CIRP Annals - Manufacturing Technology 73 (2024) 723–749

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: https://www.editorialmanager.com/CIRP/default.aspx

Artificial Intelligence in manufacturing: State of the art, perspectives, and future directions

Robert X. Gao (1)[a,*], Jörg Krüger (1)[b], Marion Merklein (1)[c], Hans-Christian Möhring (2)[d], József Váncza (1)[e]

[a] Case Western Reserve University, Cleveland, OH, USA
[b] Technical University of Berlin, Germany
[c] Friedrich-Alexander University of Erlangen-Nürnberg, Germany
[d] University of Stuttgart, Germany
[e] Institute for Computer Science and Control, Hungarian Research Network, Hungary

ARTICLE INFO

Article history:
Available online 22 July 2024

Keywords:
Artificial intelligence
Smart manufacturing
Machine learning

ABSTRACT

Inspired by the natural intelligence of humans and bio-evolution, Artificial Intelligence (AI) has seen accelerated growth since the beginning of the 21st century. Successful AI applications have been broadly reported, with Industry 4.0 providing a thematic platform for AI-related research and development in manufacturing. This paper highlights applications of AI in manufacturing, ranging from production system design and planning to process modeling, optimization, quality assurance, maintenance, automated assembly and disassembly. In addition, the paper presents an overview of representative manufacturing problems and matching AI solutions, and a perspective of future research to leverage AI towards the realization of smart manufacturing.

1. Introduction

Artificial intelligence (AI) is often referred to as "the science and engineering of making computers behave in ways that, until recently, we thought required human intelligence" [89]. The research field of AI evolves not only as the community of researchers builds on top of one another's work, but also as inspirations are taken from natural intelligence.

1.1. From natural intelligence to artificial intelligence

For 99% of the known history of the human species, humans were wanderers and hunters [219]. With trial and error, humans developed more sophisticated strategies that increased the yield of hunting while enhancing self-protection. Observing and learning from the behaviors of animals, humans gained insights that complemented their survival skills and problem-solving strategies. Through collaboration and information-sharing, humans accomplished what each alone could not. And using the tools that they developed, humans started to domesticate animals, developed agriculture, created homes and cities, and abandoned the nomadic lifestyle. Although chasing food and self-protection from predators no longer constituted the highest priority in life, curiosity for understanding the environment and propensity for self-betterment persisted, thereby continually sharpening the evolution of humans intelligence [218].

As the need for understanding the law of nature is essential to the survival of the species, humans started to recognize patterns embedded in the world and created logical and mathematical tools to synthesize and generate knowledge, describe its functioning, learn to break down complex problems into smaller ones to optimize solution strategies, and see the similarities among different problems for which existing knowledge can be adapted to [191]. Throughout this process, learning from animal intelligence played a pivotal role in the advancement of natural intelligence as a key element to human survival.

From the era of manual production to a series of industrial revolutions over the past centuries, human intelligence has not only guided the perfection of craftsmanship in the production of goods but also led to the mechanization that freed humans from hard labor, as characterized by the creation of steam engines, automobiles, electric motors, industrial robots, and computers [129]. As a precursor to the era of AI, the idea of mechanization of human natural intelligence was also a topic of exploration during the late 19th century [274]. In the 1950s, Turing [251] devised Turing test to argue the plausibility of "thinking machine". Shortly after, the term "artificial intelligence" was coined by McCarthy [184] at the Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI) in 1956, and Minsky [175] presented the envisioned steps towards AI: search, pattern recognition, learning, planning, and induction. These resemble the main characteristics of natural intelligence and catalyzed the AI research until today. This keynote paper aims to present advances in AI in manufacturing since the beginning of the 2000s, with related AI techniques being the result of cumulation of research and development during the past decades.

* Corresponding author.
E-mail address: Robert.Gao@case.edu (R.X. Gao).

https://doi.org/10.1016/j.cirp.2024.04.101

1.2. AI technologies

The development of AI has been characterized by two distinct trajectories, leading to what is known as model-based vs. data-driven methods, as illustrated in Fig. 1. Such distinction can be traced back to the 1960s, as symbolic vs. connectionist AI.

Symbolic AI consists of implementing interpretable, high-level rules and calculating symbols that have concrete semantic meanings [232]. The design of programming language Lisp in 1958 to manipulate symbols was widely regarded as AI's first contribution [171]. By contrast, connectionist AI envisions large-scale calculation of low-level functions distributed across a neural network, with meaningful behavior appearing as the collective effect of all elementary operations. Such bio-inspired concept first emerged from neurophysiologists and logicians [172] before evolving into Rosenblatt's perceptron [212]. Intuitively, the symbolic AI is a deductive machine, which relies on human-designed model (rule, knowledge) to compute an output from a given input. Connectionist AI is an inductive machine which learns the model from observed samples of desired input and output. However, the 1st generation AI systems did not take off due to their limited capability and practicality [176]. In the early 1970s, research was put on hold in what is known as the 1st AI winter [29].

A significant upgrade to the symbolic AI under the name of expert systems led the 1st AI revival around 1980 [29]. It was made possible by more powerful computers that allow bigger set of rules, which consist of structured lists of "IF … THEN", to be stored in the memory [242], leading to successful diagnosis of blood diseases [229], identification of locomotive breakdowns [22], and detection of geological deposits [73]. In addition, the systems can break down the reasoning process into blocks of "agents" which could independently utilize rules and infer consequences, inspired by the concept of modularity [4]. However, researchers soon realized that creating repositories for realistic and diverse rules to convey the subtleties in reasoning became increasingly inefficient. As research effort in expert systems slowed down and supporting hardware (e.g., Lisp machines) diminished in the mid 1980s, the 2nd AI winter began [29].

During the same period, several algorithmic and theoretical advances foreshadowed the revival of connectionist AI. Backpropagation was developed in 1986 to allow the weights of any type of neural network to be optimized [214], propelling a creative period that saw the invention of some of the most widely used networks today: convolutional (CNN) and recurrent neural network (RNN) [92,142]. Also, universal approximation theorem was proven, indicating that neural networks are universal function approximators that theoretically can fit any functions [96].

Interspersed in the evolution of symbolic vs. connectionist AI is the notable development of a series of techniques on both model-based and data-driven fronts. For example, the evolutionary algorithms developed in the 1960s use rules inspired by biological evolution, such as genetic reproduction, mutation, recombination, and selection, to search for optimal solutions [95]. In the 1990s, kernel methods such as support vector machines (SVM) were developed [39]. The convex nature of these methods provided a means to bypass the local minima constraint and their effectiveness in leveraging small datasets made them more attractive choices than the neural networks [93]. In 1997, Mitchell [177] provided a formal definition of machine learning (ML) to describe algorithms such as neural networks that can learn from data and generalize to unseen data, without explicit instructions. (A related concept is soft computing, generally describing techniques such as evolutionary algorithms and neural networks that exploit tolerance for imprecision, uncertainty, and partial truth to achieve tractability in problem-solving [31].)

Given the specifics of ML tasks, the learning process can be generally categorized as supervised, unsupervised, and reinforced based on the interaction [19]: supervised learning trains a ML model on a labeled dataset such that the model learns to predict outcomes based on the input data. The goal is to minimize the difference between the predicted and actual outcomes. Unsupervised learning, in contrast, trains a model on data without labeled responses, and the goal is to identify patterns, clusters, or relationships within the data. Reinforcement learning (RL) is a paradigm where an agent learns to make decisions by performing actions in an environment to maximize certain cumulative reward, based on the feedback from its actions and experiences rather than from direct labeled instructions.

In mainstream AI research, the concept of task-oriented, rational agent made a breakthrough around the turn of the century. By definition, a rational agent is embedded in an environment to perform tasks. After making observations, it changes the environment with its actions in a way that optimizes its utility criterion with bounded computational resources [217]. Shortly, the rational agent concept defined a dominant and most successful approach to AI [216,217] as it was pragmatically task-oriented and inclusive regarding all available AI technologies. The notion of rational agent re-initiated research in ML which was essentially aimed at improving agent performance based on accumulated experience [110]. By populating the environment with other agents, it could intuitively be extended to multi-agent systems and distributed intelligence. Finally, sensing and acting are the essence of robotics, and the agent concept brought robotics back into the realm of AI, with natural extension to the applications such as the human-robot collaboration (HRC).

The beginning of the past decade has witnessed significant shifts in the AI landscape. Digital transformation across the globe had necessitated a new generation of AI methods to tackle new challenges such as spam/fraud detection, prediction of supply chain volatility, impacts of customer response and behavior on decision-making, etc. These new AI methods require the use of "big data", which bears little resemblance to the relatively small and calibrated data known to the AI community [42].

The challenge to learning from these massive data, after a brief period of stagnation, has ultimately been met by the advancement of deep neural networks and the growth in computational infrastructure such as graphic processing unit (GPU) and cloud computing [130]. Increasingly, researchers have realized the capability of deep neural networks in exploiting the hierarchical structure embedded in

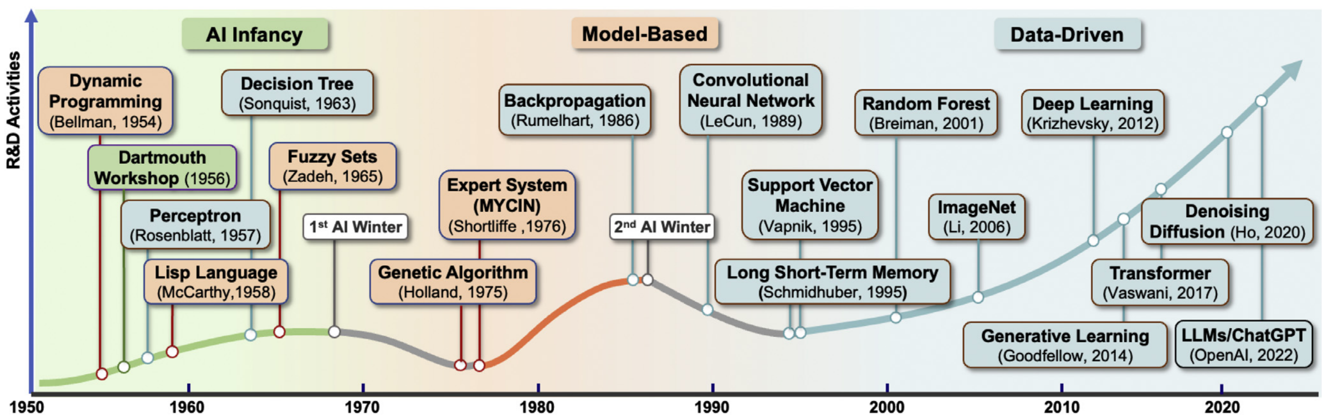

Fig. 1. Evolution of AI technologies, [16,26,185,195,229,260,283]; LLMs: large language models.

data, making them not only capable of ingesting raw data without manual feature engineering (a typical step in other ML techniques such as SVM), but also more effective in processing features as compared to their "shallow" counterparts [141]. Additionally, empirical evidence suggested that the lack of theoretical guarantee in neural networks due to the non-convexity is far outweighed by their superior performance in handling large, real-world datasets, including image processing and speech recognition, when compared to the traditional ML [58,90]. During this period, the term "deep learning" (DL), which specifically describes ML using deep neural networks without manual feature engineering, begins to gain popularity [141].

During the past decade, accelerated development of DL has brightened the prospect of fulfilling the longer-awaited promise of neural networks being universal function approximators, enabled by a series of innovations in terms of the network structures. For example, rectified linear unit (ReLU) activation function was rediscovered to replace the traditional smoother functions such as hyperbolic tangent, enabling much faster learning of deep neural networks [76]. Around the same time, a new regularization technique termed dropout has emerged, which randomly drops neurons from the network during training [234]. Intuitively, dropout can be considered as training an ensemble of networks, each with a different subset of neurons. Dropout has seen wide success in preventing model overfitting and quantifying model uncertainty [69,234]. Another major innovation is the attention mechanism [11]. Rather than having fixed neural network weights after training, the goal of attention mechanism is to allow weights generation based on the specific input and thereby capturing the dynamic relationship between the input and output. Originally conceived as solutions to one narrow and specific problem-language translation, attention mechanism has since become the cornerstone of general-purpose neural networks, propelled by the transformer architecture [254] and its variants [205,206].

The development of DL during the past decade has also facilitated the conceptualization and/or maturation of next class of learning-based AI techniques such as generative learning and deep reinforcement learning (DRL). The main idea of the former is to use deep neural networks to approximate the transformation between a known distribution where samples can be taken (e.g., standard Gaussian) and a (unknown) distribution of desired output, which subsequently enables the synthesis of output data by sampling from the known distribution [78,91]. Similarly, DRL relies on deep neural networks to directly approximate the value functions (which are associated with the cumulative reward) for decision-making. Compared to the traditional RL, DRL allows for decision-making from complex input such as images. These advances have quickly led to new results in a number of applications such as design and robotics [255,284].

1.3. AI for smart manufacturing

To advance the state of manufacturing, continued effort has been made by engineers and researchers to:

- efficiently search through variable space of manufacturing systems to optimize scheduling and planning [236];
- accurately model process dynamics and optimize process parameters in production to improve part property [65];
- timely detect defects and forecast future performance of equipment for quality assurance and maintenance [271];
- delegate repeated, laborious tasks to robots and explore seamless interaction between human and machine [261].

The increasing difficulty of meeting these objectives due to the multitude of productivity, efficiency, and sustainability requirement stemming from the growing product and process complexity, as well as the variability in customer preferences, provide opportunities for manufacturers to investigate AI capabilities. Due to the rising relevance of AI in manufacturing, several contributions have been presented as keynotes in CIRP Annals by the end of the past century. Markus and Hatvany [170] outlined the structure of subfields in manufacturing, such as design, planning, monitoring and control, and matched AI tools to the corresponding tasks. Rowe *et al.* [213] detailed AI implementation in grinding, with domain-specific applications such as wheel selection and parameter selection. Teti and Kumara [244] defined a functional view of manufacturing system consisting of design, planning, production, and system-level activities and mapped AI technologies to each functional element.

After a brief hiatus, the digital transformation of manufacturing has accelerated, enabled by massive deployment of sensors and Industrial Internet of Things and the resulting large amounts of data produced by machines, controllers, and system records, etc. The need to properly decipher the information embedded in the data brought AI research back to the spotlight in manufacturing, which offers complementary understanding of the physical characteristics of a system or process [228]. Unsurprisingly, the comeback of AI has been led by the emergent ML, and especially DL technologies. A review of ML techniques and their applications in manufacturing is found in [228]. Several other review articles are focused on more specific aspects in manufacturing, providing analysis at a more granular level. For example, Wang *et al.* [261] provides a comprehensive review of AI for HRC, while Gao *et al.* [71], Krüger *et al.* [134], and Möhring *et al.* [179] addressed the implementation of AI from the perspective of data life cycle, control, and self-optimization in their respective keynotes. Recent development of model-based AI, such as agent-based system and ontology, was also summarized in [183] and [282].

Diverging from these recent review articles that are primarily focused on specific applications and AI technologies, this keynote paper (Fig. 2) aims to present the state-of-the-art of advances in AI in manufacturing with an integrated view of system (Ch. 2), process (Ch. 3), quality (Ch. 4), and (dis)assembly (Ch. 5) from both model-based and data-driven perspectives, with the integrated view exemplified in case studies (Ch. 6).

The complementary nature between domain knowledge and data will be emphasized. Highlights will also be given to the research avenues that have been completely transformed due to the latest AI technologies, such as generative learning and DRL, which have led to the development that was unimaginable before. This keynote also picks up the baton from Markus and Hatvany [170] and Teti and Kumara [244] and provides an updated mapping between AI technologies and manufacturing tasks to bridge the gap of AI advances in manufacturing over the past 20 years. Correspondingly, newly emerged challenges and directions in AI, such as model physical consistency, real-life AI, generative AI, and generalist AI models will also be discussed (Ch. 7).

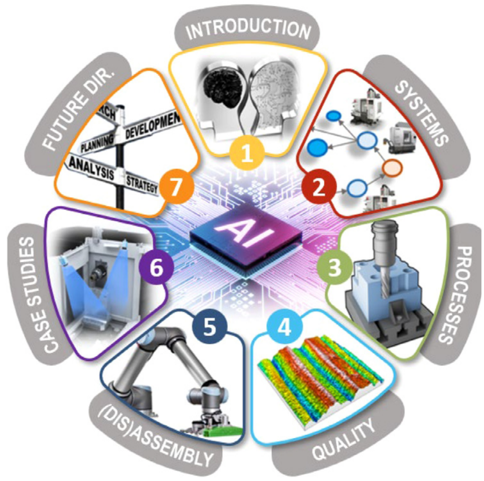

Fig. 2. Chapter structure of the keynote.

2. AI-assisted production system design and planning

Production systems, consisting of intricate *networks* of machinery and human resources, technological and logistical processes, material, information and financial flows, are among the most complex man-made systems. These systems have inspired AI research on automated problem solving, reasoning, and learning [36]. In the context of production, whether it be the synthesis of complex *objects* (i.e., production systems) or the dense network of their *actions* (i.e., production plans and schedules), the structural and functional relationships among system components must be maintained on top of which decisions will be made [54]. The AI research in design and planning of the production systems primarily revolves around modelling the systems while considering these relationships or constraints, and optimizing design, scheduling, and operations in general (Fig. 3).

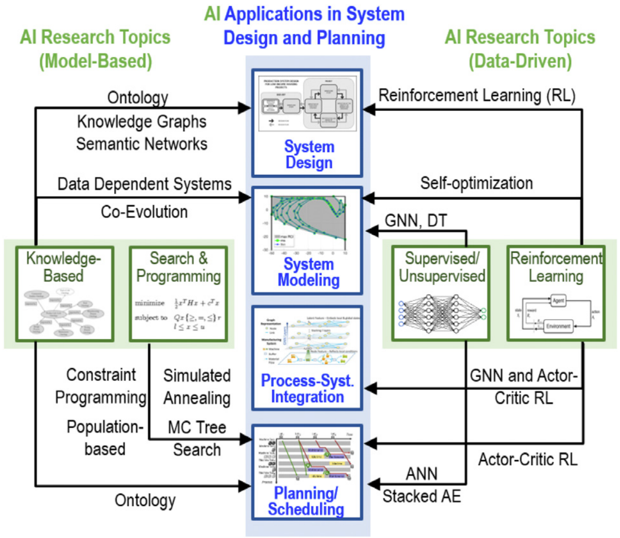

Fig. 3. Overview of AI for design of and planning in production systems.

2.1. Design of production systems

The design of production systems is primarily a configurational problem typically requiring combination- and modification-based design methods [249], commonly starting with domain knowledge before being abstracted into mathematical forms for computation.

2.1.1. From ontologies to graph neural networks

One of the primary ways that logic-based *knowledge* and *reasoning*, as well as their acquisition and management in AI research is through *ontologies* [208,243]. An ontology is a formal, well-structured vocabulary that captures a consensus set of terms to represent the entities in a domain along with their relationships. It also provides axioms that guide the interpretation and reasoning about terms, enabling verification of data validity and consistency, and inference of new knowledge. Ontologies, and more recently, *knowledge graphs* [46] have been used to capture the conceptual structure of manufacturing organizations, encompassing related products, technologies, resources, and business aspects [46]. Ontologies and knowledge graphs facilitate design standardization, semantic interoperability, principled engineering of complex software systems, and knowledge reusability.

Nowadays, the de-facto "assembly language" of ontology building is OWL (Web Ontology Language). Among others, the National Institute of Standards and Technology (NIST) uses OWL for product lifecycle modelling and promoting an Industrial Ontology Foundry (IOF) [6]. IEEE also uses this form of logic-based knowledge representation to define the field of automation and robotics [101]. To meet requirements of agility and changeability of manufacturing, a decomposed ontology structure is typically used, where one part is general and stable, representing the *core ontology*, while the other parts may be domain, application and/or company specific. Such ontologies have been developed for additive manufacturing (AM) (see Fig. 4) [162,221], steelmaking [28], and robotic assembly [111,128].

