A Survey and Perspective on Industrial Cyber-Physical Systems (ICPS): From ICPS to AI-Augmented ICPS

Jiyeong Chae, Sanghoon Lee, Junhyung Jang, Seohyung Hong, and Kyung-Joon Park, *Senior Member, IEEE*

***Abstract*—Digital Transformation integrates information technology across a broad spectrum of industrial sectors. Industrial Cyber-Physical Systems (ICPS) play a vital role in this transformation by harmonizing machinery, production, logistics, and societal needs through innovative information technologies. This article investigates the adoption of industrial artificial intelligence (industrial AI) as a methodology for effective ICPS design, introducing AI-Augmented ICPS (AICPS). The study conducts a survey, focusing on the components and interactions of AICPS. We propose design considerations for the implementation of AICPS and investigate the application of cutting-edge industrial AI techniques in each interaction. From the standpoint of AI augmentation, this article offers insights by identifying key perspectives, including uncertainty of information, safety of AI, explainability of AI, human-societal interactive ICPS, and standardization of industrial AI. This article aims to enhance understanding of AICPS and lay the groundwork for integrating independent industrial AI techniques into ICPS.**

***Index Terms*—Digital transformation, industrial artificial intelligence, industrial cyber-physical systems, industry 4.0, smart agriculture, smart city, smart factory.**

I. Introduction

With the proliferation of Internet of Things (IoT) and wireless network technologies, the concept of Digital Transformation has emerged. This paradigm aims to incorporate Information Technology (IT) across a multitude of industrial sectors, effecting structural and operational changes. At the core of Digital Transformation lies the emphasis on the utilization of industry-specific Cyber-Physical Systems (CPS) [1]. CPS are the systems that combine physical systems comprising machinery and operating environments with cyber systems consisting of computing devices and software via networks [2]. Although the term CPS can be applied to various domains, this article assumes that the term CPS pertains to industrial Cyber-Physical Systems (ICPS) within the industrial sector.

ICPS integrate diverse industrial sectors and societal requirements through novel information technologies to implement IoT and Internet of Services (IoS). The application of ICPS is accelerating in the fields of smart manufacturing, smart cities, and smart agriculture. In smart manufacturing, ICPS optimize complex production processes and facilitate the customization and diversification of product offerings [3]. In smart cities, ICPS manage resources for social, environmental, and economic sustainability for utilities and infrastructures [4]. In smart agriculture, ICPS focus on adaptive monitoring and management to boost food productivity [5].

One methodology for developing effective ICPS design is the adoption of industrial artificial intelligence (Industrial AI). The substantial amount of data generated within industrial systems makes it possible to leverage the use of industrial AI. Industrial AI is a sub-field that focuses on developing, verifying and deploying diverse AI methods for industrial use with sustainable performance. This encompasses the role of providing apt solutions customized to various industrial systems. Furthermore, it ensures that AI models developed through research are proficiently utilized to align with the demands of industrial practitioners [6].

Industrial AI is characterized by its focus on real-time processing to meet stringent security and reliability criteria, and its capacity to manage diverse, high-volume data from various industrial systems. It also integrates multiple forms of knowledge, mandates rigorous uncertainty management, and seeks specific value through strategic integration of industrial components [7]. The introduction of industrial AI in ICPS design can be regarded as an intelligent approach for ICPS, and we refer to ICPS with Industrial AI as AI-Augmented ICPS (AICPS). AICPS present an opportunity to effectively utilize the massive amounts of data generated in industrial environments. It primarily provides accurate information and insights compared to traditional approaches based on human intuition and experience, and it significantly aids in real-time decision-making and automation of industrial processes through data-driven predictive analysis.

Manuscript received 5 July 2023; revised 22 September 2023; accepted 26 September 2023. Date of publication 13 October 2023; date of current version 24 October 2023. This work was supported in part by the National Research Foundation of Korea (NRF) and in part by Korea government (MSIT) under Grant 2023R1A2C2003901. *(Jiyeong Chae and Sanghoon Lee contributed equally to this work.) (Corresponding author: Kyung-Joon Park.)*

Jiyeong Chae, Sanghoon Lee, and Kyung-Joon Park are with the Department of Electrical Engineering and Computer Science, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, South Korea (e-mail: cowldud3@dgist.ac.kr; leesh2913@dgist.ac.kr; kjp@dgist.ac.kr).

Junhyung Jang and Seohyung Hong are with the School of Undergraduate Studies, DGIST, Daegu 42988, South Korea (e-mail: jsy3535@dgist.ac.kr; tjgud1906@dgist.ac.kr).

Digital Object Identifier 10.1109/TICPS.2023.3323600

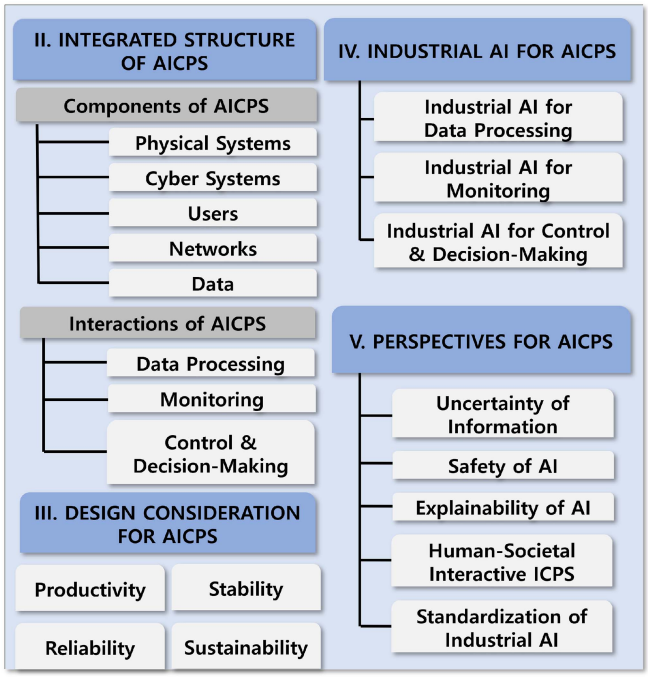

Fig. 1. Structure and relations of the sections.

AICPS cover a wide range of industrial sectors, and there are diverse applications of industrial AI. In the context of ICPS, physical and cyber systems and networks that constitute ICPS are typically developed independently. Similarly, the industrial AI techniques applied to industrial sectors are also being studied and developed independently. Therefore, a systematic investigation of the combination of ICPS and industrial AI is crucial to introduce AICPS, which integrates independent industrial AI techniques. In this context, this article presents the following key contributions:

- We propose an integrated structure that encompasses the components and interactions of AICPS, along with a set of design considerations essential for the effective implementation of AICPS.
- We conduct a review of the state-of-the-art research in industrial AI and offer insights into the techniques deployed for each interaction within AICPS to meet diverse design considerations.
- We provide substantive guidance to researchers for the successful deployment of AICPS. This guidance is informed by insights derived from five critical perspectives and aims to fulfill essential design considerations.

Fig. 1 illustrates the overall structure and relationships of the sections in this article. In Section II, we analyze the components of AICPS and their interactions to introduce the integrated structure of AICPS. In the following Section III, we explain the design considerations for AICPS. In Section IV, we review the state-of-the-art industrial AI research. We provide insights into the industrial AI techniques applied to each interaction of AICPS. In Section V, we outline research perspectives crucial for AICPS deployment, targeting essential design considerations and future research milestones. Finally, in Section VI, we conclude the research.

II. Integrated Structure of AI-Augmented Industrial Cyber-Physical Systems

In this section, we describe the components of AICPS and analyze their interactions. We present a schematic of the integrated structure of AICPS. AICPS can be structured by its components and their interactions. The integrated structure of AICPS is shown in the following Fig. 2. AICPS are composed of physical systems, cyber systems, users, networks and data, which interact and function together in an organic manner. Physical systems and cyber systems communicate through the networks, exchanging data. The data is processed by users (experts or intelligent agents). Users monitor physical systems based on the information obtained from the data and cyber systems, and make control or decision for other components.

A. Components of AICPS

This section describes the components of AICPS and categorizes them into physical systems, cyber systems, users, networks, and data. Additionally, the main examples of each component are provided and explained.

1) Physical Systems: In AICPS, physical systems refer to a type of machine or device that operates according to physical laws. These physical systems generate data through various sensors and exchange data with cyber systems through networks. Physical systems are primarily the targets of control or decision-making by users.

1) *Control systems:* In industrial environments, control systems include sensors, actuators, and controllers. Sensors gather data and interface with controllers, which regulate actuators. These systems have modernized to incorporate IoT and intelligent computing, becoming central to Industry 4.0.
2) *Infrastructures:* In industrial systems, infrastructures refer to the physical facilities that facilitate their operation. Physical infrastructures include roads, electricity, telecommunications, water supply, and sewage systems. Digital infrastructures consist of hardware such as servers, storage, and networks. These infrastructures must offer stability and reliability, often operating in tandem with sensor networks.
3) *Environments:* In industrial contexts, environments encompass the external conditions in which control systems and infrastructures operate. These may range from the climatic and geographical factors affecting the systems, to the specific needs of specialized applications such as agriculture or livestock management. Environments primarily serve as the subjects of monitoring through sensor networks.

2) Cyber Systems: Cyber systems are composed of combinations of software and hardware components within AICPS. These cyber systems collect and analyze sensor data from physical systems as well as process and management data, providing monitoring information to users. Based on data analysis

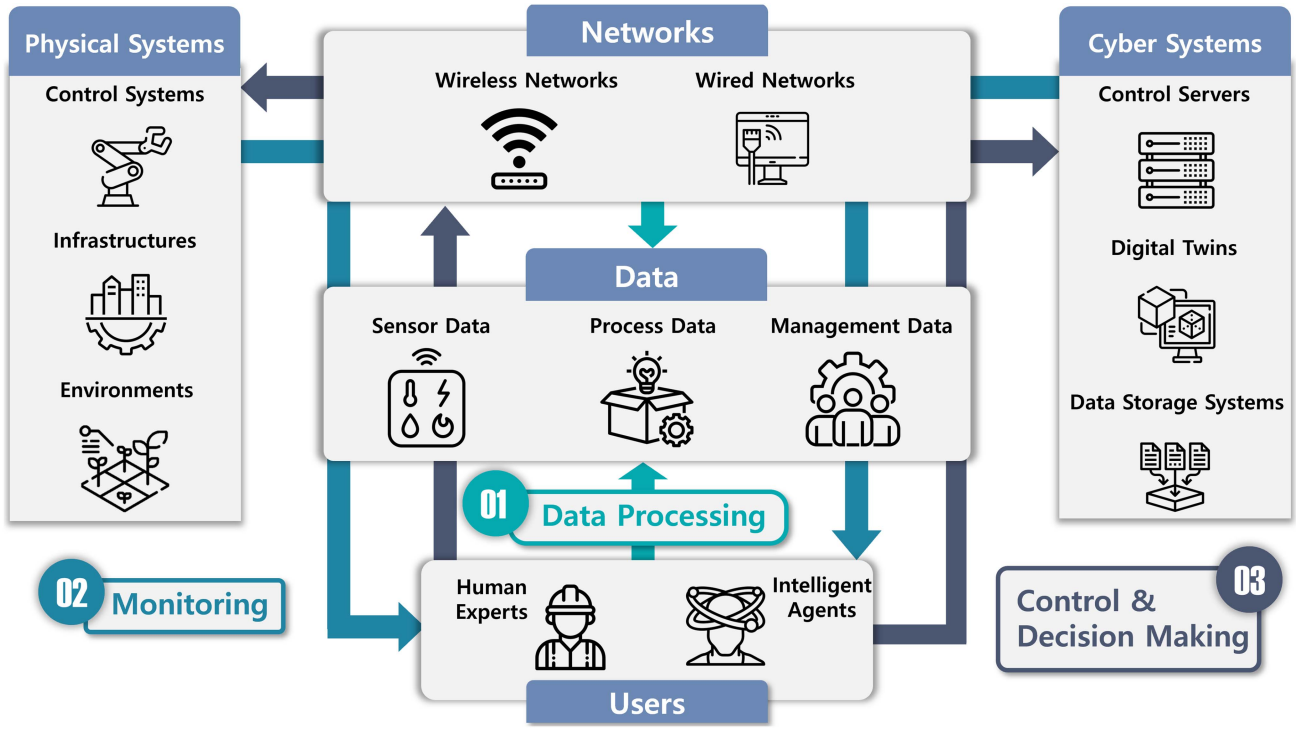

Fig. 2. Integrated structure of AI-augmented industrial cyber-physical systems.

and user decisions, cyber systems generate appropriate control commands and transmit them to the physical systems. Key components of cyber systems include control servers, digital twins, and data storage.

