

Received October 31, 2020, accepted December 3, 2020, date of publication December 7, 2020,
date of current version December 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042874

Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook

RICARDO SILVA PERES^{1,2}, (Member, IEEE), XIAODONG JIA³, JAY LEE³, KEYI SUN⁴,
ARMANDO WALTER COLOMBO⁵, (Fellow, IEEE), AND JOSE BARATA^{1,2}, (Member, IEEE)

¹Centre of Technology and Systems, UNINOVA Instituto Desenvolvimento de Novas Tecnologias, 2829-516 Caparica, Portugal

²Department of Electrical Engineering, School of Science and Technology, NOVA University of Lisbon, 2829-516 Caparica, Portugal

³Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, OH 45221, USA

⁴Foxconn Industrial Internet, Milwaukee, WI 53177, USA

⁵Department of Electrotechnical and Industrial Informatics, University of Applied Sciences Emden/Leer, 26721 Emden, Germany

Corresponding author: Ricardo Silva Peres (ricardo.peres@uninova.pt)

This work was supported by the FCT/MCTES (UNINOVA-CTS) under Grant UIDB/00066/2020.

ABSTRACT The advent of the Industry 4.0 initiative has made it so that manufacturing environments are becoming more and more dynamic, connected but also inherently more complex, with additional inter-dependencies, uncertainties and large volumes of data being generated. Recent advances in Industrial Artificial Intelligence have showcased the potential of this technology to assist manufacturers in tackling the challenges associated with this digital transformation of Cyber-Physical Systems, through its data-driven predictive analytics and capacity to assist decision-making in highly complex, non-linear and often multistage environments. However, the industrial adoption of such solutions is still relatively low beyond the experimental pilot stage, as real environments provide unique and difficult challenges for which organizations are still unprepared. The aim of this paper is thus two-fold. First, a systematic review of current Industrial Artificial Intelligence literature is presented, focusing on its application in real manufacturing environments to identify the main enabling technologies and core design principles. Then, a set of key challenges and opportunities to be addressed by future research efforts are formulated along with a conceptual framework to bridge the gap between research in this field and the manufacturing industry, with the goal of promoting industrial adoption through a successful transition towards a digitized and data-driven company-wide culture. This paper is among the first to provide a clear definition and holistic view of Industrial Artificial Intelligence in the Industry 4.0 landscape, identifying and analysing its fundamental building blocks and ongoing trends. Its findings are expected to assist and empower researchers and manufacturers alike to better understand the requirements and steps necessary for a successful transition into Industry 4.0 supported by AI, as well as the challenges that may arise during this process.

INDEX TERMS Artificial intelligence, Industry 4.0, digital transformation, guidelines, systematic review, framework, manufacturing.

NOMENCLATURE

AI	Artificial Intelligence
AR	Augmented Reality
CA	Context Awareness
CE	Continuous Engineering
CPS	Cyber-Physical System
CPPS	Cyber-Physical Production System
CV	Computer Vision
EC	Eligibility Criteria
FAIR	Findable, Accessible, Interoperable and Reusable

GAN	Generative Adversarial Network
HL	Human in the Loop
ICT	Information and Communication Technology
IOT	Internet of Things
LOA	Level of Autonomy
MAE	Mean Absolute Error
MAS	Multi-Agent System
MASE	Mean Absolute Scaled Error
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MSE	Mean Squared Error
NLP	Natural Language Processing
PRISMA	Preferred Reporting Items for Systematic review and Meta-Analysis

The associate editor coordinating the review of this manuscript and approving it for publication was Her-Terng Yau¹.

PPV	Positive Predicted Value
RQ	Research Question
RMSE	Root Mean Squared Error
SMOTE	Synthetic Minority Over-Sampling Technique
SOA	Service-Oriented Architecture
SC	Screening Criteria
TRL	Technology Readiness Level

I. INTRODUCTION

The recent shift towards customer-driven, highly customized manufacturing as part of the interconnected environment of the Industry 4.0 strategy is making it more and more important for manufacturers to strive for higher agility, productivity and sustainability [1]. Smart manufacturing has appeared as a way to apply advanced intelligent systems to enable a dynamic response to variable product demand, along with a real-time optimization along the entire value chain.

With the recent developments in ICT technologies, particularly regarding IoT, big data and CPPS, it is now feasible to implement the necessary flexibility, responsiveness and intelligence to face these challenges. CPPS in particular target the implementation of autonomous and collaborative manufacturing entities with advanced self-capabilities such as self-optimization, self-awareness and self-monitoring. In this context of the Industry 4.0 paradigm, AI is being regarded as one of the key technologies to achieve these capabilities and to disruptively redefine the way manufacturing processes and business models are structured.

AI can be generally defined as sub-discipline of computer science dealing with the development of data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement [2]. Still, there is not yet any generally accepted, unambiguous, and exact definition of the term. Due to the emphasis on learning, ML is considered one of the central sub-areas of AI (albeit not the only one), with the terms being sometimes used interchangeably. From an industrial point of view, AI technologies can be seen as enablers for systems to perceive their environment, process the data they acquire and solve complex problems, as well as to learn from experience in order to improve their capability to solve specific tasks.

A. INDUSTRIAL ARTIFICIAL INTELLIGENCE

In the context of this work, a less restrictive adaptation of the definition for Industrial AI provided in [3] is proposed. In this setting, Industrial AI can be defined as a systematic discipline focusing on the development, validation, deployment and maintenance of AI solutions (in their varied forms) for industrial applications with sustainable performance [4]. Hence, Industrial AI is an interdisciplinary area of research, encompassing fields such as ML, NLP and robotics. Considerable research efforts have been made over the last few years on how to combine and embed these concepts into existing Industry 4.0 manufacturing value chains [5]–[8].

The combination of these fields imbues the system with the ability to adapt and solve problems within pre-defined system boundaries through a certain degree of autonomous action.

Industrial AI distinguishes itself within the field of AI in five particular dimensions:

- **Infrastructures:** Concerning hardware and software, there is a large emphasis on real-time processing capabilities, ensuring industrial-grade reliability with high security requirements and interconnectivity;
- **Data:** Industrial AI requires data characterized by its large volume, high velocity variety, originating from various units, products, regimes, etc.
- **Algorithms:** It requires the integration of physical, digital and heuristic knowledge. High complexity derived from model management, deployment and governance.
- **Decision-making:** Given the industrial setting, tolerance for error is generally very low, with uncertainty handling being extremely important. Efficiency is of special importance for large-scale optimization problems.
- **Objectives:** Industrial AI addresses mostly concrete value creation through a combination of factors such as scrap reduction, improved quality, augmented operator performance or accelerated ramp-up times.

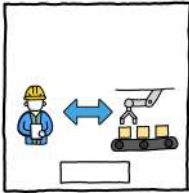
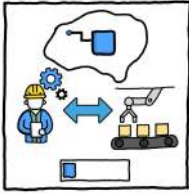
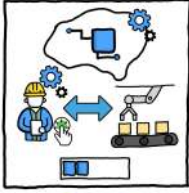
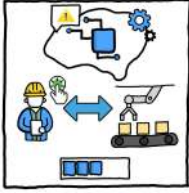
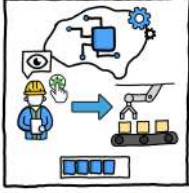
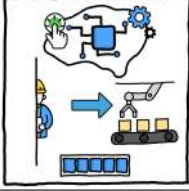
Taking into account the heterogeneous nature of industrial systems and their applications, it is useful to describe autonomous actions through a graduated model of autonomy, given that different LOA can be considered depending on the requirements of the application area and particular use case. To this extent, the taxonomy of system autonomy based on AI adopted by the Plattform Industrie 4.0 [9] can be used, defining a six-level model of automated decision-making (akin to the classification used applicable to autonomous driving [10]) on the basis of industrial processes. A representation of this model contextualized with industrial scenarios for each level is provided in Table 1.

These LOA can be used to describe not only the current state of a system or one of its parts, but also desired states to be achieved in the future. To attain a particular LOA, industrial systems need to be imbued with additional intelligence, which in turn is based on knowledge acquired through experience. Thus, Industrial AI can be seen as a core technology driving the pursuit of higher degrees of autonomy in industrial systems. Nevertheless, it is important to note that currently Industrial AI is mainly leveraged to augment human performance rather than fully replace them, which will likely still hold true even in more autonomous scenarios in the future.

B. DIGITAL TRANSFORMATION ROADBLOCK

Despite the high expectations held by the industry regarding AI, its actual prevalence in industrial enterprises is still quite low. A detailed survey of AI in manufacturing has suggested that the majority of research in the field is performed at most in laboratory environments [11]. The reason for this

TABLE 1. Taxonomy of system autonomy based on AI, defining a six-level model of automated decision-making for industrial processes. Each level of autonomy is defined, along with an illustrative industrial scenario representative of the corresponding level. Adapted from [9].

Level	Industrial Scenario	
Level 0 - No autonomy: Human operators have full control without any assistance from the AI system		A robot performing pick and place operations in pre-defined, fixed positions within fixed system boundaries. The robot is programmed with a pre-set behaviour by humans, who select and prioritise its rules
Level 1 - Assistance with respect to select functions: Human operators have full responsibility and make all decisions		The robot functions similarly to Level 0. However, at autonomy level 1, a robot assistance system programmed using AI can suggest goal-oriented improvements, such as process optimizations concerning cost, energy or time. These suggestion require the approval of a human supervisor to take effect.
Level 2 - Partial autonomy: in clearly defined areas, human operators have full responsibility and define (some) goals		At level 2 the robot is still predominantly programmed in pre-set manner by humans. However, the self-improvements go beyond level 1, with the AI programming allowing the robot to improve its behaviour within specified system boundaries and goals. An example of this behaviour would be the robot being capable of recognising and picking parts which are not in the exact pre-set position. Humans retain decision-making power and intervene when/if necessary.
Level 3 - Delimited autonomy: In larger sub-areas, the AI system warns if problems occur, human beings validate the solutions recommended by the system		The robot is only partially programmed in a pre-set manner by humans. On top of being capable of adjusting its own behavior, the robot can make and implement plans within the specified system boundaries, including for instance autonomous path control. This can be done in collaboratively with other entities in its environment. For this purpose the robotic system should be equipped with sensors necessary to perceive the environment, its context and to learn skills. Humans oversee the system's decisions, assist in resolving unforeseen disturbances and intervene in case of emergency.
Level 4 - System is adaptable and functions autonomously: Within defined system boundaries, human operators can supervise or intervene in emergency situations		At this level the system behaves as an adaptive, autonomous system in larger sub-areas within known system boundaries. Self-optimization within these boundaries is enabled through continuous learning phases and defined (partial) goals, leading to improved predictions and problem-solving capability. Humans relinquish control of a specific part of the system, shifting to a monitoring role and intervening only in emergency cases. If the human fails to intervene, the robotic system is capable of handling some situations according to its own perception of adequate corrective action.
Level 5 - Full autonomy: The AI system operates autonomously in all areas, including in cooperation and in fluctuating system boundaries. Human operators do not need to be present		At level 5 the robotic system acts with full autonomy and in collaboration with other autonomous systems within system boundaries specified by humans. In case of disturbances or fluctuating working parameters, the system is capable of dynamically adapting the plan and communicate it to other autonomous entities. In emergency cases, the system independently puts itself in secure mode.

lies in the enormous changes and expenditures needed to integrate AI applications into corporate structures and along entire value chains. To ensure the realization of a successful and complete digital transformation in the manufacturing industry, companies need to understand not only the potential impact of these disruptive technologies (which in most cases they do to some extent), but also their main requirements and consequently the organizational changes required to realize their full potential. Yet, most manufacturers do not have a comprehensive roadmap and framework to guide the integration of AI into their existing business models and processes [3].

This lack of understanding, guidance and common practices represents a critical roadblock for the data-driven digital transformation commonly associated with the Industry 4.0 vision. While several research efforts have pushed towards the deployment of AI solutions in the industry, companies often fail to follow through after their conclusion, missing the opportunity to reap their full benefits. Due to the lack of sufficient evidence of successful industrial applications of AI, the industrial adoption of the technology is thus hindered. On the research side, this exacerbates the issues with data availability and quality, as manufacturers continue to report and log their data in a variety of non-standard

formats with varying degrees of quality, making it difficult to break this cycle without systematic approaches for the implementation, deployment and management of Industrial AI solutions.

