Khenchela, Algeria

³LIRE Laboratory, Constantine, Algeria

⁴Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

Abstract

The amazing growth and advancements reached in information and communication technologies in recent years allow easily the integration of intelligent components and systems into the traditional manufacturing industry. Enabling new challenges and applications in Industry 4.0 (I4.0) new systems. Cyber Physical Systems (CPSs) are a new generation of systems composed of a set of collaborative cyber and physical components with computation capabilities, generating and exchanging data in a loop between digital and physical worlds in a highly interconnected network. These enormous large amounts of data produced in or between CPSs are heterogeneous in terms of format and type due to different data sources, which leads to errors and malfunction in these systems due to the lack of interoperability between their components. As a result, sharing and exchanging data in CPSs is a challenging task to do. On other hand, modeling digital systems that reflect the current state of the physical entities and their behavior is a complex task to achieve. Especially, when communicating and processing data in real time to extract useful information. To overcome these challenges, semantic data models and knowledge representation when applied with Machine Learning (ML) techniques can enable solutions to interoperability problems in CPS, making it possible to mirror the physical reality and monitor it through cyberspace without misinterpretation and miscommunication in the system.

This paper aims to provide a survey on the state of the art of available solutions to the semantic interoperability problem in CPS, integrating semantic models, ML, or both technologies combined in a reference architecture to achieve visionary Interoperable CPSs.

Keywords


Interoperability, Cyber Physical Systems, Machine Learning, Digital Twin, Semantic models, Ontology


SIoT-2022: International Workshop on Semantic IoT (SIoT-2022), Co-located with the KGSWC-2022, November 21-23, 2022, Madrid, Spain.


*Corresponding author.

†These authors contributed equally.

✉ hafidi.mohamedmadani@univ-khenchela.dz (H. M. Madani); meriem.djezzar@univ-khenchela.dz (M. Djezzar); mounir.hemam@univ-khenchela.dz (H. Mounir); zianou.ahmed.saghir@univ-khenchela.dz (A. S. ZIANOU); moufida.maimour@univ-lorraine.fr (M. MAIMOUR)

 0000-0003-2431-4816 (H. M. Madani)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

1. Introduction

The need to increase the product effectiveness, quality, and customization of artificial products made humans develop and produce new technologies suitable for the elaboration of production processes that humans are experiencing today [1]. Since the 1980s, and with the third industrial revolution, the integration of Artificial Intelligence (AI) and manufacturing made the rise of intelligent manufacturing [2], introducing new technologies such as robots, Computer Numeric Control (CNC) machines, and industrial and electronic automation [3].

As AI evolves, smart technologies such as the IoT, semantics and data models, Machine Learning (ML) algorithms, CPSs, and DTs are taking intelligent manufacturing to a new digital age known as the fourth industrial revolution, also called Industry 4.0 (I4.0) [4], which is an initiative that started in Germany to automate production systems efficiently. The main core of I4.0 is the CPS concept, represented as a multidimensional complex distributed system integrating a cyber and a dynamic physical world interconnected in an intensive network providing many functionalities and services such as: real-time sensing, monitoring, and real-time interaction between entities from both worlds, through the collaboration of computing, communication and control [5].

The main focus of CPS is to network devices of I4.0 by handling the interaction between physical reality and computing infrastructures with the use of a communication interface. DT is the virtual representation of a physical component, it mirrors the status and the activity of the physical entity by collecting data from the physical environment using IoT devices and sending it to software systems in order to be analyzed. AI-ML techniques are one of the most promising and advanced technologies used to create and deploy DTs that mimic the behavior of physical components.

With the use of communication technologies, DTs and IoT, CPS can be deployed and implemented with respect to the vision of I4.0. But, the massive number of communication between interconnected physical components and digital systems in Cyber Physical Systems (CPSs) implies a huge exchange of data, mostly generated, sensed, and shared from different physical and digital entities in an autonomous way which results in heterogeneity and interoperability problems, making it hard and barely impossible to implement this new generation of systems in a full scale without solving and filling the requirements of interoperability and heterogeneity in communication.

Semantic data models technologies were proposed and used across academia in order to solve semantic interoperability problems faced at the deployment of CPSs, combined with the use of ML techniques to model and mirror the physical components of the system in the digital space.

This article is organized as follows. Section 2 provides a brief presentation on I4.0 and CPS, listing their key concepts and technologies, and focusing on the CPS reference architectures. Section 3 outlines different CPS proposed architectures and solutions integrating semantic models and ML algorithms and performs a brief discussion for each solution. Section 4 concludes with open issues and future direction for researchers.

2. Background

This section covers the different necessary concepts related to the topic to better understand the proposed survey. It presents a brief definition of the CPS concept, going through its most known reference architectures in academia, and then a comparison between reference architectures is discussed. Finally, related concepts of I4.0 and semantic heterogeneity are defined.

2.1. Cyber Physical Systems

The CPS concept was presented for the first time in 2006 by Helen Gill from the National Science Foundation (NSF) in United States [6]. The concept was introduced as a prospective shift towards the next generation of engineered systems using future *networking and information technology* (NIT) [7]. Many funding programs were created to promote and accelerate the evolution in this field, CPS-VO is a US organization created to link between CPS professionals in academia, government, and industry all around the world [7]. Another foundation was created in Europe named ARTEMIS, its main goal is to develop and accelerate research on the CPS concept, by creating a partnership between European countries and Industry companies with the goal to achieve a smart and physical aware world connecting all machines, objects, and systems in a network with the use of digital information around them to communicate and collaborate together [7].

Cyber Physical Systems (CPSs) were defined differently by the scientific community. In [8], CPS is defined as: “integration of computation with physical processes”.

[9] describes CPS as: “physical and engineered systems, whose operations are monitored, coordinated, controlled, and integrated by a computing and communicating core”.

From Gill’s perspective [10], CPS are: “physical, biological, and engineered systems whose operations are integrated, monitored, and/or controlled by a computational core. Components are networked at every scale. Computing is deeply embedded into every physical component, possibly even into materials. The computational core is an embedded system, usually demands a real-time response, and is most often distributed”.