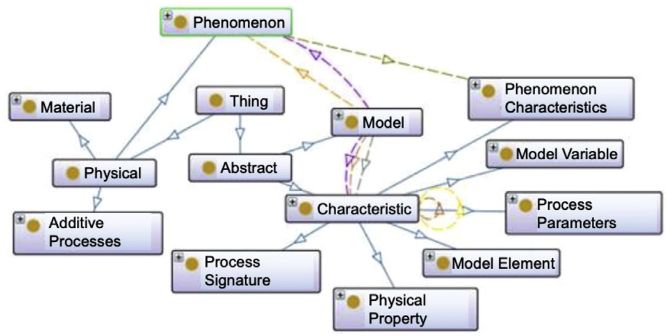

Fig. 4. OWL for additive manufacturing, adapted from [162].

However, relying solely on symbolic reasoning methods built upon the ontologies [208] and knowledge graphs [94], either individually or in combination, falls short of fully capturing the essence of production system design. The need of transforming the system representation into a computable form to modeling the increasing system complexity and realizing functions such as design synthesis and system planning and scheduling have led contemporary research to rely on recent advances in ML and data analytics [108,197].

Focusing on a single manufacturing step, Wang *et al.* [268] investigated spindle power and tool wear conditions during hard milling using the stochastic modeling and analysis technique of Data Dependent Systems (DDS). The spindle power data was decomposed into different frequency regions using DDS, and the correlation between spindle power and tool wear in the frequency domain was quantified. In the context of *self-optimizing machining systems,* Möhring *et al.* [179] presented a broad array of ML methods that were used for linking sensory characteristic features to workpiece, tool, machine, and process states. These AI-based models can be used in decision-making processes when predicting part quality, tool and machine conditions are essential elements of a system design problem.

For systems consisting of multi-step manufacturing processes, graph neural networks (GNNs) [276] have been increasingly attracting research attention. GNNs are adept at capturing complex relationships and interactions between different components in a manufacturing system. By representing machines (processes) and intermediate products as nodes, and interactions or dependencies between them as edges, GNNs capture the entire manufacturing system through its graph structure while maintaining the flexibility of neural networks in learning from data.

Fig. 5 illustrates a typical multi-layer GNN representation of a manufacturing system by Huang *et al.* [98]. The system consists of multiple types of machines corresponding to various manufacturing

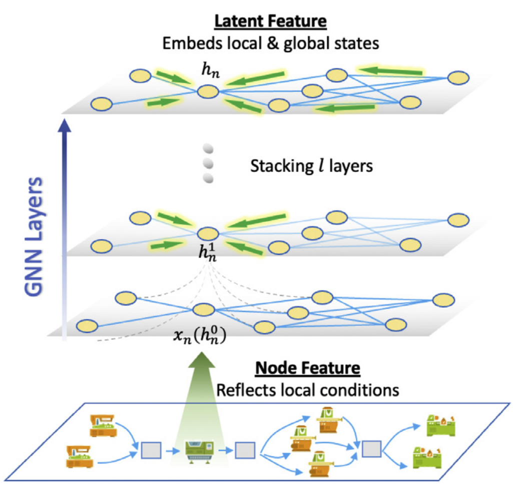

Fig. 5. GNN representation of a manufacturing system [98].

processes. In GNN, each node corresponds to a machine while the inter-node connections are determined through the domain knowledge of whether any pair of manufacturing steps must be done adjacently. At the input level, each GNN node contains a feature representing its local conditions (e.g., machine conditions). Then, for each node in the first GNN layer, its feature is determined not only from local conditions of the corresponding node, but also the conditions from all its 1st order neighbors (i.e., nodes that are one edge away). By progressively adding GNN layers, the feature corresponding to each node accumulates information regarding local conditions as well as correlations from the nodes that are further away. These high-level features have shown to be useful in system design synthesis (Section 2.1.2) and optimization of system planning and scheduling (Section 2.2).

2.1.2. Design synthesis

In the design synthesis of production systems, humans still retain their central role, while advanced simulation, process mining, data analytics, and visualization methods offer broad support for the analysis of solutions. Providing early feedback and learning from the evaluation and analysis of *partial* designs is a key to success. One example that builds upon a GNN-modeled manufacturing system is demonstrated by Klar *et al.* [123].

In this work, the main objective is to synthesize factory layout such that material transportation load is minimized. The authors formulated this using an RL-based approach where each machine of the system is sequentially added to the shop floor. At each RL step, the decision about placement is determined through (1) material flow among different machines, which is modeled using a GNN., (2) shop floor layout at the current step (i.e., partial design), which is represented by an image and is processed by a convolutional neural network (CNN), and (3) features of upcoming machine to be placed, which are extracted using a multilayer perceptron (MLP). The outputs from these three neural networks (i.e., states in RL) are then fused to determine the “goodness” or action values corresponding to different candidate x- and y-locations as well as the rotations (i.e., actions in RL) of placement. These action (or Q) values can be considered a proxy for the material transportation load and the optimal placement comes with the combination of x, y, and rotation with the highest Q values. At the initial stage of RL, the neural networks may predict the Q values randomly. However, by sequentially assigning machines and recording feedback on a trial-and-error basis, RL can improve its accuracy in Q value prediction by network weight update. This specific RL variant is also known as deep Q learning (DQN) [239]. The authors demonstrated the effectiveness of the method using a case study of a shop floor with 43 machines.

The above example provides a template of how design synthesis can be achieved through the integration of data learning and simulation and can be expanded to problems with higher complexity. Production system design is a teamwork which can be facilitated by agent-based or distributed AI solutions. As discussed in [295], the multi-agent approach to production system design is gaining momentum because it facilitates the collaboration of various engineering branches related to the problem. Furthermore, it fits well with the workflow of concurrent [161] and life cycle [86] engineering. One example is from Zhang and Lin [290] that investigated multi-agent RL (MARL) for shop floor layout design optimization. The essence of MARL is to assign an agent for each machine such that RL can be done in a distributed way [239]. In contrast to [123], their objective is to minimize the connection cost among different machines. The method is validated using a case study of a pure water manufacturing system.

Design of complex systems also necessitates better simulation methods. For this purpose, the concept of *Digital Twins* (DTs), which captures both the models and realizations of engineered systems, together with a bidirectional, continuous interaction between the virtual and the physical counterparts [190], is envisioned to play a central role through (1) consolidating design decisions from different sources, (2) supporting high-fidelity simulation and evaluation, and 3) facilitating life cycle engineering [17,222]. During the engineering phase, a DT can be applied to test and validate system design alternatives, while at run time production data can be gathered to update and improve a DT (e.g., fine-tune DT parameters to close the sim-to-real gap) to complete the bidirectional interaction (Fig. 6). Jaensch *et al.* [103] present a combined approach of these model-based and data-driven DTs that allows continuous self-improvement.

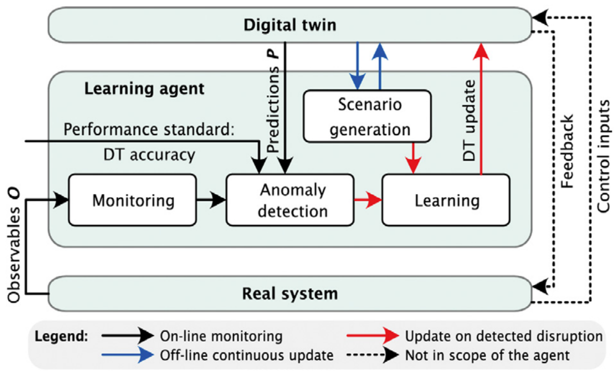

Fig. 6. Communication between digital twin and physical system [257].

The combination of simulation and learning-based approaches often can generate new and creative results. This can be extended to a co-evolution where a design problem's specification also evolves over time [248,249, 270]. Recognizing and formulating new knowledge brought up by creative design can be challenging, highlighting the importance of human involvement [53,59]. Early efforts have been focusing on the concept of explainable AI (XAI) [100]. For example, Klar *et al.*, implemented saliency maps [230] to interpret the shop floor regions of interest used by RL in determining machine placement. While certain patterns emerge in the beginning, the authors admit that the interpretation quickly becomes difficult as the problem scales up [123]. Addressing the challenges of XAI in manufacturing requires concerted efforts across various manufacturing domains (see Section 7.10).

2.2. Production planning and scheduling

Production planning and scheduling (PPS) involves strategical, tactical and operational decisions in setting and achieving production targets. It is expedient to see the problem as orders competing for finite production resources in a dynamic environment. Solutions consist of temporally interlinked actions assigned across diverse resources, adhering to multiple constraints while optimizing criteria related to cost, time, quality, resource, and energy utilization, etc. Due to their inherent combinatorial complexity, PPS problems of practical relevance are difficult to solve [64]. In a production network, this situation is further aggravated by information asymmetry [139].

2.2.1. From mathematical programming to search

For PPS, operations research (OR) has traditionally utilized mathematical programming for strategic and tactical production planning [202], while AI has rather been applied to operational production and logistics scheduling [297]. The research of AI has contributed to scheduling by enhancing constraint-based modeling and constraint programming (CP), addressing representational adequacy to ensure compliance with production constraints [138,187]. These models, which differentiate between essential hard constraints and soft, preference-based constraints, offer a clear, incrementally developable approach, critical in industrial AI applications (see also Sect. 7.10) [51]. Advanced CP enhances this with its powerful constraint propagation and search-based solution techniques, increasing flexibility and adaptability over more rigid algorithms [138]. CP's capability for incremental adjustments and its use of approximate methods for heuristic guidance further underline its utility in handling complex PPS scenarios [41,289].

Additionally, PPS considers various performance criteria, including newer aspects like robustness and environmental efficiency [140], with CP allowing intuitive representation and exploitation of these factors for improved solution efficiency [138]. Moreover, integrating production planning with process planning and system configuration, despite increasing solution complexity, offers a comprehensive approach [250]. This integration captures a broad spectrum of scheduling knowledge, as evidenced by early and subsequent scheduling ontologies [207,231].

Closely related to CP are the variety of pure search methods. Most of them use *heuristics* to estimate the value of partial solutions or their distance from the goal, striking a balance between completeness and computational efficiency. By now, *meta-heuristics* and in particular *local search* techniques such as simulated annealing, tabu search, and various population-based search approaches are routinely applied for solving production planning and scheduling problems [1]. For instance, Nonaka *et al.* [193] proposed a method for optimizing the efficiency of a job shop by exploiting the potential of alternative routings made available by flexible CNC machines (Fig. 7). To solve this complex scheduling problem, the authors combine mathematical programming with tabu search. Stricker *et al.* [237] solved the task of scheduling in matrix production of different product variants with various cycle times. This work introduces a method for identifying and autonomously adjusting high-performing solutions to the scheduling problem employing multi-objective Monte Carlo tree search (MCTS). Meta-heuristics have recently been developed further to *hyper-heuristics* which interchange different solvers while at work [220]. In parallel, *anytime algorithms* became indispensable for delivering solutions when response time was of the essence, as in real-time production scheduling [241]. Recent search techniques like *large neighborhood search* (LNS) [138] have proven to be successful in solving highly complex problem formulations, which provide intuitive models for high-level production system configuration [250] and planning [202].

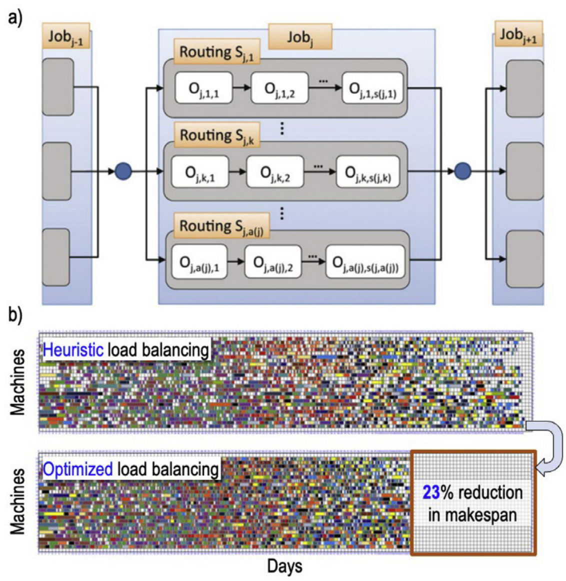

Fig. 7. Schedule optimization: a) alternative routings; b) schedules generated with heuristic and optimized load balancing [193].

2.2.2. Role of learning-based methods

Compared to design synthesis, a key differentiating factor of PPS is its dynamic nature. In contrast to design synthesis that can often be done offline, it is critical that PPS is done online in response to the dynamic changes likely to occur. This temporal scale makes it challenging for decision-making in PPS as the predictive model of production may not exist and simulation (that search methods need) becomes computationally expensive and likely to deviate from reality. Solving these challenges constitutes the first core application of ML in PPS.

One important aspect of decision-making in PPS is to forecast the expected future progression of production. Compared to the traditional time-series analysis, ML, in particular DL has demonstrated superior performance in finding highly nonlinear association between the variables indicating current production status to its future progression, allowing accurate estimation of metrics such as future demand [71] and lead times [81]. As an example, Fang *et al.* [57] developed a stacked sparse autoencoder (S-SAE) to predict the remaining time for ongoing manufacturing jobs during production. This work first utilizes the capability of S-SAE to distill essential features embedded in data related to production task (e.g., task composition), production status (e.g., waiting sequence), and machine status (e.g., machine utilization rate). This is done by using the contraction-expansion structure of S-SAE 's encoder-decoder pair. Then, for prediction, only the encoder is retained, and it is attached to an MLP for job remaining time estimation. The authors noted significantly reduced prediction mean absolute percentage error by S-SAE (5.6%) compared to linear regression (19.2%), deep belief network, or DBN (16.9%), or using MLP only (13.7%). Similarly, ML can be used to create models directly from event log data. In production planning, Kádár *et al.* suggested a process mining method for re-constructing the model of a semiconductor factory from partial, noisy and at times contradictory data [113]. Thanks to the tight coupling of physical processes and their digital representations, the model could be adapted to changing conditions automatically.

To a certain extent, these works represent how ML is incorporated into DTs in production management [226,257] through surrogate modeling. In addition, the role of ML can also encompass replacing simulation directly. For PPS, the efficiency of optimization methods depends fundamentally on the fast evaluation of the solution candidates. In the case of complex, real-world problems, this can rarely be achieved by the application of an easily evaluated objective function. Instead, computationally expensive simulations must be performed. ML techniques can act as surrogate models to replace simulations for fast evaluation of solution candidates generated by meta-heuristics [108,182] (see Fig. 8). Running simulations on well-chosen scenarios derived from historical data using unsupervised learning is another approach to optimizing efficiency [77]. This work combines model- and data-driven analysis to support scheduling in a high-mix low-volume production environment.

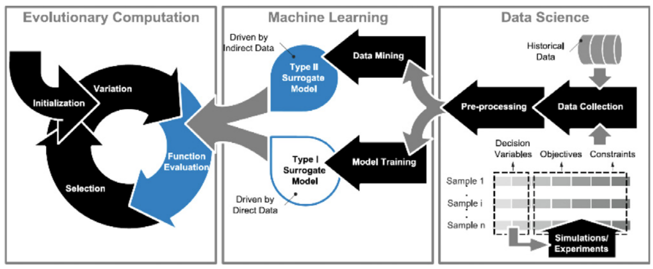

Fig. 8. Data-driven evolutionary optimization [108].

In addition to surrogate modeling, ML has also shown capable of reducing the sim-to-real gap (or deviation) between the digital and the physical representation of the system to improve the reliability of decision-making. One example is provided by Vrabic *et al.* [257]. The idea is to first determine the source of the deviation and formulate a response strategy through generating and simulating what-if scenarios for various disruptions using the DT. A neural network then encodes the discovered association and predicts the parameters of the DT based on past and present observables. Specifically, the network input contains observables at subsequent times of observation, while the output includes DT parameters [257]. The network can then determine how the DT should be updated so that its behavior matches the physical systems. The authors validated the effectiveness of the method in improving system resilience (Fig. 9) after disruption

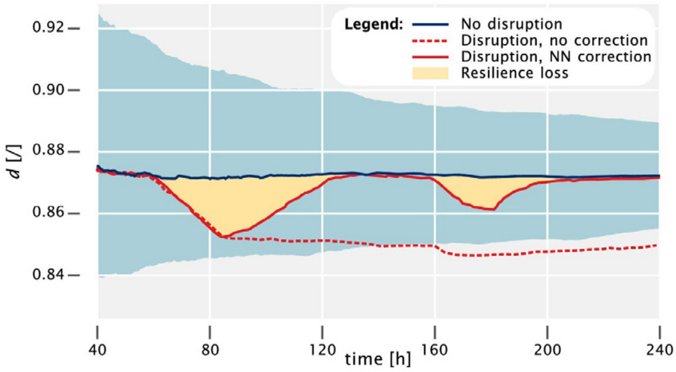

Fig. 9. Reduction of resilience loss after disruption through NN-based DT parameter update [257] (vertical axis represents accuracy of DT).

using a case study of cell and gene therapy (CGT) secondary manufacturing facilities.

Once a reliable evaluation method is available, RL has also been widely investigated for PPS [40,117,151,197] by learning a control policy for sequential decision-making from interactions with an uncertain, dynamic environment that provides feedback in the form of rewards [135,239]. For example, Epureanu *et al.* investigated [55] an RL-based method to determine the optimal strategy for handling machine breakdown. Specifically, a deep convolutional Q learning neural network as shown in Fig. 10 has been developed. The input includes encoded production information such as suspect modules and swappable stages. This information is processed by convolutional layers for prediction of Q values corresponding to three repair strategies. Simulation results using a production system consisting of seven stages show that the developed RL significantly reduces the production capacity loss as compared to a random strategy selection.

Huang *et al.* [98] built upon a GNN-model production system and investigated MARL for maximizing throughput while minimizing defects, taking into consideration the variability of each machine. The MARL is based on the method of advantage actor-critic or A2C [238]. The A2C synergizes an actor, responsible for choosing actions (i.e., setting parameters of each machine) based on the current policy, and a critic, which evaluates the chosen actions by estimating the value function. The actor's policy, represented by GNN, dictates the probabilities of taking specific actions in given states. In parallel, the critic assesses the expected return from these states, aiding in the computation of the advantage function. This function reflects the relative benefit of each action compared to the average, guiding the actor toward more rewarding choices. Learning occurs as the actor adjusts its policy to maximize rewards, using the advantage function as a directional signal, while the critic refines its value predictions to align with actual returns. Compared to two rule-based baselines using a simulated case study, the MARL is able to incur three times fewer defects while maintaining similar throughput and could double the throughput while maintaining similar defect-to-throughput ratio. Other reported work includes weighted Q-learning [269], double DQN [83], distributed policy search [67], and policy-iteration type actor-critic models [157]. Recently, policy gradient RL also proved to be successful in efficiently managing resources of biological material production in a highly uncertain environment [181].

As exemplified by MARL, in production systems, the operation is a result of the collaboration among numerous autonomous decision-makers, each driven by its own objectives within a network of inter-dependent entities [139], naturally fitting multi-agent systems that offer a robust model for representing these autonomous units and their interactions, significantly enhancing the analysis of system-level behavior [183]. Despite the efficiency of multi-agent systems in visualizing and analyzing such interactions, ensuring consistent, efficient, and goal-oriented operations in line with production expectations necessitates some type of centralized control mechanisms, a challenge yet to be fully addressed in industrial applications [194,253].

3. AI in process modeling, management and optimization

As the core step to transform raw materials into finished parts, efficient and anomaly-free manufacturing processes have been the goal for manufacturing researchers and engineers since the very beginning. As manufacturing processes are getting increasingly complex to be designed, modeled, and optimized using domain knowledge alone [150], AI techniques are increasingly investigated to compensate for this limitation in modeling and parameter optimization, as shown in Fig. 11.

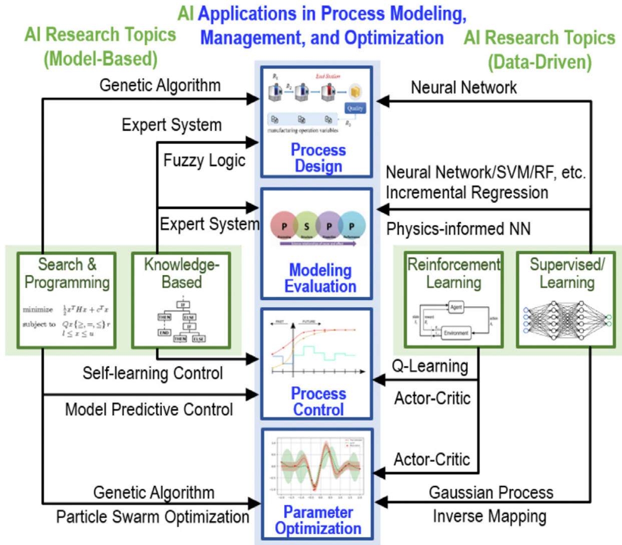

Fig. 11. AI research for process design, modeling, and optimization.