In an attempt to fill this gap, this study will present a systematic literature review of Industrial AI to derive common design principles, core technologies and the main challenges to overcome in order to achieve successful and sustainable deployments of Industrial AI at a high TRL. Furthermore, based on this assessment, the goal is to propose a strategic roadmap to guide both researchers and manufacturers alike in the transition towards digital data-driven industrial processes in the manufacturing industry. To this end, the following research questions were formulated:

RQ1 What is the current status of Industrial AI in manufacturing?

RQ2 What are the main design principles of Industrial AI?

RQ3 What are the main challenges and future research directions?

The remainder of this paper is organized as follows: Section II describes the methods employed to carry out the systematic literature review and subsequent assessment of the selected body of literature to identify key design principles in Industrial AI applications. Afterwards, Section III aims to characterize the current Industrial AI landscape through the systematic literature review, identifying its main trends, design principles and application areas. Following this, Section IV proposes a conceptual framework to guide industrial implementations of Industrial AI systems based on the results of the literature review. Section V explores the challenges and main opportunities for future research. Finally, Section VI discusses the limitations of this study, with Section VII providing its conclusions and closing remarks.

II. METHODS

A. STUDY IDENTIFICATION, SCREENING AND ELIGIBILITY

The systematic literature review was conducted through a mixed-methods approach [12] (including both qualitative and quantitative research methods), following the guidelines outlined in the PRISMA statement [13]. The PRISMA flow chart reporting the different phases of the systematic review is shown in Figure 1.

To assist in the process of constructing a search string to collect records from different digital databases, a network graph for the concept of AI was built using the Google search engine. By using as an input the string “Artificial Intelligence vs” to prompt typical direct comparisons with the term, we extracted the autocomplete suggestions of the search engine and used the ranking (ordering) of the list as the weight of the edge. By repeating this step for each of the results with a depth of three, the network graph presented in Figure 2 was generated.

From the analysis of the network graph presented in Figure 2, a search string was constructed based on the core concepts associated with AI, constrained to the

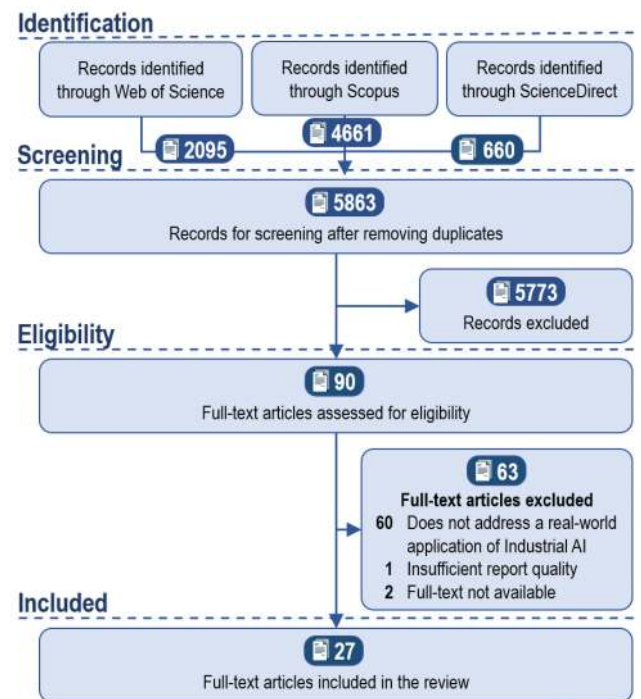


FIGURE 1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart of study inclusions and exclusions for the systematic literature review.

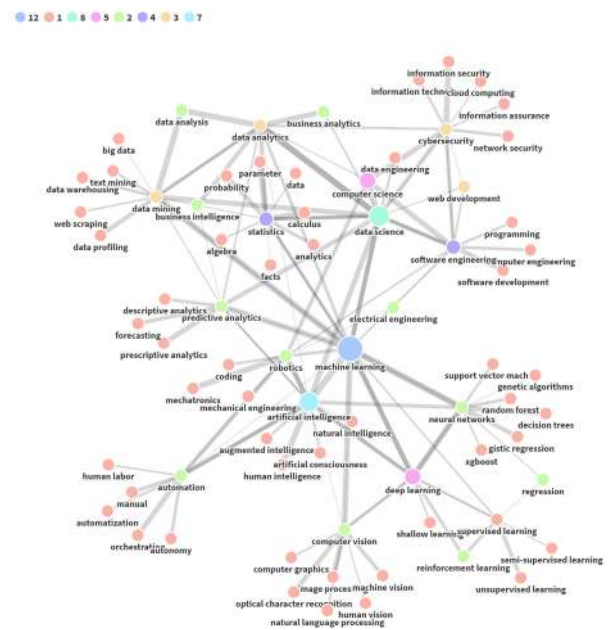


FIGURE 2. Ego graph for AI based on Google's Autocomplete queries.

manufacturing domain. This was achieved by ensuring that at least one element of each of the groups listed in Table 2 is present, using a combination of OR and AND operators.

This search string was then adapted to each of the three electronic databases included in this study, namely Web of Science, Scopus and ScienceDirect. The search was conducted on 12 July 2020, including academic research that

TABLE 2. General search string to be adapted for each of the selected digital repositories (Web of Science, ScienceDirect and Scopus).

Group 1	Group 2
artificial intelligence deep learning machine learning data science predictive analytics	manufacturing industr* 4.0

was: (1) published between 2016 and July 2020; contained at least one term from each group (2) in either the abstract, title or keywords; (3) published in peer-reviewed journals or conference proceedings; (4) written in the English language. After this the resulting records were aggregated and the duplicates removed.

Following the identification phase, a first screening process was carried out by two independent reviewers following the criteria listed in Table 3. Any discrepancies were resolved through discussion.

TABLE 3. Exclusion criteria for the screening phase of the systematic literature review.

ID	Screening Criteria
SC1	Record must include at least the Title, Year, Source, Abstract and DOI
SC2	Must be in English
SC3	Must have been published in a Q1 journal (according to SCImago [14])
SC4	Abstract must address the application of I-AI in a real environment
SC5	Must not be solely a survey, review or roadmap

Regarding SC4, a second screening was performed for records that passed the first round by first automatically searching the abstracts for mentions of “real data”, “real environment” or similar variants and then briefly reviewing the corresponding record by reading the title, abstract and keywords. A similar process was performed for SC5. This second screening was employed in an effort to exclude publications which did not focus on the technical aspects and challenges of Industrial AI in real industrial environments.

Finally, all remaining articles had their full text analyzed in further detail based on the eligibility criteria described in Table 4. No further restrictions were imposed on the setting, application, methods or outcomes reported.

TABLE 4. Exclusion criteria for the eligibility phase of the systematic literature review.

ID	Eligibility Criteria
EC1	Does not address the deployment of I-AI in a real environment
EC2	Insufficient report quality (missing or ill-reported information on one or more sections of the article, making it impossible to assess the real-world application)
EC3	Full-text not available

B. DATA COLLECTION

For each of the articles eligible to be included in the study, two types of data were extracted. Firstly, the basic information

about the publication was collected, including (1) publication title, (2) authors, (3) abstract, (4) year, (5) journal title and (6) journal classification according to SCImago [14].

Building on these, the second part deals specifically with attempting to answer the research questions listed in Section I-B:

- For RQ1 ‘What is the current status of Industrial AI in manufacturing?’, the data extracted from the eligible publications are: (1) the application domain, references to (2) software or hardware employed by the authors, along with an assessment of the (3) TRL and (4) level of autonomy of the proposed application.
- For RQ2 ‘What are the main design principles of Industrial AI’: (1) text descriptions (sentences in the full text of the publication) highlighting design decisions taken by the authors (e.g. service orientation, continuous engineering).
- For RQ3 ‘What are the main challenges and future research directions?’: (1) research objects (e.g. production data, logistics or workers), as well as (2) research purposes (e.g. predictive maintenance, energy optimization or ergonomics) addressed by each publication.

C. DATA ANALYSIS

As previously mentioned, the data analysis was performed using a mixed-approach combining both qualitative and quantitative methods. To complement the quantitative analysis, which addresses both statistical and graphical data descriptions, it is possible to take advantage of semi-automated techniques [15] using NLP to qualitatively pre-process and analyse selected literature in a much shorter amount of time and at a larger depth when compared to more traditional manual methods [16].

To this effect, the each abstract was pre-processed and cleaned following a sequence of steps through using a Python script. Initially, stop-words (e.g. “a”, “the” and “in”) and punctuation were removed given their low significance in this context. Following this, token n-grams were constructed consisting of one to three words stemmed using Porter’s stemming algorithm [17]. In this instance, stemming refers to the process of breaking a word down to its roots, meaning as an example “challenged”, “challenges” and “challenging” would correspond to the root “challenge” [18]. Finally, the frequency is counted to find the most frequent n-grams in the corpus.

III. CHARACTERIZATION OF CURRENT INDUSTRIAL AI RESEARCH

Following the PRISMA guidelines and the steps described in Section II, a total of 5863 unique records were found in the chosen electronic databases matching the search string from Table 2, out of which 90 were selected for further assessment based on the screening criteria listed in Table 3.

In an effort to ensure the quality and reliability of the publications included in the analysis, as well as to better comprehend how recent research as addressed the challenges

inherent to deploying Industrial AI solutions in real environments, only journal articles classified as Q1 (according to SCImago [14]) were considered for the analysis. The top 25 journals in terms of publication count resulting from the database search are listed in Figure 3. Along with the application of the remaining criteria mentioned in Table 4, this resulted in 27 articles being eligible for further data collection and analysis in the study.

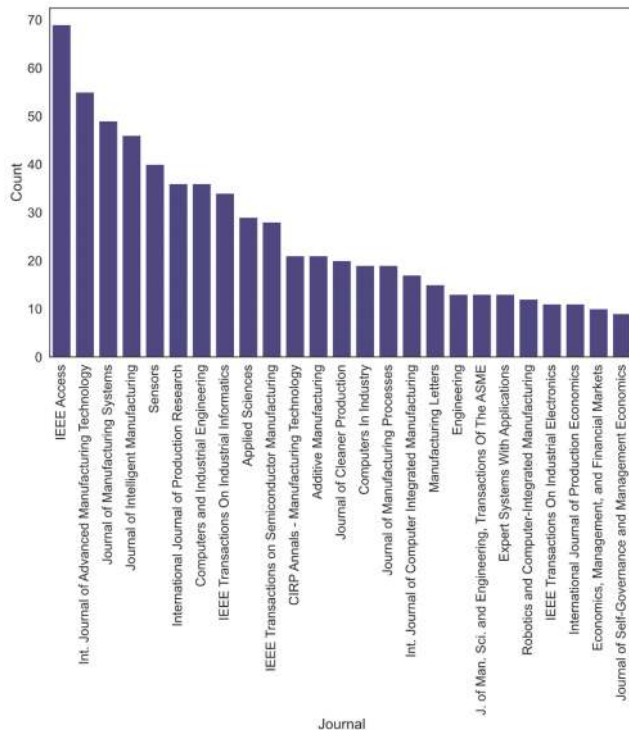


FIGURE 3. Top 25 Q1 journals encompassed in the database search of Industrial AI research, ordered by publication count.

This section presents the results from this process, with the aim of providing an overall characterization of Industrial AI research. Hence, it covers the maturity level of the solutions found in current literature, the main trends and their common design principles, serving as the basis to answer the research questions devised in Section I-B.

A. RESEARCH TRENDS

Due to the large volume of research data readily available in online repositories which provide access to digital publications, it is possible to take advantage of automated techniques [15] to search a wider range of the existing literature in a much shorter amount of time and at a much larger depth when compared to more traditional manual methods.