CPS is defined in a generic way as software networks of computer systems and physical processes, with feedback loops, used to monitor and control the physical process [6]. CPS links each physic component to its DT, enabling knowledge sharing and fast decision-making. through the integration of computing, communication, and control, making it possible to have an interactive industrial environment [11].

Unlike embedded systems, CPS focus on the interaction between various digital systems and their physical twins, by handling with a control unit sensors and actuators that influence the physical space, and process the data produced and exchanged through a communication interface [11].

An important characteristic of CPS is the ability to send and obtain information and services from different devices to other systems autonomously. So it is necessary to ensure its reliability, efficiency, and security in communication [12]. That is why one of the main requirements of I4.0 is achieving high interoperability between CPS components.

2.2. IoT, Industry 4.0 and Cyber Physical Systems

Although CPS was presented in 2006, it became a popular trend due to the advancement in information technology and IoT, which play a big role in collecting real-time data from the physical world using sensor technology and sharing it with the digital world. The physical-to-digital connection uses communication technology that enables transmitting data of the physical components to its virtual twin, by storing it in databases, with the use of WIFI and cellular networks and their protocols ...[13]. The virtual twin converts the data stored using information and data technologies such as knowledge representation and reasoning to extract explicit-implicit information and use it to mirror the functioning of the physical twin, which makes it possible to monitor and predict the status of the physical environment from the virtual world [14].

The digital-to-physical connection is represented by the information circulating from the virtual to the physical world, this information may influence the state of the physical components by changing its parameters or by executing a task to achieve a certain goal (prognostic, diagnosis, optimization). Both data and Information circulating between the physical and digital twins are stored in historical databases in order to be processed and analyzed to take final decisions about optimization, prognostic and diagnosis with the use of different machine learning algorithms, and semantic web technologies with reasoning [14].

Many benefits can be obtained from implementing IoT and CPS in Industry 4.0, like achieving better use and performance of assets, minimization of the time and cost of producing assets, optimization of the product life cycle, faster decision making, improving mass customization and production in manufacturing [15, 16], monitoring and connecting several manufacturing machines in a smart environment (software networks) [17].

Despite all the advantages, implementing IoT combined with CPSs in the vision of Industry 4.0 stays a complex mission because of the huge number of different architectures and approaches of digitalization and networking adopted by enterprises that makes communication and systems interoperability problems rise [18].

2.3. 5C Architecture

The 5C Architecture of CPS was proposed by Lee in [20]. It consists of a unified 5-level architecture that serves as a step-by-step guide for developing and implementing CPS in manufacturing. It is considered a well-known reference data-driven model widely adopted in developing CPS since it focuses on adopting and defining the data flow in these systems from the initial data acquisition, to analytics, until the decision making, as shown in Figure 1. The different levels of the architecture are outlined in Table 1.

2.4. RAMI 4.0

RAMI 4.0 stands for Reference Architectural Model and it consists of a three-dimensional model that describes all aspects of I4.0 space [21]. In this way, complex interrelations can be broken down into smaller and simpler clusters. The RAMI 4.0 is composed of:

- The “Hierarchy Levels” axis: on the right horizontal axis in Figure 2 shows the hierarchy levels from IEC 62264 and IEC 61512, the international standards series for enterprise IT

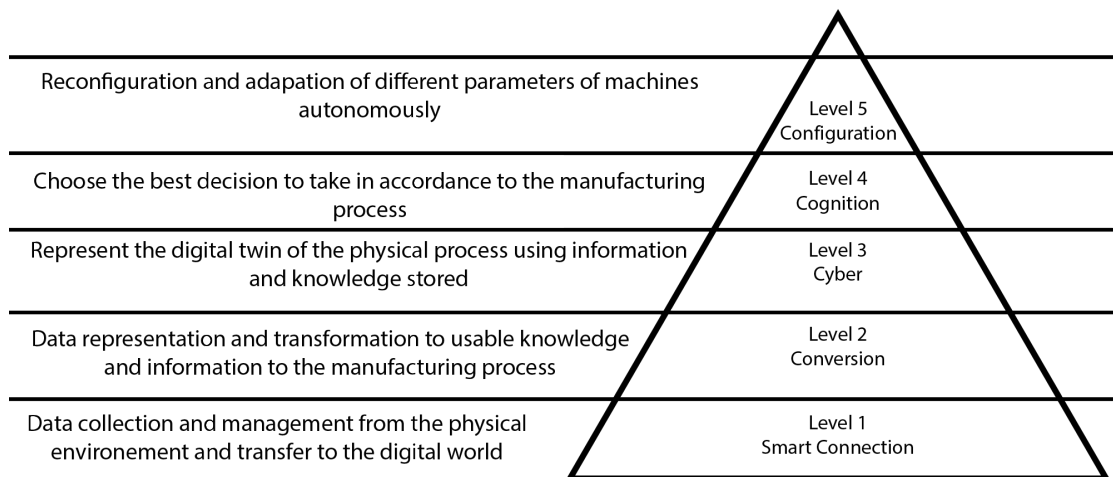


Figure 1: 5C Architecture of CPS adapted from [19]

Table 1

Different levels of 5C architecture.

5C Level	Role
Smart Connection	This level is responsible for data acquisition from different interconnected physical world components with effective management and transfer to the central server. Physical component identification and specification are required.
Data to Information Conversion	This level manages the conversion of data collected from the previous step to information. This step brings self-awareness to machines.
Cyber	This level uses inferred information from the previous step to create Digital Twins (virtual representation) of physical components of the real world. It is considered as the center that guarantees communication between physical assets since all inferred information of collected data is sent to it.
Cognition	This level uses information gathered to acquire proper knowledge to monitor the system and make prognostics for failure prediction and maintenance optimization.
Configuration	at this level, operations are sent as feedback from cyberspace to the physical world to control machines, making the physical space self-configure and self-adaptive

and control systems. These hierarchy levels (Enterprise, Work Centers, Station, Control Device) represent the different functionalities within factories or facilities. Other layers (product, field device) were added to the hierarchy levels to support the representation of the I4.0 space. The top layer labeled “connected world” serves to connect and reach external partners through service networks.