1) *Control Servers:* Control servers serve as interfaces between cyber and physical systems, monitoring the statuses of physical systems and executing control algorithms to transmit control commands to the physical systems. Recently, control servers have been implemented using cloud services, allowing for dynamic allocation of server resources to achieve fast data processing speeds and large-scale data processing capabilities.
2) *Digital Twins:* Digital twins are cyber copies of physical systems. They integrate the latest information from both physical and digital models in real-time through bi-directional data exchange [8]. Digital twins enable real-time monitoring and continuous interaction with physical systems, leading to informed decision-making [9]. They find applications in areas like product design, process optimization, monitoring, maintenance planning, fault detection, and system prediction [10], [11].
3) *Data Storage Systems:* Data storage systems are crucial for storing large volumes of data in AICPS. Choosing the right data storage systems with intelligent access is vital for effective data utilization. Data lake systems, such as the Hadoop distributed file system, have gained popularity recently. These data lakes store raw data, allowing for diverse analyses with scalability, flexibility, and cost-effectiveness [12].

3) Users: Users can be construed as types of either physical systems or cyber systems. We define the term "Users" as components that serve as the principal agents of interactions within AICPS. Users are capable of furnishing the cyber systems with pertinent information by inputting or processing application or management data, thereby indirectly controlling both the physical and cyber systems. Moreover, users have the ability to monitor the entirety of the system and execute direct decision-making and control actions based on the information relayed via the cyber systems. Within the context of AICPS, users are categorically bifurcated into human experts and intelligent agents.

1) *Human Experts:* As elements of the physical systems, human experts continue to hold a pivotal role in industry. They employ their professional expertise to process information and make decisions [13].
2) *Intelligent Agents:* Owing to the deployment of intelligent algorithms, intelligent agents, as components of cyber systems, are progressively assuming critical decision-making roles in industrial processes and increasingly supplanting human experts in key industrial decisions. Intelligent agents constitute AI-based autonomous entities that acquire data from sensors, execute decisions through AI methods, and accomplish tasks to achieve objectives efficiently, independent of human intervention [14].

4) Networks: Networks are crucial components for data exchange among various components within AICPS. Compared to conventional networks, those designed for ICPS demand elevated levels of real-time performance, safety, and decentralization. The network components can be broadly categorized into wireless and wired networks.

1) *Wired networks:* Wired networks are a communication method known for their stability, security, and broad

bandwidth. They resist radio interference, bolster unauthorized access prevention, and enable rapid data transmission. Examples of wired networks include Ethernet networks and fiber optic networks.

2) *Wireless networks:* Wireless networks offer flexibility, low cost, and adaptability. They eliminate the need for wired connections, allowing for easy location adjustments and reduced maintenance costs. They are cost-effective for communication over large areas and can operate in diverse environments. Wireless networks are commonly used for wireless sensor networks and communication between IoT-based industrial equipment and servers.

5) Data: In AICPS, data are important resources generated by all components and exchanged among them. These data are utilized for the control of physical systems, monitoring interactions between physical and cyber systems, and the analysis and optimization of industrial processes. Data collected within AICPS exhibit different characteristics depending on their sources, such as physical systems, cyber systems, networks and users. Data components can be broadly categorized into sensor data, process data, and management data.

1) *Sensor Data:* Sensor data, consisting of physical measurements from various environments and equipment, are crucial across multiple sectors. In control systems, sensors collect physical measurements from machinery. In infrastructures, sensors collect consumption data for efficiency assessments. In environments, sensors measure factors such as air quality, humidity, crop growth conditions, and soil moisture.

2) *Process Data:* Process data are derived from a range of ICPS applications encompassing manufacturing, urban management, and agriculture. In the manufacturing domain, metrics related to production volume and operational stages are collated. Within the context of urban management, data pertaining to traffic patterns and energy utilization are aggregated. In the agricultural sphere, information on crop growth rates and water supply metrics are compiled.

3) *Management Data:* Management data are employed to support the operations and administration across various sectors. In manufacturing, these data include equipment maintenance records and production line performance metrics. In the context of urban management, the data cover infrastructures maintenance and public facility operational statistics, along with energy consumption and management figures. In agriculture, the data encompass crop production schedules, harvest records, and farm inventory management metrics.

B. Interactions of AICPS

This section describes the interactions of AICPS and categorizes them into data processing, monitoring, control & decision-making.

1) Data Processing: The data generated by each component of AICPS has different characteristics; therefore, data processing is necessary to analyze and integrate data from each component. Data processing involves refining, processing, and analyzing data to transform it into meaningful information. In the field of data processing research, studies applying intelligent approaches are being conducted. AI techniques can be utilized to address major challenges of industrial data, such as high dimensionality, noise, and data loss [15]. Additionally, AI techniques can be applied in authentication frameworks to enhance the security of data transmission.

2) Monitoring: Monitoring encompasses the collection of sensor data, process data, and management data from physical systems, cyber systems, networks, and users. Subsequently, data mining techniques are employed to address an array of challenges, including quality management, production time prediction, processing time prediction, and defect identification [16].

In particular, when designing holistic monitoring systems for industrial infrastructures, it's crucial to emphasize not only traditional factors like organizational structure, operational methods, and objectives but also the human and social dimension. [17]. AI techniques enable the prediction of complex states in physical systems and the early detection of faults or abnormal states. Moreover, AI can be utilized to forecast quality based on data flow and detect malicious attacks [18], [19].

3) Control & Decision-Making: Control and decision-making aim to optimize the performance of physical systems and support decision-making in the overall process to maximize the effectiveness of AICPS. Examples of control include scheduling in allocation of resources [20], routing path of industrial sensor networks [21]. In control and decision-making, intelligent approaches enable fast decision-making that meets diverse demands, optimization of control in complex systems, and prevention of disturbance in the flow of control and materials [22].

III. Design Consideration for AICPS

Several research endeavors are underway to accelerate the adoption of AICPS. To facilitate the effective design of AICPS, the following design considerations must be taken into account: productivity, stability, reliability, and sustainability. In this section, we discuss the design considerations that need to be taken into account for AICPS.

A. Productivity

The primary objective of most industrial applications is to generate economic benefits. From this perspective, productivity becomes the most critical factor in AICPS design. In AICPS, productivity refers to the reduction of economic and human costs in industrial applications, achievable through the implementation of AICPS.

This encompasses the meaning of proactive productivity through process optimization based on data analytics in industrial environments, as well as the meaning of reactive productivity aimed at facilitating smooth process operations by fault detection and minimization of defects in the manufacturing process. Researchers suggest designing AICPS with the goal

of enhancing productivity, aiming to improve product quality while reducing development costs and time [23].

B. Stability

Due to the complexity of software and physical systems, detecting defects in advanced industrial systems is challenging. Late detection of defects can result in decreased overall production and financial losses. Furthermore, incorrect monitoring results caused by data analysis or human error can lead to unnecessary maintenance costs, thus increasing maintenance and management expenses. In the context of AICPS, stability refers to ensuring continuous operability through its implementation.

This encompasses the meaning of stability in detecting, predicting, and preventing failures in industrial systems, as well as the meaning of control stability in certain components, such as physical systems. Moreover, it could include the meaning of data stability by resolving issues of data incompleteness and scarcity in industrial environments, thereby enhancing decision-making derived from data analysis. Researchers propose designing AICPS with a focus on stability to detect, predict, and prevent system defects and malfunctions [24].

C. Reliability

AICPS face various cybersecurity and physical security threats due to malicious access through computer nodes and communication networks. Since CPS handle large-scale data and have organic interconnections between components, malicious access to some components can have hazardous consequences on all components [25]. Therefore, it is crucial to recognize the security vulnerabilities of CPS and implement robust security measures. In the context of AICPS, reliability refers to enhancing the security of industrial systems through the adoption of AICPS.

This encompasses the meaning of reliability for ensuring secure and reliable data transmission against unauthorized external access in device and network environments [26], as well as the meaning of reliability imbued with resilience, as control systems and dynamic systems respond to malicious external attacks. Researchers propose that AICPS, designed with reliability in mind, can preclude malicious access and detect and prevent malicious attacks [27].

D. Sustainability

AICPS have received positive evaluations in the aspect of energy sustainability by facilitating low-carbon energy supplies, improved energy efficiency, energy storage systems, and robust energy management frameworks [28]. Furthermore, for social sustainability, user-centered engineering aimed at enhancing human convenience [29], and frameworks considering social ethical issues are proposed [30]. In the context of AICPS, sustainability refers to involving both energy sustainability and social sustainability.

This encompasses the meaning of sustainability for minimizing environmental impact, enhancing energy efficiency, and improving human quality of life and ethical considerations. Researchers propose that AICPS, when designed with a focus on sustainability, can lead to reduced environmental pollution, improved energy efficiency, enhanced human convenience, and contribute to genuine, long-term sustainability.

IV. Industrial AI for AICPS

In this section, our aim is to investigate the forefront of industrial AI techniques for AICPS. We categorize AI techniques into data processing, monitoring, and control & decision-making. We review the technical categories, specific applications, used AI methods, AICPS components, and design considerations of 51 studies published between 2016 and 2023.

The "Technique" section of the table refers to the technical category of research. The "Application" section indicates the specific application area of research. The "Method" section provides detailed AI techniques utilized in each study. In cases where the research involved a comparison and investigation of the performance of various AI techniques, the best-performing technique is selected. The "AICPS Component" and "Design Consideration" section provides details about AICPS components and design considerations pursued by each research, which are explained in Section III.