3.1. Process design

A detailed summary of AI for process design has been presented in [147], which extensively reviewed AI applications in process design and planning with a focus on AI techniques of expert systems and evolutionary algorithms. This section will first highlight research works that mainly benefit from these model-based techniques,

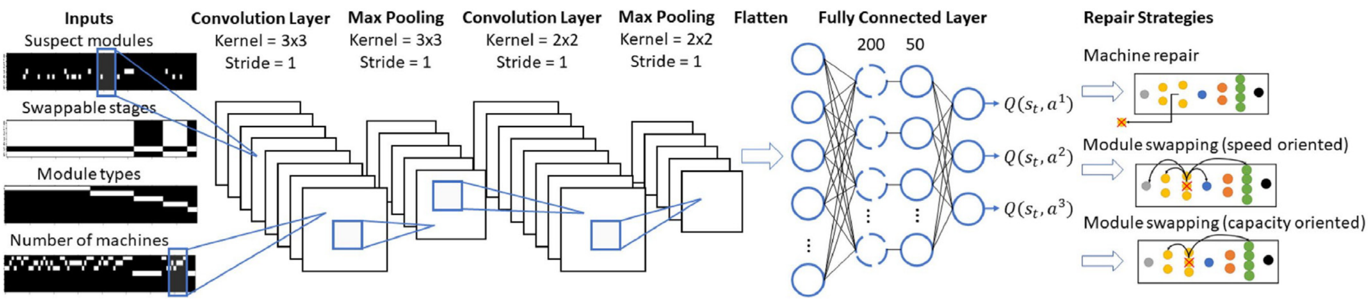

Fig. 10. Deep Q-learning for optimal strategy selection in handling machine breakdown [55].

before introducing more recent works that increasingly leverage the capability of data-driven AI.

Teti *et al*. [244] investigated in Intelligent Computing Methods for Manufacturing Systems. Future motivations for intelligent computation in manufacturing include enhancing decision-making, features, and classifications, improving performance, flexibility, and efficiency, and integrating real-time operation and automation. Achievements require enhanced knowledge bases, improved AI tools, cost reduction, and new applications. Future systems will integrate AI into various engineering processes, develop hybrid systems, and implement intelligent manufacturing systems, characterized by integration, modularity, and hybridization. The research of Molina *et al*. in [180] addresses AI's potential in machining processes, including quality and efficiency improvements, contrasts with industry readiness due to machinery and worker training deficiencies. Adoption challenges persist despite Industry 4.0′s promises.

In [5], a hybrid approach is presented by using techniques of neural networks, fuzzy logic, and rule-based systems. The research illustrates a feature-based intelligent computer-aided process planning system (CAPP) that includes (1) a standardized feature-based model in the form of STEP-based features and (2) the hybrid AI model for process planning. In addition, a digital process plan can be created, which provides the required information on the components to be manufactured. The results of the analysis show that the integration of model-based and data-driven AI techniques can make process planning more efficient. Similar work has been reported in [47] using genetic algorithm (GA) and neural network.

Beyond process planning, AI has also found applications in fixture design. For example, ML-based optimization of a clamping concept is investigated in [60] (Fig. 12). The objective of the investigation is to establish a rapid model for positioning fixture locations within an 8-second timeframe. The clamping of components is an important element in manufacturing processes which have a large influence on the dimensional accuracy. In the context of the study, an optimal clamping is determined from a multiplicity of configuration possibilities for the reduction of manufacturing errors. A milling process is chosen as the manufacturing process and the target values of the clamping optimization are the maximum workpiece deflection and the lowest natural frequency. Initially, exemplary configuration of the clamping is introduced and the generation of the input and output data for the ML models based on FE simulations are shown. Subsequently, different regression algorithms are evaluated, and a morphological box was used to identify the most promising algorithm. The research shows that XGBoost achieves good results with a small training data set and can assist designers in making decisions regarding the design of clamping system.

Another study on intelligent fixture design in high performance machining is shown in [178]. For this purpose, the influence of different workpiece-fixture setups on the natural modes is investigated. The results first show the relevance of fixture layout in the context of process-workpiece interaction for fixture design, layout, and optimization. Furthermore, the use of intelligent fixtures is examined to reduce the influence of vibrations, deformations, and positioning of thin-walled parts. It is shown that the use of model-based method in combination with process simulation enables a significant improvement of the fixture performance and process robustness.

For the identification of forming limits in sheet metal, DL algorithms have been investigated in [104]. The forming limit curve defined by the major and minor strain is used to determine the forming range. As forming behavior is difficult to investigate, since only the last process step can be mapped due to the process setup, a semi-supervised neural network is presented to detect the onset of localized necking. The studies include two steps (Fig. 13). The first is supervised feature learning where the extreme ranges of the forming sequences are considered. The second is unsupervised clustering using Student's t mixture models (SMM), which groups the remaining frames of the forming sequence. The presented approach allows location and time independent investigation and an online analysis of a distinct time point. By processing the information from the captured images, cracking of the sheet specimen can be prevented. A detailed summary of AI for process design is presented in the review article [147].

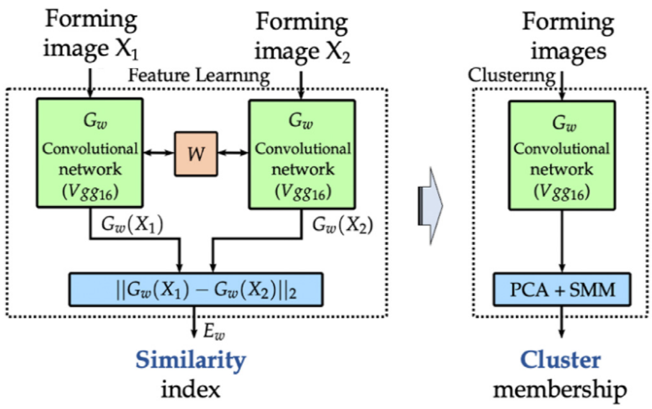

Fig. 13. Supervised feature learning and unsupervised clustering for determination of forming limits, adapted from [104].

3.2. Process modeling and evaluation

Depending on the specific application, process modeling requires associating process parameters to the final property of the produced part (e.g., process-structure-property-performance, or PSPP relationship) or revealing the mechanism underlying the time-evolution of process (e.g., process dynamics). The capability to accurately predict part property is crucial for process optimization, while understanding the underlying process mechanism serves as the technical basis for process control.

While empirical equations have been developed over the years for describing the PSPP relationship, process-to-process variation and process physics that are unaccounted for by these models inevitably cause deviation in terms of process modeling. The advent of data-

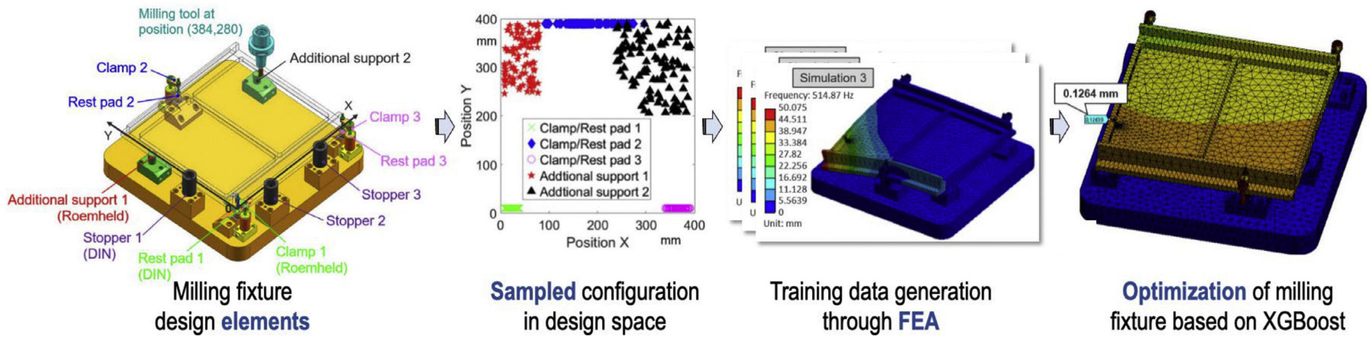

Fig. 12. Fixture design optimization for milling process using FE analysis and XGBoost, adapted from [60].

driven AI provides a promising tool to effectively use the in-process sensing data to close this gap.

Choi *et al.* [35] presented a neural network-based method to understand the relationship between the input process parameters in injection molding and process output properties, such as those associated with linear shrinkage. The idea is based on the self-organizing properties of a neural network. The system consists of three functional software groups: a user interface and command module, an optimization and synthesis module, and a computer-aided engineering (CAE) analysis software module. The authors have shown that a prediction error of 0.5% has been achieved [35]. Bak *et al.* [12] also investigated a neural network model for die-casting process. First, an optimal set of dominant manufacturing parameters for high product quality in a die casting process is determined using the minimal redundancy and maximum relevance (MRMR)-based approach. With the selected parameters, a prediction accuracy of 99.6% has been achieved by the neural network for process yield prediction.

The applicability of AI techniques to flexible rolling process for customized semi-finished products is shown in [121]. Kirchen *et al.* established fundamental correlations between process and quality parameters using data-driven AI methods. Here, the predictive model is set up using incremental regression modeling and subsequently evaluated with the aid of process and quality data. The quality variable describing the homogeneity of the sheet thickness of the semi-finished product can be predicted with a maximum deviation of only 5%. The predictive model allows to derive adapted parameter settings between process steps for a product, which offers the possibility for process optimization. Similarly, in [264], a DBN is investigated to model the complex relationship between material removal rate (MRR) and the underlying process parameters in chemical-mechanical polishing. The outcome shows that the DBN can predict the MRR with less than 3% error, which is an order of magnitude smaller than the results from the physical equation.

In addition to analyzing part properties, Brillinger *et al.* [27] have explored the application of AI in predicting the energy consumption of CNC machines during processes. For this purpose, a training part is first processed, and high frequency measurements and the NC instructions are collected. Based on this data, a ML model is trained (Fig. 14). After that, the validation part is processed. The NC instructions of the validation part are passed to the already trained model. The model then predicts energy consumption. For this purpose, three ML algorithms were trained "random forest", "decision tree" and "AdaBoost + decision tree" [27]. Although there are exceptions for certain measures and aggregates, the most accurate predictions can be obtained with the random forest technique. Accordingly, the energy demand curve of the machining process is accurately predicted.

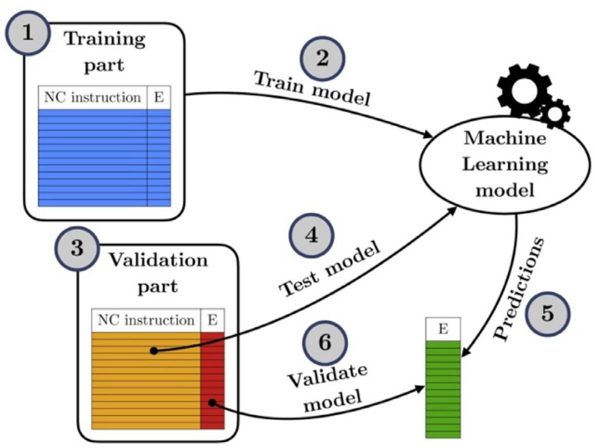

Fig. 14. Energy prediction for CNC machining with ML [27].

Brecher *et al.* [25] presented a knowledge-based approach to evaluate the quality of toolpaths by reducing process data to information about machining deficiencies. The approach can be used for post-process and planning-integrated diagnosis as a first step towards optimization. The method allows NC planners to obtain feedback from virtual or real processes to improve their knowledge of the current planning state and identify existing deficits. They are thus supported in deriving possible means of optimization from aggregated information without having to perform time-consuming analyses of raw data.

One of the most recent research trends in AI-based process modeling is the metal AM process. Due to its multi-physics nature, predicting part property and modeling process dynamics for AM is challenging. As a result, researchers have turned to data-driven AI for establishing the complex relationship between process dynamics and part property. For metal AM, the characteristics of the melt pool are known to play a crucial role in determining the process behavior and outcome. For the evaluation of the melt pool, a methodology based on edge image templates combined with Bayesian inference is demonstrated by Lindenmeyer *et al.* [156]. Specifically, high-speed X-ray images of the melt pool area were analyzed. The developed detection method has a 60% accuracy for identifying the dimensions and shape of the melt pool.

The AM process induces process heat that results in local varying mechanical properties and is site dependent. To predict mechanical properties of the parts, infrared measurements were performed on several thin-walled components at the selected positions and converted to wavelet-based scalograms by Xie *et al.* [277]. Subsequently, a CNN is used to predict mechanical properties obtained from miniaturized tensile tests (Fig. 15). Furthermore, through a random forest algorithm, an infrared thermography parameter can be used to relate the mechanical properties to the temperature ranges of the component [277].

With the recent development of physics-informed ML [115], researchers have started to integrate physical knowledge about the AM process and data-driven method, to accurately predict its time-dependent evolution. The main idea of physics-informed ML is that, in addition to the prediction accuracy in the ML loss function, a physical-consistent term is also added such that any deviation from the physical equation will be penalized. As a result, the predicted output from the ML is expected to be physically consistent. Early proof-of-concept has been reported by Zobery *et al.* [296] and Liao *et al.* [154]. Both studies integrated the heat equation for prediction of temperature evolution in AM.

3.3. Process optimization

All processes in manufacturing are subject to variations. As an example, scattering of the geometrical or mechanical properties of materials can cause disruptions during series production and must therefore be detected, evaluated, and ultimately controlled and optimized in time. Conventional methods using process simulations are often time-consuming and not flexible enough to react quickly enough to changes for in-situ optimization. AI offers new opportunities to exploit further optimization potential.

For process control, researchers have generally followed the vast knowledge of control theory and integrated it with AI-based modeling of process dynamics. For example, in [8], an integrated method of model predictive control (MPC) and Gaussian process (GP)-based AM model is developed. The objective of GP is to predict the time-evolution of the melt pool width given the laser power and other process parameters. Once the predictive model is obtained, it is linearized locally to be compatible with MPC. Simulated results have shown that the integrated control method is effective in controlling the melt pool width from deviating from a reference trajectory, which is widely considered critical to ensure AM part property. Similarly, in [210], a PID controller is investigated for AM melt pool depth control, with the depth information directly predicted from sensor images.

Process optimization can be considered an inverse problem where suited process parameters need to be determined to arrive at the optimized output. One of the methods is through sensitivity analysis,

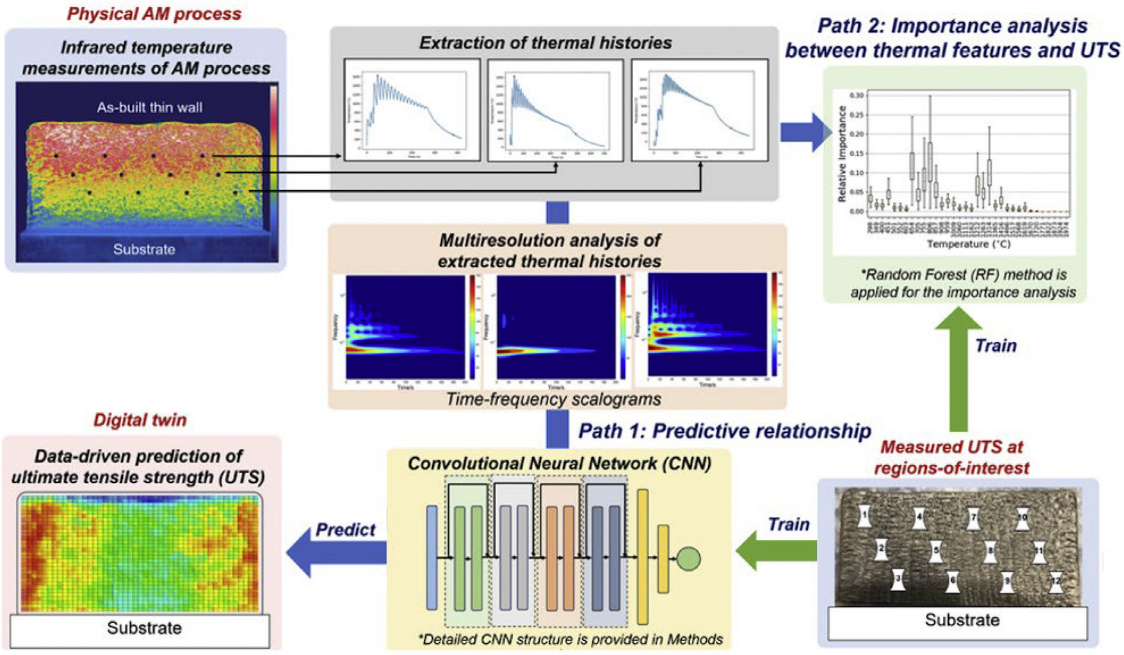

Fig. 15. Mechanistic data-driven method for tensile property prediction in AM, adapted from [277].

as it indicates how the output changes when a unit of change happens at the input (e.g., process parameters). In [245], an ML-based methodology for predicting and improving the energy requirements of battery production was developed (Fig. 16). The approach is intended to highlight the interdisciplinary nature of battery production and to be applicable to other sectors. Once an AI-model is established at the third phase of the process, "modeling & evaluation", to identify the most influential factors, sensitivity analysis is carried out to evaluate energy efficiency potentials and derive actions for improvements. It is reported that energy savings of up to 9% have been achieved.

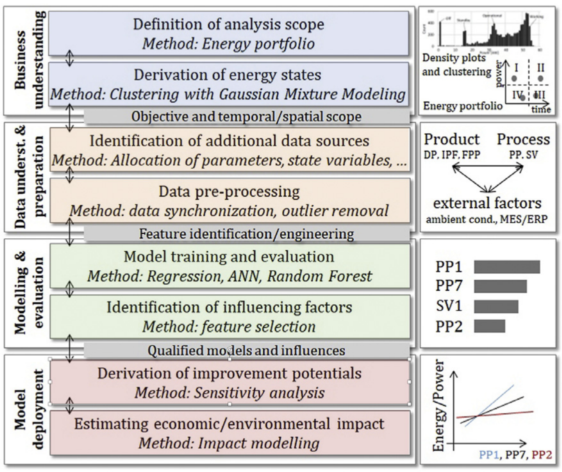

Fig. 16. ML based optimization of energy efficiency, adapted from [245].

Other researchers tackle optimization by directly modeling the inverse mapping using ML, e.g., for the production of composite materials. For this purpose, ML models were developed in [99] for the rapid evaluation of a wide range of boundary conditions (Fig. 17). By comparing thermocouple data with the predictions using these boundary conditions, all plausible solutions can be identified. To this end, two long short-term memory (LSTM) networks were developed to predict workpiece and mold temperatures for a given thermal stack and air temperature profile. In addition, a neural network was developed for multi-objective optimization of temperature cycle. The method mitigates the risk associated with unknown boundary conditions, and the system can be used for real-time optimization of the curing process with active adjustment of an oven.

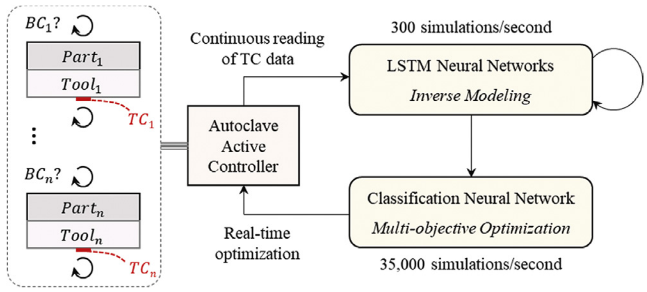

Fig. 17. Schematic of ML framework for inverse modeling of composites processing, and optimization of air temperature profile [99].

For ML techniques with probabilistic nature, such as Gaussian process, Bayesian optimization provides an intuitive way of parameter optimization while considering process model uncertainty [233]. The idea is to first model the output variables of interest with respect to the process parameters. Then, Bayesian optimization is carried out sequentially to determine the parameter point to conduct the subsequent experiment that has the highest probability of improving the output. ML combined with Bayesian optimization has shown to be particularly advantageous in situations where acquiring experimental data is a time-consuming process. The effectiveness of the method is demonstrated by Maier *et al.* [166] and Khosravi *et al.* [119] for grinding, where feed rate and cutting speed, and gain of PID controller are optimally tuned, respectively.