- The “Life Cycle and Value Stream” axis: The left horizontal axis represents the life cycle of facilities and products, based on IEC 62890 for life-cycle management. A distinction is made between the two levels “types” and “instances”. When the product is in design and development phase. We are in the “type” level. When designing the actual product have

been completed and production starts, we are in the “instance” level.

- The “6 Layers” axis: represented by the vertical axis serves to describe physical machines and objects from real world: The first layer labeled “Asset” consists of defining the physical objects and their properties, such as parts, documents, diagrams, humans, etc.... The second layer named “integration” maps the physical objects to the digital world components, it serves as a link between physical and digital worlds where transformations and easy processing of information happens. The third layer “Communication” is responsible for providing standardized communication between the integration layer and the information layer, the standardization uses a unified data format that serves to represent data in an organized way in the information layer, and to process and transform the data in the integration layer so it can link physical and digital worlds together. The fourth layer “Information Layer” has the role to store and hold data in an organized way, its main purpose is to provide relevant information about products from the data stored (such as statistics on the number of sales, and the number of objects produced). It can also give the machines used in a production of a product or information about customers and their feedback...etc. The “functional” layer has the goal to take actions, coordinate, and select components to proceed with tasks, it involves various activities that can be made in the system like authentication, user inputs, remote access...etc. The last layer named “business layer” contains the business strategy, business environment, and management. It deals with several business activities in order to achieve the business plan model.

2.5. IIRA

The Industrial Internet Reference Architecture (IIRA) is an open cross-industry architecture developed by IIC based on IIoT standards [23], emphasizing interoperability among industries such as manufacturing, energy, healthcare, and transportation. This model is organized in five Viewpoints [24] presented as layers in IIRA, each layer is linked to certain functionalities:

- Business layer: Functions that enable end-to-end tasks in an industrial system (e.g., enterprise resource planning, life cycle management, planning, and scheduling);
- Application layer: Functions that enable business operations by implementing application logics (e.g., activity/workflows, application programming interface, user interface).
- Information layer: Functions that handle data gathering and deployment (e.g., data collection and storage, semantics, quality processing).
- Operations layer: Functions that enable operations related to component monitoring and diagnosis throughout their life cycle (e.g., provisioning and deployment, optimization).
- Control layer: Functions that enable the control of an industrial system (sensing and actuation, communication, abstraction, digitalization, analytics, asset management).

2.6. Comparison between Architectures

This subsection consists of analyzing the similarities and differences between the architecture reference models described in previous subsections, highlighting their main goals and relations

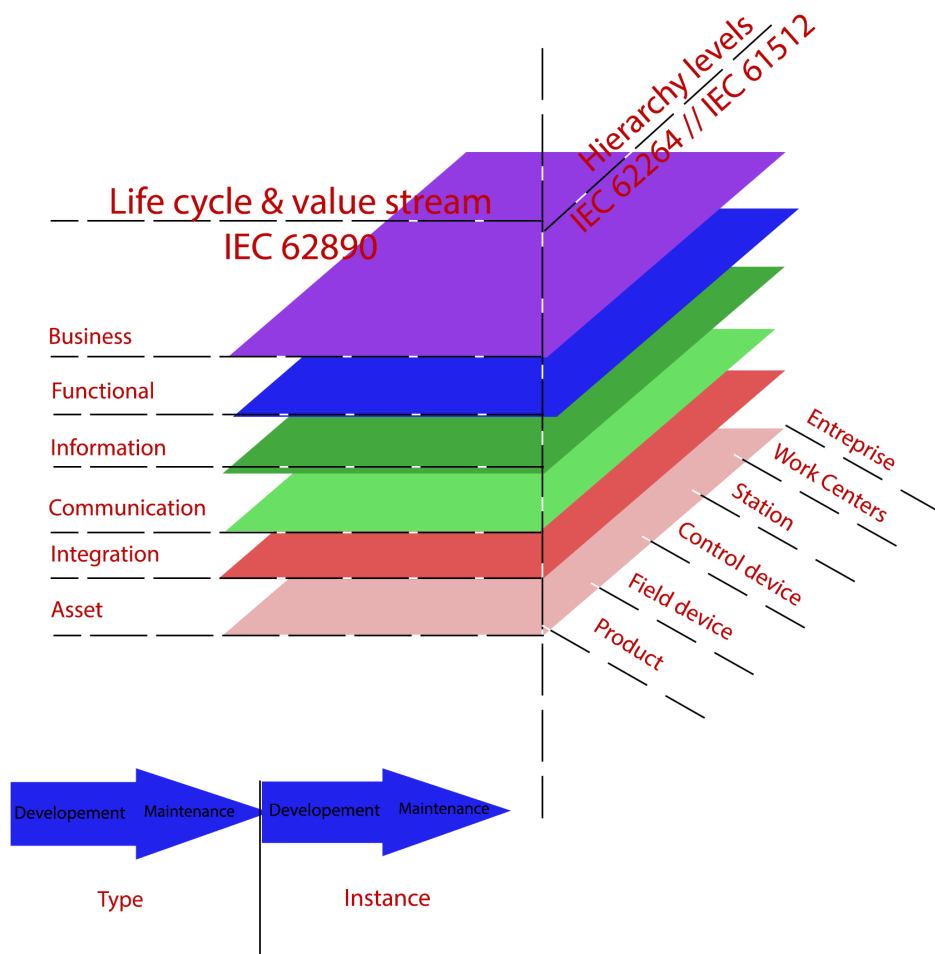


Figure 2: RAMI 4.0 Architecture adapted from [22]

that make them interoperable with each other. All these architecture reference models define and represent CPS concepts, but their proposals lie in targeting different goals [24].