To establish a baseline for the thorough assessment carried out later in this study, NLP techniques were used to find the 30 most frequent bigrams and trigrams occurring in the corpus. First, these were applied to the corpus of 5863 abstracts resulting from the initial stage of the database search (see Figures 4 and 5). This was done in order to obtain

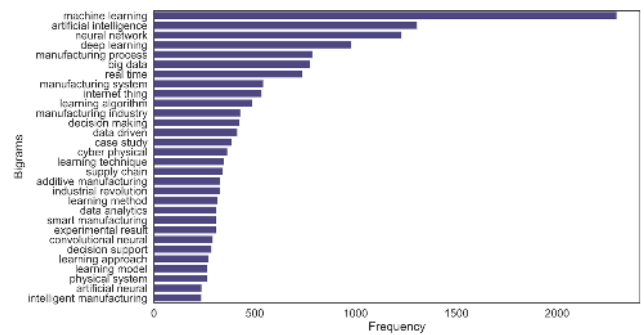


FIGURE 4. Top 30 most frequent bigrams resulting from the analysis of the corpus of 5863 abstracts.

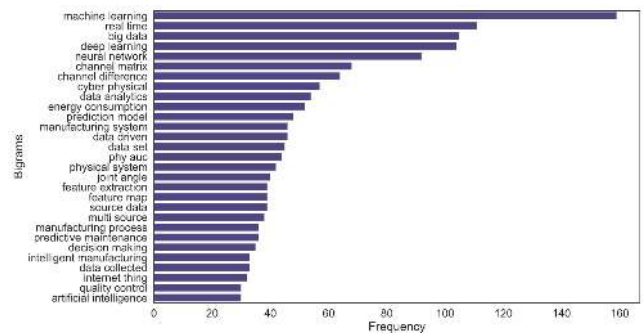


FIGURE 5. Top 30 most frequent bigrams resulting from the analysis of the corpus of 27 full-text articles.

a general sense of the publications and their main topics. Then, the same approach was applied to the full-text of the 27 articles included in the study (see Figures 6 and 7) in order to enable the comparison between the overall landscape of Industrial AI and the research efforts which are specifically discussing the deployment in real industrial environments.

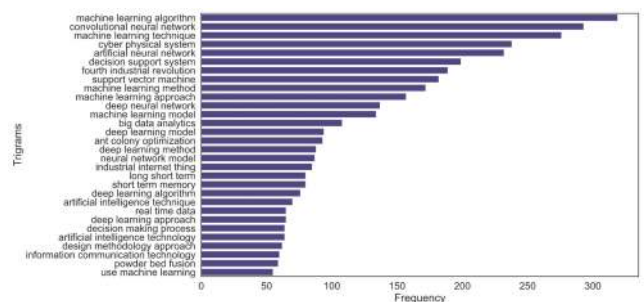


FIGURE 6. Top 30 most frequent trigrams resulting from the analysis of the corpus of 5863 abstracts.

Aligned the constraints imposed during the initial search, it is possible to verify that before narrowing down the study, the abstracts appear to frequently mention key concepts related with Industrial AI, namely *Machine Learning*, *Industry 4.0*, *Big Data*, *Industrial IoT* and *CPPS*. Additionally some broad areas of interest are identified as well, including *supply chain*, *additive manufacturing* and *decision-support* in general.

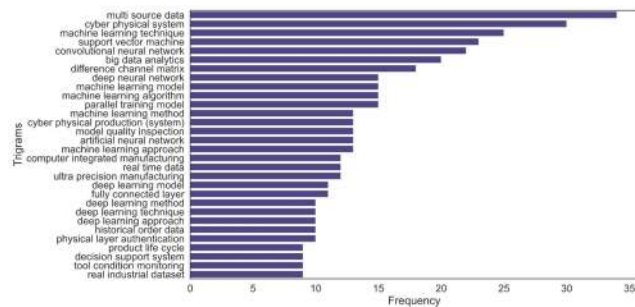


FIGURE 7. Top 30 most frequent trigrams resulting from the analysis of the corpus of 27 full-text articles.

In contrast, it is interesting to observe that the systematic approach was successful in narrowing down the focus of the selected publications, from the more general topics to real data-driven industrial applications still aligned with the general scope. This is useful to assist and guide the more thorough manual assessment of the full-texts, later presented in Section III-B.

From this, the main application areas of interest appear to *energy optimization*, *predictive maintenance* and *quality control*, largely dominated by *deep learning* in terms of the methods employed. Furthermore, it is also useful to identify some of the design principles that one can hypothesize to be inherent to the most frequent terms. More specifically, *interoperability* (e.g. “multi source data”), *real-time capability* (e.g. “real time data”), *cybersecurity* (e.g. “physical layer authentication”), *scalability* and *decentralization* (e.g. “big data analytics”, “parallel training”).

B. STUDY OF THE INDUSTRIAL AI LANDSCAPE

Going beyond the semi-automated analysis, a more thorough manual assessment of the 27 publications included in the study in their full-text form was carried out. Table 5 summarizes the findings from this assessment, extracting from each article the key information required to address the research questions of the study. This includes the application setting, data and methods reported by the respective authors, as well as a qualitative assessment of the perceived TRL and LOA levels of the research work in question, providing further insight into the current status of the Industrial AI landscape. The design principles are identified as explicitly stated by the respective authors as core concepts of the reported research, among those identified as the common fundamental principles among current Industrial AI literature.

From the analysis of Table 5 one can derive several observations in regards to the current state of Industrial AI research. A summary of the findings concerning the TRL, LOA, and publication years of the articles included in the study can be found in Figure 8. Concerning the assessment of the TRL for the research work reported in each of the publications, it is clear that there is a major roadblock around TRL 5-6 (pilot stage), with the vast majority of publications (70.3%) standing at or below this level. This is in line with the

discussion from Section I-B, as the nonexistence of common practices and guidelines for the long-term deployment of Industrial AI, along with the lack of substantial evidence of industrial success makes it difficult to increase the market uptake of such solutions beyond this stage.

It is also interesting to look further into the coverage of the different design properties highlighted in this study. Figure 9 summarizes the results extracted from Table 5 in this regard.

Given the emphasis on decision-making support and human involvement associated with the reported applications, it is natural to see *real-time capability* and *human in the loop* as the most common design properties among the publications included in the study. It is however concerning to see that there is still such a gap in terms of the employment of *explainability* and *cybersecurity* practices. Such a gap might explain the resistance towards a higher industrial uptake of these solutions, as their low interpretability by non-expert personnel and lack of coverage in terms of privacy and security issues can easily result in a perceived lack of trust and reliability by industrial stakeholders. Additionally, there is no common, formalized way to report results, with considerable variance between different reports.

It is also worth noting that authors rarely discuss limitations of their approaches. However, data availability, quality and related issues (e.g. scarcity, contamination, drift) could be considered limitations for nearly all publications included in the study. In this regard, the formulation and adoption of common Industrial AI reporting guidelines, akin to the TRIPOD statement [48] in healthcare could contribute to mitigate this issue, improving publication quality and reducing the risk of bias.

The perceived level of autonomy of the applications is still generally quite low. Applications are typically limited to very specific and tight boundaries providing decision-making support to human supervisors, hence the emphasis on the human in the loop.

In addition, it can be observed that solutions showcasing higher maturity and TRL tend to share a substantial number of the design principles, while lower TRLs tend to focus on more specific, narrow aspects. Interoperability for instance is considered in most solutions above TRL 6 due to the heterogeneous and multi-source nature of data in Industry 4.0 manufacturing systems, but it typically comes at the cost of considerable additional pre-processing and cleaning effort.

The following subsections provide an in-depth analysis of the topics covered in Table 5, going into further detail regarding the core enabling technologies and the main design principles that guide industrial implementations of AI.

C. CORE ENABLING TECHNOLOGIES

While the theory and concepts behind Industrial AI are paramount for its understanding and scientific advancement, it is through the tooling that its potential can be fully realized to solve real world problems. In this regard this field presents

TABLE 5. Overview of eligible applications of Industrial AI reported between 2016-2020. Public datasets used in the studies are included when available. Limitations are listed as reported by the authors. Legend: RT - Real-time; HL - Human in the Loop; RB - Robustness; IO - Interoperability; DC - Decentralization; SO - Service-orientation; M - Modularity; CA - Context-Awareness; S - Scalability; CE - Continuous Engineering; IP - Interoperability; C - Cybersecurity.

Author	Application Domain	Data	Methods	RT	HL	RB	IO	DC	SO	M	CA	S	CE	IP	C	Limitations	TRL	LOA
Tannous et al. [19]	Collaborative Robots, Welding	Real	Haptic Touch, Statistical	✓	✓	-	-	-	-	-	-	-	-	-	-	Data Quality	5-6	2
Lai et al. [20]	Workforce Training, Manual Assembly	Synthetic, Real	AR, DL Object Detection	✓	✓	-	-	-	-	✓	-	-	-	-	-	-	5-6	1
Bergamini et al. [21]	Collaborative Robots, CV, Manual Assembly	Cornell dataset [22]	Deep Learning, ROS	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	5	2
Ojer et al. [23]	Quality Control, CV, Manual Assembly, Electronics	Synthetic, Real	Segmentation, Supervised Learning	✓	✓	✓	-	-	-	-	-	✓	-	-	-	-	5-6	1
Chien et al. [24]	Demand Forecasting, Supply Chain Management	Real	Deep Reinforcement Learning	-	✓	✓	-	-	-	-	-	-	-	-	-	-	6-7	1
Zan et al. [25]	Pattern Recognition, Quality Control	Synthetic, Real	Deep Learning	-	✓	-	-	-	-	-	-	-	-	-	-	-	4-5	1
Li et al. [26]	Quality Control, CV, Additive Manufacturing	UB-Moog dataset [27]	Semi-Supervised Deep Learning	✓	-	-	-	-	-	-	-	-	-	-	-	Interpretability, Data Availability	3-4	1
Romeo et al. [28]	Design Engineering	Real	Supervised Learning	-	✓	-	-	-	✓	-	-	-	✓	✓	-	Data Availability, Data Quality	5-6	1
Yu et al. [29]	Fault Detection, Predictive Maintenance	Real	Unsupervised Learning	✓	✓	✓	-	✓	-	-	-	✓	✓	-	✓	Data Quality	8-9	1
Ruiz-Sarmiento et al. [30]	Predictive Maintenance, Steel Industry	Real	Supervised Learning	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	5-6	1
Huang et al. [31]	Production Progress Prediction, IoT	Real	Deep Learning	✓	-	-	✓	✓	-	-	-	-	-	-	-	Data Availability	6-7	1
Pan et al. [32]	Physical Layer Authentication	Real	Supervised Learning	✓	-	-	-	-	-	-	-	-	-	-	✓	Data Availability, Robustness	5-6	2
Maggipinto et al. [33]	Virtual Metrology, Semiconductor Industry	Real	Deep Learning	✓	-	-	-	-	-	-	-	-	✓	-	-	-	4-5	2
Abobakr et al. [34]	Posture Analysis, CV, Ergonomics	Synthetic, Real	Deep Learning	-	✓	✓	-	-	-	-	-	-	-	-	-	Complexity, Comp. Cost	5-6	1
Juez-Gil et al. [35]	Lifetime Prediction, Steel Plates	Real (Public [35])	Supervised Learning	-	✓	-	-	-	-	-	-	-	-	✓	-	Data Availability	5-6	1
Ansari et al. [36]	Prescriptive Maintenance, Automotive Industry	Real	Supervised Learning, NLP	✓	✓	-	✓	✓	-	✓	✓	✓	✓	-	-	Data Availability, Data Quality, Cybersecurity	6-7	1
Shi et al. [37]	Condition Monitoring, Ultra-Precision Process	Real	Deep Learning	-	-	✓	-	-	-	-	-	-	-	-	-	Variability in setup parameters	4-5	1
Stoyanov et al. [38]	Qualification Testing, Electronics Industry	Real	Supervised Learning	✓	-	✓	-	-	-	-	-	-	-	-	-	-	4-5	1
Zhu et al. [39]	Fault Detection, CV, Chemical Industry	Real	Deep Learning, Statistical	✓	✓	-	-	-	-	-	-	-	-	-	-	-	4-5	1
Peres et al. [40]	Quality Control, Automotive	Real	Supervised Learning, MAS	✓	✓	-	-	✓	✓	✓	-	-	✓	-	-	Concept Drift	6	1
Qin et al. [41]	Energy Optimization, Additive Manufacturing	Real	Deep Learning	-	-	-	✓	✓	-	-	✓	-	-	-	-	-	4-5	1
Lee et al. [42]	Quality Control, Metal Casting	Real	Supervised Learning	✓	✓	-	✓	✓	✓	✓	✓	-	-	-	-	Data Availability, Cybersecurity	7-8	2-3
Schmitt et al. [43]	Quality Control, Electronics Industry	Real	Supervised Learning	✓	✓	-	✓	✓	✓	-	✓	✓	-	-	-	Automated data pipelines, CE	7-8	2-3
Hwangbo et al. [44]	Robotics, Legged Locomotion	Synthetic, Real	Reinforcement Learning	✓	-	✓	-	-	-	✓	-	-	-	-	-	Human expertise per task, Task generalization	5-6	2-3
De Vita et al. [45]	Fault Prediction, Data Fusion, IoT	Real	Deep Learning	✓	✓	-	✓	✓	-	✓	✓	✓	-	-	-	-	6	1
Woo et al. [46]	Predictive Analytics, Energy Efficiency	Real	Supervised Learning, Statistical, MAS	✓	-	-	✓	✓	✓	✓	✓	✓	✓	-	-	Disturbance handling (Robustness)	5-6	2-3
Lee et al. [47]	Predictive Analytics, Condition Monitoring	-	-	✓	-	✓	✓	-	✓	-	✓	-	-	-	-	Data Quality, Cybersecurity	2-3	2-3

a rich ecosystem of technologies, tools, frameworks and libraries that are fueled by a large community riding on the hype of recent AI advancements (in addition to pre-existing technology in operations research).