A. Industrial AI for Data Processing

Table I presents a summary of 11 recent studies on industrial AI for data processing. We have categorized data processing techniques into four main areas: authentication, data compression, data imputation, and soft sensing.

1) Authentication: Ensuring secure communication is a growing concern for data privacy. AI approaches are being utilized in the field of authentication frameworks to strengthen the security of data transmission. In [32], a deep learning-based framework is proposed for enhancing the security of industrial wireless sensor networks through physical layer authentication. It utilizes channel state information (CSI) values and higher-layer protocol authentication labels like EAP and AKA to train a deep learning model. Experimental evaluations involving three deep learning methods show that the Prepossessed Convolutional Neural Network (CNN) method exhibits superior authentication performance. Furthermore, in [31], a lightweight intelligent authentication approach is introduced to address isolated IoT security design across different layers. It employs an Support Vector Machine (SVM)-based classifier to verify the compatibility of access time slots, access frequency bands, and designed codes of IoT devices.

2) Data Compression: In industrial environments, data collected often exhibits high dimensionality, which can hinder real-time applications and increase substantial computational and communication costs. Data compression refers to the techniques used to reduce the size of data while preserving essential information. Recently, AI techniques for data compression have been under investigation, aiming to enhance the efficiency of transmission in industrial settings.

The research by [34], [35] utilizes Variational Auto-Encoder (VAE) and Generative Adversarial Network (GAN) techniques for data encoding and decoding. VAE is a probabilistic generative model known for its ability to capture the underlying

TABLE I
INDUSTRIAL AI FOR DATA PROCESSING

Technique	Application	Method	AICPS Component	Design Consideration
Authentication	Holistic authentication [31] PHY-layer authentication [32]	CNN [32] SVM [31]	Wireless networks [31], [32]	Reliability [31], [32]
Data Compression	Compressed sensing [33] Data encoding/decoding [34], [35]	CAE [33] GAN [35] VAE [34]	Wireless networks [33]–[35]	Stability [33]–[35]
Data Imputation	Electric motors [36] Power consumption [37], [38] Water quality [39]	KNN [37] LSTM [38], [39] Ridge Regression [36]	Process data [37], [38] Sensor data [36], [39]	Productivity [36] Stability [36]–[39] Sustainability [37], [38]
Soft Sensing	Plastic process [40] Sulfur recovery unit [41]	CNN [40] LSTM [41] RNN [41]	Process data [40] Sensor data [41]	Productivity [40], [41]

structure of data and generate new data samples that resemble the input data. On the other hand, GAN consists of two neural networks, a generator and a discriminator, engaged in a game where the generator aims to produce data that is indistinguishable from real data, while the discriminator tries to tell real from fake. This adversarial training process results in the generation of realistic data samples. In contrast, [33] applies Convolutional Auto-Encoder (CAE) to the compressed sensing technique, which aims to recover the original signal with fewer samples. This approach outperforms conventional methods, demonstrating superior results with reduced data samples.

3) Data Imputation: Data loss is a common problem in industry fields and can occur due to communication failures, malfunctioning equipment, errors in data recording, or insufficient data collection personnel. Data imputation refers to the techniques of predicting and substituting missing data to construct a complete dataset.

[36], [37], [38], [39] investigate data imputation for addressing the issue of data loss. In [37], K-Nearest Neighbors (KNN) is used to estimate missing power data in a consumption system by creating feature vectors from historical and missing power data differences, considering patterns. Deep learning-based data imputation, such as the sequence-to-sequence imputation model (SSIM) in [39], utilizes bidirectional Long Short Term Memory (LSTM) networks and variable-length sliding window algorithms for generating training samples. LSTM networks excel at capturing information from both past and future time indices. Additionally, [38] proposes bidirectional imputation based on LSTM and transfer learning, replacing missing data using models trained with data from other systems.

In [36], a comparison of machine learning-based data imputation methods is conducted on missing vibration and current sensor data from electric motors. The study assesses models like Support Vector Regression, Decision Tree Regression, Ridge Regression, KNN, MissForest, and XGBoost Regression, with Ridge Regression proving the most effective.

4) Soft Sensing: Data scarcity poses challenges in terms of imbalanced data and accuracy of data-driven analysis. In industrial processes, quantity variables are typically sampled at a fast rate, while quality variables are measured infrequently. Soft sensor refers to the techniques that estimates difficult-to-measure variables using the dependency of easy-to-measure variables.

Industrial process datasets frequently comprise time series data underutilized by traditional soft sensors. Deep learning-based soft sensor methods have been developed to better capture these temporal characteristics. In [41], a soft sensor model based on Recurrent Neural Network (RNN) and LSTM is proposed. This model exhibits high prediction accuracy for data measured at different time periods using transfer learning techniques. Simulation results with sulfur recovery unit data indicate that the RNN technique outperforms LSTM in typical situations, but LSTM performs better for limited dataset. On the other hand, [40] introduces the Gated Convolutional neural network-based Transformer (GCT). GCT encodes short-term patterns in time series data to filter important features. When applied to industrial processes like polypropylene and purified terephthalic acid, GCT outperforms traditional approaches combining LSTM and CNN.

B. Industrial AI for Monitoring

Table II presents a summary of 26 recent studies on industrial AI for Monitoring. We have categorized monitoring techniques into six main areas: defect detection, fault detection, fault prediction, human activity recognition, malicious attack detection, and quality prediction.

1) Defect Detection: Numerous research studies focus on the detection of defects or faults within industrial systems. We classify these into two categories: component-level defect detection and system-level fault detection. Defect detection pertains to the identification of flaws occurring in specific components of the overall system. Such defects can result in reduced productivity, increased costs, and, in severe cases, catastrophic failures leading to potential fatalities. The application of industrial AI in real-time defect detection is presently an active field of research.

In [42], an artificial neural network (ANN) is employed to detect insulation failure in stator winding, a fault that contributes to 37% of all machine failures. This early detection is crucial as insulation failure can lead to stator inter-turn faults, resulting in performance degradation or motor failure. Similarly, in [43], a CNN-based defect inspection method is proposed for the early identification of injection molding faults. Injection molding, which uses heated polymers to form shapes, can affect the quality of the product, making precise process control and defect inspection essential. Furthermore, [44] introduces a physics-based

TABLE II
INDUSTRIAL AI FOR MONITORING

Technique	Application	Method	AICPS Component	Design Consideration
Defect Detection	Stator winding [42] Injection molding [43] Rolling element bearing [44]	MLP [42] CNN [43], [44]	Control systems [42]–[44]	Productivity [42], [43] Stability [42], [44]
Fault Detection	Nuclear power plant [45], [46] Electric arc system [47] Wastewater treatment [48] Oil&Gas pipeline [49], [50]	DNN [45] CNN [46], [47], [49] Deep Clustering [48] LSTM [49], [50] SVM [50]	Control systems [45] Infrastructures [46]–[50]	Stability [45]–[50]
Fault Prediction	Conveyor operation [51] Relative humidity [52] Reliability of a cylinder [53] Power outage [54]	CNN [51] DNN [53] Collaborative NN [54] RF [52]	Control systems [51], [53], [54] Environments [52], [54]	Stability [52]–[54] Productivity [51], [53]
Human Activity Recognition	Worker activity [55], [56]	CNN [55], [56] SVM [55]	Human experts [55], [56]	Sustainability [55], [56]
Malicious Attack Detection	DDoS/ DoS attacks [57], [58] Spoofing attacks [59] Advanced persistent threat [60]	MLP [57] CNN [58]–[60]	Wireless networks [57]–[60]	Reliability [57]–[60]
Quality Prediction	Gas emission [61]–[63] Product quality [64]–[66] Water quality [67]	CNN [66] DeepFM [65] SVM [64], [67] RF [61]–[63]	Control systems [64]–[66] Environments [61], [62], [67]	Productivity [64]–[67] Sustainability [61], [62], [64], [67]

CNN technique for diagnosing defect in rolling element bearings. This method utilizes vibration signal data and incorporates the physical characteristics of the bearings to effectively detect defects.

2) Fault Detection: Fault detection is more critical than detecting internal defects in products or equipment. The malfunction of physical systems poses potential risks to human life, the environment, and property. Therefore, research is being conducted on alarm systems for fault detection in physical systems. In [45], an online fault monitoring system is proposed for nuclear power plants. This system employs Deep CNN and the sliding window technique to enhance fault monitoring during plant operation. The sliding window technique prioritizes current data over past data streams, allowing for dynamic error diagnosis using faster simulators to predict the plant's actual status.

In [47], researchers introduce a CNN-based model for detecting arc faults, known for their electrical hazards due to high temperatures. This detection model is designed to classify normal and abnormal states of load currents without the need for additional transformation. On the other hand, in [48], researchers propose a system for detecting abnormal data in wastewater treatment using deep clustering. The method employs a self-supervised deep clustering network capable of extracting nonlinear features and identifying normal patterns from unlabeled data.

Researchers in [46], [49], [50] propose various methods for pipeline leak detection. In [46], they implement a CNN-based approach using trajectory-based image features derived from time-series acoustic data. Similarly, in [49], effective techniques are presented, including 2D CNN and LSTM-AE, which convert time-series data into spectrograms for improved accuracy. In [50], a semi-supervised method combines LSTM-AE with a one-class SVM to address data scarcity challenges, enabling precise leak detection by learning essential pipeline features.

3) Fault Prediction: Beyond fault detection, predicting and preventing faults can reduce the time and costs associated with maintenance. This concept is commonly referred to as predictive maintenance. Numerous researchers employ AI techniques to predict and prevent faults in various industries.

In [51], a predictive maintenance framework is proposed for conveyor systems. The framework utilizes time-series imaging and CNN for data classification, accurately predicting the three levels of faults (integrity, minor fault, critical fault) in the conveyor. In [52], researchers propose a predictive maintenance system to prevent system failures caused by high humidity. High humidity can lead to various problems in electrical and mechanical systems, such as metal corrosion, moisture condensation, and bacterial growth. The system utilizes data collected from IoT and applies the Random Forest (RF) to predict relative humidity. In [53], researchers proposes a system to evaluate the reliability of factory equipment using DNN techniques applied to time-series equipment data collected in factories. The method are tested on the reliability of a cylinder, a critical component of a trolley in the automotive assembly line. The study in [54] proposes a power outage prediction system for industrial infrastructures considering unpredictable weather conditions, which can result in malfunctions and failures in power supply devices within industrial environments. The proposed system utilizes the Collaborative Neural Network (NN) to address the power outage prediction problem by transforming it into two separate sub-problems that can be simultaneously solved.

4) Human Activity Recognition: In the field of industrial manufacturing, human activity recognition is one of the key technologies from a social sustainability perspective. In this context, [55] proposes a solution for activity recognition and detection by predicting work-related features and objects in recorded videos using CNN and SVM. The proposed solution demonstrates its utility through the verification of task accuracy in solid fuel boiler equipment. [56] employs a multi-modal

approach using both Inertial Measurement Unit (IMU) data from smart armbands and video data from thermal cameras, predicting worker activities via CNN. In the evaluation of six common activities in assembly work, the approach exhibits a recognition accuracy of 97%.