At first, the 5C architecture targets modeling and describing assets data collection and processing, mainly found in small smart industries and IoT environments. It is considered the first architecture published in the literature and it is adapted only for horizontal integration.

RAMI 4.0 was proposed to describe and model CPS systems in the I4.0 scenario. It focuses on defining manufacturing assets operations and it focuses on describing in detail CPS design, control, communication, and business by integrating the value chain company and product life-cycle.

With IIoT proposal as a highlight, IIRA is based on the ISO/IEC/ IEEE 42010 and describes the development plan to create the IIoT system. It deeply focuses on the IIoT system as a core concept in all sectors, such as product life-cycle from design to maintenance and control[24].

As we can see from the architecture reference models, several layers from RAMI 4.0 architecture are similar in functionalities with levels in the 5C and IIRA architectures. These layers and levels can be mapped together ensuring possible interoperability establishment between different architectures.

2.7. Semantic Models

Semantic models have been considered an important technology to achieve interoperability between different systems [25]. They have the power to model and describe the properties and relationships between concepts and entities. One of the most used semantic models in research is ontology.

An ontology is defined as an ‘explicit specification of a conceptualization of a domain’ [26]. It is used to represent in a generic way a domain knowledge, by encoding it in the form of axioms, natural language labels, synonyms, definitions, and other types of annotation properties making it possible to achieve an agreed understanding between applications [27]. Ontology itself is relatively complex concept that requires a special set of expertise to involve and maintain it [28]. Most of the ontologies are encoded using the Ontology Web Language (OWL) since it’s more expressive than other ontology languages like Resource Description Framework Schema (RDFS). OWL is a part of the Semantic Web stack and it is based on Description Logic [29, 30].

2.7.1. RDF

RDF is a model for encoding semantic relationships between items of data so that these relationships can be interpreted computationally. It is considered the primary foundation for the Semantic Web.

2.7.2. RDFS

RDFS is a set of classes with certain properties using the RDF extensible knowledge representation data model, providing basic elements for the description of ontologies. It uses various forms of RDF vocabularies, intended to structure RDF resources.

2.7.3. OWL

OWL is a semantic markup language designed to represent complex knowledge as concepts and their relationships. It is used to publish and share ontologies. It is developed as a vocabulary extension of RDF. [31]

2.7.4. Semantic Heterogeneity

The semantic heterogeneity concept does not have a unique definition [32]. Several definitions that provide a certain degree of understanding of the term were found in the literature :

- According to Merriam-Webster dictionary, semantic heterogeneity is defined as a quality or a state of being made up of parts that are different — related to the meanings of words and phrases.

- From [33], the term is presented as differences in the meaning and use of data that make it difficult to identify the various relationships that exist between similar or related objects in different components.
- Semantic heterogeneity can be defined also as differences in the real-world interpretation of context, meaning, and use of data [34].

2.8. Digital Twin

DT concept refers to the virtual representation of a physical entity, including all its properties and functioning. The term was first coined in 2011 by John Vickers. The concept was defined by its pioneers Grieves and Vickers [35] as "a set of virtual information constructs that fully describes a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level.

At its optimum, any information that could be obtained from inspecting a physically manufactured product can be obtained from its DT". It is mainly used to mirror the physical states of a manufactured product and to help users predict properties or take action from the digital space.

We can define Digital Twinning as a process that involves a physical entity, a cyber twin that mirrors the physical entity and a physical connection (communication interface) used to share and translate data between them, as shown in Figure 3.

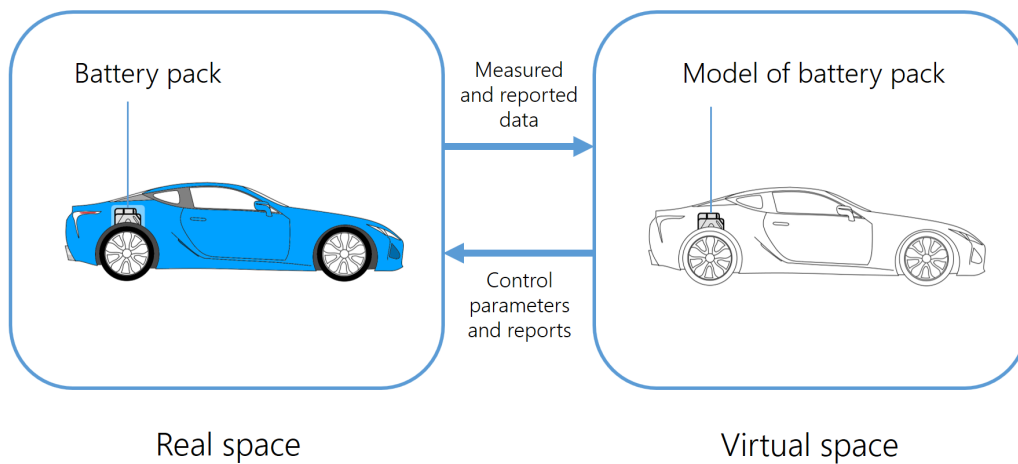


Figure 3: Digital Twinning concept schema drawn from [36]

2.9. AI-ML

AI can be defined as the ability to reproduce intelligence reasoning by computer systems for decision-making purposes.

Machine Learning (ML) is a subset of AI. ML can be defined as the set of techniques that allow a machine to learn, solve or perform a task without having to program it explicitly. This set of techniques concerns the analysis, design, development, and implementation of methods allowing the machine to follow a systematic process to solve a problem, where it will be difficult or impossible to do so by a classical algorithmic method [37].

The ability to make a machine learn to perform a task requires a set of data that contains a lot of information specific to the latter. The techniques of ML analyze and process all these data in order to extract the knowledge and features found and then apply and reuse them on new data to solve real-life problems. ML techniques are usually categorized into three methods namely supervised, unsupervised, and Reinforcement Learning (RL) [37].

2.9.1. Supervised Learning

Supervised learning is employed when the training dataset is labeled. Each sample of data is associated with the desired result (output). The goal of the ML algorithm is to find a function that maps the input data to the output. Unseen data can be fed as input to a trained model in order to predict and map it to the most relevant output.