Given the prevalence of ML in current Industrial AI literature, one efficient way to kickstart the overview of its core enabling technologies is to build an Ego graph based on the term “scikit-learn” [49], currently one of the most popular and open-source ML libraries available in Python.

This step was carried out following the same process described in Section II-A, the result of which can be seen in Figure 10.

A widely adopted basic tooling for traditional ML applications consists in a set of three main open source libraries. Beyond the aforementioned *scikit-learn*, which provides various tools for model fitting, data preprocessing, model selection and evaluation, *Pandas* [50] facilitates data manipulation and *Numpy* for scientific computing. To complement these,

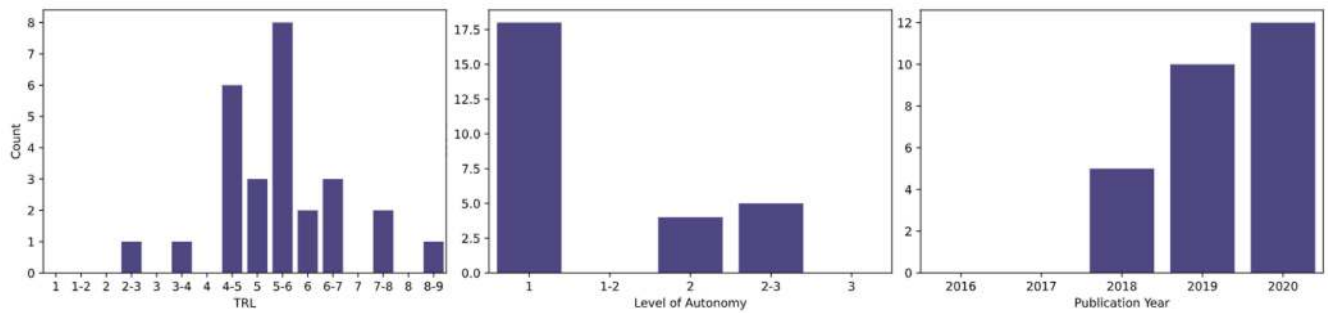


FIGURE 8. Distribution of technology readiness level, level of autonomy and publication year for the articles included in the study.

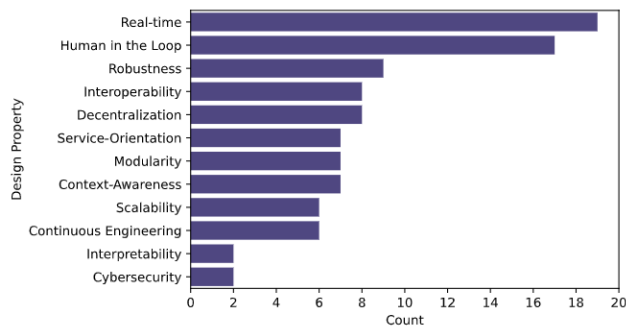


FIGURE 9. Coverage of the different design properties across the body of articles included in the study. Further explored in Table 6.

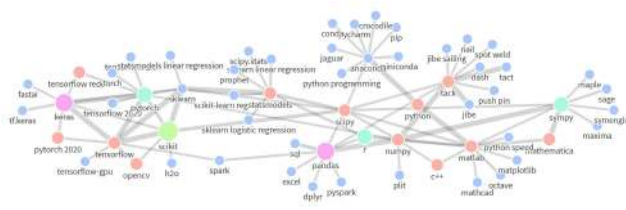


FIGURE 10. Ego graph for AI/ML technology landscape based on Google's autocomplete queries originating from "scikit-learn", a popular ML framework. Search Depth = 3.

matplotlib and *seaborn* are commonly used for visualization purposes.

For Deep Learning, *Tensorflow* [51] (alternatives include *Pytorch* and *CNTK*) is a core open source, end-to-end ML platform for Deep Learning provided by Google. *Tensorflow* provides a large ecosystem of tools, from graph modules to visualization with *TensorBoard*. It is also available through *Google Colab*, a cloud service providing free access to GPU and TPU computing with minimal effort and resources required on local machines, making it also interesting for training and teaching environments. Higher-level APIs such as *Keras* and *fastai* provide user-friendly interfaces for rapid-experimentation and flexibility, which in conjunction with *CUDA* [52] can facilitate distributed training in clusters of GPUs or TPUs.

Moving towards big data, scalability and decentralization, open-source processing engines such as *Apache Spark*, *Flink* and *Storm* enable distributed computation with low latency and high throughput [53]. These engines are easily integrated with multiple data sources, which in this context can include distributed storage like *HDFS*, data warehouses such as *Apache Hive* and NoSQL databases such as *Cassandra* and *MongoDB*. Clusters using this engines can be easily set up in cloud-based services such as *Microsoft Azure* and *Amazon AWS* which enable the deployment of such solutions at scale.

Another concept which has been gaining popularity recently is the automation of ML (AutoML) processes through libraries such as *H2O* and the low-code *Pycaret*, making the barrier of entry of these technologies lower and lower. Also it enables more customized experiences at scale, with automatic feature engineering, modeling and even deployment. On the one hand, AutoML is a step towards the democratization of AI, making it more widely accessible instead of being available only to a select few. On the other, one downside of this is the false sense of expertise it can often convey, with practitioners lacking fundamental skills to go beyond prototyping and to address real-world challenges in data quality, scalability and integration, which in turn can further contribute to a lack of trust and interest from industrial stakeholders.

Finally, going beyond ML, it is also worth it to look into technologies that assist in addressing the challenge of interoperability and modularity in heterogeneous and complex real world industrial environments. In this direction, *IEC Fieldbus (Profibus)* [54] and the *OPC-UA* [55], [56] machine-to-machine communication protocol are seen as industry standards for interoperability in data exchange contemplated in the communication layer of the Reference Architectural Model Industrie 4.0 [57]. In the case of modularity and robustness, *Industrial MAS* [58] have played a pivotal role in the development of CPPS, allowing them to be designed in a decentralized manner through the distribution of functionality among autonomous and cooperative agents [59]. Through the combination with other AI methods such as ML, industrial agents have provided CPPS with the capacity to learn and self-adapt with additional flexibility [60] and reconfigurability [61].

D. DESIGN PRINCIPLES

Naturally, several of the common design principles found in Industrial AI literature are shared with those of the Industry 4.0 vision [62]. These include *decentralization*, *modularity* and *real-time capability*. However, others are more specific to the context of applied AI, as it is the case for the aspects of *interpretability*, *robustness* and *cybersecurity* for instance from the standpoint of privacy-preserving AI. Table 6 indicates the relevance of each design principle for Industrial AI systems, framing it in the context of application examples to facilitate their comprehension. Figure 11 highlights the design principles identified through the systematic literature review, along with the respective application areas which will be further explored in Section III-E.

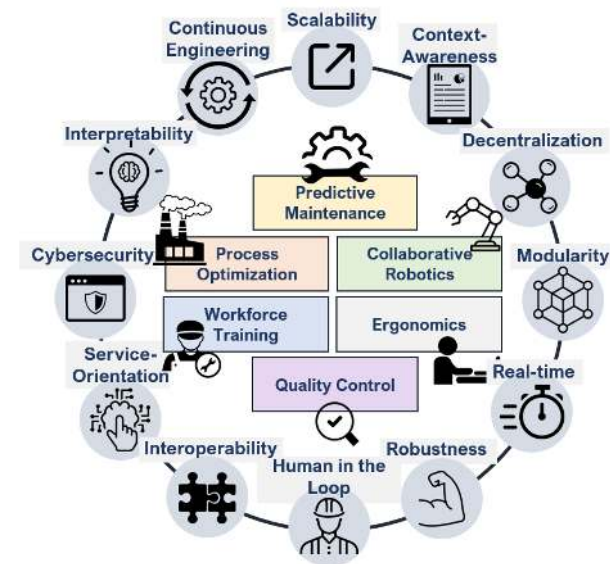


FIGURE 11. The main design principles acting as pillars of Industrial AI in the context of its application areas.

Generally, human operators and engineers currently play a central role in most Industrial AI applications found in real manufacturing environments. While the adoption level for this involvement of the human in the loop is quite high, we are still far from a fully symbiotic relationship between the human and AI, with the former mostly taking full responsibility of the action and the latter acting only as a decision-support system with limited autonomy within clearly defined boundaries. To progress further in this direction, a better understanding of the reasoning and mechanisms behind AI-based decisions is necessary on the stakeholders' side, demystifying the AI "black box".

Such an understanding can be achieved through the consideration of AI interpretability. While it is difficult to find a clear-cut definition of the term, in the context of this work we have adopted the formalization used in [68], describing it as the degree to which an observer can understand the cause of a decision. The terms interpretability and explainability will be used interchangeably. Thus, an explanation can be seen as the

mode through which an observer can obtain understanding, or simply put, the answer to a "why" question. Another important aspect is that explanations are social, as in part of an interaction between the explainer and the receiver of said explanation. Hence, context plays a large role in determining the content and nature of the explanations.

In order to enable the interpretation of information, not only the context but also the semantics and grammatical structure of the information are crucial. The more explicit these definitions become, the easier it is for the different actors to communicate and collaborate in an interoperable way. In the context of Industrial AI, functions and consequences must be unambiguously defined, with created and explicit knowledge being constantly validated by domain knowledge experts. In line with this, robust integration with legacy IT systems (such as ERP, PLM and MES applications) should be addressed proactively as these typically encompass various data sources which can provide valuable inputs to successfully deploy Industrial AI applications at scale. Standardization and the adoption good common practices in a company-wide data-driven culture can play a major role in this direction, easing the replication and scaling of these applications as part of complex CPPS beyond the first implementation.

Effective design of complex CPPS is also dependent on modularity [69], which in this case implies that AI modules should have clear interfaces and easily allow their composition within the CPPS, providing it with additional capabilities and functionalities as needed in a flexible way.

Interestingly, the added flexibility of the plug & produce paradigm in modular systems makes the already difficult aspect of security even worse, as cybersecurity is typically seen as a characteristic rather than a design principle. As stated in [70], this misconception has led to the development of several insecure systems since this principle is not something that can be easily bought or added onto an existing system. It is a continuous and iterative process starting from the design stage, encompassing most if not all aspects of the respective system.

E. MAJOR APPLICATION AREAS

1) PROCESS OPTIMIZATION

The employment of Industrial AI towards process optimization in manufacturing is gaining rapid traction, enabling smarter, more efficient data-driven decision-making by leveraging both historical and real-time data. In this regard, the main emphasis has been put into energy consumption prediction and optimization problems [41], production efficiency [71] and demand forecasting [24]. Thus, the application of Industrial AI for process optimization can contribute to make manufacturing processes more profitable, while also being more sustainable and efficient.

2) QUALITY CONTROL

The inherent complexity of multistage manufacturing processes such as assembly and machining, along with their

TABLE 6. Key design principles and their relevance to Industrial AI.