5) Malicious Attack Detection: With the increasing likelihood of cyber attacks, vulnerabilities in industrial communication systems can lead to critical malfunctions or system-wide paralysis. Industrial AI research focuses on developing effective malicious attack detection systems for industrial communications.

In industrial communication systems, Distributed Denial of Service (DDoS) attacks can have a significant negative impact. In [57], the authors propose an Multi-Layer Perceptron (MLP)-based method to address IoT network intrusions caused by DDoS/DOS attacks. They evaluate its performance by training MLP on internet packets and successfully identifying different attack types. In [58], the authors use ResNet, which is a type of CNN, to detect DDoS/DoS attacks in 5G networks.

Additionally, researchers are exploring anomaly detection systems for various malicious attacks. MAC spoofing is an ID-based attack in wireless networks. The emergence of virtual MAC spoofing has made detecting such attacks more challenging. [59] presents a system using CSI and deep CNN techniques to detect virtual MAC spoofing. It extracts physical information like amplitude and phase from CSI acquired during packet transmission and trains CNN to classify devices in the same location with high accuracy. To detect Advanced Persistent Threats, which persistently infiltrate systems and steal information, [60] employs CNN. They propose applying ResNet to consortium block-chain, ensuring secure data transmission and maintenance in industrial systems, for effective APT attack detection.

6) Quality Prediction: AI technique can be utilized to predict information about quality that cannot be directly measured through sensors in complex industrial processes or is not immediately available. Various industrial AI research has been conducted with the aim of quality prediction. To adhere to regulatory guidelines and mitigate pollutant emissions, studies such as [61], [62], [63] employ RF algorithm to predict the concentration of specific elements in gas emissions.

The research in [61] focuses on predicting methane concentrations in shale gas fields for greenhouse gas emission measurement. [62] monitors odor concentrations and grades at wastewater treatment plants. Similarly, [63] predicts odor concentrations based on quantitative data from compounds emitted in urban areas, identifying odor emission sources. Predicting pollutant emissions is essential for both sustainability and productivity improvement. In [67], SVM is used to predict water quality in industrial aquaculture for systematic feed supply and water quality management. Additionally, in [64], SVM predicts NOx emissions in the air to classify coal combustion quality in thermal power plants.

Traditionally, the assessment of product quality has been dependent on offline laboratory analysis, posing challenges for real-time impact on production processes. Hence, real-time product quality prediction can significantly enhance production quality. The research by [65], [66] focuses on predicting product quality in complex manufacturing systems. [65] employs Deep Factorization Machine (FM) to predict refined quality from mineral purification process data. Moreover, [66] predicts steel industry product quality using a 1D-CNN-based Multi-Objective Ensemble Learning approach.

C. Industrial AI for Control & Decision-Making

Table III presents a summary of 14 recent studies on industrial AI for control & decision-Making. We have categorized control & decision-making techniques into four main areas: parameter optimization, production scheduling, data-driven control, and task & resource allocation.

1) Parameter Optimization: To maximize efficiency and maintain stability in systems and processes, optimal parameter settings are essential. Research on parameter optimization through AI in the industrial sector has become increasingly active. Industrial AI can tune control parameters of controllers of industrial systems. [68] propose an AI-based optimization technique for finding optimal parameters for controlling an industrial robot arm. The researchers use ANN to model a Genetic Algorithm (GA) and apply the GA to calculate optimized parameters for gimbal joints. Gimbal joints are joints that can be adjusted directly by users to create fine movements of an industrial robot arm, and are expected to replace existing rotational joints. Simulation results demonstrate that ANN-based optimization can improve the performance of complex robot joint control.

The research conducted by [69], [70] focuses on optimizing the parameters of 3D printing using AI techniques. [69] utilizes RF to enhance production efficiency while ensuring desired interfacial performance. On the other hand, [70] integrates various machine learning methods with the GUI of a 3D printer to assist users in selecting the optimal parameters when using the printer. A comparative study of parameter optimization performance across eight ML algorithms revealed that the Gradient Boosting (GB) method exhibited the best performance.

2) Production Scheduling: Recently, attempts are made to dynamically solve scheduling problems using AI technique and real-time data. Studies of [71], [72], [73], [74] utilize industrial AI technology for scheduling in manufacturing systems. The use of industrial AI enables flexible production scheduling by effectively processing dynamic events that are difficult to predict in real-time. In [71], an AI scheduler employing Q-learning and composite reward functions is proposed to achieve real-time production scheduling of manufacturing operations. Real-time status tracking of the order system and machine processing system is achieved through an internet-supported sensor network in a smart factory, which is then used by the AI scheduler to efficiently schedule production schedules. In [72], the scheduling of systems that consider various variables such as processing time, priority, and transportation time is more accurately and efficiently calculated using Fuzzy Inference System, an AI technique based on fuzzy logic.

In [73], an adaptive scheduling system using closed-loop adaptive scheduling for manufacturing systems is proposed to improve adaptability. Compared to existing dynamic scheduling, a scheduling solution composed of offline learning through GA

TABLE III
INDUSTRIAL AI FOR CONTROL & DECISION-MAKING

Technique	Application	Method	AICPS Component	Design Consideration
Parameter Optimization	Gimbal joint controller [68] 3D-printing [69], [70]	ANN [68], [69] GA [68] GB [70] RF [69]	Control systems [68]–[70]	Stability [68] Productivity [69], [70]
Production Scheduling	Flexible manufacturing system [71]–[74]	Q-learning [71] Fuzzy Logic [72] GA [73] KNN [73] DDQN [74]	Control systems [71]–[74]	Productivity [71]–[73] Stability [73], [74]
Data-driven Control	Distillation column system [75] pH control system [76] Power system [77]	AC [76] ANN [77] GA [75] RNN [75]	Control systems [75]–[77]	Productivity [75] Stability [75] Sustainability [76], [77]
Task & Resource Allocation	Multichannel access [78] Cloud edge service [79], [80] Wireless sensor network [81]	DDPG [78] DQN [79] A3C [80] DNN [81]	Wireless networks [78], [81] Control servers [79], [80]	Productivity [78]–[80] Stability [81]

and online adjustment through KNN is used to autonomously adjust scheduling rules and enable flexible production scheduling. It can also adapt to various disturbance scenarios, such as machine failures, urgent orders, and changes in uncertain processing and delivery times. In [74], a distributed and hierarchical approach for real-time dynamic scheduling is proposed. After training the scheduling agent with the Double Deep Q-Network (DDQN) technique, real-time scheduling decisions are made by utilizing the difference between production information and scheduling objectives.

3) Data-Driven Control: Data-driven control is a control approach that has been studied in recent years and is being used to control complex nonlinear systems. [75] propose a data-driven control system for a batch distillation column that uses a RNN to design an emulator for physical systems and tunes the plant controller using GA and Particle Swarm Optimization techniques, unlike conventional PID control or human manual operation for controlling physical plants. The system showed significant benefits in product quality and energy consumption compared to the conventional PID control in a closed-loop simulation.

The study by [76] proposes a data-driven autonomous pH controller using the Actor-Critic (AC) algorithm for neutralizing acidic pH wastewater generated in the electroplating industry. The proposed controller outperforms conventional PID controllers in stabilizing the pH of effluents within a neutral range across various scenarios. [77] aims to reduce the computational demand in real-time control of power converters by applying ANN. Experimental results indicate that the ANN-based model predictive control approach maintains control performance while significantly reducing computational complexity.

4) Task & Resource Allocation: In Industry 4.0, cloud computing or edge computing is used to increase network efficiency, but as the number of channels increases, accessibility and reliability decrease in edge computing. In [78], a multi-channel access and task offloading algorithm is proposed using multi agent deep reinforcement learning to reduce computation latency and increase channel access success rate. Experimental results show that the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm outperforms other techniques like DQN and AC in terms of channel access success rate and channel utilization rate.

Resource allocation scheduling can increase the sustainability of AICPS and improve productivity by effectively managing resource utilization. Studies such as those by [79], [80], [81] employ AI techniques for resource allocation in sensor networks or distributed edge systems. Given limited computing and storage resources, efficient service placement in edge clouds is necessary for processing large amounts of data. Especially in real situations where service demand is uncertain, determining which service to place on each edge node for optimal resource allocation is crucial. In [79], convex optimization and DQN based algorithms are proposed for joint optimization of service placement, workload scheduling, and resource allocation in industrial environments. In [80], the Asynchronous Advantage Actor-Critic (A3C) technique is employed to optimize resource allocation and minimize waiting time in hybrid network paradigms that integrate cloud computing with edge computing.

Furthermore, energy-efficient wireless sensor networks are required for efficient data exchange between IoT devices. Consequently, [81] proposes a resource allocation system using deep learning that meets energy efficiency in wireless sensor networks. The whale-optimization-based DNN technique is utilized for network energy efficiency, improving the performance of power allocation optimization by reducing the amount of total transmission power.

V. PERSPECTIVES FOR AICPS

The integration of industrial AI and ICPS can yield benefits of productivity, stability, reliability, and sustainability to various industrial processes. However, several challenges remain to be addressed for the practical application of AICPS in industrial fields. In this section, we outline the research perspectives that must be addressed for the successful establishment of AICPS in the industrial sector. Fig. 3 represents the perspectives for AICPS design. We identify five main perspectives: uncertainty of information, safety of AI, explainability of AI, human-societal interactive ICPS, and standardization of industrial AI.

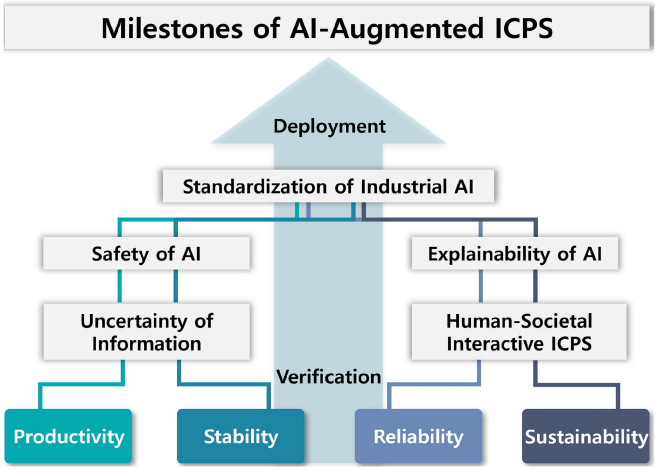

Fig. 3. Perspectives for AICPS design.

A. Uncertainty of Information

Despite the active research on industrial AI, lack of stability and reliability due to uncertainty of information continue to pose research challenges. In this section, we describe the factors that create uncertainty of information that must be overcome for the adoption of AICPS, and investigate research aimed at resolving this uncertainty.

Lack of reliability in data is a common problem in industrial environments. Industrial environments sometimes face difficulties due to data scarcity. The more complex the equipment, the harder it becomes to explain the relationships through existing knowledge, making a data-driven approach more effective [82]. However, obtaining sufficient data can be challenging for some complex equipment due to their low digital degree and complexity. Furthermore, obtaining data for fault prediction can be difficult due to the long-term sustained normal state and occasional occurrence of faults or defects in the system [83]. Data scarcity makes the use of AI challenging. For example, if there are not enough samples for training DNN, it is easy to encounter high variance and overfitting problems [84]. For high-dimensional data with a small number of samples, machine learning performance estimates can be biased [85]. Dropout techniques [83] and batch normalization [86] are two well-known regularization methods used to prevent overfitting when training deep AI models. In addition, research is underway to address the fundamental issue of data scarcity, such as data augmentation and transfer learning.