2.9.2. Unsupervised Learning

Unsupervised learning is used in situations where the training data that contains the different examples are not labeled in advance. Unsupervised learning consists in partitioning the examples of the training data into categories based on chosen similarity criteria. It allows the automatic construction of classes without any intervention, but it requires a good estimation of the number of classes.

2.9.3. Reinforcement Learning

RL is a framework for solving control tasks and decision-making by building agents that learn from the environment by interacting with it through trial and error and receiving positive and negative rewards for correct and incorrect performances respectively as feedback. The ultimate goal of the agent is to maximize its reward in any given situation.

3. Semantic Models, AI-ML, DT Integration in Cyber-Physical Systems

In this section, we discuss the current deployments and implementation of CPS developed with the integration of Semantic models technologies, AI-ML techniques, and DT creation. Table 2 outlines the selected works from academia to be discussed and the technologies integrated into their deployments.

Table 2

Selected papers and the integrated technologies.

Ref	Semantic models	AI-ML	DTs integration
[38]	✓		✓
[16]	✓		✓
[39]	✓		✓
[40]	✓		✓
[41]	✓		✓
[15]	✓		✓
[42]	✓		✓
[43]	✓		✓
[44]		✓	
[45]		✓	✓
[46]		✓	✓
[47]		✓	✓
[48]		✓	
[49]		✓	✓
[50]		✓	
[51]		✓	
[52]		✓	✓

3.1. Semantic Models Integration in CPS

Semantic models can serve to bridge the gap between different project participants in CPSs by assuring shared understandable information.

In [38], the authors show how machine-to-machine communication is important for collaborative manufacturing automation (smart manufacturing). The article introduced a concept of semantic aware CPS based on industrial machines that can perform semantic machine-to-machine communications. The authors proposed a semantic and communication layer on top of the physical and cyber layers of a CPS. The semantic layer provides logical communication between CPSs from different organizations without the need to be attached together but acts as if they were by converting the messages exchanged into semantic expressions. The communication layer then transfers the semantic expressions between CPSs as web requests using HTTP network protocols.

Different engineering models are created to illustrate different components and tools in CPSs where each engineering model uses local terminologies related to its engineering environment. Semantic models can be used as an integration tool of these different models together by creating ontologies and aligning them at terminological and instance levels to achieve a good representative DT of the physical system. Each ontology brings specific semantics about each physical component aspect and provides it to its DT to fulfill and reach a certain goal.

Semantic models can be classified and differentiated according to the semantics they define as:

- **Geometric model:** describes the geometric properties (shape, size) of the physical entity.

- **Physical model:** describes the different parts or the abilities of the physical entity (composition, capacity...).
- **Behavioral model:** refers to the behavior of the physical entity communicating with other entities.
- **Rule Model:** defines the relation between domain knowledge concepts (constraints, associations, deduction, negation...).
- **Process Model:** describes the underlying process and function in which the physical entity takes part in the CPS.

In the operational phase of CPSs. The DT receives collected data from sensing the physical environment. This collected data comes from different components and describes different properties. Semantic models are used to describe and represent the collected data appropriately, making it possible to visualize and monitor dynamic systems.

A semantic model based on an extension of IoT-Lite ontology is proposed in [16] to represent and integrate digital twins of devices in an IIoT system. The semantic model focuses on describing data information of the physical abilities collected from IoT devices and the relationships between them in a high-level form so it can be represented and interpreted by digital twins of the IoT system.

The semantic model offers the ability to show information to end users in a better manner. Moreover, semantic models can extract implicit information about physical objects by reasoning about the relations between objects in the knowledge base. This kind of information can help to detect early faults and error occurrences [53]. For example, a DT for part assembly is created based on an ontology that describes geometrical information about parts and constraints in assembly units [15], the expressiveness of the ontology made it possible to reason and infer relations between concepts of the domain knowledge and define assembly requirements between parts.

In Smart manufacturing, [39] proposed a 4-layer architecture based on ontology to manage and reconfigure resources in smart manufacturing. The ontology describes manufacturing resources and their properties using OWL. A rule base model is used to reason about manufacturing resources ontology and to reconfigure their status. A reconfiguration of an intelligent manipulator is implemented as a use case to show the feasibility of this architecture.

Another research in the assembly field [40] presented an ontology-based model of a DT for assembly workshop. The ontology of the DT assembly workshop is built to describe the objects, attributes, and relationships that exist in the assembly workshop and participate in the assembly behaviors. A model of the assembly process is defined as an occurrence of several events in a range of time, a representation of event ontology is presented as an event-oriented description logic to act as a logical base of event ontology language. The ontology of the DT assembly workshop with the assembly process model permits mirroring and monitoring the assembly workshop using the event ontology language.

The study in [41] presented a DT modeling architecture for manufacturing processes using multi-agent systems, Material-Process-Functions-Quality (MPFQ) model, and semantic models to manage knowledge. Semantic models were used to properly define the knowledge and help multi-agent systems to better interact and interpret shared information, enabling semantic interoperability. Another study [42] discussed the challenges faced in dealing with DT data

management (data variety, data mining) and their influence on DT dynamics, it proposes a novel concept of DT ontology model and methodology to address these data management challenges. The DT ontology model models the conceptual knowledge of the DT domain. Using the proposed methodology, such domain knowledge is transformed into a minimum data model structure to map, query, and manage databases for DT applications. The research is tested using a case study based on Condition Based Monitoring (CBM). From another perspective of modeling approaches, the research in [43] proposes a semantic modeling approach based on high-level architecture and GOPRR for DT integration. The semantic modeling approach serves as a tool to create DT ontology that describes different heterogeneous properties of the physical twin to the digital model.

3.2. ML and DT integration in CPS

ML techniques are a key enabler technology in CPSs, they can be applied in both the engineering and operational phase.

DT can mirror and replicate any physical object in cyberspace, it provides also a feedback mechanism to control and monitor the physical object. ML algorithms are known for their high performance in decision-making applications. In monitoring CPSs, anomaly detection is a required task to control the physical object's state and its behavior.