Design Principle	Relevance to Industrial AI	Adoption	Application Example
Real-time Capability	For data-driven applications, real-time capability is crucial to turn new AI predictions and insights into actionable knowledge at both the level of the processes and of the overall smart factory operations in a timely manner, adequate for the increasingly real-time economy.	Very High	Industrial AI can be used for inline monitoring or control due to low inference times, thus managing to cope with the constraints imposed by the takt time. Examples include inline sorting, quality control [43] and assisting workers in manual assembly tasks [20].
Human in the Loop	Humans can take on either a collaborative or supervisory role ensuring the safety and possibly continuous improvement of AI operations within specific boundaries.	Very High	Provide labelling for supervised learning based on domain knowledge, monitor/validate suggestions/decisions from the AI system depending on the desired LOA.
Robustness	As shop-floor conditions can change considerably over time, applications should be robust to cope with unexpected disturbances and changes in either the data, the system or the environment itself.	Good	CV applications should be robust to geometric or photometric changes in shop-floor conditions. Supervised learning models using structured data from production should be robust to changes or drifts caused by disturbances, maintenance, material changes or other events.
(Semantic) Interoperability	Interoperability is the means through which machines, people and data can be connected in a meaningful way. It enables heterogeneous assets to effectively exchange information, harmonizing different communication standards, protocols and data representations.	Good	IoT devices at the edge can publish data through MQTT, PLC I/O data can be aggregated in an OPC-UA server, while a CPPS can consume both, then transform the data, annotate it with context in a common representation format and send it to historical cloud storage to be later used for analysis or model training by a different module or micro-service.
Decentralization	Autonomous and semi-autonomous operation and decision-making distributed between the edge, fog and cloud layers through IoT, CPPS and Cloud Computing.	Good	Big data can be collected from multiple sensors and stored in a distributed file system. Time consuming ML model training based on the collected data can be performed in the cloud, with low-latency inference carried out by models deployed at the edge.
Service Orientation	As part of the adoption of *-as-a-service revenue models, Industrial AI can be structured as a collection of loosely coupled and self-contained micro-services. This increases flexibility and facilitates integration and continuous engineering.	Fair	AI-as-a-service has been gaining traction with commercial solutions being made available by major players such as Microsoft ¹ , Google ² and Amazon ³ . Through simple APIs AI services can be integrated into existing CPPS without requiring considerable development effort or resources [63].
Modularity	Industrial AI solutions can be designed as modular building blocks that can flexibly be added or removed to the overall CPPS providing specific capabilities as necessary.	Fair	Modular Industrial AI components should be able to be dynamically deployed as the CPPS deems necessary, not only adapting to changes in system topology but also to accommodate new data sources.
Context Awareness	Context is information used to characterize the situation of an entity [64]. Contextualization is crucial to realize the full potential of raw data, making sense of it and transforming it into actionable knowledge.	Fair	IoT systems use multiple sensors to capture low-level contextual data of industrial entities. The heterogeneous, multi-source nature of the data provides crucial information to understand and reason about the its situation within the environment.
Scalability	Beyond the proof of concept stage, manufacturers should be able to scale Industrial AI solutions once data governance strategies are in place to enable easier, repeatable and faster rollouts. As more data becomes available, solutions can be expanded to cover fleets of assets or across multiple sites.	Low	Scalability should be considered early in the design stage. For instance, models can simply grow large enough that they are unable to fit in the working memory of a single training device. It should be considered across different stages, from data warehousing to model training and deployment, being closely related with decentralization and modularity.
Continuous Engineering	Continuous monitoring and improvement of deployed solutions to cope with the dynamic and ever-changing nature of real industrial environments.	Low	Monitor the model's online performance after deployment, re-training in the cloud when it degrades beyond a given threshold and then re-deploying to the edge.
Interpretability	The higher the interpretability of an Industrial AI application, the easier it is for industrial stakeholders to comprehend the reasoning behind certain predictions or decisions. This builds trust, improves transparency and assists in promoting market uptake.	Very Low	Knowing <i>why</i> a given output is produced can provide further insight into the model, the problem and the data itself. As an example, it can assist domain experts in identifying if the milling machine is more likely to fail due to the feed rate, the spindle RPM or the current tool supplier.
Cybersecurity	As sensitive data-driven systems in highly decentralized, flexible and connected environments, Industrial AI systems are vulnerable to cyber-attacks like eavesdropping (compromising privacy), data poisoning (adversarial contamination of training data) and denial-of-service [65].	Very Low	Applying encrypted computation, allowing Industrial AI systems to be trained and run on encrypted data without allowing them to see, leak or abuse data in its unencrypted form [66]. Examples include homomorphic encryption, secure multi-party computation and functional encryption [67].

unforeseen disturbances and uncertainties, make it challenging to guarantee the desired quality of the product in industries like automotive and aerospace [72]. Thus, effective methods to enable the automated and early detection of potential defects during production using real-time data are highly desirable to manufacturers. Emerging applications include automated visual inspection using deep learning methods [23], [27], defect prediction to mitigate multistage propagation (aligned with the zero-defect manufacturing paradigm) [40] and online quality prediction [43].

3) PREDICTIVE MAINTENANCE

Given the steep costs associated with unplanned downtime in the manufacturing industry, a large portion of existing

applications of Industrial AI are focused on increasing machine operability and uptime by detecting possible problems before they occur. On the one hand, this mitigates the risk of breakdown events occurring with catastrophic consequences, on the other it can be seen as a way to reduce unnecessary inspection and maintenance operations resulting from time-based policies, thus resulting in cost and resource optimization.

Generally, such approaches can be modelled based on the degradation severity of machine performance and the processing of multiple heterogeneous data sources. This can be done for a particular type of machine or based on the information of a fleet of machines based on similarity. Afterwards, through the assessment of maintenance effectiveness of different maintenance policies, relevant costs, resources and the particular context at hand, the optimal maintenance strategy can be determined [73].

¹<https://azure.microsoft.com/en-us/services/>

²<https://cloud.google.com/ai-platform>

³<https://aws.amazon.com/machine-learning/ai-services/>

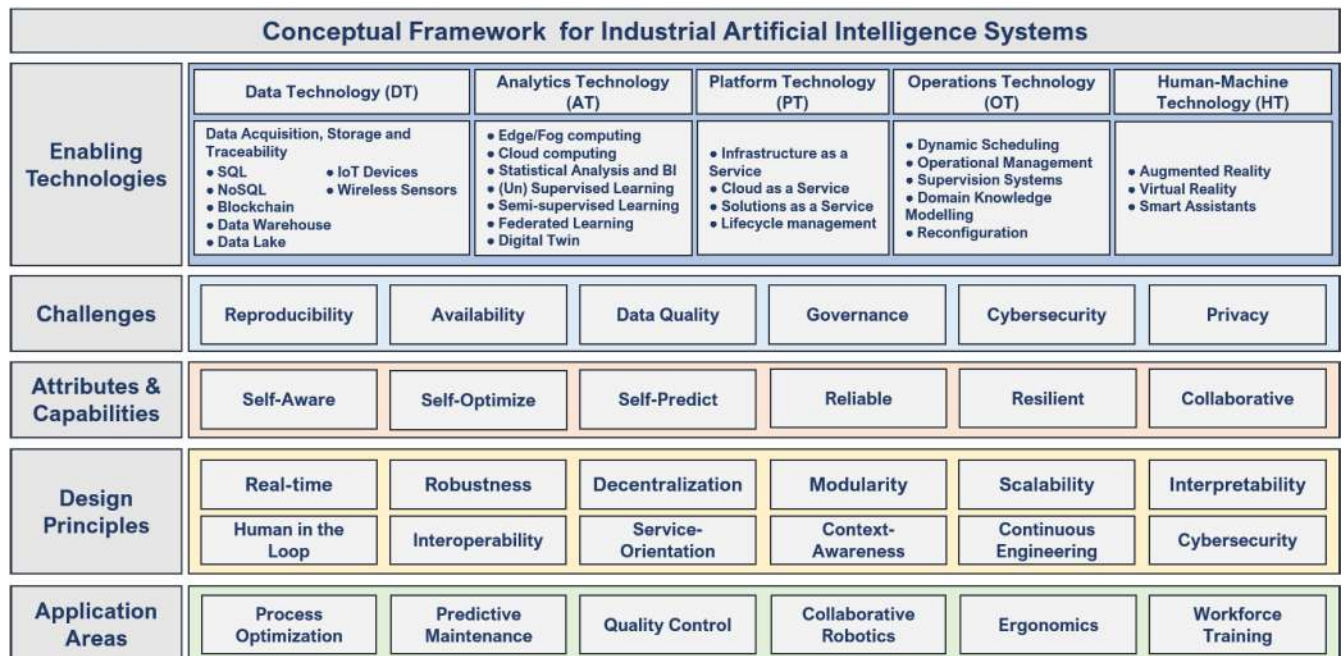


FIGURE 12. Conceptual framework for Industrial AI Systems. Adapted from [4].

4) HUMAN-ROBOT COLLABORATION AND ERGONOMICS

Industrial AI presents a tremendous opportunity to empower human-robot collaboration and provide support to current human-centric tasks on the shop-floor, be it by improving the welfare and safety of operators or by making their tasks easier and more efficient. Therefore, current opportunities for the application of Industrial AI in this domain include workforce training and task support [20], [23], collaborative robotics [19], [21] and ergonomics [34], [74].

IV. CONCEPTUAL FRAMEWORK FOR INDUSTRIAL AI

The aim of this section is to provide the guidelines for manufacturers to overcome the roadblock discussed in Section I-B. For this purpose, these guidelines are clearly defined within a framework for industrial implementation (Figure 12) adapted from [4], bringing together the findings originating from the systematic literature review in terms of the applications, challenges, principles and technologies driving Industrial AI.

The proposed conceptual framework highlights the capabilities and attributes these systems should encompass, based on the design principles defined previously, to meet the common requirements of manufacturing environments in the Industry 4.0 setting. These include not only self-awareness and self-optimization for continuous improvement, but also the resilience and reliability expected for industrial applications. Such characteristics ensure that the system is capable of handling equipment failures or other unexpected disturbances by quickly restoring its normal operation state. As discussed in Table 1, it is expected that with additional autonomy, the system's intelligent entities should be able to collaboratively resume normal operation by dynamically rescheduling and implementing new plans in a coordinated manner.

The enabling technologies behind Industrial AI can be broadly categorized into five dimensions, namely data, analytics, platform, operations and human-machine technology. These five dimensions are described below:

- **Data Technology (DT):** The digitalization of manufacturing is making it so that larger and larger volumes of data are generated at each step. These data can be structured, unstructured or mixed, originating from multiple sources at different levels of abstraction (e.g. component, machine or shop-floor level). Thus, to extract value from the data it is imperative to enrich it with context and improve standardization in data acquisition and exchange processes.
- **Analytics Technology (AT):** In order to transform the raw data acquired through DT into knowledge and, consequently, added value for enterprises, adequate AT is required. This includes data processing at different levels, from the edge to the cloud, leveraging real-time data streams and ML to enable continuous improvement through self-learning mechanisms and self-optimization. New challenges in this area push towards the exploration of novel methods such as federated learning and semi-supervised learning approaches.
- **Platform Technology (PT):** Platform technologies act as enablers for the remaining technologies, facilitating the interconnection between different elements be it at the edge, fog, or cloud levels. Given the requirements for added flexibility and agility, it is important for PT to support self-reconfiguration and self-organization capabilities. Furthermore, the higher degree of connectivity makes it so that cybersecurity becomes a critical point to address to ensure the privacy, availability and integrity of the system.

- **Operations Technology (OT):** Operations technology is crucial to value creation, moving from analytics to actionable knowledge provided by decision-making support systems. Through the combination of OT with the previous technologies, a shift can be made from experience-driven to data-driven production with optimized operational maintenance and management.
- **Human-Machine Technology (HT):** Industrial AI carries the disruptive potential to profoundly change the role of the human in modern manufacturing and the way these systems interact with people. Hence, HT must be explored to ensure that stakeholders can be empowered to effectively, intuitively and seamlessly interact with these systems to fully reap the benefits provided by Industrial AI. HT can assist personnel through virtual or augmented reality to improve and facilitate operations such as maintenance and assembly, or even in cases of remote diagnosis. Additionally, while HT can be used to train the workforce in these tasks, it is also important that enterprises invest in proper training and acquisition of talent to ensure the full potential of these systems can be harnessed.