Data augmentation is the process of supplementing a dataset with generated similar data derived from the information in the dataset itself [87]. It is used for oversampling, where minority class data is replicated to address data imbalance when there is a scarcity of data for all labels or a discrepancy in data balance across labels [88]. Data augmentation includes basic approaches such as adding noise [89], rotating [90], or cropping/flipping [91] data (primarily for image data), as well as more complex approaches such as decomposition methods, statistical generative methods, and learning-based methods [92]. Decomposition methods feature using decomposition techniques such as empirical mode decomposition to generate new data that preserves the information of the original data [93]. Statistical generative methods employ modeling the dynamics of data using statistical models based on the data [94]. Learning-based methods feature modeling the dynamics of data, primarily through AI techniques such as GAN [95].

The traditional machine learning methodology assumes that the data used for training and testing are from the same domain. However, in some industrial environments, collecting enough data for training may be impossible or difficult for various reasons. Transfer learning is a technique that allows the use of models trained on data from a different but related domain to be utilized for applications in the target domain where the data is limited or difficult to obtain [96]. Transfer learning is commonly used in industrial scenarios such as fault diagnosis [97], image recognition [98], and quality prediction [99].

B. Safety of AI

The primary aim of ideal AICPS is to develop the optimal AI model by leveraging data acquired from industrial environments and subsequently apply it to real-world industrial processes. However, the dynamic nature of the industrial environments and model instability, the intrinsic characteristics of AI, can lead to machine malfunctions or pose risks to human safety during application process. Within this context, ensuring the stability of AI models emerges as a critical task that needs to be addressed to facilitate the successful establishment of AICPS [100].

The definition of stability for AI models has not yet been firmly established; however, current research is exploring various methods ensuring the stability of AI models themselves. In the context of the study discussed in [100], the proposal suggests additional normalization processes or the application of constraints to guarantee the model's stability. During the training process, incorporating multiplicative factors related to safety elements or safety margins can reduce the uncertainty between the distributions of training and testing data. Furthermore, in situations where the model encounters unpredicted circumstances, the safe fail approach can be employed by opting for rejection or manual review options with human intervention in the prediction process.

If the model designer lacks expertise in regularization and constraint domains, they may encounter difficulties in applying regularization and constraint techniques. To address this, in [101], a new framework for constraints is proposed. When the designer specifies the constraint conditions in the framework, the framework designs suitable constraint algorithms. This allows the designer to easily apply constraint algorithms that satisfy the constraint domain, even without being an expert in constraint domains. Recently, research has also been conducted to integrate AI and mathematical programming in a synergistic manner to reduce model uncertainty. By combining machine learning in the upper stream and mathematical programming in the lower stream through a closed-loop-based data-driven optimization approach, stable decision-making for uncertainty models has been made possible [102].

Ensuring safety while not compromising efficiency is an important issue. Even in the field of reinforcement learning, which has focused more on cost optimization and paid relatively less attention to stability aspects, research is underway to secure the stability of models. In [103], the study introduces the use of permissive schedulers to enable controller design that guarantees both stability and optimality. Recently, the introduction of a reactive system called "shield" has also been applied to enhance the stability of reinforcement learning models. Shield monitors the behavior of the model and modifies its actions only when violating certain conditions to ensure stability [104]. Additionally, a framework called "FoRShield" has been proposed for the safety of control systems. It filters out risky choices of the model while guiding it towards correct actions, providing feedback to the learning system [105].

C. Human-Societal Interactive ICPS

Despite the advancement in automation achieved through the introduction of ICPS, the role of human interaction within industrial processes remains irreplaceable. As such, an approach focusing on human-friendliness is essential in the design of industrial systems. The concept of human-friendliness in ICPS aims to enhance the user experience by prioritizing the user's needs during human-machine interactions. Originating from user-centric engineering, this approach adapts the system configuration based on insights gathered from extensive monitoring of human-computer interactions. Such insights are increasingly serving as foundational elements for the successful evolution of ICPS [29].

Specifically, [106] discusses the advent of adaptive production systems that extend beyond mere "collaborative work" between humans and machines. These systems "support" the enhancement of human physical abilities, sensing functions, and cognitive capabilities through automation. The aim is to optimally adjust the types and levels of automation in the system to amplify human capacities, emphasizing and valuing the human role within industrial environments [106]. Here, automation functions not as a substitute but as a facilitator for more proficient human labor. By ameliorating human limitations and enhancing end production objectives, automation contributes to creating a human-friendly work environments [107].

Not only is human-friendliness pivotal, but ethical considerations also remain indispensable for the effective integration of ICPS into human and societal contexts. While ICPS efficiently augments productivity through technological innovations like automation, it simultaneously poses ethical challenges. These challenges encompass risks related to large-scale data collection–such as the potential for personal data leakage–accountability in AI decision-making processes, and ethical dilemmas arising from the unintended consequences of automation techniques. To address these ethical complexities engendered by ICPS, a framework is proposed specifically aimed at identifying and preempting ethical issues in the design phase of ICPS [30].

As the complexity and autonomy of ICPS escalate, the range of potential ethical dilemmas correspondingly expands. From development to commercialization, ICPS involves interactions with a diverse array of stakeholders, necessitating rigorous ethical considerations throughout entire lifecycle of ICPS. Specifically within this lifecycle, it is essential to empower stakeholders to engage in self-directed ethical questioning as a foundational approach for mitigating ethical dilemmas. Such ethical inquiry activities serve to reinforce awareness of ethical considerations among all parties involved–ranging from designers and researchers to engineers [108].

[109] introduces a novel approach to engineering ethics by constructing ethical controllers designed for ethically appropriate decision-making in real-world scenarios. These ethical controllers employ a phased strategy based on the strengths and weaknesses of both deontological and consequentialist theories. This approach enables the system to flexibly adapt to a wide range of ethical dilemma scenarios, thereby enhancing the ethical decision-making capabilities of autonomous systems [109].

D. Explainability of AI

AI models base their decisions on probabilistic values, which can raise trust concerns when dealing with unexpected accidents, faults, or attacks in industrial environments. Particularly in actual systems, it is unfeasible to train models for all possible scenarios due to the imperfections in the training data. Furthermore, even with comprehensive training, AI models cannot achieve a zero-percent error rate; they can only minimize it [110].

Moreover, utilizing AI methods for controlling physical systems remains an area of uncertainty. It is also difficult to determine the degree of control stability in physical systems using AI models. Within the spectrum of AI technologies, controllers employing reinforcement learning excel in terms of adaptability; however, they also remain unable to ensure unequivocal stability in control operations [111]. Especially in industrial contexts, the financial burden of development for technique implementation is considerable, compounded by the difficulties of conducting empirical tests on actual systems.

AI systems are frequently characterized as enigmatic "black boxes". A variety of machine learning techniques, such as SVM, RF, RL and DL, are employed for their superior performance attributes. However, these algorithms are largely opaque when it comes to explainability. Consequently, unless the decision-making processes and predictive outcomes produced by these AI models are rendered comprehensible to human operators, the reliability of the system remains questionable. This contextual landscape necessitates the introduction of Explainable AI (XAI), a conceptual paradigm committed to elucidating AI-generated decisions and predictions in a form that is both understandable and explainable to humans [112].

XAI systems should be able to explain not only their functions, but also what they have done, what they are currently doing, and what will happen in the future [112]. XAI has two primary tasks: transparency design and post-hoc explanation [113]. Transparency design aims to make the operation of AI models transparent from the developer's perspective. Post-hoc explanation, on the other hand, explains the reasons behind the results inferred by an AI model from the user's perspective.

TABLE IV
INDUSTRIAL AI STANDARDS

Reference	Abbreviation	Name	Institution	Design Consideration
[121]	ISO/IEC TS 4213	Assessment of machine learning classification performance	ISO/IEC	Productivity
[122]	ISO/IEC WD TS 25058	Guidance for quality evaluation of AI systems	ISO/IEC	Productivity
[123]	ISO/IEC PRF 25059	Quality model for AI systems	ISO/IEC	Productivity
[124]	ISO/IEC CD 5259	Data quality for analytics and machine learning (ML)	ISO/IEC	Productivity
[125]	ISO/IEC TR 24027	Bias in AI systems and AI aided decision making	ISO/IEC	Stability
[126]	ISO/IEC TR 24028	Overview of trustworthiness in artificial intelligence	ISO/IEC	Stability
[127]	ISO/IEC CD TS 12791	Treatment of unwanted bias in classification and regression machine learning tasks	ISO/IEC	Stability
[128]	ISO/IEC CD TR 5469	Functional safety and AI systems	ISO/IEC	Stability
[129]	ISO/IEC AWI 27090	Guidance for addressing security threats and failures in artificial intelligence systems	ISO/IEC	Reliability
[130]	NIS AI RMF	NIST AI Risk Management Framework	NIST	Reliability

In [114], two forms of explainability for AI models are proposed. Firstly, "Mathematical explainability" emphasizes predicting the model's outcomes based on mathematical concepts. This approach relies on mathematically analyzing the model's parameters or weights to provide explainability. It serves as an aspect of transparency design by applying mathematical concepts to enhance the explainability of the model's workings and decisions. Secondly, "Perceptual explainability" highlights an approach that allows for the interpretation of the model in a way that is intuitively understandable to humans. This involves visualizing the key features of the model to facilitate easy comprehension for users. It serves as an aspect of post-hoc explanation by visualizing the model's results to make them interpretable. [115] discusses various explanation techniques used to enhance explainability for models from a post-hoc explainability perspective. Examples of such techniques include text explanations, visual explanations, explanations by example, and explanations by simplification. To aid in the understanding of the model's interpretation, text explanations feature generating text, while visual explanations visualize the behavior of the model. Explanations by example showcase the model's generated results as examples to facilitate easier understanding, and explanations by simplification involve designing simplified models based on the trained model to enhance user understanding with lower complexity [115].

While AI models offer a multitude of advantages in the industrial sector, it is equally crucial to ensure that human experts can comprehend these models. In particular, it is essential to explain to the workers the control stability and equipment maintenance that the AI model is responsible for. To enhance user's comprehension of the AI model, simplifying the model may result in a trade-off with performance. Striking a delicate balance between the model's performance and its explainability becomes imperative.

E. Standardization of Industrial AI

A standard is a technical specification that provides detailed requirements, specifications, rules, guidelines, and procedures for a specific operation, product, system, or service. These specifications are developed through the consensus of industry and market actors, approved by accredited bodies, and published as documents [116]. Standardization of industrial AI is particularly crucial since the integration of industrial AI and CPS is mandatory. To achieve standardization, multiple frameworks need to be consolidated into a single standard, allowing developers to utilize validated frameworks. Numerous studies investigate the integration of industrial AI and CPS within a unified framework [117], [118], [119], [120].