In [45], ML techniques were used to detect anomalies in CPSs. A novel approach named Anomaly DeTection with digiTAL twIN (ATTAIN) is presented in this research paper. A Timed Automation Machine (TAM) is built to represent the DT of the CPS. Generative Adversarial Network (GAN) techniques are used to detect anomalies in the CPS. A generator model is used to capture the characteristics and features of the input data and learn to generate realistic unlabeled samples with the same features. TAM labels the samples produced by the generator and feed them to a discriminator with real labeled samples. The discriminator is trained to distinguish normal data from anomalous data. The use of DT with ML in this approach gave better results in anomaly detection compared to approaches not using DT.

ML techniques are showing impressive performance at detecting cyber-security attacks in CPSs, the research in [44] discusses and implements a comparative study on the use of supervised ML techniques to detect cyber-attacks in CPSs. A case study using a dataset of a secure water treatment plant is selected to perform cyber attack detection using ML techniques, results showed impressive accuracy on the trained dataset.

Sensor data collected from the environment and the machines can be used also to detect failure and errors by training a ML model from historical diagnosis data of the system or from historical machine data to predict failures and errors in the machines and the system. But, training an ML model with unbalanced or insufficient data can lead to errors when the model is used with real-world data. The performance of the ML model depends on the quality of training data and its volume. In [46]. An architecture that uses a discrete physics-based computational model with ML techniques to create a DT for investigating several damaged structure scenarios is presented. A ML classifier that represents the DT of the damaged structure is trained with data taken and generated from a stochastic computational model that simulates the damaged structure scenarios, the ML classifier learns to detect damaged structures and serves to warn the user of the location of the damage. The DT is then connected to the physical entity to enable the

use of real-time decision-making. A use case of a wind turbine is modeled with physics-based models and data generated from it was used to train a ML classifier to detect damages in the turbine structure. Results showed a low accuracy in detecting damages.

In [47], a study discusses the creation of a DT with the use of the petrochemical industrial IoT, ML connected in a loop with the physical factory to exchange information in real-time, realizing production control optimization. This approach optimizes DT models by applying real-time big data analysis with ML, it supports petrochemical processes and other manufacturing systems to dynamically adapt to the changes happening in the environment, taking into account time lags between time series data and reduction of data dimensionality. Several ML algorithms were tested and trained using data gathered from the industrial IoT system. A case study in a real petrochemical factory was used to examine the effectiveness of this approach. Moreover, the authors in [48] presented a compositional falsification framework for Signal Temporal Logic (STL) specifications against CPS-ML models based on a decomposition between the analysis of ML components and the system containing them. A ML analyzer was developed that can abstract feature spaces, and approximate ML classifiers, providing some miss-classified feature vectors to be used in the falsification process. A case study of autonomous driving cars is implemented, the proposed framework using DNN showed effective results. In another context, a DT architecture reference model using the state machine technique to design cloud-based CPSs is proposed [49]. Every physical thing is represented by a cloud-based DT of seven elements (sensors, actuators, functions, events, data storage, network, power unit). Sensing or actuating the physical thing is considered an event. All the data gathered by the smart things are stored at different levels of storage from mobile stationary to the cloud-based data center. Bayesian networks and fuzzy logic were used in this article to create a system that selects system modes of interaction based on event-triggered and physical things status.

In [50], The authors investigated and tested the Inductive Conformal Prediction (ICP) framework in CPS with ML components for assurance monitoring in real time. ICP framework provides predictions with well-calibrated confidence, these predictions are combined with a monitor that ensures a small error rate and limits the number of high-dimensional inputs at Deep Neural Network (DNN) model in cases where an accurate prediction cannot be made. Tests were made on two different datasets, different neural network architectures were used and different non-conformity functions were implemented. Results showed that the proposed architecture of using the ICP framework combined with DNN model can minimize the number of alarms due to predictions with multiple classes.

In Task learning, AI-ML algorithms such as RL can provide generalized task policy by training an agent to learn the task through different knowledge domains bases demonstrating the task. The authors in [51] proposed a neural network as a control strategy for Connected Vehicles (CV), the neural network model adjusts the speed of the following vehicle taking into account the distance, deceleration, and speed of the leading vehicle.

In [52], the authors proposed an algorithm based on RL that uses data from the manufacturing system to fix errors and differences in data while representing the DT of the CPS. The DT acts as the agent of the RL algorithm to ensure minimal policy. The algorithm proposed was tested in sheet metal assembly.

4. Conclusion

Given the different architecture proposals that integrate semantic models to ensure interoperability in communication between Cyber-Physical spaces and ML techniques to create virtual replicas of the physical world components, CPSs are a complex concept to design and implement, semantic models such ontologies can simplify the representation of data in a good manner and facilitate the construction of a DT of the physical reality.

The use of DT can increase the performance of CPSs effectiveness and simplifies their monitoring to end users by providing an appropriate digital representation of the physical system.

The use of ontology makes data management easy and communication easier, data and information about concepts are well-defined by properties and information flow between concepts is highlighted with relationships that ensure good communication and interpretation of all the data in the digital space. Several tasks can be accomplished such as anomaly detection with the use of an Ontology, a rule base can be used to reason about the ontology and extract implicit information about the physical system.

ML techniques serve in both controlling and monitoring the CPS. They can be a tool for constructing a DT by training a model from data gathered or generated from cyber-physical spaces. The DT-trained model predicts the behavior of the physical system from learning patterns using historical data gathered.

Generating data from simulation using physics-based computational models and feeding it to ML models can bring significance to results obtained from these ML models, by analyzing and mapping the results obtained from the trained model to physics-based parameters.

One of the relevant challenges in implementing ML in CPS is that data generated from the simulation are not sufficient. The use of a hybrid way to generate samples with the same features characteristic of real data gathered from reality and similar in distribution is required, at Last, ML techniques can approximate the behavior of the real system but there will be always a difference between real data and simulated data from DT.