One of the promising ML techniques that has attracted much attention recently for process control and optimization is RL [239]. As RL commonly learns the association between process state and control/optimization adjustment, it is considered model-free. Several studies on RL-based process control and optimization have been reported [52,126]. For example, Dornheim *et al.* [52] investigated RL to find the optimal blank holder force in deep drawing to minimize the tear of the produced part. Their approach is based on DQN that uses a neural network to predict the goodness associated with each blank holder force adjustment, and the one with the highest value is selected. Training of the neural network is carried out in simulated environment. The authors reported that after 200 iterations, the RL-based method is able to outperform rule-based methods. Besides DQN, the method of actor-critic has also been reported for control/optimization of welding [109].

Finally, the methods of search have also contributed to process optimization. These methods can effectively explore the parameter space and converge to the optimal solutions. For example, evolutionary algorithms have shown effective in optimizing tool path for milling, with different optimization properties such as time, straightness, and cutter engagement [189]. In addition, GA and particle swarm optimization have been successfully implemented for parameter optimization for cutting [136] and turning [18], respectively.

4. AI for quality assurance and maintenance

Quality assurance and maintenance have been related to AI since the early 1980s [146]. Over the years, the development of model-based AI techniques, such as those based on defect-induced signal features and the physics-based models has contributed to safe operations and provided technological basis for fault and defect identification [105], machine/tool degradation and remaining useful life (RUL) prediction [112]. When it comes to data-driven AI, the field of quality assurance and maintenance has some unique challenges as compared to the other aspects in manufacturing, such as data imbalance and domain knowledge integration [116]. Collectively, these challenges started to reshape the research of AI-based quality assurance and maintenance, leading to new development of data-driven AI that is robust, interpretable, and consistent with physical knowledge. The AI techniques and the corresponding applications in quality assurance and maintenance are summarized in Fig. 18.

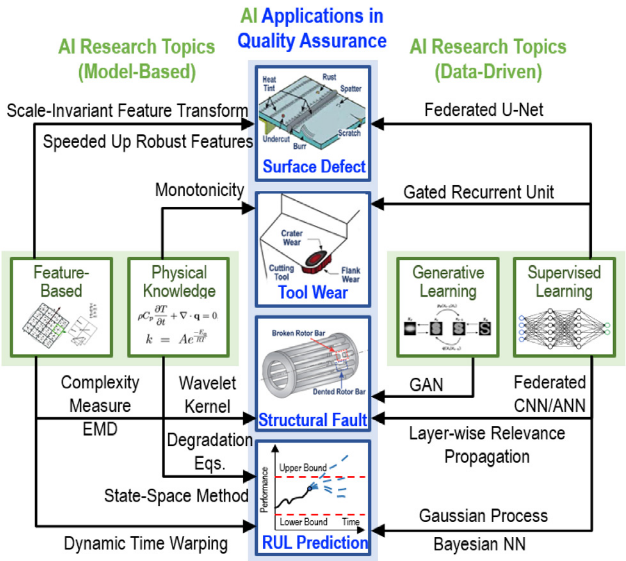

Fig. 18. AI-related research for quality assurance and maintenance.

4.1. AI-enhanced condition monitoring

As the first step towards quality assurance, condition monitoring refers to monitoring the quality attributes of a part, machine, or tool to identify deviations that are indicative of potential defects or faults [102]. Traditionally, this task relies primarily on inspection by human experts, which is typically a time-consuming process conducted off-line. The consequence is delays in defect detection and interruption to the manufacturing processes. The development of AI, especially the advances in image analysis, has accelerated research towards automated online condition monitoring.

Images have been widely investigated in AI for manufacturing. For example, time-frequency images of vibration signal for machine condition monitoring [48] and optical images for surface roughness estimation [34]. In these applications, the information contained in the image has been distilled into a scalar prediction while its spatial information is largely discarded. However, for part surface inspection, it would be desirable that the AI algorithm can pinpoint the image region that deviates from normal conditions, such that the outcomes are more easily understood by human.

Early success of image-based defect detection often relies on point-of-interest extraction enabled by scale-invariant feature transform (SIFT) [160] and speeded-up robust features (SURF) [14]. The motivation is that surface defect can exhibit noticeable change in terms of patterns of pixel intensity as compared to the non-defective region. As a result, both SIFT and SURF rely on models for local extrema detection in image pixel variation, such as the difference in Gaussian (DoG) model and the Hessian matrix model. Coupled with multi-scale analysis, SIFT and SURF have shown to be effective in finding the points-of-interest to support defect detection on steel surface [240] and PCB board [85].

The recent development of DL techniques, especially the CNNs have further sparked research in detection of surface defect whose characteristics cannot be described using features such as local extrema of pixel densities [223]. As a data-driven AI technique, the working principle of CNN is fundamentally different than model-based method as it does not depend on pre-defined models for feature extraction. Instead, image features that are most relevant for defect detection are learned through training images.

Mehta and Shao [173] presented a CNN-based approach for surface defect detection and segmentation in AM. For this purpose, a U-Net structure, a variant of CNN is considered (Fig. 19) [211]. U-Net is advantageous over other CNN variants (such as fully convolutional networks) in defect detection where defects can vary in size and shape, due to the network's unique architectural features [211]. Its design, which includes a contracting path and an expanding path, allows for effective feature analysis at multiple levels that are adaptive to both large and small defects as well as different defect complexities (e.g., shapes). Moreover, the skip connections connecting the contracting and expanding paths facilities the integration of multi-level image features to enable accurate segmentation of defects at the U-Net's output.

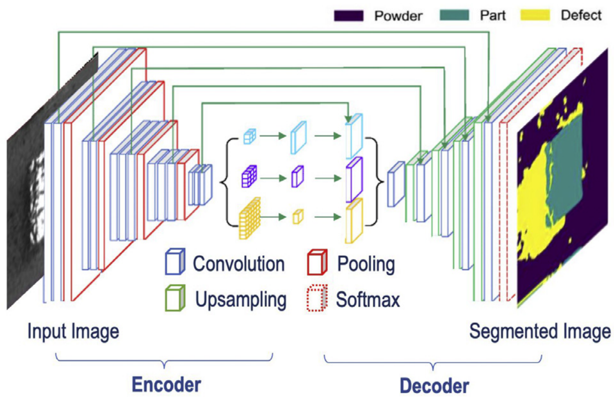

Fig. 19. U-Net for defect segmentation for AM, adapted from [211].

To overcome the limitation that training images containing defects can be insufficient from a single source (e.g., manufacturer) to fully optimize the U-Net, the method of federated learning is considered by the authors [173]. Federated learning refers to a *collaborative* data-driven method that works under the premise that limitation in data quantity can be overcome by pooling data information from various sources (known as clients). To avoid sharing data directly and violating data privacy, federated learning uses a single global model while each client only provides a model parameter update that is computed using its own data (Fig. 20) [127,173]. As a result, data is never shared across different clients, while the information embedded in this data contributes to the construction of a global model. The authors have shown that by using federated learning, defect

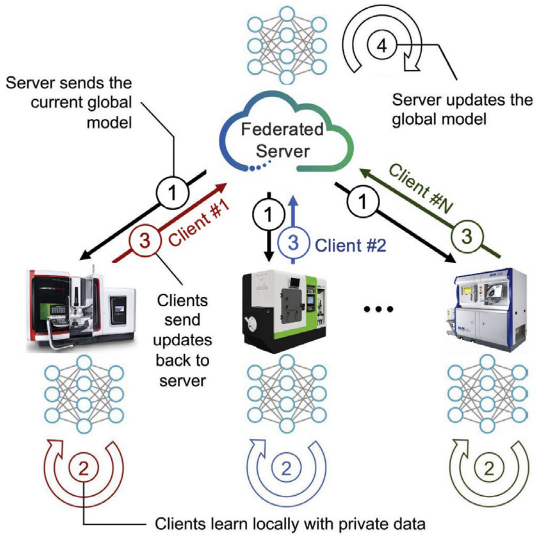

Fig. 20. Schematic of federated learning, adapted from [127,173].

segmentation accuracy of the global model has outperformed individual models that have been trained using siloed data from each client by 20%.

Beyond images, time series data are also commonly collected to infer the snapshot of machine/tool conditions using AI. As the temporal information embedded in time series data often comes directly from physics, the integration of physical knowledge and data-driven method has been investigated to improve the physical consistency of AI-models. Wang *et al.* [259] presented a gated recurrent unit (or GRU [37]) network for milling tool wear prediction. This study features physics-informed network training to ensure that the predicted tool wear level is monotonically increasing as the cutting cycle increases. By using physics-informed training, the network prediction logic will be penalized if the predicted tool wear does not increase monotonically. As a result, network weight update is guided to achieve maximum physical consistency during training. The authors have shown that by integrating physics-informed training, not only physical inconsistency in tool wear prediction has been eliminated, but also the tool wear predictive error is consistently lower (up to 50% in terms of root mean square error, RMSE) as compared to the scenario without physics-informed training.

4.2. Structural fault identification

As the key components in power transmission, smooth and fault-free operation of rotary machines, such as induction motors, gearboxes, and bearings are critical for manufacturing processes such as cutting, grinding, and metal forming. AI-based diagnosis allows to identify information related to hidden structural faults from sensing data collected from these machines, leading to informed decision-making on predictive maintenance to prevent unexpected interruption to production.

Model-based AI for fault identification traditionally relies on time-frequency analysis of sensing signals, such as the wavelet-based method, to reveal the structural fault as manifested at the characteristic frequencies computed using physical models [61]. Beyond time-frequency analysis, researchers have also identified correlation between machine structural fault and sensing signal that is based on complexity measures in information theory. For example, Yan and Gao observed that fault severity level in bearings is highly correlated to complexity measures of the vibration signal such as approximate entropy, permutation entropy and Lempel-Ziv complexity, leading to reliable fault identification [279,280]. Additionally, empirical signal decomposition, such as empirical mode decomposition (EMD) has also shown capable of revealing the composition of sensing signal and extracting fault related information from the decomposed intrinsic modes [70,145].

The advent of data-driven AI, especially DL, has opened a new avenue for fault identification in rotary machines. The main advantage is the elimination of the need to compute and select a priori features or measures that are relevant to the structural fault [70]. However, the collection of faulty data is often limited due to production and safety constraints [71]. As a result, data augmentation has been one of the main research focuses.

Recently, data synthesis based on generative adversarial network, or GAN [78], to alleviate the lack of data from faulty conditions for model training has shown great potential. The structure of GAN typically consists of a generator and a discriminator (both as neural network, Fig. 21). The objective of the generator is to learn to transform samples from a known high-dimensional distribution into samples from the underlying distribution of faulty data. The objective of the discriminator is to learn to distinguish synthetic samples (from the generator) from real samples collected from the faulty machines. The performance of both is improved through adversarial training, in which the generator is trained to improve the data synthesis quality and reduce the discriminator's accuracy, while the discriminator is trained to improve its capability of detecting synthesized data. Such adversarial training is expected to arrive at an equilibrium in which the discriminator can no longer distinguish synthetic data samples from real ones. At this point, the generator can be used for high-fidelity data synthesis and data augmentation.

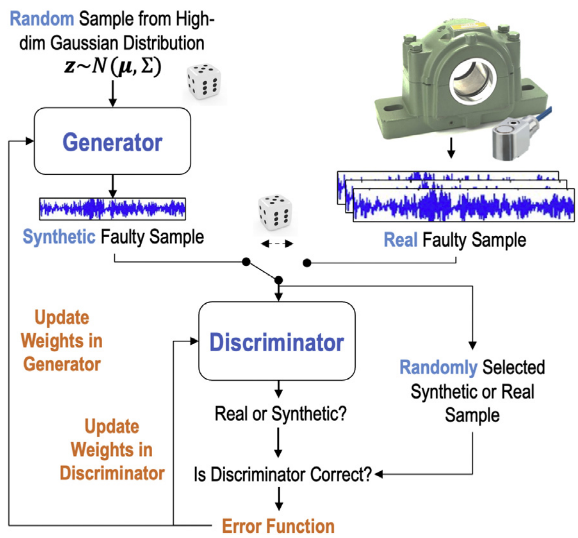

Fig. 21. Schematic of GAN for data synthesis, adapted from [78].

The GAN-based data synthesis has been validated in several publications over the past years, with consistent conclusion that the data-driven AI models trained using synthetic data have outperformed the ones trained using unbalanced data in fault identification for induction motor [227], tool wear [38] and bearing [158]. For example, Shao *et al.* [227] have shown that the fault identification accuracy improvement can be as high as close to 50% (from 50% to 99.3%) when the imbalance ratio is 2:1 between the healthy and faulty motor data. Cooper *et al.* [38] investigated synthesis of wavelet spectrums using GAN for non-compliant tool detection in milling. Different from the previous works in which the classifier is either constructed separately or incorporated with the discriminator, the generator of the GAN is inverted to perform non-compliance detection in this work, resulting in a 25% improvement in detection accuracy for the dataset with 2:1 imbalance ratio. Liu *et al.* [158] examined the waveform and the corresponding Fourier spectrum between the real and synthetic bearing vibration signal (Fig. 22). The authors noted that beyond visual similarities, the frequency components in Fourier spectrum are well preserved in synthetic data,

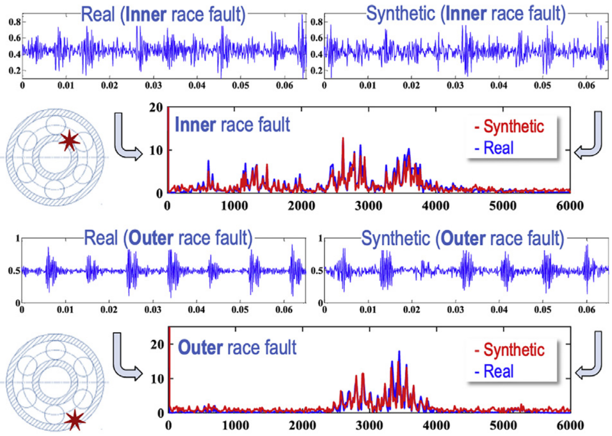

Fig. 22. Consistency between real and synthetic vibration data in time and frequency domains, adapted from [158].

critical for vibration-based fault identification and indicative of good performance from GAN.

Besides GAN-based data synthesis, federated learning has also attracted research interest to bypass the limitation in faulty data quantity. Good fault identification accuracy has been reported in several publications [75,292]. Zhang *et al.* [286] investigated a scenario in which data from each client is not guaranteed to be available during the training process. The motivation is that as each client has full control of its own data, it is possible that data from certain clients may not participate during certain stages of the training process due to issues such as scheduling conflicts. Such partial participation constitutes one of the main differences between federated and non-federated learning. In [286], the authors evaluated a federated learning scenario with 50 clients, each having a dataset of distinct imbalance levels between data of healthy and various faulty bearing conditions. Each training epoch has a client participation rate from 80% to 100%. The authors demonstrated that federated learning is robust to partial participation and achieved fault identification accuracy comparable to the one that would have been achieved using centralized training (around 96%). This demonstrates federated learning as a reliable technique in addressing limitations in data quantity for fault identification of rotary machines.

For critical applications such as fault identification, once the decision is made by AI, it is imperative to evaluate the logic behind the decision against the existing physical knowledge to avoid spurious findings. That is, to know exactly how each element in the input contributes to the final decision. Research on such post-analysis methods represents the first step towards opening the black-box of AI models [9], especially DL models, and facilitates broader acceptance of AI-enabled applications [230,285].

As an example, Grezmak *et al.* [79] investigated layer-wise relevance propagation (LRP) to determine the prediction logic of CNN-based motor fault identification with wavelet time-frequency spectrum of the vibration signal as the CNN input. In contrast to the fault identification that transforms the input into a discrete probability distribution of motor conditions, LRP works backwards by redistributing the final probability distribution as relevant score until it reaches the input (Fig. 23). The score redistribution follows two rules: (1) the score assigned to each neuron is proportional to the multiplication of the activation of the neuron and the weights connecting it to the next CNN layer, and (2) all neuron scores within each CNN layer add up to one [9]. Based on these two rules, each pixel in the spectrum with a positive score can be considered as contributing to the CNN decision. The authors observed that the scores in the wavelet spectrum image exhibit alternating positive and negative bands, with positive bands largely falling on the characteristic frequencies and their harmonics, indicating consistency between the CNN prediction logic and human knowledge for motor fault identification.

Beyond the post-analysis techniques such as LRP, researchers also investigated network structures that constrain the prediction logic to

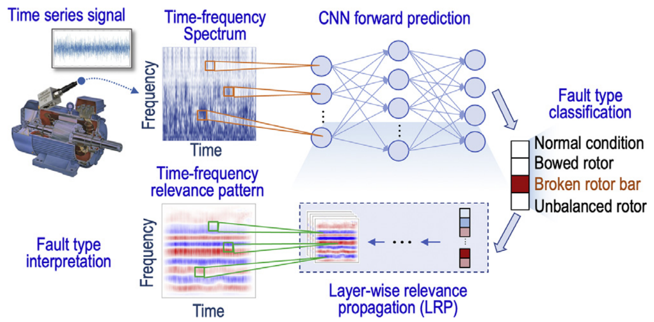

Fig. 23. LRP for interpretable CNN-based motor fault identification [79].

be physically consistent, for example, a continuous wavelet convolutional layer (CWConv) for bearing fault identification [153]. Different from standard CNN kernel that is unconstrained, the wavelet kernel is parameterized by two wavelet parameters: scaling and translation. As a result, the behavior of the kernel is consistent with the known physical properties of wavelet, leading to improved model interpretability as compared to standard CNN.

4.3. Machine remaining life prognosis

Prognosis aims at predicting the progression of machine performance from its current status to its functional failure. Accurate RUL estimation provides the technological basis for predictive maintenance [72].

Model-based method in RUL estimation relies on physical degradation models, such as the Paris' law or the Arrhenius equation for characterizing damage propagation [198]. As the parameters in these equations are often undetermined, calibration is needed to update model parameters using sensor data. The degradation models with updated parameters then carry out the estimation of RUL. Such a combination of physical model-based prediction and data-driven parameter update can be considered a hybrid AI approach. Among the techniques under this category, a state-space model using particle filter has attracted much research interest due to the root in Bayes' theorem [7].

Bayes' theorem allows to calibrate the degradation model parameters using both physical relationship and sensing data [7]. Specifically, the physical relationship can be considered as the prior knowledge about the parameters, while sensing data allows to compute the likelihood of the prior knowledge given the real-world observations. Bayes' theorem fuses this information to arrive at a more accurate, posterior estimation of the model parameters. In practice, particle filter utilizes weighted particles to represent the uncertainty in model parameters, making it suited to characterize any distributions underlying parameters. Recent development of particle filter includes a multi-modal particle filter [267] that can accommodate different degradation modalities and a local search particle filter that improves the convergence of the algorithm in characterizing the model parameters [263].

Beyond the state-space model, a hybrid AI method has also been developed by integrating physical equations and ML. The main idea is to use ML to learn the difference between the physical equation and the real-world observation and thereby, complementing the physical knowledge. For example, Zhang *et al.* [287] developed physics-guided GP (PGGP) by embedding a degradation equation into its mean function (Fig. 24). GP is a ML method where machine performance at any time step in the future is predicted as a Gaussian distribution that is conditioned on its past performance [287]. Furthermore, uncertainty analysis is inherently incorporated. The authors demonstrated that the hybrid AI outperforms the pure data-driven approach in long-term estimation of RUL of HVAC systems and Li-Ion batteries by up to 75% in terms of prediction error. A similar work for bearing RUL prognosis is reported in [107].