Standardization efforts related to industrial AI for intelligent design are actively ongoing. Standardization of industrial AI is crucial in ensuring the performance and reliability of products and services that use AI, facilitating the adoption of AI technology in industries where reliability and stability are of utmost importance. Moreover, interoperability issues may arise among companies that develop products and services using AI technology. Therefore, standardization can resolve such problems by unifying AI technology in industrial domains. Ultimately, standardization can enable the more efficient use of AI technology, resulting in the development of superior products and services. Table IV provides a summary of references, abbreviations, names, institutions and design considerations of standards about AI technology related to AICPS design.

Standards related to AI techniques for enhancing productivity are currently being developed and published. ISO/IEC TS 4213, ISO/IEC WD TS 25058, and ISO/IEC PRF 25059 are standards for evaluating the performance of AI techniques. ISO/IEC TS 4213 proposes standards for AI techniques used in classification [121]. ISO/IEC WD TS 25058 and ISO/IEC

PRF 25059 are currently under development as standards for guidelines and models for evaluating the performance of AI systems [122], [123]. ISO/IEC CD 5259 is a set of standards for data quality for machine learning, which is currently being studied [124]. It provides terms for data quality, requirements and guidelines for quality management, a process framework for quality measurement, and a framework for quality visualization.

The stability of AI technology is also a crucial factor that needs to be addressed to ensure its efficient utilization across diverse industries. In this regard, two standards, namely ISO/IEC TR 24027 and ISO/IEC CD TS 12791, are currently under development to address the issue of bias in AI. Bias in AI often arises from personal and unnecessary assumptions. These biases may be inherent in the training data or the model itself [131]. Such biases can potentially lead to inaccurate predictions by the AI model and incorrect decisions for some or all of the data. ISO/IEC TR 24027 aims to provide standards pertaining to the concept and causes of bias, types of bias, bias evaluation and measurement, as well as bias management and mitigation [125]. On the other hand, ISO/IEC CD TS 12791 is focused on developing mitigation techniques that can be applied across the entire life cycle of an AI system to address undesired biases in classification and regression machine learning tasks [127].

Ensuring the reliability of AI models against cybersecurity threats is a crucial aspect of AI technology [132]. In this regard, ISO/IEC AWI 27090 is a standard currently being developed to provide guidelines for organizations to address security threats and errors in AI systems [129]. The objective of this standard is to furnish organizations with information to better comprehend the consequences of security threats to AI systems and to provide explanations on how to detect and mitigate such threats. Moreover, NIS AI RMF proposes an approach for organizations and individuals to enhance the reliability of AI systems and promote sustainable and responsible design, development, deployment, and use of AI systems [130]. It outlines the characteristics of trustworthy AI systems, which encompass the following attributes: validity and reliability; safety; security and resilience; accountability and transparency; explainability and interpretability; enhanced privacy; and fairness, with harmful biases appropriately managed.

VI. Conclusion

In light of the escalating global attention on AI, there is an increased effort to develop AI techniques tailored for industrial processes. Accordingly, the importance of integrating ICPS and industrial AI is being emphasized, and it is expected that efforts to strengthen this integration will intensify in future research. In this context, we propose the establishment of AICPS to optimize the management of industrial processes. The adoption of AICPS provides several technological and economic benefits, such as real-time monitoring of machinery, efficient maintenance management, and effective scheduling planning.

In this article, we suggest a set of design considerations and analyze the components and interactions of AICPS. In addition, we also present cutting-edge industrial AI technologies for AICPS and identify research challenges that need to be addressed for the successful implementation of AICPS. These challenges, which include uncertainty of information, safety of AI, explainability of AI, human-societal interactive ICPS, and standardization of industrial AI suggest that there are still issues to be resolved for the establishment of AICPS. In the majority of the research studies we surveyed, the validation of AI techniques was conducted within simulators and testbeds. Out of the 51 papers examined in our paper, only five were validated on actual industrial systems. This highlights a "lack of assurance" stemming from the fact that the aforementioned research challenges have not been fully addressed. Consequently, for the effective utilization of industrial AI in ICPS, there needs to be rigorous research into effective and systematic verification and deployment processes that can ensure the safety of both users and industrial systems.

We outline the direction of AICPS research that can address research challenges considering five design considerations. This serves as a foundational framework for future research, particularly in the context of emerging technologies and trends. Therefore, this article not only advocates for an intelligent approach to industrial process management and emphasizes the criticality of integrating industrial AI with ICPS, but it also delineates milestones for future research trajectories that scholars ought to consider.

References

[1] C. Klötzer, J. Weißenborn, and A. Pflaum, "The evolution of cyber-physical systems as a driving force behind digital transformation," in *Proc. IEEE 19th Conf. Bus. Informat.*, 2017, pp. 5–14.

[2] K.-J. Park, R. Zheng, and X. Liu, "Cyber-physical systems: Milestones and research challenges," *Comput. Commun.*, vol. 36, no. 1, pp. 1–7, 2012.

[3] S. Lee, J. Kim, G. Wi, Y. Won, Y. Eun, and K.-J. Park, "Deep reinforcement learning-driven scheduling in multijob serial lines: A case study in automotive parts assembly," *IEEE Trans. Ind. Informat.*, early access, Aug. 08, 2023, doi: 10.1109/TII.2023.3292538.

[4] K. Zhang, Y. Shi, S. Karnouskos, T. Sauter, H. Fang, and A. W. Colombo, "Advancements in industrial cyber-physical systems: An overview and perspectives," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 716–729, Jan. 2023.

[5] P. Skobelev, A. Tabachinskiy, E. Simonova, and O. Goryanin, "Development of crop-simulation multiagent system for smart digital twin of plant," in *Proc. IEEE 8th Int. Conf. Inf. Technol. Nanotechnol.*, 2022, pp. 1–8.

[6] J. Lee, H. Davari, J. Singh, and V. Pandhare, "Industrial artificial intelligence for industry 4.0-based manufacturing systems," *Manuf. Lett.*, vol. 18, pp. 20–23, 2018.

[7] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, "Industrial artificial intelligence in industry 4.0-systematic review, challenges and outlook," *IEEE Access*, vol. 8, pp. 220121–220139, 2020.

[8] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE Access*, vol. 8, pp. 108952–108971, 2020.

[9] K. T. Park et al., "Design and implementation of a digital twin application for a connected micro smart factory," *Int. J. Comput. Integr. Manuf.*, vol. 32, no. 6, pp. 596–614, 2019.

[10] Q. Qi and F. Tao, "Digital twin and Big Data towards smart manufacturing and industry 4.0: 360 degree comparison," *IEEE Access*, vol. 6, pp. 3585–3593, 2018.

[11] Y. Lu, C. Liu, I. Kevin, K. Wang, H. Huang, and X. Xu, "Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues," *Robot. Comput.- Integr. Manuf.*, vol. 61, 2020, Art. no. 101837.

[12] W. Yu, Y. Liu, T. Dillon, W. Rahayu, and F. Mostafa, "An integrated framework for health state monitoring in a smart factory employing IoT and Big Data techniques," *IEEE Internet Things J.*, vol. 9, no. 3, pp. 2443–2454, Feb. 2022.

[13] O. Niggemann, G. Biswas, J. S. Kinnebrew, H. Khorasgani, S. Volgmann, and A. Bunte, "Data-driven monitoring of cyber-physical systems leveraging on Big Data and the Internet-of-Things for diagnosis and control," in *Proc. DX*, pp. 185–192, 2015.

[14] S. J. Russell, *Artificial Intelligence a Modern Approach*. London, UK.: Pearson Educ. Inc., 2010.

[15] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, "Implementation of industrial cyber physical system: Challenges and solutions," in *Proc. IEEE Int. Conf. Ind. Cyber Phys. Syst.*, 2019, pp. 173–178, doi: 10.1109/ICPHYS.2019.8780271.

[16] Y. Cheng, K. Chen, H. Sun, Y. Zhang, and F. Tao, "Data and knowledge mining with Big Data towards smart production," *J. Ind. Inf. Integration*, vol. 9, pp. 1–13, 2018.

[17] Y. Jiang, S. Yin, and O. Kaynak, "Performance supervised plant-wide process monitoring in industry 4.0: A roadmap," *IEEE Open J. Ind. Electron. Soc.*, vol. 2, pp. 21–35, 2021.

[18] T. Wang et al., "An intelligent edge-computing-based method to counter coupling problems in cyber-physical systems," *IEEE Netw.*, vol. 34, no. 3, pp. 16–22, May/Jun. 2020.

[19] S. Ponomarev and T. Atkison, "Industrial control system network intrusion detection by telemetry analysis," *IEEE Trans. Dependable Secure Comput.*, vol. 13, no. 2, pp. 252–260, Mar./Apr. 2016.

[20] A. Nayak, R. R. Levalle, S. Lee, and S. Y. Nof, "Resource sharing in cyber-physical systems: Modelling framework and case studies," *Int. J. Prod. Res.*, vol. 54, no. 23, pp. 6969–6983, 2016.

[21] H.-S. Park, S. Moon, J. Kwak, and K.-J. Park, "CAPL: Criticality-aware adaptive path learning for industrial wireless sensor-actuator networks," *IEEE Trans. Ind. Informat.*, vol. 19, no. 8, pp. 9123–9133, Aug. 2023.

[22] M. Glatt, C. Sinnwell, L. Yi, S. Donohoe, B. Ravani, and J. C. Aurich, "Modeling and implementation of a digital twin of material flows based on physics simulation," *J. Manuf. Syst.*, vol. 58, pp. 231–245, 2021.

[23] P. Hehenberger, B. Vogel-Heuser, D. Bradley, B. Eynard, T. Tomiyama, and S. Achiche, "Design, modelling, simulation and integration of cyber physical systems: Methods and applications," *Comput. Ind.*, vol. 82, pp. 273–289, 2016.

[24] M.-C. Chiu, C.-D. Tsai, and T.-L. Li, "An integrative machine learning method to improve fault detection and productivity performance in a cyber-physical system," *J. Comput. Inf. Sci. Eng.*, vol. 20, no. 2, 2020, Art. no. 021009.

[25] M. Pajic, J. Weimer, N. Bezzo, O. Sokolsky, G. J. Pappas, and I. Lee, "Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators," *IEEE Control Syst. Mag.*, vol. 37, no. 2, pp. 66–81, Apr. 2017.

[26] Y. Jiang et al., "Secure data transmission and trustworthiness judgement approaches against cyber-physical attacks in an integrated data-driven framework," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 12, pp. 7799–7809, Dec. 2022.

[27] S. Kim, K.-J. Park, and C. Lu, "A survey on network security for cyber–physical systems: From threats to resilient design," *IEEE Commun. Surv. Tut.*, vol. 24, no. 3, pp. 1534–1573, Thirdquarter, 2022.

[28] O. Inderwildi, C. Zhang, X. Wang, and M. Kraft, "The impact of intelligent cyber-physical systems on the decarbonization of energy," *Energy Environ. Sci.*, vol. 13, no. 3, pp. 744–771, 2020.

[29] M. Broy and A. Schmidt, "Challenges in engineering cyber-physical systems," *IEEE Comput.*, vol. 47, no. 2, pp. 70–72, Feb. 2014.