Ontology modeling with ML techniques can be used to analyze and represent the semantic mismatch between real data and simulated data by calculating similarities and comparing their features and distribution. We believe that semantic models, DT, and ML together can serve as a bridge to create the visionary CPS and ensure knowledge interoperability when all three technologies are applied in a good way and in a refined architecture.

This survey performed a review of the CPS reference architectures (5C, RAMI 4.0, and IIRA) with a comparison between them. Furthermore, this article proposed a literature review of CPS architecture projects and an analysis of the use of semantic models and ML employed in CPS engineering and implementation.

For future works, combining semantic models and ML techniques together following I4.0 reference architectures design as an inspiration to implement CPS is considered, since it could bring a high level of integrity to these systems. In addition, an exhaustive and detailed architecture that takes into account all aspects of designing CPS using ML and Ontology together must be explored and defined as a unified standard solution to ensure interoperability between different industrial systems and achieve I4.0 vision.

References

- [1] F. Yang, S. Gu, Industry 4.0, a revolution that requires technology and national strategies, *Complex & Intelligent Systems* 7 (2021) 1311–1325.
- [2] J. Zhou, P. Li, Y. Zhou, B. Wang, J. Zang, L. Meng, Toward new-generation intelligent manufacturing, *Engineering* 4 (2018) 11–20.
- [3] G. Schuh, T. Potente, R. Varandani, C. Hausberg, B. Fränken, Collaboration moves productivity to the next level, *Procedia CIRP* 17 (2014) 3–8. URL: <https://www.sciencedirect.com/science/article/pii/S2212827114003709>. doi:<https://doi.org/10.1016/j.procir.2014.02.037>, variety Management in Manufacturing.
- [4] P. Dallasega, E. Rauch, C. Linder, Industry 4.0 as an enabler of proximity for construction supply chains: A systematic literature review, *Computers in industry* 99 (2018) 205–225.
- [5] F. Tao, Q. Qi, L. Wang, A. Nee, Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: Correlation and comparison, *Engineering* 5 (2019) 653–661.
- [6] E. A. Lee, S. A. Seshia, Introduction to embedded systems: A cyber-physical systems approach, Mit Press, 2016.
- [7] V. Gunes, S. Peter, T. Givargis, F. Vahid, A survey on concepts, applications, and challenges in cyber-physical systems, *KSII Transactions on Internet and Information Systems (TIIS)* 8 (2014) 4242–4268.
- [8] E. A. Lee, Cyber physical systems: Design challenges, in: 2008 11th IEEE international symposium on object and component-oriented real-time distributed computing (ISORC), IEEE, 2008, pp. 363–369.
- [9] R. Rajkumar, I. Lee, L. Sha, J. Stankovic, Cyber-physical systems: the next computing revolution, in: Design automation conference, IEEE, 2010, pp. 731–736.
- [10] H. Gill, A continuing vision: Cyber-physical systems, in: Fourth annual Carnegie Mellon conference on the electricity industry, 2008.
- [11] N. Carvalho, O. Chaim, E. Cazarini, M. Gerolamo, Manufacturing in the fourth industrial revolution: A positive prospect in sustainable manufacturing, *Procedia Manufacturing* 21 (2018) 671–678.
- [12] G.-J. Cheng, L.-T. Liu, X.-J. Qiang, Y. Liu, Industry 4.0 development and application of intelligent manufacturing, in: 2016 international conference on information system and artificial intelligence (ISAI), IEEE, 2016, pp. 407–410.
- [13] A. Redelinghuys, A. Basson, K. Kruger, A six-layer digital twin architecture for a manufacturing cell, in: International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing, Springer, 2018, pp. 412–423.
- [14] R. van Dinter, B. Tekinerdogan, C. Catal, Predictive maintenance using digital twins: A systematic literature review, *Information and Software Technology* (2022) 107008.
- [15] Q. Bao, G. Zhao, Y. Yu, S. Dai, W. Wang, Ontology-based modeling of part digital twin oriented to assembly, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 236 (2022) 16–28.
- [16] C. Steinmetz, A. Rettberg, F. G. C. Ribeiro, G. Schroeder, C. E. Pereira, Internet of things ontology for digital twin in cyber physical systems, in: 2018 VIII Brazilian symposium on computing systems engineering (SBESC), IEEE, 2018, pp. 154–159.