Naturally, these elements must be integrated into new or legacy systems, at which point CPPS can play a major role acting as the glue that brings it all together, creating the necessary bridges between the Industrial AI components and the shop-floor. As such, future research in this field will remain a truly interdisciplinary effort, requiring the combination of multidisciplinary expertise and domain knowledge.

From here, Section V will further explore the challenges portrayed in the conceptual framework, discussing potential opportunities for the research agenda of Industrial AI.

V. CHALLENGES AND OPPORTUNITIES FOR FUTURE RESEARCH

Concisely, Industrial AI challenges for future research can be categorized in three fronts:

- **FAIR Data for Industrial AI:** Ensuring data is easily findable, accessible, interoperable and reusable is crucial to serve the best interests of the research and industrial communities alike, promoting the advancement of science to the benefit of society [75].
- **FAIR Models for Industrial AI:** Beyond data, to enable the implementation of FAIR frameworks digital assets such as models should be reliably found and re-used when appropriate, through persistent identifiers linked to rich metadata (including for example provenance) represented through common, standardized and secure formats.
- **Cyber-infrastructures:** Proper infrastructures are crucial to ensure the level of quality, security and reliability required to improve the industrial uptake of Industrial AI solutions. These include remote operations, cybersecurity, privacy-preserving mechanisms, 5G technology and collaborative prognostics, among others.

The remainder of this section will address some of the main specific challenges within these major fronts in further detail.

A. DATA AVAILABILITY

Industrial AI has shown tremendous potential in a wide array of manufacturing applications, from defect classification in quality control, to fault detection in predictive maintenance, ergonomics and human assistance in manual tasks. Still, data availability remains as a major challenge to overcome in order to go beyond the pilot stage, as a considerable volume of research work has been based on the assumption that sufficient and adequate data is available to successfully train and validate models.

ML and Deep Learning in particular require a very large amounts of (mostly labelled) data to achieve proper generalization and avoid overfitting. However, in real manufacturing environments data from different settings, conditions and configurations is often scarce (e.g. different failures, defects, energy consumption), given that these typically represent undesired states of the system and acquiring said data with currently adopted practices tends to be unfeasible from both an economic and operational standpoint. Moreover, labelling raw data is a time consuming and costly endeavour which in this context frequently requires expertise and domain knowledge.

In the industry, given sufficient time and guidance organizational changes and alignment of business models towards Industrial AI can ensure that data is collected, curated and stored adequately with the intent of facilitating data-driven AI applications. In the meantime, from a research perspective there are currently two main venues being explored to address this challenge.

1) DATA SYNTHESIS

The first is related with the generation of synthetic data which closely resembles data from real operational environments and thus allows the generalization of Industrial AI solutions based on it to real scenarios. This approach has been successfully employed in roughly 20% of the publications included in this study as documented in Table 5. For instance regarding classification problems, traditional approaches to handle class imbalances due to data availability issues usually involve artificially re-sampling the data set. One way to achieve this is by under-sampling the majority class [76], resulting in a balanced distribution. However, it also implies that some valuable data can potentially be lost, which can be particularly critical for small datasets. Another way is oversampling the minority class [77], with one simple method being the randomized replication of instances of the minority to once again achieve a more balanced distribution, at the risk of potential overfitting.

A popular approach is SMOTE [78], which creates synthetic samples by interpolating among neighboring minority class instances in feature space. While SMOTE and its extensions have achieved considerable results in recent years [79],

studies have shown that for high-dimensionality problems its effects are underwhelming [80].

With the advent of GANs [81], new opportunities for the generation of reliable synthetic data in manufacturing have arisen. The GANs can be used to learn the distributions of the original data and the generate fake, yet realistic samples to expand the training dataset. Some promising examples of this can be found in the literature, addressing the class imbalance problem in applications in which the data of faulty operational conditions is scarce and difficult to obtain [82]–[84]. Recent efforts have also shown promising empirical results on semi-supervised learning [85]–[87] for cases in which labelling the entire dataset is unfeasible or too costly.

2) TRANSFER LEARNING

The motivation behind the concept of transfer learning lies in the fact that humans can intelligently apply previous knowledge to solve new problems either faster or with better solutions [88]. For the purposes of Industrial AI, it consists in the improvement of learning in a new problem for which data is scarce, through the transfer of knowledge from a related task in a given source domain for which sufficient data is available.

The applicability of transfer learning in manufacturing can generally be split into three main scenarios, namely the transfer between different (1) working conditions, (2) different machines/stations, (3) different types of machine faults or product defects [11]. A recent application can be found in [89], where the authors applied transfer learning from a model trained on non-manufacturing data to manufacturing condition monitoring, as well as transferring between different working conditions and machines. In another example [90], transfer learning is applied for bearing fault diagnosis under different working conditions categorized by imposing different motor speeds on the test bearings.

B. DATA QUALITY

Following the common saying of “garbage in, garbage out”, Industrial AI and ML models in particular rely heavily on accurate, clean and often appropriately labelled (in supervised approaches) training data to produce useful results, making data quality a critical factor for the industrial success of these solutions. Data quality can be organized into four main dimensions [91]: (1) *Intrinsic*, which refers to the characteristics that are native to the data itself, including timeliness, completeness, accuracy and consistency [92]; (2) *Contextual*, meaning the attributes that are dependent on the context of the task at hand such as relevancy and quantity; (3) *Representational*, addressing the fact that systems should present data in a way that is interpretable and is represented concisely and consistently; (4) *Accessibility*, emphasizing that systems should store and provide access to the data in a way that it becomes easy to manipulate in a way that is accessible but also reliable, secure and privacy-preserving.

While some approaches can be found in the literature to improve specific dimensions of data quality [93], the assessment and measurement of these dimensions have historically

relied on self-report surveys and user questionnaires [94] due to their association with subjective and situational judgments for quantification. Thus, further research is needed to jointly improve the way the different dimensions of data quality can be monitored and optimized, as this will likely have a direct impact towards improving the performance of Industrial AI applications leveraging these data.

C. CYBERSECURITY AND PRIVACY

It is evident that the typical Industry 4.0 combination of multiple data sources and emerging technologies such as IoT, cloud computing, AI and blockchain enhance operation efficiency of entire manufacturing processes. Nevertheless, this also comes at the cost of possible cybersecurity threats as discussed in Table 6, especially in the context of collecting large volumes of data for centralized processing, which poses severe privacy concerns [95].

Recently, federated learning approaches have started to emerge as a way to mitigate the aforementioned privacy and scalability issues, distributing the training process across multiple industrial nodes. Through federated learning, these nodes can collaboratively build a model without sharing sensitive private data samples, only local parameters.

While this represents a large step-up in terms of addressing critical issues of data privacy and security, recent studies have shown that even in federated learning scenarios several risks can be found, particularly regarding reverse-engineering attacks that can extract sensitive information about the datasets directly from the model [96]. As such, it is imperative that future research addresses privacy-preserving constructs for AI, with some examples including secure multi-party computation schemes and differential privacy [97].

D. GOVERNANCE

1) INTERPRETABILITY AND TRUST

The demystification of the “blackbox” nature of some Industrial AI solutions is a crucial factor to enable stakeholders to better comprehend the technology and facilitate its widespread industrial adoption. It is considerably easier to convince stakeholders that a given solution should be adopted to improve the bottom-line if it can be easily broken-down and its processes validated by domain experts. To this end, explainable AI emerges as one of the main research directions to drive the industrial adoption of Industrial AI with interpretability tools as the main catalyzer.

In this direction, some approaches have gained considerable traction in the field over recent years, particularly those based on model-agnostic methods [98]. These include Local interpretable model-agnostic explanations, or LIME [99], and Shapley Additive Explanations [100]. An important aspect of model-agnostic methods is that they provide increased model flexibility, since the interpretation method can work with most ML models. This avoids having to limit the range of models to those that are inherently interpretable, avoiding a potential lost of predictive performance compared to

other types of models. Another alternative would be to use model-specific interpretation methods, with the disadvantage being that such an approach is binding to a specific model type, making it more difficult to switch later on if needed [101].

Therefore, the modularity of model-agnostic interpretation methods makes them highly desirable for AI research in the long-term, particularly for the automation of interpretability at scale. Since the interpretations are decoupled from the underlying ML model, replacing either side of the process is easier, following the trend of automated feature engineering, hyper-parameter optimization, model selection and ultimately model interpretation.

2) ALGORITHMIC FAIRNESS AND BIAS

The topic of algorithmic fairness and bias is particularly challenging as there is no universally accepted notion of fairness. On top of this, bias can originate from several sources other than the data itself, including the data pipeline and its pre-processing steps, people involved and their respective actions (whether intentional or not).

Several mitigation actions can be taken, including the collection of additional data, adapting data-processing and post-processing (e.g. thresholding). Regardless, one of the first steps towards the mitigation of AI unfairness and bias should be auditing. Some efforts towards facilitating this process are currently being made, including the Aequitas open-source library [102] which intends to empower both practitioners and policymakers to audit ML models for discrimination and bias, thus being able to make equitable decisions regarding the development and deployment of predictive solutions.

Furthermore, often it is possible to come across scenarios in which there is a trade-off between performance and fairness. Models that are fairer or less discriminatory across different subsets of the population often do so at the cost of global performance. Hence, it is important to assess in which application contexts the trade-off is worth it, for instance depending on whether or not human welfare is directly impacted by the outcome of the system.

3) COMMON MISCONCEPTIONS

Interpretability, fairness and bias are definitely important aspects that should be taken into account when approaching the topic of Industrial AI. However, this does not necessarily mean that algorithms and models should always be fully or even equally interpretable or fair. In this regard, one should first consider the domain and intended application. Could the solution adversely impact human welfare? Perhaps, if factory workers are directly impacted by the outcome of an Industrial AI-based high stakes decision-making process then this is crucial, yet such is not always the case.

Moreover, the potential trade-off with other important metrics must be considered. Would improving the interpretability or fairness of the solution significantly impact its performance, security or privacy in a negative manner? These

are questions which do not necessarily have an immediate answer, thus making it imperative for the future research agenda on Industrial AI to include ways to assess, quantify and audit such characteristics in a continuous way, ensuring that stakeholders have all the knowledge required to take informed and conscious business actions.

VI. LIMITATIONS

Due to the exclusion criteria used while performing the retrieval of identified research from the electronic databases, it is possible that some relevant publications might have been left out of the study. This is particularly true regarding the specific focus on the tailoring and deploying Industrial AI solutions in real industrial environments, as records were included on the basis of mentioning a variation of “real world/environment/scenario/use case” within the respective abstract. Furthermore, the inclusion of articles published solely in the English language naturally implies the exclusion of non-English documents and possible relevant content contained therein.

Additionally, the conceptual framework presented in Section IV has been developed based on the systematic review of existing Industrial AI literature and empirical evidence in the context of manufacturing technology development and strategic management. It provides an holistic view of the core elements manufacturing organizations and technology providers should consider to overcome current Industrial AI roadblocks and successfully transition towards data-driven business models aligned with Industry 4.0. Hence, some adaptation and tailoring to specific manufacturing settings and other industries is expected to be necessary.

Lastly, despite being a well-established research field, some of the more recent facets of Industrial AI are still in their infancy. Due to this aspect of novelty, some of the publications cited in this work are still in their pre-print form and can thus be subjected to changes in the future.

VII. CONCLUSION

A systematic review of journal publications indexed to the Web of Science, Scopus and ScienceDirect databases was conducted following PRISMA guidelines to clearly define and characterize the current landscape of the Industrial AI research in manufacturing. To address the first RQ “*What is the current status of Industrial AI in manufacturing?*”, the study identified the main research trends, enabling technologies and application areas of Industrial AI, with particular emphasis on real-world deployments of such solutions.

This made it possible to identify and characterize the core design principles that define Industrial AI applications, corresponding to the second RQ, having implications in the design, implementation and deployment of future Industrial AI solutions in manufacturing. To this end, the findings from the systematic review and characterization of the landscape were formalized into a conceptual framework to assist

manufacturers in the transition towards a data-driven culture, providing an holistic view of the Industrial AI ecosystem.

Lastly, to address the final RQ “*What are the main challenges and future research directions?*”, further considerations regarding the current status of Industrial AI applications led to the categorization of the main challenges currently being faced in the field, enabling the formulation of possible venues for future research efforts to facilitate the transition towards successful data-driven industrial deployments of Industrial AI solutions.