[30] D. Trentesaux, E. Caillaud, and R. Rault, "A framework fostering the consideration of ethics during the design of industrial cyber-physical systems," in *Proc. Int. Workshop Service Orientation Holonic Multi-Agent Manuf.*, 2021, pp. 349–362.

[31] H. Fang, A. Qi, and X. Wang, "Fast authentication and progressive authorization in large-scale IoT: How to leverage AI for security enhancement," *IEEE Netw.*, vol. 34, no. 3, pp. 24–29, May/Jun. 2020.

[32] R.-F. Liao et al., "Deep-learning-based physical layer authentication for industrial wireless sensor networks," *Sensors*, vol. 19, no. 11, 2019, Art. no. 2440.

[33] M. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Learning-based sparse data reconstruction for compressed data aggregation in IoT networks," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11732–11742, Jul. 2021.

[34] F. Kingma, P. Abbeel, and J. Ho, "Bit-swap: Recursive bits-back coding for lossless compression with hierarchical latent variables," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 3408–3417. [Online]. Available: https://proceedings.mlr.press/v97/kingma19a.html

[35] M. Tschannen, E. Agustsson, and M. Lucic, "Deep generative models for distribution-preserving lossy compression," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[36] S. Kalay, E. Çinar, and İ. Sarıçiçek, "A comparison of data imputation methods utilizing machine learning for a new IoT system platform," in *Proc. 8th Int. Conf. Control, Decis. Inf. Technol.*, 2022, pp. 69–74.

[37] M. Kim, S. Park, J. Lee, Y. Joo, and J. K. Choi, "Learning-based adaptive imputation method with KNN algorithm for missing power data," *Energies*, vol. 10, no. 10, 2017, Art. no. 1668.

[38] J. Ma, J. C. Cheng, F. Jiang, W. Chen, M. Wang, and C. Zhai, "A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data," *Energy Buildings*, vol. 216, 2020, Art. no. 109941.

[39] Y.-F. Zhang, P. J. Thorburn, W. Xiang, and P. Fitch, "SSIM–a deep learning approach for recovering missing time series sensor data," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6618–6628, Aug. 2019.

[40] Z. Geng, Z. Chen, Q. Meng, and Y. Han, "Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1521–1529, Mar. 2022.

[41] F. Curreri, L. Patanè, and M. G. Xibilia, "RNN-and LSTM-based soft sensors transferability for an industrial process," *Sensors*, vol. 21, no. 3, 2021, Art. no. 823.

[42] A. K. Verma, S. Nagpal, A. Desai, and R. Sudha, "An efficient neural-network model for real-time fault detection in industrial machine," *Neural Comput. Appl.*, vol. 33, pp. 1297–1310, 2021.

[43] H. Ha and J. Jeong, "CNN-based defect inspection for injection molding using edge computing and industrial IoT systems," *Appl. Sci.*, vol. 11, no. 14, 2021, Art. no. 6378.

[44] M. Sadoughi and C. Hu, "Physics-based convolutional neural network for fault diagnosis of rolling element bearings," *IEEE Sensors J.*, vol. 19, no. 11, pp. 4181–4192, Jun. 2019.

[45] H. A. Saeed, H. Wang, M. Peng, A. Hussain, and A. Nawaz, "Online fault monitoring based on deep neural network & sliding window technique," *Prog. Nucl. Energy*, vol. 121, 2020, Art. no. 103236.

[46] J.-H. Bae et al., "Deep-learning-based pipe leak detection using image-based leak features," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 2361–2365.

[47] Y. Wang, L. Hou, K. C. Paul, Y. Ban, C. Chen, and T. Zhao, "ArcNet: Series AC arc fault detection based on raw current and convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 77–86, Jan. 2022.

[48] H. Han, M. Sun, F. Li, Z. Liu, and C. Wang, "Self-supervised deep clustering method for detecting abnormal data of wastewater treatment process," *IEEE Trans. Ind. Informat.*, early access, Apr. 20. 2023, doi: 10.1109/TII.2023.3268777.

[49] C. Spandonidis, P. Theodoropoulos, F. Giannopoulos, N. Galiatsatos, and A. Petsa, "Evaluation of deep learning approaches for oil & gas pipeline leak detection using wireless sensor networks," *Eng. Appl. Artif. Intell.*, vol. 113, 2022, Art. no. 104890.

[50] Z. Zuo, L. Ma, S. Liang, J. Liang, H. Zhang, and T. Liu, "A semi-supervised leakage detection method driven by multivariate time series for natural gas gathering pipeline," *Process Saf. Environ. Protection*, vol. 164, pp. 468–478, 2022.

[51] K. S. Kiangala and Z. Wang, "An effective predictive maintenance framework for conveyor motors using dual time-series imaging and convolutional neural network in an industry 4.0 environment," *IEEE Access*, vol. 8, pp. 121033–121049, 2020.

[52] A. T. Prihatno, H. Nurcahyanto, and Y. M. Jang, "Predictive maintenance of relative humidity using random forest method," in *Proc. IEEE Int. Conf. Artif. Intell. Inf. Commun.*, 2021, pp. 497–499.

[53] B. Chen, Y. Liu, C. Zhang, and Z. Wang, "Time series data for equipment reliability analysis with deep learning," *IEEE Access*, vol. 8, pp. 105484–105493, 2020.

[54] A. K. Onaolapo, R. P. Carpanen, D. G. Dorrell, and E. E. Ojo, "Event-driven power outage prediction using collaborative neural networks," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 3079–3087, Mar. 2023.

[55] J. Patalas-Maliszewska, D. Halikowski, and R. Damaševičius, "An automated recognition of work activity in industrial manufacturing using convolutional neural networks," *Electron.*, vol. 10, no. 23, 2021, Art. no. 2946.

[56] W. Tao, M. C. Leu, and Z. Yin, "Multi-modal recognition of worker activity for human-centered intelligent manufacturing," *Eng. Appl. Artif. Intell.*, vol. 95, 2020, Art. no. 103868.

[57] E. Hodo et al., "Threat analysis of IoT networks using artificial neural network intrusion detection system," in *Proc. IEEE Int. Symp. Netw., Comput. Commun.*, 2016, pp. 1–6.

[58] B. Hussain, Q. Du, B. Sun, and Z. Han, "Deep learning-based DDos-attack detection for cyber–physical system over 5G network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 860–870, Feb. 2021.

[59] P. Jiang, H. Wu, C. Wang, and C. Xin, "Virtual MAC spoofing detection through deep learning," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–6.

[60] Z. Rahman, X. Yi, and I. Khalil, "Blockchain based AI-enabled industry 4.0 CPS protection against advanced persistent threat," *IEEE Internet Things J.*, vol. 10, no. 8, pp. 6769–6778, Apr. 2023.

[61] J. T. Shaw et al., "A case study application of machine-learning for the detection of greenhouse gas emission sources," *Atmospheric Pollut. Res.*, vol. 13, no. 10, 2022, Art. no. 101563.

[62] F. Cangialosi, E. Bruno, and G. D. Santis, "Application of machine learning for fenceline monitoring of odor classes and concentrations at a wastewater treatment plant," *Sensors*, vol. 21, no. 14, 2021, Art. no. 4716.

[63] Y. Choi, K. Kim, S. Kim, and D. Kim, "Identification of odor emission sources in urban areas using machine learning-based classification models," *Atmospheric Environ.: X*, vol. 13, 2022, Art. no. 100156.

[64] J. F. Tuttle, L. D. Blackburn, and K. M. Powell, "On-line classification of coal combustion quality using nonlinear SVM for improved neural network NOx emission rate prediction," *Comput. Chem. Eng.*, vol. 141, 2020, Art. no. 106990.

[65] L. Ren, Z. Meng, X. Wang, L. Zhang, and L. T. Yang, "A data-driven approach of product quality prediction for complex production systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6457–6465, Sep. 2021.

[66] X. Wang, Y. Wang, L. Tang, and Q. Zhang, "Multi-objective ensemble learning with multi-scale data for product quality prediction in iron and steel industry," *IEEE Trans. Evol. Comput.*, early access, Jun. 28, 2023, doi: 10.1109/TEVC.2023.3290172.

[67] T. Li, J. Lu, J. Wu, Z. Zhang, and L. Chen, "Predicting aquaculture water quality using machine learning approaches," *Water*, vol. 14, no. 18, 2022, Art. no. 2836.

[68] A. Azizi, "Applications of artificial intelligence techniques to enhance sustainability of industry 4.0: Design of an artificial neural network model as dynamic behavior optimizer of robotic arms," *Complexity*, vol. 2020, pp. 1–10, 2020.

[69] R. Cai, W. Wen, K. Wang, Y. Peng, S. Ahzi, and F. Chinesta, "Tailoring interfacial properties of 3d-printed continuous natural fiber reinforced polypropylene composites through parameter optimization using machine learning methods," *Mater. Today Commun.*, vol. 32, 2022, Art. no. 103985.

[70] S. R. Dabbagh, O. Ozcan, and S. Tasoglu, "Machine learning-enabled optimization of extrusion-based 3d printing," *Methods*, vol. 206, pp. 27–40, 2022.

[71] T. Zhou, D. Tang, H. Zhu, and L. Wang, "Reinforcement learning with composite rewards for production scheduling in a smart factory," *IEEE Access*, vol. 9, pp. 752–766, 2021.

[72] P. M. Kumar, G. C. Babu, A. Selvaraj, M. Raza, A. K. Luhach, and V. G. Díaz, "Multi-criteria-based approach for job scheduling in industry 4.0 in smart cities using fuzzy logic," *Soft Comput.*, vol. 25, pp. 12059–12074, 2021.

[73] F. Qiao, J. Liu, and Y. Ma, "Industrial big-data-driven and CPS-based adaptive production scheduling for smart manufacturing," *Int. J. Prod. Res.*, vol. 59, no. 23, pp. 7139–7159, 2021.

[74] R. Liu, R. Piplani, and C. Toro, "Deep reinforcement learning for dynamic scheduling of a flexible job shop," *Int. J. Prod. Res.*, vol. 60, no. 13, pp. 4049–4069, 2022.

[75] I. M. A. Nahrendra, P. H. Rusmin, and E. M. I. Hidayat, "Adaptive control of cyber-physical distillation column using data driven control approach," in *Proc. IEEE Int. Conf. Elect. Eng. Informat.*, 2019, pp. 93–98.

[76] D. A. Goulart and R. D. Pereira, "Autonomous ph control by reinforcement learning for electroplating industry wastewater," *Comput. Chem. Eng.*, vol. 140, 2020, Art. no. 106909.

[77] D. Wang et al., "Model predictive control using artificial neural network for power converters," *IEEE Trans. Ind. Electron.*, vol. 69, no. 4, pp. 3689–3699, Apr. 2022.

[78] Z. Cao, P. Zhou, R. Li, S. Huang, and D. Wu, "Multiagent deep reinforcement learning for joint multichannel access and task offloading of mobile-edge computing in industry 4.0," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6201–6213, Jul. 2020.