Additionally, DL has also been integrated with Bayes' theorem for RUL estimation [19]. The concept is that for any observed data, there

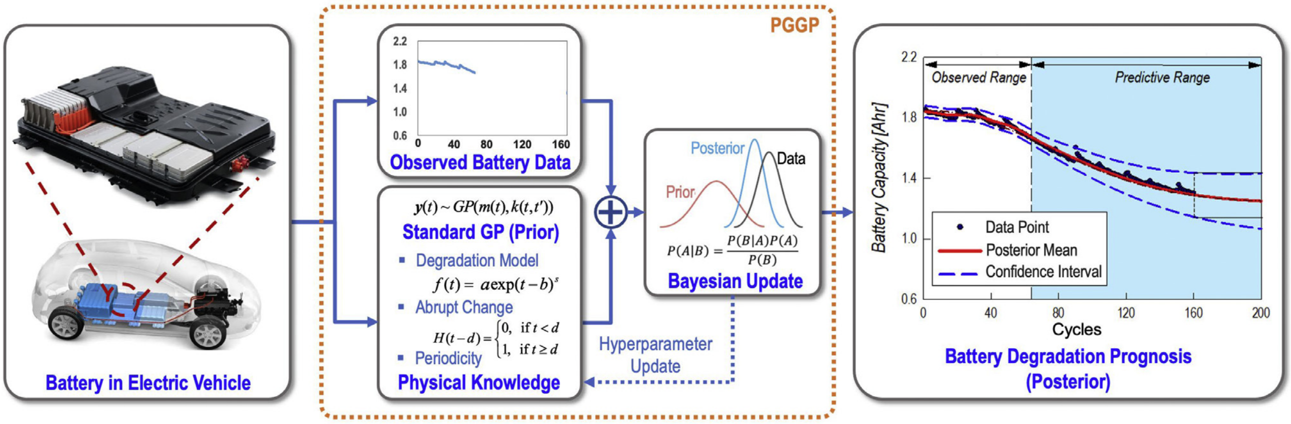

Fig. 24. Physics-guided Gaussian process for system performance prognosis (with Lithium-Ion battery as a case study), adapted from [287].

are an infinite number of possible models, and the prediction should consider the results from all of them. More specifically, each model is assigned a model probability, and the prediction result can be considered as the sum of each model's prediction weighted by the corresponding model probability [19]. The effectiveness of this method has been validated on Li-Ion battery [293]. In addition to the methods described above, other prognosis techniques have also been developed, such as the dynamic time warping method that aims to find the optimal match between the degradation trajectory to be predicted and a reference trajectory [120].

5. AI for automated and flexible assembly and disassembly

The following chapter (Fig. 25) discusses the advances and potentials of AI for assembly and disassembly. Specifically, the contribution of AI to the growing challenges of assembly planning and material flow control is first described, followed by the fundamental AI-based advances in robotics for automation of assembly and disassembly. The rapidly emerging research in the field of AI for HRC to cope with the rising demand for flexibility and changeability is also discussed, followed by an overview about future potentials of AI for assembly and disassembly.

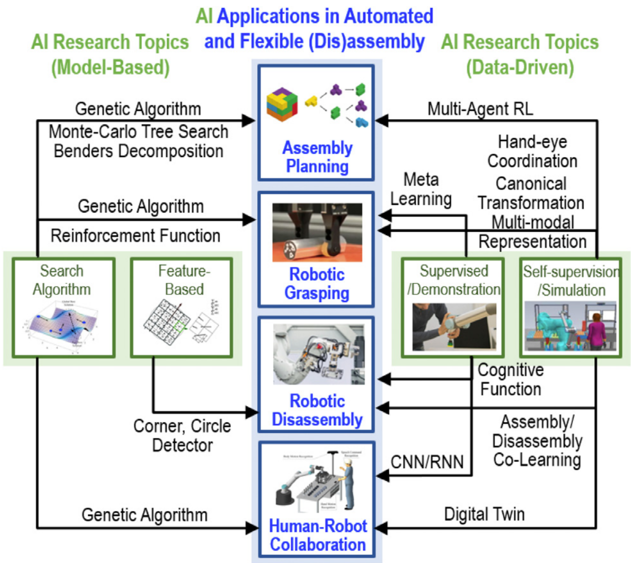

Fig. 25. AI-related research for automated assembly and disassembly.

5.1. Assembly planning and material flow control

The increasing variety of products, especially in the automotive industry, goes hand in hand with the growing demands on the flexibility, changeability and reconfigurability of assembly lines, which have seen the evolvement of lineless structures such as matrix production [237]. These growing demands, coupled with the comparatively low degree of automation of complex handling and assembly operations, demographic development, and associated increase in average age of assembly workers, have resulted in a significant increase in the complexity for assembly planning and material flow control and highlighted the need for efficient methods to search for the optimal solutions among combinatorial variety of outcomes.

5.1.1. Assembly planning & optimization

Introduced in 1975, GA, a subdomain of evolutionary computation, is regarded as a means for solving a broad range of optimization problems for assembly planning and scheduling. For example, Dini *et al.* [49] proposed a GA-based method for generating and evaluating assembly sequences. Also based on GA, Raatz *et al.* [204] proposed a method for optimizing the interlocking of human and robot operations in collaborative robot-assisted assembly. Kardos *et al.* [114] suggested a constraint programming approach to assembly planning based on boundary-aware decomposition to account for process complexity.

Complementing the GA-based method are approaches for knowledge-based solutions generation through ontologies. Ahmad *et al.* [3] described a new method in which inferences are generated based on explicit knowledge mapped in modular ontologies, based on which product and process requirements are mapped to available resources. Additionally, self-learning and self-optimizing assembly systems have been proposed for adaptive assembly by Kluge *et al.* [125]. Planning for general manufacturing systems has been described in Ch. 2.

5.1.2. Material flow control

Due to the flexibilization of assembly based on the dissolution of rigid line structures and transition to matrix production, the needs for flexible solutions for material transport using mobile robot platforms are growing, opening the potential of AI for increasing the degree of automation of autonomous mobile robot (AMR). Among a great wealth of research for AMR, learning unknown environments and automatic navigation forms a major focus. Various methods of supervised learning, self- and semi-supervised learning, unsupervised learning, and RL are developed to identify movements of AMR relative to its environment. These methods can be combined and assigned to artificial perception. A review of vision-based navigation is provided by [43].

For research of material flow control in known environment, major research focuses have been placed on "obstacle avoidance" [23], "indoor navigation" [87], and more recently with the rise of automated warehouse, "multi-robot cooperation" [155]. As an example, Malus *et al.* [168] combined perception and navigation-related abilities of AMRs with MARL, where AMR agents learn to bid for orders to realize self-organizing order dispatching. Specifically, agents are given order specification and learn to form a bid based on their location and immediate plans. The reward function is designed to reward all agents upon order completion by any individual agent, which stimulates agent learning towards cooperation. MARL training is carried out first using a simplified simulation before transferred to a physics-based simulation of AMR fleet for validation (Fig. 26). The authors demonstrated that the learned policy outperforms "closest-first" policy by learning to cooperate and adapt to the workspace layout.

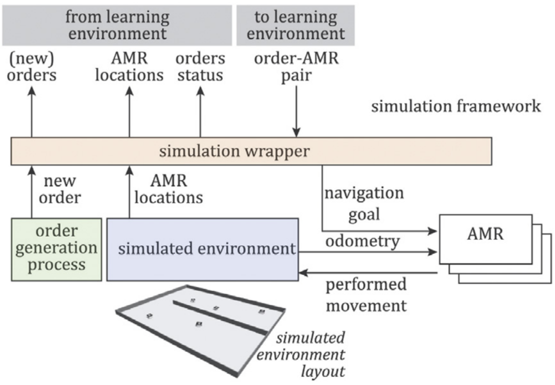

Fig. 26. MARL-based order dispatching, adapted from [168].

5.2. Potentials of AI enhanced robotics for assembly

Industrial robots are essential for automation of assembly operations and have benefited from the considerable progress in fundamental research in robotics through the development of AI methods. Their contribution to robotic object handling and manipulation plays a crucial role in enabling robotic assembly.

5.2.1. AI-enhanced "Pick and Place"

Handling operations in assembly often have a degree of complexity, which requires the combination of several cognitive and sensory motor abilities of a human such as learning/ reasoning, visual cognition, flexible mechanics, and haptic cognition [131]. As a result of continued progress in ML for robot-assisted manufacturing, the field of "robot learning" as a subdomain of ML has gained increasing attention in recent years. In [159], Liu *et al.* presented a comprehensive review of this field.

Advances in the field of visual object recognition and detection through CNN form the basis to achieve hand-eye coordination during object grasping and manipulation. In early works, Levine *et al.* [149] demonstrated the robot's ability to independently learn to grasp unknown objects based on 800,000 gripping cycles of 14 robots working in parallel. The idea is to establish a neural network-based mapping between workspace sensing images and robot action through interaction between robot and workspace such that grasping success probability is maximized.

One of the limitations in [149] is data collection with physical experiment that lasts more than 2 months. This motivates other researchers to explore methods with improved learning efficiency. For example, Finn *et al.* [62] demonstrated that one-shot imitation learning could reduce needs for training data. One-shot imitation learning can be considered a variant of meta-learning where the objective is to optimize meta-parameters such that a small number of gradient steps can produce good performance under new tasks. Additionally, simulation-based methods have been investigated to replace physical robots and facilitate the learning process. To bridge the sim-to-real gap such that the grasping algorithm learned in simulation can be translated into the real world, Rao *et al.* [209] have investigated cycle-generative adversarial network (Cycle-GAN) to synthesize real-looking texture for the simulation (Fig. 27), and reduced learning time from months to days.

More recently, Wen *et al.* proposed a method to learning category-level, task-compatible grasping using simulation data [272]. The method achieved category-level generalization through a CNN-based canonical transformation. Then, grasping heatmaps are generated for different part categories based on grasping stability and task compatibility (Fig. 28). A survey on ML vision-based robotic grasping and manipulation is presented in [124].

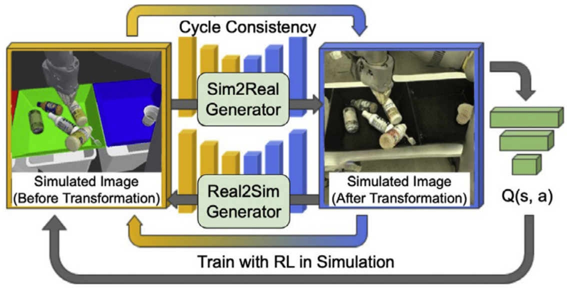

Fig. 27. Transformation of simulated image to support simulation-based robotic grasping learning [209].

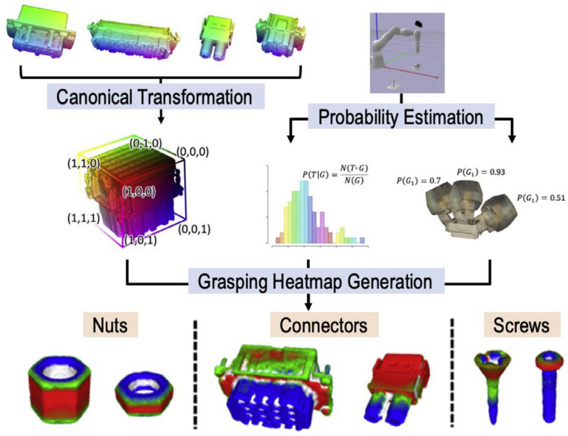

Fig. 28. Robotic grasping learning based on canonical transformation and simulation-based grasping probability estimation, adapted from [272].

In addition to learning algorithms, Gabriel *et al.* [68] optimized object grasp points based on constrained GA, which evaluates both part's mass distribution and the holding force each gripper can provide. The developed method first identifies potential grasp points based on a point cloud of the geometry. Then, part segmentation and identification of center of mass are carried out. Finally, GA-based optimization is implemented with weighted fitness criteria for evaluation and selection of optimal grasp points.

To ensure a robust, fast, and accurate stacking process, Bobka *et al.* [21] presented a deviation compensation strategy which increases accuracy through modeling of process-specific deviations. Potential multidimensional regression methods for modeling the deviations are compared. Supported by ANN, placing operations with significantly increased precision are performed by a robot-based fuel cell stacking system. The contribution of the work is the model simplicity where the authors designed a feed forward ANN (FF-ANN) with only 3 hidden layers and 40 neurons to achieve the significantly optimized stacking accuracy of limp fuel cell components. In [284], a learning-based method for joint object picking and placing has been developed to assist in autonomous object assembly using height maps of both object and assembly kit. Specifically, three neural networks are developed to predict picking location, placing location, and the needed rotation of the object before placing, as shown in Fig. 29. The novelty of the study is a self-supervised approach to use disassembly as a means of data collection for network training that minimizes human intervention. The authors demonstrated that after 10 h of training, the method can achieve a 94% assembly success rate.

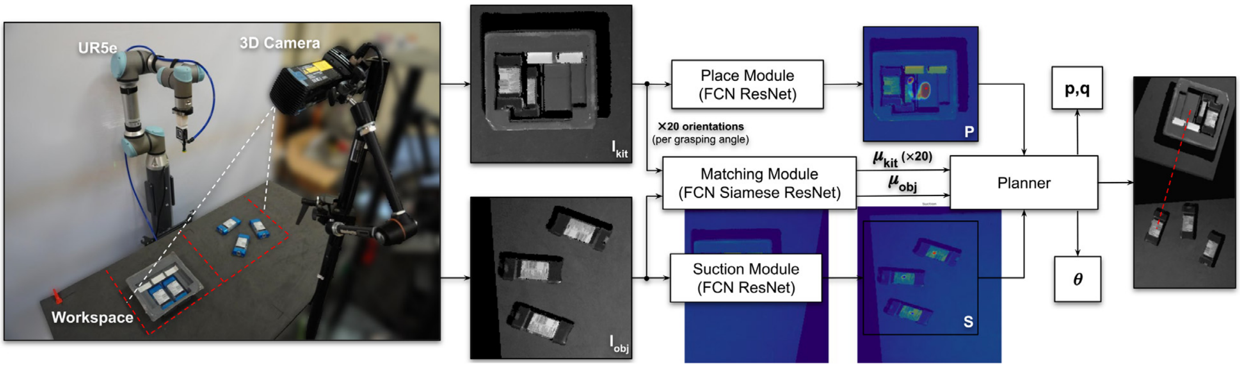

Fig. 29. Joint learning of object picking and placing for robotic assembly [284].

5.2.2. AI methods for contact-rich joining tasks

A human's ability to manually perform complex joining processes with small tolerances such as peg-in-hole insertion is based on a fusion of visual and sensory motor abilities in combination with cognitive skills for motor control as well as dexterity and compliance of the human arms and hands. However, translating such capability into complex robot-based assembly operations, especially the transition from free motion when approaching the workpiece to the moment of contact between robot and stiff environment, is challenging [133].

For these types of contact processes, two control strategies are commonly applied, indirect control methods such as compliance or impedance control and direct force control. In these operations, one of the fundamental challenges is to control the position and force of the end effector at the same time, where the stiffness as well as position and shape of contact object often are not exactly known. Qiao *et al.* [203] introduced a learning mechanism to compensate for this lack of knowledge. Through a "reinforcement" function, regarded as a principle for parameter identification and adaptation, the method optimizes a joint position and force control of the robot in contact with the unknown constraint environment.

Other studies of control of contact-rich tasks were focused on ML [148] to combine visual data as generated from camera data in hand-eye coordination scenarios with other sensing modalities such as force/torque and proprioception. For example, Lee *et al.* [144] proposed a multimodal representation learning approach, where camera images, force/torque signals and data from the robot encoders for current position and velocity are jointly encoded into one model comprising MLP and CNN as shown in Fig. 30. The authors also proposed a self-supervision approach to avoid time-consuming data labeling. By investigating model-free RL to determine subsequent robot action, the need for an accurate model of process dynamics can be avoided, which is difficult to obtain for contact rich tasks. Haninger *et al.* [84] demonstrated the fusion of image and force/torque data through variational autoencoder (VAE) for peg-in-hole insertion. The VAE learns the representation of the data by minimizing the difference between each sample of the real data used for encoding and the predicted state generated from the decoder. Based on the mutual information between these states, the controllability of the model is evaluated. In addition to torque/force, Pfrommer *et al.* [201] proposed a ContactNets-based method, which learns the contact parameters including friction without contact or force sensing.

5.2.3. AI enhanced assembly of deformable objects

Handling and assembly of deformable objects traditionally can only be carried out manually whereas robot-based handling and assembly is mostly limited to rigid objects. This is due to the heterogeneous geometrical and mechanical properties in combination with non-linear dynamics that are challenging for the classical modeling and control methods [167].

The potential of AI-based approaches to identify the control parameters or learn the control policy without an explicit model, e. g. by teleoperation or kinesthetic teaching, is high, as shown in [143] for complex operations such as force controlled bimanual robotic handling operations of deformable linear objects (DOL) and weaves. In contrast to teleoperation or kinesthetic teaching, Wu *et al.* [275] introduced an approach for learning of deformable object manipulation without demonstration through model-free visual RL. In this work, the pick and place policies are learned separately to avoid challenges such as reward assignment in RL. The developed solution first trains place policy with uniformly random pick, and then selects the optimal pick point that maximizes the value function in RL. In robotics the research field of deformable object manipulation (DOM) is steadily growing with new approaches often based on ML, especially DRL but also artificial visuomotor learning [148].

In [167], Makris *et al.* gave a comprehensive overview about control methods for handling of deformable objects in assembly. The overview includes different ML based methods such as SVM and

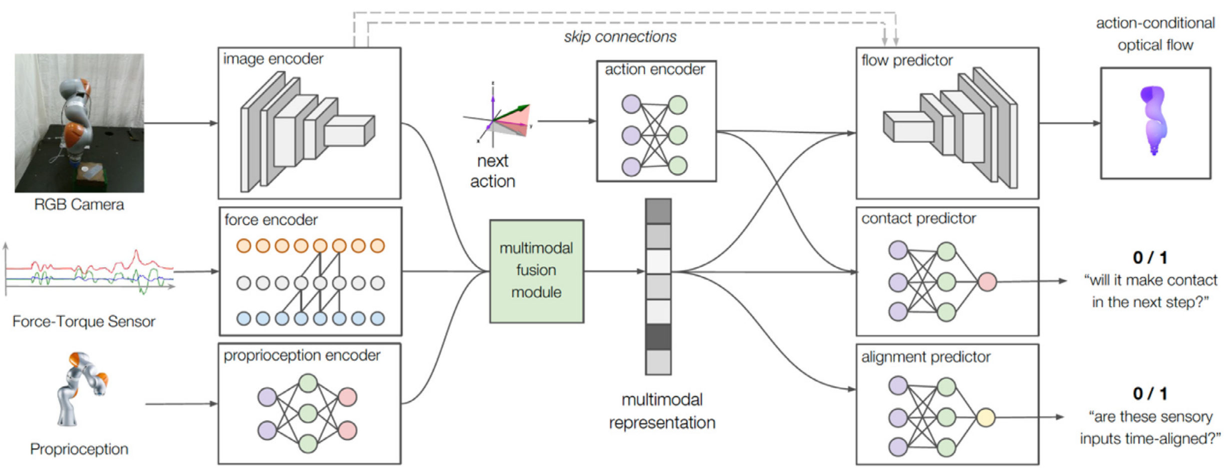

Fig. 30. Neural network architecture for self-supervised multimodal representation learning of a robotic peg in hole insertion task [144].

neural networks for deformation control, extraction of grasping poses, pose estimation of grasping garment or the estimation of deformations of soft objects with unknown mechanical properties. It also refers to RL based methods for learning of manipulation of ropes, clothes, and fluids.

5.3. AI based disassembly

Compared to the assembly of products, automated disassembly represents a greater challenge due to the often-unknown internal structure. Also, the evaluation of the form of separation of inter-component connection can be more complex than the planning and design of the connection itself. Vongbunyong *et al.* [256] introduced a cognitive robotics-based system for disassembly of LCD screens. The system is equipped with cognitive functions for reasoning, execution monitoring, learning and revision. It stores relevant information from successful disassembly processes of a product. The results show that the system is flexible enough to deal with any product models without prior information.

Physics-based simulation and search algorithms have also been investigated for disassembly planning. One example is reported in [247], where a progressive Breadth-First Search (BFS) is developed and implemented in combination with a rigid-body simulator ensuring axis-aligned torques and forces for each component. Specifically, given assembled states of all components, the method iteratively searches for an ordered sequence of disassembly paths that connect the assembled state and a disassembled state for each component subject to the precedence relationship. At each iteration, the developed method tries to disassemble each of the remaining components until all are disassembled. The authors demonstrated the effectiveness of the method by solving complex assemblies such as electric motor in the simulated evaluation.

Other researchers have been investigating computer vision for identification of components in the assembled part, which serves as the basis for disassembly. These techniques include CNN [169] and its variants such as you-only-look-once, YOLO [15] and region-based CNN (RCNN) [63,281] that have shown capable of segmenting components such as screwhead in the assembled part.