REFERENCES

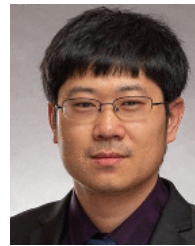
- [1] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, and S. D. Noh, “Smart manufacturing: Past research, present findings, and future directions,” *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 3, no. 1, pp. 111–128, Jan. 2016.
- [2] *Systems and Software Engineering—Vocabulary*, Standard 24765:2017, International Organization for Standardization, Geneva, Switzerland, Sep. 2017.
- [3] J. Lee, J. Singh, and M. Azamfar, “Industrial artificial intelligence,” *Intell. Maintenance Syst. (IMS)*, Dept. Mech. Eng., Univ. Cincinnati, Cincinnati, OH, USA, 2019.
- [4] J. Lee, *Industrial AI*. Singapore: Springer 2020.
- [5] J. Lee, B. Bagheri, and H.-A. Kao, “A cyber-physical systems architecture for industry 4.0-based manufacturing systems,” *Manuf. Lett.*, vol. 3, pp. 18–23, Jan. 2015.
- [6] R. S. Peres, A. Dionisio Rocha, P. Leitao, and J. Barata, “IDARTS—Towards intelligent data analysis and real-time supervision for industry 4.0,” *Comput. Ind.*, vol. 101, pp. 138–146, Oct. 2018.
- [7] X. Zhang, X. Ming, Z. Liu, D. Yin, Z. Chen, and Y. Chang, “A reference framework and overall planning of industrial artificial intelligence (I-AI) for new application scenarios,” *Int. J. Adv. Manuf. Technol.*, vol. 101, nos. 9–12, pp. 2367–2389, Apr. 2019.
- [8] A. Clark, N. A. Zhuravleva, A. Siekelova, and K. F. Michalikova, “Industrial artificial intelligence, business process optimization, and big data-driven decision-making processes in cyber-physical system-based smart factories,” *J. Self-Governance Manage. Econ.*, vol. 8, no. 2, pp. 28–34, 2020.
- [9] *Plattform Industrie 4.0, Technology Scenario Artificial Intelligence in Industrie 4.0*, Federal Ministry Economic Affairs Energy (BMWi), Berlin, Germany, 2019.
- [10] C. Rödel, S. Stadler, A. Meschtscherjakov, and M. Tscheligi, “Towards autonomous cars: The effect of autonomy levels on acceptance and user experience,” in *Proc. 6th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, 2014, p. 1–8.
- [11] J. F. Arinez, Q. Chang, R. X. Gao, C. Xu, and J. Zhang, “Artificial intelligence in advanced manufacturing: Current status and future outlook,” *ASME J. Manuf. Sci. Eng.*, vol. 142, no. 11, Nov. 2020, Art. no. 110804, doi: 10.1115/1.4047855.
- [12] L. A. Curry, I. M. Nembhard, and E. H. Bradley, “Qualitative and mixed methods provide unique contributions to outcomes research,” *Circulation*, vol. 119, no. 10, pp. 1442–1452, Mar. 2009.
- [13] K. Knobloch, U. Yoon, and P. M. Vogt, “Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias,” *J. Cranio-Maxillofacial Surg.*, vol. 39, no. 2, pp. 91–92, Mar. 2011.
- [14] SCImago. *SJR—SCImago Journal and Country Rank*. Accessed: Jul. 12, 2020. [Online]. Available: <http://www.scimagojr.com/>
- [15] H. Yasin, M. M. Yasin, and F. M. Yasin, “Automated multiple related documents summarization via Jaccards coefficient,” *Int. J. Comput. Appl.*, vol. 12, no. 3, pp. 12–15, 2011.
- [16] K. Crowston, E. E. Allen, and R. Heckman, “Using natural language processing technology for qualitative data analysis,” *Int. J. Social Res. Methodol.*, vol. 15, no. 6, pp. 523–543, Nov. 2012.
- [17] P. Willett, “The Porter stemming algorithm: Then and now,” *Program, Electron. Library Inf. Syst.*, vol. 40, no. 3, pp. 219–223, 2006, doi: 10.1108/00330330610681295.
- [18] H. Yasin, M. M. Yasin, and F. M. Yasin, “Automated multiple related documents summarization via jaccards coefficient,” *Int. J. Comput. Appl.*, vol. 12, no. 3, pp. 12–15, Jan. 2011.
- [19] M. Tannous, M. Miraglia, F. Inglese, L. Giorgini, F. Ricciardi, R. Pelliccia, M. Milazzo, and C. Stefanini, “Haptic-based touch detection for collaborative robots in welding applications,” *Robot. Comput.-Integr. Manuf.*, vol. 64, Aug. 2020, Art. no. 101952.
- [20] Z.-H. Lai, W. Tao, M. C. Leu, and Z. Yin, “Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing,” *J. Manuf. Syst.*, vol. 55, pp. 69–81, Apr. 2020.
- [21] L. Bergamini, M. Sposato, M. Pellicciari, M. Peruzzini, S. Calderara, and J. Schmidt, “Deep learning-based method for vision-guided robotic grasping of unknown objects,” *Adv. Eng. Informat.*, vol. 44, Apr. 2020, Art. no. 101052.
- [22] Cornell University. *Robot Learning Lab: Learning to Grasp*. Accessed: Jul. 19, 2020. [Online]. Available: http://pr.cs.cornell.edu/grasping/rect_data/data.php
- [23] M. Ojer, I. Serrano, F. Saiz, I. Barandiaran, I. Gil, D. Aguinaga, and D. Alejandro, “Real-time automatic optical system to assist operators in the assembling of electronic components,” *Int. J. Adv. Manuf. Technol.*, vol. 107, pp. 2261–2275, 2020, doi: 10.1007/s00170-020-05125-z.
- [24] C.-F. Chien, Y.-S. Lin, and S.-K. Lin, “Deep reinforcement learning for selecting demand forecast models to empower industry 3.5 and an empirical study for a semiconductor component distributor,” *Int. J. Prod. Res.*, vol. 58, no. 9, pp. 2784–2804, May 2020.
- [25] T. Zan, Z. Liu, H. Wang, M. Wang, and X. Gao, “Control chart pattern recognition using the convolutional neural network,” *J. Intell. Manuf.*, vol. 31, no. 3, pp. 703–716, Mar. 2020.
- [26] X. Li, X. Jia, Q. Yang, and J. Lee, “Quality analysis in metal additive manufacturing with deep learning,” *J. Intell. Manuf.*, vol. 31, pp. 2003–2017, Feb. 2020.
- [27] B. Zhang, P. Jaiswal, R. Rai, P. Guerrier, and G. Baggs, “Convolutional neural network-based inspection of metal additive manufacturing parts,” *Rapid Prototyping J.*, vol. 25, no. 3, pp. 530–540, Apr. 2019.
- [28] L. Romeo, J. Loncarski, M. Paolanti, G. Bocchini, A. Mancini, and E. Frontoni, “Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0,” *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112869.
- [29] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, “A global manufacturing big data ecosystem for fault detection in predictive maintenance,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 183–192, Jan. 2020.
- [30] J.-R. Ruiz-Sarmiento, J. Monroy, F.-A. Moreno, C. Galindo, J.-M. Bonelo, and J. Gonzalez-Jimenez, “A predictive model for the maintenance of industrial machinery in the context of industry 4.0,” *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103289.
- [31] S. Huang, Y. Guo, D. Liu, S. Zha, and W. Fang, “A two-stage transfer learning-based deep learning approach for production progress prediction in IoT-enabled manufacturing,” *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10627–10638, 2019.
- [32] F. Pan, Z. Pang, H. Wen, M. Luvisotto, M. Xiao, R.-F. Liao, and J. Chen, “Threshold-free physical layer authentication based on machine learning for industrial wireless CPS,” *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6481–6491, Dec. 2019.
- [33] M. Maggipinto, A. Beghi, S. McLoone, and G. A. Susto, “DeepVM: A deep learning-based approach with automatic feature extraction for 2D input data virtual metrology,” *J. Process Control*, vol. 84, pp. 24–34, Dec. 2019.
- [34] A. Abobakr, D. Nahavandi, M. Hossny, J. Iskander, M. Attia, S. Nahavandi, and M. Smets, “RGB-D ergonomic assessment system of adopted working postures,” *Appl. Ergonom.*, vol. 80, pp. 75–88, Oct. 2019.
- [35] M. Juez-Gil, I. N. Erdakov, A. Bustillo, and D. Y. Pimenov, “A regression-tree multilayer-perceptron hybrid strategy for the prediction of ore crushing-plate lifetimes,” *J. Adv. Res.*, vol. 18, pp. 173–184, Jul. 2019.
- [36] F. Ansari, R. Glawar, and T. Nemeth, “PriMa: A prescriptive maintenance model for cyber-physical production systems,” *Int. J. Comput. Integr. Manuf.*, vol. 32, nos. 4–5, pp. 482–503, May 2019.
- [37] C. Shi, G. Panoutsos, B. Luo, H. Liu, B. Li, and X. Lin, “Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultra-precision manufacturing,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3794–3803, May 2019.
- [38] S. Stoyanov, M. Ahsan, C. Bailey, T. Wotherspoon, and C. Hunt, “Predictive analytics methodology for smart qualification testing of electronic components,” *J. Intell. Manuf.*, vol. 30, no. 3, pp. 1497–1514, Mar. 2019.
- [39] W. Zhu, Y. Ma, M. G. Benton, J. A. Romagnoli, and Y. Zhan, “Deep learning for pyrolysis reactor monitoring: From thermal imaging toward smart monitoring system,” *AIChE J.*, vol. 65, no. 2, pp. 582–591, Feb. 2019.