- [17] M. Kherbache, M. Maimour, E. Rondeau, When digital twin meets network softwarization in the industrial iot: Real-time requirements case study, *Sensors* 21 (2021) 8194.
- [18] Z. Huang, C. Jowers, D. Kent, A. Dehghan-Manshadi, M. S. Dargusch, The implementation of industry 4.0 in manufacturing: from lean manufacturing to product design, *The International Journal of Advanced Manufacturing Technology* 121 (2022) 3351–3367.
- [19] R. M. de Salles, F. A. Coda, J. R. Silva, D. J. dos Santos Filho, P. E. Miyagi, F. Junqueira, Requirements analysis for machine to machine integration within industry 4.0, in: 2018 13th IEEE International Conference on Industry Applications (INDUSCON), IEEE, 2018, pp. 1237–1243.
- [20] J. Lee, B. Bagheri, H.-A. Kao, A cyber-physical systems architecture for industry 4.0-based manufacturing systems, *Manufacturing letters* 3 (2015) 18–23.
- [21] K. Schweichhart, Reference architectural model industrie 4.0 (rami 4.0), An Introduction. Available online: [https://www.plattform-i40.de I 40](https://www.plattform-i40.de/I40) (2016).
- [22] M. Hankel, B. Rexroth, The reference architectural model industrie 4.0 (rami 4.0), *ZVEI* 2 (2015) 4–9.
- [23] S.-W. Lin, B. Miller, J. Durand, R. Joshi, P. Didier, A. Chigani, R. Torenbeek, D. Duggal, R. Martin, G. Bleakley, et al., Industrial internet reference architecture, *Industrial Internet Consortium (IIC), Tech. Rep* (2015).
- [24] M. Moghaddam, M. N. Cadavid, C. R. Kenley, A. V. Deshmukh, Reference architectures for smart manufacturing: A critical review, *Journal of manufacturing systems* 49 (2018) 215–225.
- [25] M. Djezzar, M. Hemam, M. Maimour, F. Z. Amara, K. Falek, Z. A. Seghir, An approach for semantic enrichment of sensor data, in: 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), IEEE, 2018, pp. 1–7.
- [26] T. R. Gruber, Toward principles for the design of ontologies used for knowledge sharing?, *International journal of human-computer studies* 43 (1995) 907–928.
- [27] F. Ortiz-Rodriguez, S. Tiwari, R. Panchal, J. M. Medina-Quintero, R. Barrera, MEXIN: Multidialectal Ontology Supporting NLP Approach to Improve Government Electronic Communication with the Mexican Ethnic Groups, 2022, p. 461–463.
- [28] A. Nikiforova, S. Tiwari, V. Rovite, J. Klovins, N. Kante, Evaluation and visualization of healthcare semantic models, *Evaluation* 323 (2020) 91773–5.
- [29] J. Domingue, D. Fensel, J. A. Hendler, *Handbook of semantic web technologies*, Springer Science & Business Media, 2011.
- [30] M. Hemam, M. Djezzar, Z. Boufaïda, Multi-viewpoint ontological representation of composite concepts: a description logics-based approach, *International Journal of Intelligent Information and Database Systems* 10 (2017) 51–68.
- [31] M. M. Taye, Understanding semantic web and ontologies: Theory and applications, *arXiv preprint arXiv:1006.4567* (2010).
- [32] V. Jirkovský, M. Obitko, V. Mařík, Understanding data heterogeneity in the context of cyber-physical systems integration, *IEEE Transactions on Industrial Informatics* 13 (2016) 660–667.
- [33] J. Hammer, D. McLeod, An approach to resolving semantic heterogeneity in a federation of autonomous, heterogeneous database systems, *International Journal of Intelligent and Cooperative Information Systems* 2 (1993) 51–83.

- [34] D. George, Understanding structural and semantic heterogeneity in the context of database schema integration, *Journal of the Department of Computing, UCLAN* 4 (2005) 29–44.
- [35] M. Grieves, J. Vickers, Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems, in: *Transdisciplinary perspectives on complex systems*, Springer, 2017, pp. 85–113.
- [36] E. Fontes, Digital twins and model-based battery design, 2019.
- [37] M. HAFIDI, Contribution au dépistage intelligent du cancer du sein basé sur la thermographie médicale (2020).
- [38] Y. Lu, M. R. Asghar, Semantic communications between distributed cyber-physical systems towards collaborative automation for smart manufacturing, *Journal of manufacturing systems* 55 (2020) 348–359.
- [39] J. Wan, B. Yin, D. Li, A. Celesti, F. Tao, Q. Hua, An ontology-based resource reconfiguration method for manufacturing cyber-physical systems, *IEEE/ASME Transactions on Mechatronics* 23 (2018) 2537–2546.
- [40] C. Zhang, W. Xu, J. Liu, Z. Liu, Z. Zhou, D. T. Pham, A reconfigurable modeling approach for digital twin-based manufacturing system, *Procedia Cirp* 83 (2019) 118–125.
- [41] X. Zheng, F. Psarommatis, P. Petrali, C. Turrin, J. Lu, D. Kiritsis, A quality-oriented digital twin modelling method for manufacturing processes based on a multi-agent architecture, *Procedia Manufacturing* 51 (2020) 309–315.
- [42] S. Singh, E. Shehab, N. Higgins, K. Fowler, D. Reynolds, J. A. Erkoyuncu, P. Gadd, Data management for developing digital twin ontology model, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 235 (2021) 2323–2337.
- [43] H. Li, J. Lu, X. Zheng, G. Wang, D. Kiritsis, Supporting digital twin integration using semantic modeling and high-level architecture, in: *IFIP International Conference on Advances in Production Management Systems*, Springer, 2021, pp. 228–236.
- [44] P. Semwal, A. Handa, Cyber-attack detection in cyber-physical systems using supervised machine learning, in: *Handbook of Big Data Analytics and Forensics*, Springer, 2022, pp. 131–140.
- [45] Q. Xu, S. Ali, T. Yue, Digital twin-based anomaly detection in cyber-physical systems, in: *2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST)*, IEEE, 2021, pp. 205–216.
- [46] T. Ritto, F. Rochinha, Digital twin, physics-based model, and machine learning applied to damage detection in structures, *Mechanical Systems and Signal Processing* 155 (2021) 107614.
- [47] Q. Min, Y. Lu, Z. Liu, C. Su, B. Wang, Machine learning based digital twin framework for production optimization in petrochemical industry, *International Journal of Information Management* 49 (2019) 502–519.
- [48] T. Dreossi, A. Donzé, S. A. Seshia, Compositional falsification of cyber-physical systems with machine learning components, *Journal of Automated Reasoning* 63 (2019) 1031–1053.
- [49] K. M. Alam, A. El Saddik, C2ps: A digital twin architecture reference model for the cloud-based cyber-physical systems, *IEEE access* 5 (2017) 2050–2062.
- [50] D. Boursinos, X. Koutsoukos, Assurance monitoring of cyber-physical systems with machine learning components, *arXiv preprint arXiv:2001.05014* (2020).
- [51] A. Sargolzaei, C. D. Crane, A. Abbaspour, S. Noei, A machine learning approach for fault

- detection in vehicular cyber-physical systems, in: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2016, pp. 636–640.
- [52] C. Cronrath, A. R. Aderiani, B. Lennartson, Enhancing digital twins through reinforcement learning, in: 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), IEEE, 2019, pp. 293–298.
- [53] M. Sabou, S. Biffl, A. Einfalt, L. Krammer, W. Kastner, F. J. Ekaputra, Semantics for cyber-physical systems: A cross-domain perspective, *Semantic Web* 11 (2020) 115–124.