5.4. AI for symbiotic human robot collaboration

HRC is regarded as a means for increased flexibility of assembly lines and as a contribution to the change from mass production to mass customization [261]. Among various elements of HRC, the adaptability of robots to the workspace as well as human worker actions are fundamental to realize seamless HRC [132].

In recent years, AI-based approaches for enhancing robot's perception of human action have been introduced. Wang *et al.* [266] described a CNN-based approach for human action recognition, which serves as the basis for human action prediction. Specifically, the authors investigated transfer learning to resolve the limitation in manufacturing data collection. The developed method has achieved an action recognition accuracy of 96.6% in a car engine assembly scenario. As a further step towards HRC, human trajectory prediction is investigated in [288], which is critical to enable robot to deliver part/tool to the desired location to realize collaboration. Specifically, trajectory prediction is realized through an RNN. The novelty of the work includes two functional units for parsing the evolutionary pattern of human body joints and their coordination. In a car engine assembly case study, the developed method allows robots to synchronize with humans to realize pro-active part/tool pick-up and handover.

Wang *et al.* [262] presented a novel form of communication between human and robot for HRC through analysis of brainwave signals. In this approach brainwaves are first converted to time-frequency images which serve as an input to a VGG16 (a variant of CNN). The network then determines the human command through predicting the subject, predicate, and object, such as “robot assembly block” (Fig. 31). The predicted command is then translated to robot action via function blocks. The feasibility of brainwave-based robot control is demonstrated by car engine assembly. This example underlines the high potential of AI for task-oriented robot programming [134] as a future means for reduction of robot programming effort in assembly. Recent studies of HRC have been summarized by Semeraro *et al.* [225].

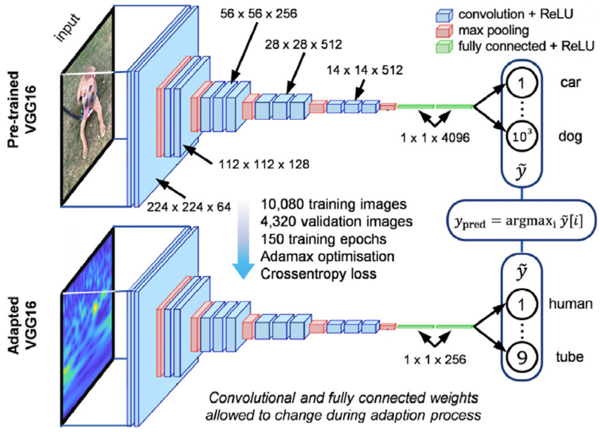

Fig. 31. VGG16 structure and adaption to time-frequency images [262].

5.5. The future of AI-based assembly and disassembly automation

The ongoing research shows a high potential of AI for the improvement of robot capabilities to support assembly and disassembly. A potential future research direction is high payload applications, as current AI-based robotic research has restricted payload limits. For these applications, further improvements of cognitive robotic skills to ensure safe interaction between human and robot are required. The steady increase of exoskeleton support systems in assembly will also benefit from the development of AI methods for improved detection of human intentions and states as well as the adaptive control of robot kinematics (Fig. 32) [137]. As a result, the adaptation of the support to the worker and the respective assembly operation can be individually controlled in the sense of a symbiotic connection between human and exoskeleton.

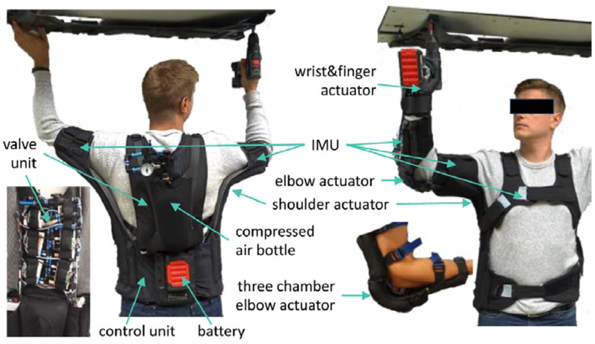

Fig. 32. Soft robotic assist system PowerGrasp. Left: dual arm system, right: single arm system with elbow and wrist/finger actuator [137].

A lot of promising research work in recent years has shown the feasibility of AI-based approaches to enable robotic object handling for assembly/disassembly. Their broad application in industrial scenarios will generally require an integration of intelligent planning methods and a high in-process adaptivity of the robot. The increase of cognitive skills of the robot such as tasks recognition and automated generation of control parameters also shows a high potential for the reduction of effort for robot programming which so far is regarded as one of the major obstacles for the application of robots.

6. Industrial case studies

The ultimate goal of AI in manufacturing is to have various AI techniques developed successfully translated into realization of smart manufacturing. To this end, representative industry case studies are presented in this chapter.

6.1. Optimization of production scheduling

In general, manufacturing is about turning raw materials into products. For manufacturers that make different types of products, each production line is often focused on producing one single type of product at any time. The production scheduling problem is therefore to determine how long each line should take one type of product before switching to another type, with the objective of fulfilling the maximum number of orders while minimizing the cost. The scheduling problem is usually constrained by factors such as the availability of production lines, pending orders, predicted incoming orders, inventory levels, and production rates of each product type. Additionally, the cost of switching is a significant factor to consider, which can encompass scenarios such as the production of off-grade products sold at reduced margins and the associated downtime of production lines.

Similar to the research efforts described in Ch. 2, Dow Chemical, which has 50+ plants worldwide with each manufacturing 20–50 products, has developed a RL-based system named AlphaDow to tackle challenges in production scheduling [118]. AlphaDow is based on the actor-critic RL framework as shown in Fig. 33. At high-level, it ingests the state of the current production schedule and determines the schedule change that is needed for the subsequent operations (such as product switching). In AlphaDow, the state consists of manufacturing data (such as production rates of each product type), customer demand (such as open orders, predicted orders), current inventory levels, as well as maintenance schedule. The objective of RL includes maximized on-time shipments, minimized inventory levels, minimized off-grade products, etc. Additionally, the technique of masking is incorporated at the RL output to suppress product switching that is incompatible with the production system.

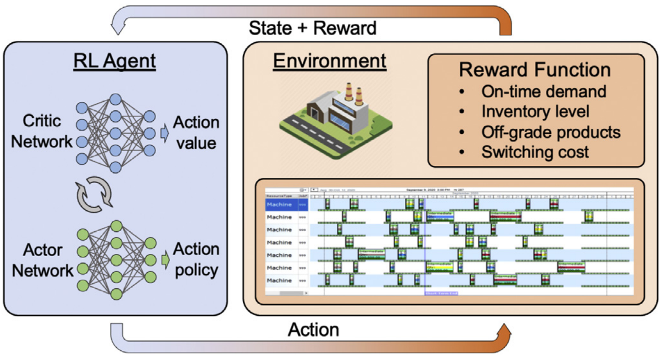

Fig. 33. Actor-critic RL for production scheduling, adapted from [118].

The training of RL is carried out using the method of population-based bandits (PB2) [199], which is based on a multi- agent approach that leverages GP-based Bayesian optimization to improve the search efficiency of optimal RL hyperparameters. The outcome has shown that the RL-based approach is able to reduce overall cost as compared to heuristic rule-based scheduling method over a period of 12 months in a simplified product scenario consisting of 5 different products. Dow Chemical expects to expand the method into all its plants in the future.

6.2. Automated chip detection and removal for machining

Cutting metals often results in chips that are prone to accumulate around the tool and workpieces, leading to degraded processing. As a result, operators are traditionally required to remove the chips regularly. The cleaning process takes away valuable machine operation time and hinders automation. DMG Mori has developed “AI chip removal”, an AI-based system for automated chip accumulation analysis and chip removal path generation as shown in Fig. 34 [50].

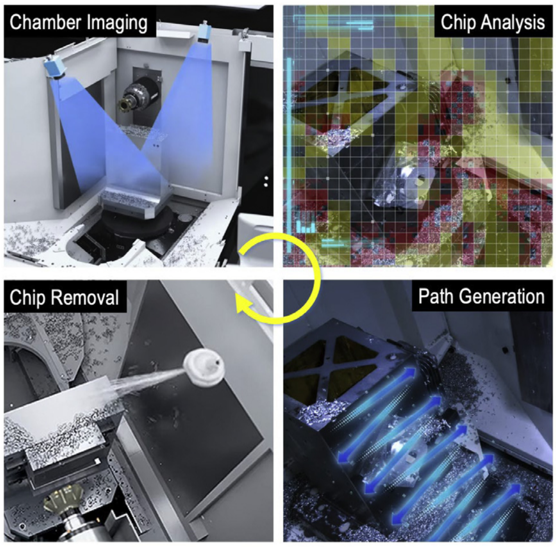

Fig. 34. “AI chip removal” system from DMG Mori [50].

The “AI chip removal” system consists of four steps. First, the machining chamber is equipped with two cameras to take high-resolution images of the entire chamber. The cameras are enhanced with water-repellent films and air blow to prevent chips and coolant from adhering that can degrade the image quality. The images are then analyzed using image recognition techniques from computer vision to evaluate the location and level of accumulation of chips, resulting in heatmaps that provide comprehensive information about chips in the machining chamber.

The heatmaps serve as the basis for cleaning path generation, which can be considered a process optimization problem similar to those described in Ch. 3. Specifically, the necessity of coolant cleaning and the amount of coolant discharging are computed based on the chip locations as well as the level of accumulation. Finally, chip removal is executed through chip flush nozzles with a wide movable range based on the determined cleaning path. Compared with the conventional fixed nozzle system that is limited when chip accumulation pattern or workpiece type changes, the “AI chip removal” system enables automatic angle adjustment that is adaptive to different accumulation patterns and workpiece types, suited for machining of high-mix products and contributing to improved operating rates of machining system.

6.3. Natural language processing for machine maintenance

Various AI techniques have been presented for machine condition monitoring, fault diagnosis and RUL prognosis in Ch. 4. These techniques are mainly based on sensor data, such as time series and images. However, this data does not contain information that is equally important for predictive maintenance, such as describing what fixes were implemented to resolve the discovered quality issues. The vast majority of such information is buried in the text data, such as maintenance records.

Compared to time series and image data, text poses unique challenges for pattern recognition. For example, text may contain typos, acronyms, abbreviations, non-standard concatenations, etc., making it especially difficult for pattern recognition. As one of the early works, researchers at Boeing have developed the PArts Name Discovery Analytics method, or PANDA, that leverages linguistics domain knowledge and natural language processing (NLP) to extract part names from maintenance records as a first step towards effective text pattern recognition [192] (Fig. 35).

The method starts by user entering a few basic “seed” part name heads, such as fan, valve, relay, and switch. This is followed by the construction of tree data structures, such that the first level nodes are the entered part heads with the descendants being the tokens that precede them in the dataset. The tree is then traversed in depth-first

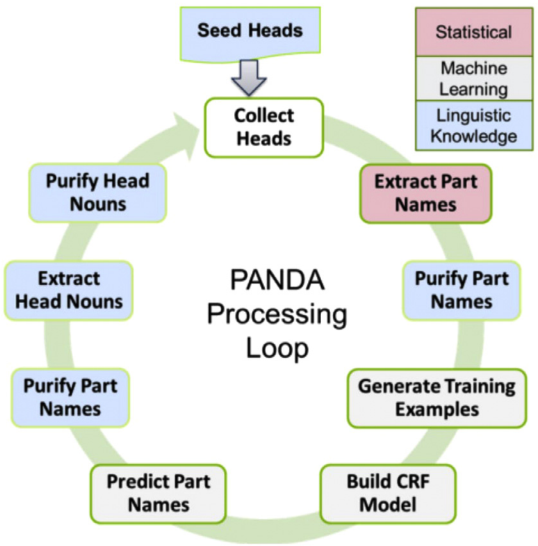

Fig. 35. PANDA method for part name extraction from noisy text [192].

manner as long as the minimum frequency criteria is satisfied, resulting in candidate part names.

These candidates are purified using a series of empirical filters before generating training samples for ML training. Rather than using the part names alone, their k-previous and k-next word tokens and their part-of-speech tags are also used as features for ML, which provide each part name a context in the maintenance record. The training samples are then used to train sequence ML models that can predict sequences of tokens, such as conditional random field (CRF) or LSTM. The part name predicted from the trained ML model will subsequently pass through a few human-in-the-loop steps for validation. In evaluation, PANDA scored an 81% accuracy for part name extraction, demonstrating its capability of analyzing text data and supporting maintenance activities.

More recently, the research of NLP has exploded with the development of generative AI and large language models (LLMs). These developments have rapidly transformed the state-of-the-art NLP and opened new possibilities to advance smart manufacturing. These new developments are described in Ch. 7.

6.4. Human activity recognition for assembly

As manufacturing is transitioning from mass production to mass customization, it is becoming increasingly difficult for workers to achieve a high level of assembly quality without making errors due to the high variety of operations that they have to carry out. While innovation in robot technologies as described in Ch. 5 can be leveraged to assist human workers, human workers are still needed in the foreseeable future to handle operations that require flexibility and dexterity, such as picking up small objects (e.g., screws) and installing them in a constrained space. The increasing variety of operations also makes it more difficult to assess worker's performance and timely detect error.

To tackle these challenges, the German Research Center for Artificial Intelligence (DFKI) and Hitachi have jointly developed AI-based technology for human activity recognition of workers through multimodal sensor fusion, as shown in Fig. 36 [45].

The primary sensors consist of an eye-tracking glass and an armband, generating gaze point and muscle activation signal as sensing data, respectively. Human action recognition is carried out by training two deep neural networks to analyze gaze point for object recognition (e.g., screw) and muscle activation signal for action recognition (e.g., twist). Additionally, this system can also assist workers in identifying correct types of parts/tools during pick up (through gaze point), assessing quality of assembly, such as tightening level of screws, by analyzing the muscle activation signal, and evaluating ergonomics by utilizing the human pose and the muscle activation

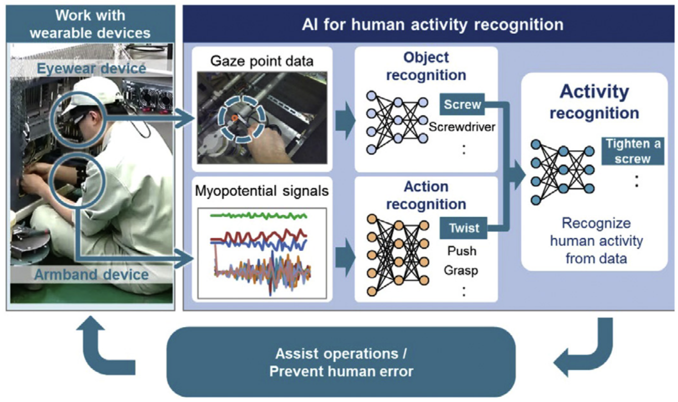

Fig. 36. Flowchart of AI-based human activity recognition [45].

data (Fig. 37). DFKI and Hitachi have demonstrated the effectiveness of the system in manufacturing operations and preventing human error in assembly.

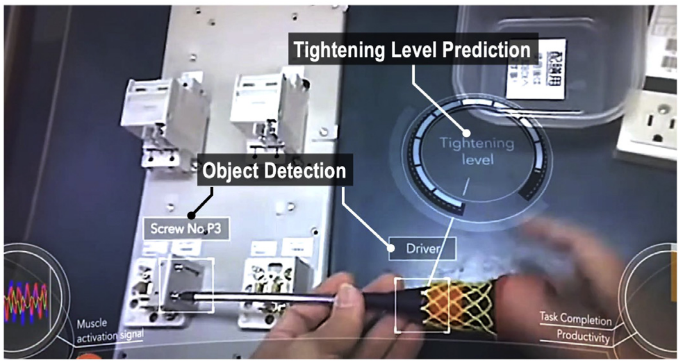

Fig. 37. Object detection and screw tightening level prediction [44].

7. Challenges and future directions

Advancement in manufacturing has continued to drive the pursuit for enhancing the performance, robustness, and trustworthiness of AI methods to transform data into actionable insights. At the same time, the continued surge in data acquired from manufacturing processes and systems has presented new challenges for researchers and practitioners to more efficiently harness the potential of AI in realizing smart manufacturing. This section outlines ten recommendations for future research.

7.1. Unsupervised learning for unlabeled data

Improving data availability is one of the central focuses in data-driven AI. However, manual labeling of manufacturing data can be a daunting task. Most of the reported research of AI in manufacturing has been focused on supervised learning. For datasets used to develop supervised learning algorithms, data labeling is either pre-determined before data collection (such as pre-seeding structural fault into machine) or carried out manually after data is collected.

To improve the flexibility and efficiency in data labeling, one of the future research topics in AI in manufacturing can be unsupervised learning. The focus is on characterizing patterns embedded in data such that these patterns can be used for a variety of tasks and in combination with supervised learning. The goal is to utilize unlabeled data to not only develop data-driven AI models, but also improve the models' capability to generalize as compared to those developed using a limited number of labeled data. Early work on unsupervised learning for model generalization has demonstrated its effectiveness, for example, for motor fault diagnosis in [215]. The authors demonstrated that the features learned from unlabeled data of two motor conditions using unsupervised learning can be well generalized to additional motor conditions, as reflected by a 9.6% higher diagnostic accuracy when compared to supervised method that is trained using a limited amount of labeled data.

7.2. Integrating physics with AI

Integrating physical laws and principles with AI has long been regarded as a crucial milestone to enable the utilization of physical knowledge and information extracted from sensor data to effectively solve manufacturing problems [80]. Promising results have been achieved in applications such as AM part property prediction and tool wear prognosis [246,259].

Nevertheless, AI architecture and physical components have largely been treated independently from each other, with the output from one side serving as the input to the other. Many AI models, particularly deep learning models, remain decoupled from physical domain knowledge. A potential solution lies in designing AI structures that possess a physically interpretable behavior, allowing better understanding and optimization of the properties of related AI models. An example of early effort in this direction is to replace the first layer in a CNN that is randomly generated with a continuous wavelet convolutional layer [153]. The resulting new layer mimics the behavior of wavelet transform, which is grounded in mathematical principles and decomposes the input signal into the time and frequency domains, thus ensuring physical interpretability of the layer output. At the same time, parameters of the wavelet layer can be continually optimized by the incoming data. To accelerate physics-AI integration and enhance applications to manufacturing, further research on AI model design that incorporates physical domain knowledge is warranted.

7.3. Embedding safety constraints

While integrating physical knowledge with data-driven AI contributes to improving the interpretability and physical consistency of data-driven methods, for AI-based decision-making that involve critical operations such as real-time control of manufacturing processes and collaborative robots, an extra layer of safety constraints is needed to avoid catastrophic outcome.

As one of the main drivers for AI-based decision-making, research using reinforcement learning has shown to be able to alleviate this limitation through algorithm training in a simulated environment before fine-tuning in real-world scenarios [52]. Still, such an approach cannot eliminate the safety concerns entirely. A promising future research direction is to combine AI-based decision-making with model predictive control (MPC). MPC has seen significant success in recent decades and has established itself as the primary method for the systematic handling of safety constraints [88]. By combining the constraint satisfaction capability of MPC with the modeling capability of data-driven AI, integrated models can be developed as shown in Fig. 38. An early work is the development of MPC-based safety filter [258]. The main idea is to solve an optimization problem to find the safety-compliant control signal that is the closest to the output from the data-driven RL algorithms. This method is suited for any data-driven decision-making algorithms and has been validated in the research of self-driving vehicles.

7.4. Controlling false discovery rate and causal AI

One of the most promising aspects of AI for manufacturing is the discovery of new knowledge. For example, for complex processes such as AM [122] and semiconductor manufacturing [264], it is essential to be able to screen a large number of process parameters and determine which ones are the most influential, in order to efficiently achieve process optimization.

The primary challenge for AI-based scientific discovery is to ensure low false discovery rate (FDR), as high FDR can result in significant waste of effort in the subsequent confirmatory study.

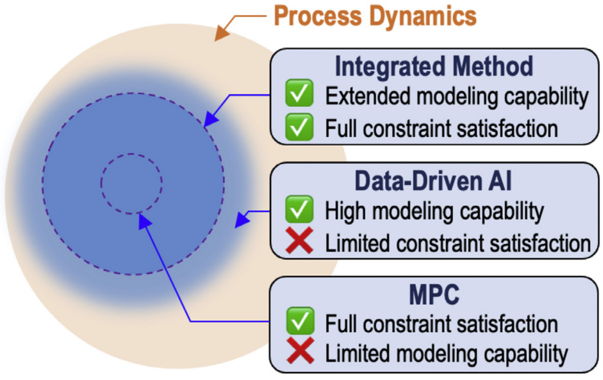

Fig. 38. Integrated method combining constraint satisfaction from MPC and modeling capability of data-driven AI, adapted from [88].