- [40] R. S. Peres, J. Barata, P. Leita, and G. Garcia, "Multistage quality control using machine learning in the automotive industry," *IEEE Access*, vol. 7, pp. 79908–79916, 2019.
- [41] J. Qin, Y. Liu, and R. Grosvenor, "Multi-source data analytics for AM energy consumption prediction," *Adv. Eng. Informat.*, vol. 38, pp. 840–850, Oct. 2018.
- [42] J. Lee, S. Noh, H.-J. Kim, and Y.-S. Kang, "Implementation of cyber-physical production systems for quality prediction and operation control in metal casting," *Sensors*, vol. 18, no. 5, p. 1428, May 2018.
- [43] J. Schmitt, J. Bönig, T. Borggräfe, G. Beiting, and J. Deuse, "Predictive model-based quality inspection using machine learning and edge cloud computing," *Adv. Eng. Informat.*, vol. 45, Aug. 2020, Art. no. 101101.
- [44] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Sci. Robot.*, vol. 4, no. 26, Jan. 2019, Art. no. eaau5872.
- [45] F. De Vita, D. Bruneo, and S. K. Das, "On the use of a full stack hardware/software infrastructure for sensor data fusion and fault prediction in industry 4.0," *Pattern Recognit. Lett.*, vol. 138, pp. 30–37, Oct. 2020.
- [46] J. Woo, S.-J. Shin, W. Seo, and P. Meilanitasari, "Developing a big data analytics platform for manufacturing systems: Architecture, method, and implementation," *Int. J. Adv. Manuf. Technol.*, vol. 99, nos. 9–12, pp. 2193–2217, Dec. 2018.
- [47] 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, Oct. 2018.
- [48] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement," *Circulations*, vol. 131, no. 2, pp. 211–219, 2015.
- [49] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
- [50] W. McKinney, "Data structures for statistical computing in python," in *Proc. 9th Python Sci. Conf.*, Austin, TX, USA, vol. 445, 2010, pp. 51–56.
- [51] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.
- [52] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with CUDA," *Queue*, vol. 6, no. 2, pp. 40–53, Mar. 2008.
- [53] B. Cheng, J. Zhang, G. P. Hanc, S. Karnouskos, and A. W. Colombo, "Industrial cyberphysical systems: Realizing cloud-based big data infrastructures," *IEEE Ind. Electron. Mag.*, vol. 12, no. 1, pp. 25–35, Mar. 2018.
- [54] E. Tovar and F. Vasques, "Real-time fieldbus communications using profibus networks," *IEEE Trans. Ind. Electron.*, vol. 46, no. 6, pp. 1241–1251, Dec. 1999.
- [55] S. Leitner and W. Mahnke, "OPC UA—Service-oriented architecture for industrial applications," Softwaretechnik-Trends, ABB Corporate Res. Center, Ladenburg, Germany, Tech. Rep., 2006, vol. 26.
- [56] R. Henssen and M. Schleipen, "Interoperability between OPC UA and AutomationML," *Procedia CIRP*, vol. 25, pp. 297–304, Dec. 2014.
- [57] M. Hankel and B. Rexroth, "The reference architectural model industrie 4.0 (RAMI 4.0)," in *Proc. ZVEI*, vol. 410, Apr. 2015, p. 2.
- [58] P. Leitão and S. Karnouskos, *Industrial Agents: Emerging Applications of Software Agents in Industry*. San Mateo, CA, USA: Morgan Kaufmann, 2015.
- [59] P. Leitão, S. Karnouskos, L. Ribeiro, J. Lee, T. Strasser, and A. W. Colombo, "Smart agents in industrial cyber-physical systems," *Proc. IEEE*, vol. 104, no. 5, pp. 1086–1101, May 2016.
- [60] R. S. Peres, A. D. Rocha, A. Coelho, and J. B. Oliveira, "A highly flexible, distributed data analysis framework for industry 4.0 manufacturing systems," in *Proc. Int. Workshop Service Orientation Holonic Multi-Agent Manuf.* Springer, 2016, pp. 373–381.
- [61] A. M. Farid and L. Ribeiro, "An axiomatic design of a multiagent reconfigurable mechatronic system architecture," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1142–1155, Oct. 2015.
- [62] M. Ghobakhloo, "The future of manufacturing industry: A strategic roadmap toward industry 4.0," *J. Manuf. Technol. Manage.*, vol. 29, no. 6, pp. 910–936, Oct. 2018.
- [63] M. Ribeiro, K. Grolinger, and M. A. M. Capretz, "MLaaS: Machine learning as a service," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 896–902.
- [64] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggle, "Towards a better understanding of context and context-awareness," in *Proc. Int. Symp. Handheld Ubiquitous Comput.* Springer, 1999, pp. 304–307.
- [65] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztin, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the Mirai botnet," in *Proc. 26th USENIX Secur. Symp. (USENIX Secur.)*, 2017, pp. 1093–1110.
- [66] M. Brundage *et al.*, "Toward trustworthy AI development: Mechanisms for supporting verifiable claims," 2020, *arXiv:2004.07213*. [Online]. Available: <http://arxiv.org/abs/2004.07213>
- [67] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 201–210.
- [68] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019.
- [69] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber-physical systems," *Proc. IEEE*, vol. 100, no. 1, pp. 13–28, Jan. 2011.
- [70] N. Tuptuk and S. Hailes, "Security of smart manufacturing systems," *J. Manuf. Syst.*, vol. 47, pp. 93–106, Apr. 2018.
- [71] Y. C. Liang, W. D. Li, X. Lu, and S. Wang, "Fog computing and convolutional neural network enabled prognosis for machining process optimization," *J. Manuf. Syst.*, vol. 52, pp. 32–42, Jul. 2019.
- [72] Y. Ding, D. Ceglarek, and J. Shi, "Fault diagnosis of multistage manufacturing processes by using state space approach," *J. Manuf. Sci. Eng.*, vol. 124, no. 2, pp. 313–322, May 2002.
- [73] J. Yan, Y. Meng, L. Lu, and L. Li, "Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance," *IEEE Access*, vol. 5, pp. 23484–23491, 2017.
- [74] M. Massiris-Fernández, J. Á. Fernández, J. M. Bajo, and C. A. Delrieux, "Ergonomic risk assessment based on computer vision and machine learning," *Comput. Ind. Eng.*, vol. 149, Nov. 2020, Art. no. 106816.
- [75] S. Collins, F. Genova, N. Harrower, S. Hodson, S. Jones, L. Laaksonen, D. Mietchen, R. E. Petrauskait, and P. Wittenburg, "Turning fair into reality: Final report and action plan from the European commission expert group on fair data," Eur. Commission, Directorate Gen. Res. Innov., Brussels, Belgium, Tech. Rep., 2018.
- [76] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 2, pp. 539–550, Apr. 2009.
- [77] J. A. Sáez, B. Krawczyk, and M. Wo niak, "Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets," *Pattern Recognit.*, vol. 57, pp. 164–178, Sep. 2016.
- [78] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
- [79] A. Fernandez, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, Apr. 2018.
- [80] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinf.*, vol. 14, no. 1, Dec. 2013.
- [81] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [82] D. Cabrera, F. Sancho, J. Long, R.-V. Sánchez, S. Zhang, M. Cerrada, and C. Li, "Generative adversarial networks selection approach for extremely imbalanced fault diagnosis of reciprocating machinery," *IEEE Access*, vol. 7, pp. 70643–70653, 2019.
- [83] W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019.
- [84] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104837.
- [85] Z. Dai, Z. Yang, F. Yang, W. W. Cohen, and R. R. Salakhutdinov, "Good semi-supervised learning that requires a bad gan," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6510–6520.
- [86] A. Kumar, P. Sattigeri, and T. Fletcher, "Semi-supervised learning with GANs: Manifold invariance with improved inference," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5534–5544.
- [87] H. Di, X. Ke, Z. Peng, and Z. Dongdong, "Surface defect classification of steels with a new semi-supervised learning method," *Opt. Lasers Eng.*, vol. 117, pp. 40–48, Jun. 2019.
- [88] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

- [89] P. Wang and R. X. Gao, "Transfer learning for enhanced machine fault diagnosis in manufacturing," *CIRP Ann.*, vol. 69, no. 1, pp. 413–416, 2020.
- [90] J. Zhu, N. Chen, and C. Shen, "A new deep transfer learning method for bearing fault diagnosis under different working conditions," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8394–8402, Aug. 2020.
- [91] Y. W. Lee, D. M. Strong, B. K. Kahn, and R. Y. Wang, "AIMQ: A methodology for information quality assessment," *Inf. Manage.*, vol. 40, no. 2, pp. 133–146, Dec. 2002.
- [92] B. T. Hazen, C. A. Boone, J. D. Ezell, and L. A. Jones-Farmer, "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications," *Int. J. Prod. Econ.*, vol. 154, pp. 72–80, Aug. 2014.
- [93] Z. Song, Y. Sun, J. Wan, and P. Liang, "Data quality management for service-oriented manufacturing cyber-physical systems," *Comput. Electr. Eng.*, vol. 64, pp. 34–44, Nov. 2017.
- [94] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–52, Jul. 2009.
- [95] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu, "Efficient and privacy-enhanced federated learning for industrial artificial intelligence," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6532–6542, Oct. 2020.
- [96] G. Xu, H. Li, S. Liu, K. Yang, and X. Lin, "VerifyNet: Secure and verifiable federated learning," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 911–926, Jul. 2020.
- [97] T. Ryffel, A. Trask, M. Dahl, B. Wagner, J. Mancuso, D. Rueckert, and J. Passerat-Palmbach, "A generic framework for privacy preserving deep learning," 2018, *arXiv:1811.04017*. [Online]. Available: <http://arxiv.org/abs/1811.04017>
- [98] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," 2016, *arXiv:1606.05386*. [Online]. Available: <http://arxiv.org/abs/1606.05386>
- [99] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you? explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144.
- [100] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4765–4774.
- [101] C. Molnar. (2019). *Interpretable Machine Learning*. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>
- [102] P. Saleiro, B. Kuester, L. Hinkson, J. London, A. Stevens, A. Anisfeld, K. T. Rodolfa, and R. Ghani, "Aequitas: A bias and fairness audit toolkit," 2018, *arXiv:1811.05577*. [Online]. Available: <http://arxiv.org/abs/1811.05577>



RICARDO SILVA PERES (Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical and computer engineering with a focus on the application of AI in smart manufacturing from the NOVA University of Lisbon, Portugal, in 2015 and 2019, respectively. He is currently an Invited Professor with the Department of Electrical Engineering, School of Science and Technology, NOVA University of Lisbon. Since 2014, he has been a Researcher with the Centre of Technology and Systems, UNINOVA Instituto Desenvolvimento de Novas Tecnologias, focusing on the development of intelligent and predictive manufacturing systems. He has participated in several national and international research projects, including FP7 PRIME, H2020 PERFoRM, H2020 OpenMOS, H2020 GOOD MAN, and H2020 AVANGARD. He has authored several publications in high-ranked international scientific journals and conference proceedings (peer reviewed). His research interests include predictive manufacturing, AI, cyber-physical systems, and multi-agent systems. He has been a member of the IEEE IES Technical Committee on Industrial Agents since 2018 and the IEEE Standards Association P2805.1/2/3 Edge Computing Nodes Working Group since 2019.



research interests include prognostics and health management, data mining, and ML.



JAY LEE is currently an Ohio Eminent Scholar, the L. W. Scott Alter Chair Professor, and a Distinguished University Professor with the University of Cincinnati, Cincinnati, OH, USA. He is also the Founding Director with the National Science Foundation (NSF) Industry/University Cooperative Research Center on Intelligent Maintenance Systems, which is a multi-campus NSF Industry/University Cooperative Research Center consisting of the University of Cincinnati (lead institution), The University of Michigan, Missouri University of S&T, and The University of Texas at Austin. Since its inception in 2001, the Center has been supported by more than 100 global companies and was ranked with the highest economic impacts (270:1) by NSF Economics Impacts Report in 2012. He is currently on leave serving the Vice Chairman of the Foxconn Technology Group. He has authored a book *Industrial AI* (Springer, in 2020). He was selected by SME as 30 visionaries in smart manufacturing in 2016 and 20 most influential professors in smart manufacturing in 2020.



KEYI SUN received the Ph.D. degree in engineering from NCSU. She has more than ten years of experience in industrial big data. She is currently leading the Industrial AI Team, Foxconn Industrial Internet, to use a systematic way to develop algorithms for industrial applications. She is also an Associate Dean with the IAI Institute, which aims to cultivate the industrial big data and AI talents.



ARMANDO WALTER COLOMBO (Fellow, IEEE) received the B.Sc. degree in electronics engineering from the National Technological University of Mendoza, Mendoza, Argentina, in 1990, the M.Sc. degree in control system engineering from the National University of San Juan, San Juan, Argentina, in 1994, and the Ph.D. degree in engineering from the University of Erlangen-Nuremberg, Erlangen, Germany, in 1998. He joined the Department of Electrotechnical and Industrial Informatics, University of Applied Sciences Emden/Leer, Germany, where he became a Full Professor in 2010 and the Director of the Institute for Industrial Informatics, Automation and Robotics, in 2012. Over the past 16 years, he has been a Manager for Collaborative Projects and also an Edison Level 2 Group Senior Expert at Schneider Electric, Industrial Business Unit. From 1999 to 2000, he was an Adjunct Professor with the Group of Robotic Systems and CIM, Faculty of Technical Sciences, New University of Lisbon, Portugal. His current research interests include industrial cyber-physical systems, industrial digitalization and system-of-systems engineering, Internet-of-Services, Industry 4.0-compliant solutions.



JOSE BARATA (Member, IEEE) received the Ph.D. degree in robotics and integrated manufacturing from the Nova University of Lisbon in 2004. He is currently a Professor with the Department of Electrical Engineering, NOVA University of Lisbon, and a Senior Researcher with the UNINOVIA Instituto Desenvolvimento de Novas Tecnologias. He has participated in more than 15 international research projects involving different programs, including NMP, IST, ITEA, and ESPRIT. Since 2004, he has been leading the UNINOVIA participation in EU projects, namely, EUPASS, Self-Learning, IDEAS, PRIME,

RIVERWATCH, ROBO-PARTNER, and PROSECO. In the last years, he has participated actively researching SOA-based approaches for the implementation of intelligent manufacturing devices, such as within the Inlife project. He has authored or coauthored over 100 original papers in international journals and international conferences. His main research interests are in the area of intelligent manufacturing, with an emphasis on complex adaptive systems, involving intelligent manufacturing devices. He is a member of the IEEE technical committees on Industrial Agents (IES), Self-Organisation and Cybernetics for Informatics (SMC), and Education in Engineering and Industrial Technologies (IES). He is also a member of the IFAC technical committee 4.4 (cost-oriented automation).

...