[79] Y. Hao, M. Chen, H. Gharavi, Y. Zhang, and K. Hwang, "Deep reinforcement learning for edge service placement in softwarized industrial cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5552–5561, Aug. 2021.

[80] D. Wang, N. Zhao, B. Song, P. Lin, and F. R. Yu, "Resource management for secure computation offloading in softwarized cyber–physical systems," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9294–9304, Jun. 2021.

[81] Q. W. Ahmed et al., "AI-based resource allocation techniques in wireless sensor Internet of Things networks in energy efficiency with data optimization," *Electronics*, vol. 11, no. 13, 2022, Art. no. 2071.

[82] S. Kim, Y. Won, K.-J. Park, and Y. Eun, "A data-driven indirect estimation of machine parameters for smart production systems," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 6537–6546, Oct. 2022.

[83] Y. Wang, F. Tao, M. Zhang, L. Wang, and Y. Zuo, "Digital twin enhanced fault prediction for the autoclave with insufficient data," *J. Manuf. Syst.*, vol. 60, pp. 350–359, 2021.

[84] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[85] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS One*, vol. 14, no. 11, 2019, Art. no. e0224365.

[86] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[87] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart augmentation learning an optimal data augmentation strategy," *IEEE Access*, vol. 5, pp. 5858–5869, 2017.

[88] C. Khosla and B. S. Saini, "Enhancing performance of deep learning models with different data augmentation techniques: A survey," in *Proc. IEEE Int. Conf. Intell. Eng. Manage.*, 2020, pp. 79–85.

[89] M. Sáiz-Abajo, B.-H. Mevik, V. Segtnan, and T. Næs, "Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data," *Analytica Chimica Acta*, vol. 533, no. 2, pp. 147–159, 2005.

[90] M. M. Krell and S. K. Kim, "Rotational data augmentation for electroencephalographic data," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017, pp. 471–474.

[91] P. Chen, S. Liu, H. Zhao, and J. Jia, "Gridmask data augmentation," 2020, *arXiv:2001.04086*.

[92] Q. Wen et al., "Time series data augmentation for deep learning: A survey," 2020, *arXiv:2002.12478*.

[93] B. Li et al., "Component-mixing strategy: A decomposition-based data augmentation algorithm for motor imagery signals," *Neurocomputing*, vol. 465, pp. 325–335, 2021.

[94] J. Goodman, S. Sarkani, and T. Mazzuchi, "Distance-based probabilistic data augmentation for synthetic minority oversampling," *ACM/IMS Trans. Data Sci.*, vol. 2, no. 4, pp. 1–18, 2022.

[95] N.-T. Tran, V.-H. Tran, N.-B. Nguyen, T.-K. Nguyen, and N.-M. Cheung, "On data augmentation for GAN training," *IEEE Trans. Image Process.*, vol. 30, pp. 1882–1897, 2021.

[96] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[97] W. Li et al., "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges," *Mech. Syst. Signal Process.*, vol. 167, 2022, Art. no. 108487.

[98] B. Maschler, S. Kamm, and M. Weyrich, "Deep industrial transfer learning at runtime for image recognition," *at-Automatisierungstechnik*, vol. 69, no. 3, pp. 211–220, 2021.

[99] H. Tercan, A. Guajardo, and T. Meisen, "Industrial transfer learning: Boosting machine learning in production," in *Proc. IEEE 17th Int. Conf. Ind. Informat.*, 2019, pp. 274–279.

[100] K. R. Varshney and H. Alemzadeh, "On the safety of machine learning: Cyber-physical systems, decision sciences, and data products," *Big Data*, vol. 5, no. 3, pp. 246–255, 2017.

[101] P. S. Thomas, B. Castro da Silva, A. G. Barto, S. Giguere, Y. Brun, and E. Brunskill, "Preventing undesirable behavior of intelligent machines," *Science*, vol. 366, no. 6468, pp. 999–1004, 2019.

[102] C. Ning and F. You, "Optimization under uncertainty in the era of Big Data and deep learning: When machine learning meets mathematical programming," *Comput. Chem. Eng.*, vol. 125, pp. 434–448, 2019.

[103] S. Junges, N. Jansen, C. Dehnert, U. Topcu, and J.-P. Katoen, "Safety-constrained reinforcement learning for MDPs," in *Proc. Tools Algorithms Construction Anal. Syst.: 22nd Int. Conf., TACAS, Held Part Eur. Joint Conf. Theory Pract. Softw.*, 2016, pp. 130–146.

[104] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu, "Safe reinforcement learning via shielding," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2661–2669.

[105] H. Sibai, M. Potok, and S. Mitra, "Safe Reinforcement Learning for Control Systems: A Hybrid Systems Perspective and Case Study," in *Proc. ACM Hybrid Syst. Comput. Control*, 2019, pp. 1–9.

[106] D. Romero, P. Bernus, O. Noran, J. Stahre, and Å. Fast-Berglund, "The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems," in *Proc. Adv. Prod. Manage. Syst. Initiatives Sustain. World: IFIP WG 5.7 Int. Conf., APMS, Iguassu Falls* 2016, pp. 677–686.
[107] D. Romero, O. Noran, J. Stahre, P. Bernus, and Å. Fast-Berglund, "Towards a human-centred reference architecture for next generation balanced automation systems: Human-automation symbiosis," in *Proc. Adv. Prod. Manage. Systems: Innov. Prod. Manage. Towards Sustain. Growth: IFIP WG 5.7 Int. Conf.*, 2015, pp. 556–566.
[108] D. Trentesaux, E. Caillaud, and R. Rault, "A vision of applied ethics in industrial cyber-physical sytems," in *Proc. Int. Workshop Serv. Orientation Holonic Multi-Agent Manuf.*, 2021, pp. 319–331.
[109] D. Trentesaux and S. Karnouskos, "Engineering ethical behaviors in autonomous industrial cyber-physical human systems," *Cognition, Technol. Work*, vol. 24, no. 1, pp. 113–126, 2022.
[110] S. Shafaei, S. Kugele, M. H. Osman, and A. Knoll, "Uncertainty in machine learning: A safety perspective on autonomous driving," in *Computer Safety, Reliability, and Security*. Berlin, Germany: Springer, 2018, pp. 458–464.
[111] N. Fulton and A. Platzer, "Safe reinforcement learning via formal methods: Toward safe control through proof and learning," in *Proc. AAAI*, 2018, pp. 6485–6492.
[112] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI–explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, 2019, Art. no.eaay7120.
[113] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, "Explainable AI: A brief survey on history, research areas, approaches and challenges," in *Proc. Natural Lang. Process. Chin. Comput.: 8th CCF Int. Conf.*, 2019, pp. 563–574.
[114] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021.
[115] A. B. Arrietaet al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020.
[116] S. Ali, T. Al. Balushi, Z. Nadir, and O. K. Hussain, "Standards for CPS," *Cyber Secur. Cyber Phys. Syst.*, vol. 768, pp. 161–174, 2018.
[117] J. Ma, Q. Wang, and Z. Zhao, "SLAE–CPS: Smart Lean Automation Engine enabled by cyber-physical systems technologies," *Sensors*, vol. 17, no. 7, 2017, Art. no. 1500.
[118] I. Dumitrache, I. S. Sacala, M. A. Moisescu, and S. I. Caramihai, "A conceptual framework for modeling and design of cyber-physical systems," *Stud. Informat. Control*, vol. 26, no. 3, pp. 325–334, 2017.
[119] C. Alippi and M. Roveri, "The (not) far-away path to smart cyber-physical systems: An information-centric framework," *IEEE Comput.*, vol. 50, no. 4, pp. 38–47, Apr. 2017.
[120] A. Bousdekis, D. Apostolou, and G. Mentzas, "A human cyber physical system framework for operator 4.0–artificial intelligence symbiosis," *Manuf. Lett.*, vol. 25, pp. 10–15, 2020.
[121] "Information technology – artificial intelligence – assessment of machine learning classification performance," ISO/IEC TS 4213:2022, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2022.
[122] "Software and systems engineering – Systems and software quality requirements and evaluation (SQuaRE) – Guidance for quality evaluation of AI systems," ISO/IEC DTS 25058, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2023.
[123] "Software engineering – Systems and software quality requirements and evaluation (square) – Quality model for ai systems," ISO/IEC 25059, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2023.
[124] "Artificial intelligence – Data quality for analytics and machine learning (ML)," ISO/IEC CD 5259, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2022.
[125] "Information technology – Artificial intelligence (AI) – Bias in AI systems and AI aided decision making," ISO/IEC CD 5259, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2021.
[126] "Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence," ISO/IEC TR 24028:2020, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2020.
[127] "Information technology – Artificial intelligence – Treatment of unwanted bias in classification and regression machine learning tasks," ISO/IEC DTS 12791, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2023.
[128] "Artificial intelligence – Functional safety and AI systems," ISO/IEC DTR 5469, ISO/IEC JTC 1/SC 42, Washington, D.C., USA, 2023.
[129] "Cybersecurity – Artificial intelligence – Guidance for addressing security threats and failures in artificial intelligence systems," ISO/IEC AWI 27090, ISO/IEC JTC 1/SC 27, Berlin, Germany, 2023.
[130] "NIST AI risk management framework," NIST AI RMF, NIST, Maryland, USA, 2023.
[131] P. Turney, "Bias and the quantification of stability," *Mach. Learn.*, vol. 20, pp. 23–33, 1995.
[132] S. Kim and K.-J. Park, "A survey on machine-learning based security design for cyber-physical systems," *Appl. Sci.*, vol. 11, no. 12, Art. no. 5458, 2021.

Jiyeong Chae received the B.S. degree in computer science from the school of Undergraduate Studies, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea, in 2023. She is currently working toward the M.S. degree in computer science with the Department of Electrical Engineering and Computer Science, the Daegu Gyeongbuk Institute of Science and Technology. Her research interests include industrial cyber-physical systems and industrial artificial intelligence.

Sanghoon Lee received the B.S. degree in computer science from the school of Undergraduate Studies, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea, in 2022. He is currently working toward the Ph.D. degree in computer science with the Department of Electrical Engineering and Computer Science. Daegu Gyeongbuk Institute of Science and Technology. His research interests include industrial cyber-physical systems and industrial artificial intelligence.

Junhyung Jang has been majoring in computer science from the school of Undergraduate Studies, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea, since 2018. He is currently Student Intern with the Department of Electrical Engineering and Computer Science. His research interests include industrial artificial intelligence and software development.

Seohyung Hong has been majoring in computer science from the school of Undergraduate Studies, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea, since 2020. She is currently a Student Intern with the Department of Electrical Engineering and Computer Science. Her research focuses on artificial intelligence.

Kyung-Joon Park (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering and Ph.D. degree in electrical engineering and computer science from Seoul National University, Seoul, South Korea. From 2005 to 2006, he was a Senior Engineer with Samsung Electronics, Suwon-si, South Korea. From 2006 to 2010, he was a Postdoctoral Research Associate with the Department of Computer Science, University of Illinois Urbana-Champaign, Champaign, IL, USA. He is currently a Professor with the Department of Electrical Engineering and Computer Science, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea. His research interests include cyber-physical systems, robot operating system, and smart manufacturing.