Common post-analysis techniques such as SHapley Additive exPlanations (SHAP), while effective in quantifying the influence of process parameters on the part property, are not able to distinguish correlation from causation [74,164]. Various research efforts have been made to tackle this challenge. For example, knockoff filter [13] has been developed to control FDR for data-driven AI. The main idea of knockoff filter is to construct dummy parameters that are designed to mimic the structure found within the existing parameters while exhibiting feature importance statistics in a way to allow accurate FDR control. Another promising branch of research is the emerging causal ML, which aims to combine causal inference framework with data-driven methods to ensure validity of findings [165]. These studies should motivate manufacturing researchers to develop more rigorous and reliable procedures for continued scientific discovery using AI.

7.5. Extending NLP with embedding and transformer

While NLP has been investigated for manufacturing as described in Ch. 6.3, the process of AI algorithm development generally lacks systematic guidance and often requires empirical design. Handling text has long been a challenge to AI due to the limitation in transforming words into computable elements while retaining the semantic structure. Breakthrough came in the 2010s through embedding [174] and the transformer architecture [254].

The technique of embedding is characterized by the mapping of words to their representations in a high-dimensional space [174]. To establish domain-specific embedding, a key step involves training of the mapping to maximize the consistency between an individual word and its existing manufacturing context while minimizing the consistency with non-existing contexts. This enables the implicit encoding of word semantics. The result is that semantically similar words exhibit similar representations.

Originally designed for language translation, the structure of transformer features a set of self-attention that allows to capture and quantify the association between words in the translated sentence and the original sentence [254]. With sentences encoded using embedding, transformer enables massive parallel computing to achieve state-of-the-art performance in translation.

One promising future research direction for AI in manufacturing is to utilize transformer (and its variant) as a backbone model on top of which specific text-based analysis module can be extended. Early progress includes information parsing from maintenance logs [33], which confirms the potential of transformer as a powerful tool to support the development of language-based AI methods and applications.

7.6. Learning from human demonstration

In the era of Industry 5.0, human workers will be back in the spotlight as the concept of "human-in-the-loop" acknowledges the crucial role that human expertise plays in manufacturing settings [186,278].

The emphasis on leveraging the full potential of human expertise provides an opportunity to develop skill transfer from human demonstration to AI algorithms, with potential benefits including rapid reprogramming of robot to learn new skills without having to rely on large-scale data collections [149].

One of the promising directions is imitation learning, which enables a robot to acquire the ability in fine manipulation in assembly by learning from human demonstration via teleoperation [294]. To reduce the error compounding effect due to the difference between the training and testing scenarios, the algorithm learns to implicitly associate sensing images to the robot actions on an average basis rather than aiming at reproducing actions at each instant. Early results have shown that imitation learning through teleoperation can achieve human-like capability for object manipulation and assembly after only 10-min human demonstration. In a follow-up work, the authors further installed the manipulator on a mobile platform, drastically increasing the capability of this setup [66]. Teleoperation has also been integrated with virtual reality (VR) to enable more flexible human demonstration [291]. Future research of imitation learning in manufacturing can also be focused on extending its application beyond robotics to realize skill transfer from human to AI. Additionally, imitation learning can be integrated with curriculum learning, which brings in external expertise into the learning process by appropriate tasks sequencing and generation, and transferring skill or knowledge learned for tasks with increased complexity [188].

7.7. Adopting generative AI

Generative AI itself has been around in manufacturing for a few years as exemplified by research on GAN and its variants. However, the capability to control the generated data at a more granular level has always been challenging. The recent development of diffusion models aims to tackle this challenge.

The diffusion model consists of the diffusion process and a reverse process. The diffusion process progressively adds random noise to the data, and the reverse process, usually built on neural networks, learns to progressively remove the noise [163]. The method was originally developed for image synthesis where a sequence of well-trained denoisers represent a mapping from a known distribution to the distribution of images. New images can then be synthesized by first sampling from the known distribution, before passing through the denoiser sequence.

The diffusion model becomes significantly more impactful once it is combined with language-based instruction that can be used to tailor the synthesized data [235], with the training process of the denoiser including an additional input that is the language-based instruction, which is usually in the form of an image caption. Once the denoiser is trained, the diffusion model can synthesize realistic images that are highly customized to the user instruction.

For manufacturing, generative AI can have broad applications such as design and optimization of materials and processes. One promising approach is self-supervised learning enabled by physics-based simulation. The idea is to attach the simulation by ingesting the output of the generative AI (e.g., design parameters) and verify its effectiveness (e.g., whether desired property is achieved). Since generative AI can take the desired property as the input, the deviation between the output of the simulation and the input of the generative AI constitutes a self-supervised circle to guide the improvement of generative AI. Such ideas have been explored for design optimization of material microstructure [255] and robotic gripper [82]. Research on integrating language-based instruction into the self-supervised learning framework is expected to further enhance the utility of generative AI in design and optimization in manufacturing.

7.8. From specialist to generalist model

In 2022, ChatGPT, an LLM from OpenAI, took the world by storm [20,200]. At the heart of its popularity is what can be considered the transition from the traditional specialist AI model to a new generation of generalist model. For a specialist model, the task itself is implicitly determined when training data is collected [2]. Each model is unique to the task it is trained on and does not have capability to carry out new tasks. By contrast, a generalist model is trained with task-related instruction and therefore can perform various tasks [273]. However, building generalist LLMs for manufacturing can be challenging on several fronts:

(1) Data quantity: A generalist model is currently considered possible only when building on top of a foundation model that is pre-trained on massive data in an unsupervised manner. However, such quantity of manufacturing data is yet to be acquired.
(2) Pre-training: Even though ChatGPT is pre-trained with unlabeled data, its "Q&A" format allows formulation as predicting the next word in the sentence and thereby, minimizing the requirement of data labeling. Such formulation is, however, difficult to translate to manufacturing, as manufacturing data generally does not contain the desired output in themselves.
(3) Fine-tuning: The pre-trained model is also fine-tuned with human feedback. For example, tens of thousands of instructions are manually tuned and tens of thousands of ChatGPT's answers are evaluated by human experts to encourage more natural ones [196]. The requirement of such scale of human feedback can be beyond the capacity of any individual manufacturers.

Despite these challenges, building upon the existing generalist models has shown to be highly performant in manufacturing applications such as robotics and maintenance. One early example is shown in Fig. 39, where multi-modal instruction is utilized to control the robot to carry out distinct tasks by fine-tuning a T5 LLM [106]. More recently, researchers have found high efficiency of fine-tuning ChatGPT using manufacturing domain knowledge (such as ontology) to improve LLM's understanding of hierarchical structure of machine components, leading to accurate responses when it comes to determining components of interest and suggesting resolution based on the issues presented in the maintenance log [265]. This example illustrates that LLMs can be adapted to specific manufacturing problems by infusing domain-specific information into these models and using them in the workflows where conversational problem-solving is appropriate. Such cases could be the specification of system design or planning problems, the diagnosis and maintenance of manufacturing equipment, or supporting teamwork in HRC in assembly.

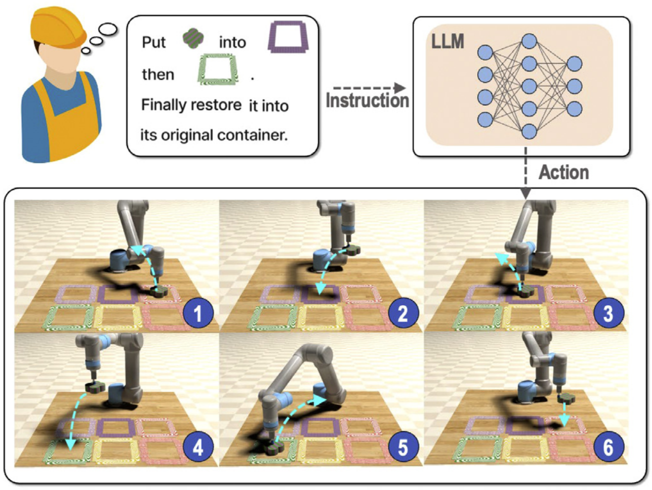

Fig. 39. Multi-modal instruction for robotic action, 1–6 illustrate actions carried out by robot to fulfill instruction [106].

Despite the initial success, LLMs' capability in approximate information retrieval is often (mis)taken in having abilities of reasoning and planning. However, LLMs are essentially language models whereas manufacturing applications such as planning requires world models [252]. As a result, synergistic integration of LLMs and world models would open a potential path to solving general planning problems. Additionally, as LLM models become larger, training often requires high computational power, leaving behind a large computational carbon footprint [224]. As a result, research should also consider efficiency and energy consumption as model evaluation metrics. To this end, researchers have started developing model evaluation metric that takes into account the computational footprint [32]. As the emphasis on sustainability is essential for the future of manufacturing, research on model energy efficiency in real-world applications should be encouraged.

7.9. Developing AI-specific hardware

While most of the AI research in manufacturing has been solely focused on software (i.e., algorithms), increasing demand for sensing, data transmission, and processing speed and accuracy in AI implementation has brought hardware limitations into the spotlight, including generic design of sensors and processing units that are not optimized for AI algorithms. Dedicated AI hardware, in particular Lisp machines, have provided a major boost to AI research before the beginning of the second AI winter (see Fig. 1). Recognizing the state of AI research, it is envisioned that advancing the state of design of specific hardware for a new generation of AI algorithms can again lead to tremendous benefits for AI in manufacturing. Three promising directions are described as recommendations:

AI-enhanced metrology. In manufacturing, often the sensor is not able to directly measure the variables of interest and only indirect measurement can be made. Examples include measuring subsurface structure using electrical capacitance tomography [56], where only capacitance between different nodes is obtained, and angle measurement is based on angle-dependent second harmonic generation (SHG) spectrum [152]. In these cases, obtaining variables of interest requires solving an inverse problem which is commonly ill-posed. The advancement of AI methods provides a new way of resolving this issue by establishing complex inverse mapping using neural networks. Early work [152] has shown that neural networks can solve the inverse problem in angle measurement based on the SHG spectrum to achieve a sub-arcsecond level of accuracy and resolution.

Sensing-AI codesign. The idea of codesign is to formulate sensing mechanism as an optimization problem such that it can be optimized in an end-to-end fashion together with the AI algorithm for manufacturing tasks. This ensures that only the information that is most relevant to the tasks is captured. While similar to AI-enhanced metrology, the uniqueness of codesign is that sensing mechanism can be tailored for each application. Sensing-AI codesign has already been progressing in the medical field, where optimal MRI strategy is needed to minimize scanning time while restoring images with the highest quality. By exploring codesign, an 8x reduction of scanning time has been achieved with minimal image quality degradation [10]. The codesign approach can be extended to sensor-rich manufacturing environment to further facilitate optimization of AI methods.

Hardware-level AI computation. Modern AI algorithms such as those based on deep neural networks require significant computation even for inference. While advances in GPU have significantly improved the computational efficiency, many applications in manufacturing still struggle with AI computation using standard GPU, such as real-time control in AM. To resolve this limitation, researchers have started exploring hardware-level implementation of AI computation. One example is to replace a digital convolution layer in CNN with optical convolution by designing specific optical elements and light pathway to mimic the convolution operation with images [30,97] (Fig. 40). The result is an orders-of-magnitude reduction of computational time. By synergistically integrating hardware design with an AI algorithm, the development of AI-specific hardware is expected to significantly improve the efficiency and scalability of AI in manufacturing.

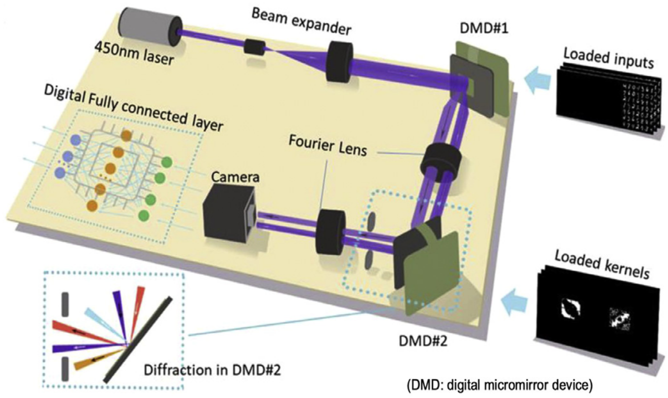

Fig. 40. Setup for optical convolution operation [97].

7.10. Real-life introduction of AI

Despite AI support, many industrial projects in production have yet to progress beyond the stage of prototype building. The lack of adoption of AI techniques developed in the research community can be attributed to various factors beyond technological and engineering advancements. Successful integration of these advanced techniques into industry requires appropriate scope setting and problem statement, change management, development of a credible business case, and last but not least, gaining the trust of potential users.

Often, there could be unrealistic expectations for AI-enhanced projects arising from a lack of understanding of the constraints of AI and/or a risk-averse, conservative position against adopting new technologies that is generally perceived as black-box. In fact, it is often difficult to forecast precisely what to expect from an AI application. Gaining industry's trust is essential. Towards this end, end users need to be provided with interpretable, physically trustworthy predictions when AI is involved, and at the same time, an option for humans to remain in control. They should also be able to change solutions without violating the underlying constraints. If requested, explanations need to be generated along with the solutions.

All this poses strict requirements towards interactive user interfaces with a reasonable response time that is consistent with the expectation for real-world operations. For domain experts on the factory floor, modelling limitations and assumptions underlying the AI algorithms need to be clearly laid out and readily accessible for fine-tuning. Based on the idea that even though "all models are wrong, but some are useful" [24], declarative AI approaches that rest upon explicit assumptions have an advantage over pragmatic purely data-driven methods in this respect. XAI attempts to open the "black box" models generated by data-driven techniques and make them amenable to human interpretation and comprehension. XAI methods characterize model accuracy, fairness and transparency, thereby promoting trust in an AI system. XAI provides an answer to concerns about the legal, security and compliance risks of using AI in an industrial environment. This also facilitates making the distinction between real domain constraints and inveterate past practices which should be rather dispensed of. For management, it is essential to build and maintain trust via, if possible, public success stories, and new business models which mitigate risks and enable sharing of benefits.

8. Conclusions

Artificial intelligence is destined to play a pivotal role in redefining the manufacturing landscape. This transformative shift will be driven by key technologies, with the goal to enhance: (1) production system design and planning, (2) process modeling, management, and optimization, (3) quality assurance and maintenance, and (4) automated assembly and disassembly. This keynote has offered a comprehensive overview of the current state-of-the-art of AI in manufacturing, illuminating its manufacturing-specific life cycle, from the initial design through process management to quality maintenance and automation.

Additionally, AI's place in manufacturing has been exemplified through the exploration of industrial case studies, which showcase practical implementations and benefits of AI in automation of machining operation with computer vision, optimization of maintenance with natural language processing, and human action recognition for process monitoring and inspection. Furthermore, specific challenges arising from issues such as data validity in real-world applications, the need for further integrating physics with AI methods, real-life concerns of AI including transparency, interpretability, and trustworthiness, the necessity for the adoption of next-generation generative AI with granular control, and the development of generalist AI promoting natural interaction with users are outlined, with the proposed future directions intending to provide a roadmap for researchers and practitioners.

With the rapidly evolving AI landscape, novel methodologies and tools continue to emerge. Embracing these advancements will not only allow manufacturers to harness the power of AI by capitalizing on the wealth of information available in a data-rich environment but also enable deeper understanding of the mechanisms underlying manufacturing processes and systems to ultimately advance the science base for smart manufacturing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Robert X. Gao: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Jörg Krüger:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Marion Merklein:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing, Resources. **Hans-Christian Möhring:** Conceptualization, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing, Project administration. **József Váncza:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgements

R. Gao acknowledges support from the National Science Foundation under grants CMMI-2040288, CNS-2125460, and EEC-2133630 (Engineering Research Center, HAMMER). J. Váncza thanks the National Research, Development, and Innovation Office for the grant TKP2021-NKTA-01. Authors are grateful for Dr. Jianjing Zhang, Department of Mechanical and Aerospace Engineering, Case Western Reserve University, for his valuable input and assistance. Support from Dr. Zsolt Kemény of Hungarian Research Network, Raphaela März, Michael Lechner and Simon Wituschek of Friedrich-Alexander University of Erlangen-Nürnberg, is also sincerely appreciated.

References

[1] Aarts EH, Lenstra JK (2003) *Local Search in Combinatorial Optimization*, Princeton University Press.
[2] Addepalli S, Weyde T, Namoano B, Oyedeji OA, Wang T, Erkoyuncu JA, Roy R (2023) Automation of Knowledge Extraction for Degradation Analysis. *CIRP Annals* 72(1):33–36.
[3] Ahmad M, Ferrer BR, Ahmad B, Vera D, Lastra JL, Harrison R (2018) Knowledge-Based PPR Modelling for Assembly Automation. *CIRP Journal of Manufacturing Science and Technology* 21:33–46.
[4] Alan FJ (1983) *Modularity of Mind: An Essay On Faculty Psychology*, MIT Press.
[5] Amaitik SM, Kiliç SE (2007) An Intelligent Process Planning System for Prismatic Parts Using STEP Features. *The International Journal of Advanced Manufacturing Technology* 31:978–993.
[6] Ameri F, Sormaz D, Psarommatis F, Kiritsis D (2022) Industrial Ontologies for Interoperability in Agile and Resilient Manufacturing. *International Journal of Production Research* 60(2):420–441.
[7] Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing* 50(2):174–188.
[8] Asadi F, Olleak A, Yi J, Guo Y (2021) Gaussian Process (GP)-Based Learning Control of Selective Laser Melting Process. *American Control Conference (ACC)*, 508–513.
[9] Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W (2015) On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PloS One* 10(7):e0130140.
[10] Bahadir CD, Wang AQ, Dalca AV, Sabuncu MR (2020) Deep-Learning-Based Optimization of the Under-Sampling Pattern in MRI. *IEEE Transactions on Computational Imaging* 6:1139–1152.
[11] Bahdanau D, Cho KH, Bengio Y (2014) Neural Machine Translation by Jointly Learning to Align and Translate. *International Conference on Learning Representations*, 1–15.
[12] Bak C, Roy AG, Son H (2021) Quality Prediction for Aluminum Diecasting Process Based on Shallow Neural Network and Data Feature Selection Technique. *CIRP Journal of Manufacturing Science and Technology* 33:327–338.
[13] Barber RF, Candès EJ (2015) Controlling the False Discovery Rate Via Knockoffs. *The Annals of Statistics* 43(5):2055–2085.
[14] Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding* 110(3):346–359.
[15] Bdiwi M, Rashid A, Putz M (2016) Autonomous Disassembly of Electric Vehicle Motors Based on Robot Cognition. *IEEE International Conference on Robotics and Automation (ICRA)*, 2500–2505.
[16] Bellman R (1954) The Theory of Dynamic Programming. *Bulletin of the American Mathematical Society* 60(6):503–515.
[17] Bergs T, Biermann D, Erkorkmaz K, M'Saoubi R (2023) Digital Twins for Cutting Processes. *CIRP Annals* 72(2):541–567.
[18] Bharathi Raja S, Baskar N (2011) Particle Swarm Optimization Technique for Determining Optimal Machining Parameters of Different Work Piece Materials in Turning Operation. *The International Journal of Advanced Manufacturing Technology* 54:445–463.
[19] Bishop CM (2006) *Pattern Recognition and Machine Learning*, Springer.
[20] Biswas SS (2023) Potential Use of ChatGPT in Global Warming. *Annals of Biomedical Engineering* 51(6):1126–1127.
[21] Bobka P, Gabriel F, Dröder K (2020) Fast and Precise Pick and Place Stacking of Limp Fuel Cell Components Supported by Artificial Neural Networks. *CIRP Annals* 69(1):1–4.
[22] Bonissone PP, Johnson HE (1984) *DELTA: An Expert System For Diesel Electric Locomotive Repair*, General Electric Corporate Research and Development.
[23] Borenstein J, Koren Y (1991) Histogramic in-Motion Mapping for Mobile Robot Obstacle Avoidance. *IEEE Transactions on Robotics and Automation* 7(4):535–539.
[24] Box GEP (1979) Robustness in the Strategy of Scientific Model Building. *Robustness in Statistics* : 201–236.
[25] Brecher C, Lohse W (2013) Evaluation of Toolpath Quality: User-Assisted CAM for Complex Milling Processes. *CIRP Journal of Manufacturing Science and Technology* 6(4):233–245.
[26] Breiman L (2001) Random Forests. *Machine learning* 45:5–32.
[27] Brillinger M, Wuwer M, Hadi MA, Haas F (2021) Energy Prediction for CNC Machining with Machine Learning. *CIRP Journal of Manufacturing Science and Technology* 35:715–723.
[28] Cao Q, Beden S, Beckmann A (2022) A Core Reference Ontology for Steelmaking Process Knowledge Modelling and Information Management. *Computers in Industry* 135:103574.
[29] Cardon D, Cointet JP, Mazières A, Libbrecht E (2018) Neurons Spike Back. *Réseaux* 211(5):173–220.
[30] Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G (2018) Hybrid Optical-Electronic Convolutional Neural Networks with Optimized Diffractive Optics for Image Classification. *Scientific Reports* 8(1):12324.
[31] Chaturvedi DK (2008) *Soft Computing Techniques and Its Applications in Electrical Engineering*, Springer.
[32] Chavan V, Koch P, Schlüter M, Briese C (2023) Towards Realistic Evaluation of Industrial Continual Learning Scenarios with an Emphasis on Energy Consumption and Computational Footprint. *IEEE/CVF International Conference on Computer Vision*, 11506–11518.
[33] Chen T, Zhu J, Zeng Z, Jia X (2021) Compressor Fault Diagnosis Knowledge: A Benchmark Dataset for Knowledge Extraction from Maintenance Log Sheets Based on Sequence Labeling. *IEEE Access* 9:59394–59405.
[34] Chen Y, Yi H, Liao C, Huang P, Chen Q (2021) Visual measurement of milling surface roughness based on Xception model with convolutional neural network. *Measurement* 186:110217.
[35] Choi GH, Lee IK, Chang N, Kim SG (1994) Optimization of Process Parameters of Injection Molding with Neural Network Application in a Process Simulation Environment. *CIRP Annals* 43(1):449–452.
[36] Chryssolouris G (2006) *Manufacturing Systems: Theory and Practice*, Springer.
[37] Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical Evaluation of Gated Recurrent Networks on Sequence Modeling. *NIPS Workshop on Deep Learning* .
[38] Cooper C, Zhang J, Gao R, Wang P, Ragai I (2020) Anomaly Detection in Milling Tools Using Acoustic Signals and Generative Adversarial Networks. *Procedia Manufacturing* 48:372–378.
[39] Cortes C, Vapnik V (1995) Support-Vector Networks. *Machine Learning* 20 (3):273–297.
[40] Csáji BC, Monostori L (2008) Adaptive Stochastic Resource Control: A Machine Learning Approach. *Journal of Artificial Intelligence Research* 32:453–486.

[41] Da Col G, Teppan EC (2022) Industrial-Size Job Shop Scheduling with Constraint Programming. *Operations Research Perspectives* 9:100249.
[42] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: A Large-Scale Hierarchical Image Database. *IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.
[43] DeSouza GN, Kak AC (2002) Vision for Mobile Robot Navigation: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(2):237–267.
[44] DFKI. *AI Technology for Human Activity Recognition of Workers Using Wearable Devices*, https://www.youtube.com/watch?v=HBfZyA7w9UA Accessed 15 March 2023.
[45] DFKI and Hitachi. *DFKI and Hitachi Jointly Develop AI Technology for Human Activity Recognition of Workers Using Wearable Devices*, https://www.hitachi.com/New/cnews/month/2017/03/170308.pdf Accessed 15 March 2023.
[46] Dietz JLG (2006) *Enterprise Ontology: Theory and Methodology*, Springer.
[47] Ding L, Yue Y, Ahmet K, Jackson M, Parkin R (2005) Global Optimization of a Feature-Based Process Sequence Using GA and ANN Techniques. *International Journal of Production Research* 43(15):3247–3272.
[48] Ding X, He Q (2017) Energy-Fluctuated Multiscale Feature Learning with Deep Convnet for Intelligent Spindle Bearing Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement* 66(8):1926–1935.
[49] Dini G, Failli F, Lazzerini B, Marcelloni F (1999) Generation of Optimized Assembly Sequences Using Genetic Algorithms. *CIRP Annals* 48(1):17–20.
[50] Mori DMG (2020) *AI Chip Removal Developed for Automatic Removal of Chips*, https://www.dmgmori.co.jp/corporate/en/news/pdf/20201026_aichip_e.pdf accessed on March 15, 2023.
[51] Dolgui A, Ivanov D, Sethi SP, Sokolov B (2019) Scheduling in Production, Supply Chain and Industry 4.0 Systems by Optimal Control: Fundamentals, State-of-the-Art and Applications. *International Journal of Production Research* 57(2):411–432.
[52] Dornheim J, Link N, Gumbsch P (2020) Model-Free Adaptive Optimal Control of Episodic Fixed-Horizon Manufacturing Processes Using Reinforcement Learning. *International Journal of Control, Automation and Systems* 18(6):1593–1604.
[53] ElMaraghy H, ElMaraghy W (2022) Adaptive Cognitive Manufacturing System (ACMS) – A New Paradigm. *International Journal of Production Research* 60 (24):7436–7449.
[54] ElMaraghy H, Monostori L, Schuh G, ElMaraghy W (2021) Evolution and Future of Manufacturing Systems. *CIRP Annals* 70(2):635–658.
[55] Epureanu BI, Li X, Nassehi A, Koren Y (2020) Self-Repair of Smart Manufacturing Systems by Deep Reinforcement Learning. *CIRP Annals* 69(1):421–424.
[56] Fan Z, Gao R (2011) Enhancement of Measurement Efficiency for Electrical Capacitance Tomography. *IEEE Transactions on Instrumentation and Measurement* 60(5):1699–1708.
[57] Fang W, Guo Y, Liao W, Ramani K, Huang S (2019) Big Data Driven Jobs Remaining Time Prediction in Discrete Manufacturing System: A Deep Learning-Based Approach. *International Journal of Production Research* 58(9):2751–2766.
[58] Farabet C, Couprie C, Najman L, LeCun Y (2012) Learning Hierarchical Features for Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1915–1929.
[59] Feigenbaum EA (2003) Some Challenges and Grand Challenges for Computational Intelligence. *Journal of the ACM* 50(1):32–40.
[60] Feng Q, Maier W, Stehle T, Möhring HC (2021) Optimization of a Clamping Concept Based on Machine Learning. *Production Engineering* 16(1):9–22.
[61] Feng Z, Liang M, Chu F (2013) Recent Advances in Time–Frequency Analysis Methods for Machinery Fault Diagnosis: A Review with Application Examples. *Mechanical Systems and Signal Processing* 38(1):165–205.
[62] Finn C, Yu T, Zhang T, Abbeel P, Levine S (2017) One-Shot Visual Imitation Learning Via Meta-Learning. *Conference on Robot Learning*, 357–368.
[63] Foo G, Kara S, Pagnucco M (2021) Screw Detection for Disassembly of Electronic Waste Using Reasoning and Re-Training of a Deep Learning Model. *Procedia CIRP* 98:666–671.
[64] Framinan J, Leisten R, Ruiz R (2014) *Manufacturing Scheduling Systems: An integrated View On models, Methods and Tools*, Springer.
[65] Freitag M, Hildebrandt T (2016) Automatic Design of Scheduling Rules for Complex Manufacturing Systems by Multi-Objective Simulation-Based Optimization. *CIRP Annals* 65(1):433–436.
[66] Fu Z, Zhao TZ, Finn C (2024) *Mobile Aloha: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation*, .
[67] Gabel T, Riedmiller M (2012) Distributed Policy Search Reinforcement Learning For Job-Shop Scheduling Tasks. *International Journal of Production Research* 50 (1):41–61.
[68] Gabriel F, Roemer M, Bobka P, Droeder K (2021) Model-Based Grasp Planning for Energy-Efficient Vacuum-Based Handling. *CIRP Annals* 70(1):1–4.
[69] Gal Y, Ghahramani Z (2016) Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *International Conference on Machine Learning* : 1050–1059.
[70] Gao Q, Duan C, Fan H, Meng Q (2008) Rotating Machine Fault Diagnosis Using Empirical Mode Decomposition. *Mechanical Systems and Signal Processing* 22 (5):1072–1081.
[71] Gao R, Wang L, Helu M, Teti R (2020) Big Data Analytics for Smart Factories of the Future. *CIRP Annals* 69(2):668–692.
[72] Gao R, Wang L, Teti R, Dornfeld D, Kumara S, Mori M, Helu M (2015) Cloud-enabled prognosis for manufacturing. *CIRP Annals* 64(2):749–772.
[73] Gaschnig J (1982) Prospector: an Expert System For Mineral Exploration. *Introductory Readings in Expert Systems (D. Michie ed.)*, Gordon and Breach Science Publishers, 47–64.
[74] Gawade V, Zhang B, Guo Y (2023) Explainable AI for Layer-Wise Emission Prediction in Laser Fusion. *CIRP Annals* 72(1):437–440.
[75] Geng D, He H, Lan X, Liu C (2022) Bearing Fault Diagnosis based on Improved Federated Learning Algorithm. *Computing* 104(1):1–19.
[76] Glorot X, Bordes A, Bengio Y (2011) Deep Sparse Rectifier Neural Networks. *International Conference on Artificial Intelligence and Statistics* : 315–323.
[77] Gödri I, Kardos C, Pfeiffer A, Váncza J (2019) Data Analytics-Based Decision Support Workflow for High-Mix Low-Volume Production Systems. *CIRP Annals* 68 (1):471–474.
[78] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 27:2672–2680.
[79] Grezmak J, Zhang J, Wang P, Loparo KA, Gao R (2019) Interpretable Convolutional Neural Network Through Layer-Wise Relevance Propagation for Machine Fault Diagnosis. *IEEE Sensors Journal* 20(6):3172–3181.
[80] Guo S, Agarwal M, Cooper C, Tian Q, Gao R, Grace WG, Guo YB (2022) Machine Learning for Metal Additive Manufacturing: Towards a Physics-Informed Data-Driven Paradigm. *Journal of Manufacturing Systems* 62:145–163.
[81] Gyulai D, Pfeiffer A, Nick G, Gallina V, Sihn W, Monostori L (2018) Lead Time Prediction in a Flow-Shop Environment with Analytical and Machine Learning Approaches. *IFAC-PapersOnLine* 51(11):1029–1034.
[82] Ha H, Agrawal S, Song S (2021) Fit2Form: 3D Generative Model for Robot Gripper form Design. *Conference on Robot Learning* : 176–187.
[83] Han BA, Yang JJ (2020) Research on Adaptive Job Shop Scheduling Problems Based on Dueling Double DQN. *IEEE Access* 8:186474–186495.
[84] Haninger K, Garcia RV, Krüger J (2020) *Towards Learning Controllable Representations of Physical Systems*, .
[85] Hassanin AA, Abd El-Samie FE, El Banby GM (2019) A real-time approach for automatic defect detection from PCBs based on SURF features and morphological operations. *Multimedia Tools and Applications* 78:34437–34457.
[86] Hauschild MZ, Kara S, Røpke I (2020) Absolute Sustainability: Challenges to Life Cycle Engineering. *CIRP Annals* 69(2):533–553.
[87] Herbert M, Zwingel M, Czapka C, Franke J (2022) A Multi-Source Localization System for Driverless Material Transport in Mixed Indoor and Outdoor Areas. *Congress of the German Academic Association for Production Technology (WGP)* : 421–429.
[88] Hewing L, Wabersich KP, Menner M, Zeilinger MN (2020) Learning-Based Model Predictive Control: Toward Safe Learning in Control. *Annual Review of Control, Robotics, and Autonomous Systems* 3:269–296.
[89] High P (1997) *Carnegie Mellon Dean of Computer Science on the Future of AI*, https://www.forbes.com/sites/peterhigh/2017/10/30/carnegie-mellon-dean-of-computer-science-on-the-future-of-ai/?sh=34e570832197 Accessed 15 May 2021.
[90] Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine* 29(6):82–97.
[91] Ho J, Jain A, Abbeel P (2020) Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems* 33:6840–6851.
[92] Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Computation* 9(8):1735–1780.
[93] Hofmann T, Schölkopf B, Smola AJ (2008) Kernel Methods in Machine Learning. *The Annals of Statistics* 36(3):1171–1220.
[94] Hogan A, Blomqvist E, Cochez M, d'Amato C, Melo GD, Gutierrez C, Kirrane S, Gayo JE, Navigli R, Neumaier S, Ngomo AC (2021) Knowledge Graphs. *ACM Computing Surveys* 54(4):1–37.
[95] Holland JH (1992) *Adaptation in Natural and Artificial systems: An introductory Analysis With Applications to biology, control, and Artificial Intelligence*, MIT Press.
[96] Hornik K, Stinchcombe M, White H (1989) Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2(5):359–366.
[97] Hu Z, Li S, Schwartz RL, Solyanik-Gorgone M, Miscuglio M, Gupta P, Sorger VJ (2022) High-Throughput Multichannel Parallelized Diffraction Convolutional Neural Network Accelerator. *Laser & Photonics Reviews* 16(12):2200213.
[98] Huang J, Zhang J, Chang Q, Gao R (2021) Integrated Process-System Modelling and Control Through Graph Neural Network and Reinforcement Learning. *CIRP Annals* 70(1):377–380.
[99] Humfeld KD, Gu D, Butler GA, Nelson K, Zobeiry N (2021) A Machine Learning Framework for Real-Time Inverse Modeling and Multi-Objective Process Optimization of Composites for Active Manufacturing Control. *Composites Part B: Engineering* 223:109150.
[100] Ibrahim R, Shafiq MO (2023) Explainable Convolutional Neural Networks: A Taxonomy, Review, and Future Directions. *ACM Computing Surveys* 55(10):1–37.
[101] IEEE Robotics and Automation Society. IEEE Standard Ontologies for Robotics and Automation. *IEEE Standard* 1872:1–60.
[102] International Organization for Standardization. *ISO 17359: Condition Monitoring and Diagnostics of Machines – General Guidelines*, .
[103] Jaensch F, Csiszar A, Scheifele C, Verl A (2018) Digital Twins of Manufacturing Systems as a Base for Machine Learning. *International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, 1–6.
[104] Jaremenko C, Ravikumar N, Affronti E, Merklein M, Maier A (2019) Determination of Forming Limits in Sheet Metal Forming Using Deep Learning. *Materials* 12(7):1051.
[105] Jiang X, Senin N, Scott PJ, Blateyron F (2021) Feature-based Characterisation of Surface Topography and Its Application. *CIRP Annals* 70(2):681–702.
[106] Jiang Y, Gupta A, Zhang Z, Wang G, Dou Y, Chen Y, Fei-Fei L, Anandkumar A, Zhu Y, Fan L (2022) VIMA: General Robot Manipulation with Multimodal Prompts. NeurIPS 2022 Foundation Models for Decision Making Workshop.
[107] Jin X, Ni J (2019) Physics-Based Gaussian Process for the Health Monitoring for a Rolling Bearing. *Acta Astronautica* 154:133–139.
[108] Jin Y, Wang H, Chugh T, Guo D, Miettinen K (2018) Data-Driven Evolutionary Optimization: An Overview and Case Studies. *IEEE Transactions on Evolutionary Computation* 23(3):442–458.
[109] Jin Z, Li H, Gao H (2019) An Intelligent Weld Control Strategy Based on Reinforcement Learning Approach. *The International Journal of Advanced Manufacturing Technology* 100:2163–2175.
[110] Jordan MI, Mitchell TM (2015) Machine Learning: Trends, Perspectives, and Prospects. *Science* 349(6245):255–260.

[111] Jorge VA, Rey VF, Maffei R, Fiorini SR, Carbonera JL, Branchi F, Meireles JP, Franco GS, Farina F, Da Silva TS, Kolberg M (2015) Exploring the IEEE Ontology for Robotics and Automation for Heterogeneous Agent Interaction. *Robotics and Computer-Integrated Manufacturing* 33:12–20.
[112] Jouin M, Gouriveau R, Hissel D, Péra MC, Zerhouni N (2016) Particle Filter-Based Prognostics: Review, Discussion and Perspectives. *Mechanical Systems and Signal Processing* 72:2–31.
[113] Kádár B, Lengyel A, Monostori L, Suginishi Y, Pfeiffer A, Nonaka Y (2010) Enhanced Control of Complex Production Structures by Tight Coupling of the Digital and the Physical Worlds. *CIRP Annals* 59(1):437–440.
[114] Kardos C, Kovács A, Váncza J (2017) Decomposition Approach to Optimal Feature-Based Assembly Planning. *CIRP Annals* 66(1):417–420.
[115] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L (2021) Physics-Informed Machine Learning. *Nature Reviews Physics* 3(6):422–440.
[116] Kaur H, Pannu HS, Malhi AK (2019) A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions. *ACM Computing Surveys* 52(4):1–36.
[117] Kayhan BM, Yildiz G (2021) Reinforcement Learning Applications to Machine Scheduling Problems: A Comprehensive Literature Review. *Journal of Intelligent Manufacturing* : 1–25.
[118] Kelloway A (2020) *AlphaDow: Reinforcement Learning for Industrial Production Scheduling*, https://ray2020.sched.com/event/aWFt/alphadow-reinforcement-learning-for-industrial-production-scheduling-adam-kelloway-dow-chemical Accessed 15 March 2023.
[119] Khosravi M, König C, Maier M, Smith RS, Lygeros J, Rupenyan A (2022) Safety-Aware Cascade Controller Tuning Using Constrained Bayesian Optimization. *IEEE Transactions on Industrial Electronics* 70(2):2128–2138.
[120] Kim S, Kim NH, Choi JH (2020) Prediction of Remaining Useful Life by Data Augmentation Technique Based on Dynamic Time Warping. *Mechanical Systems and Signal Processing* 136:106486.
[121] Kirchen I, Vogel-Heuser B, Hildenbrand P, Schulte R, Vogel M, Lechner M, Merklein M (2017) Data-Driven Model Development for Quality Prediction in Forming Technology. *IEEE International Conference on Industrial Informatics (INDIN)*, 775–780.
[122] Kladovasilakis N, Charalampous P, Kostavelis I, Tzetzis D, Tzovaras D (2021) Impact of Metal Additive Manufacturing Parameters on the Powder Bed Fusion and Direct Energy Deposition Processes: A Comprehensive Review. *Progress in Additive Manufacturing* : 349–365.
[123] Klar M, Ruediger P, Schuermann M, Gören GT, Glatt M, Ravani B, Aurich JC (2024) Explainable Generative Design in Manufacturing for Reinforcement Learning Based Factory Layout Planning. *Journal of Manufacturing Systems* 72:74–92.
[124] Kleeberger K, Bormann R, Kraus W, Huber MF (2020) A Survey on Learning-Based Robotic Grasping. *Current Robotics Reports* : 239–249.
[125] Kluge S, Riffelmacher P, Hummel V, Constantinescu C, Westhamper E (2008) Self-Learning and Self-Optimizing Assembly Systems. *CIRP Conference on Assembly Technologies and Systems* 221.
[126] Knaak C, Masseling L, Duong E, Abels P, Gillner A (2021) Improving Build Quality in Laser Powder Bed Fusion Using High Dynamic Range Imaging and Model-Based Reinforcement Learning. *IEEE Access* 9:55214–55231.
[127] Konečný J, McMahan B, Ramage D (2015) *Federated Optimization: Distributed Optimization Beyond the Datacenter*, .
[128] Kootbally Z, Schlenoff C, Lawler C, Kramer T, Gupta SK (2015) Towards Robust Assembly with Knowledge Representation for the Planning Domain Definition Language (PDDL). *Robotics and Computer-Integrated Manufacturing* 33:42–55.
[129] Koren Y (2010) *The Global Manufacturing Revolution: Product-Process-Business Integration and Reconfigurable Systems*, John Wiley & Sons.
[130] Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems* 25:1097–1105.
[131] Krüger J, Fleischer J, Franke J, Groche P (2019) AI in Production. *Whitepaper of the German Academic Association for Production Technology (WGP)* : 1–25.
[132] Krüger J, Lien TK, Verl A (2009) Cooperation of Human and Machines in Assembly Lines. *CIRP Annals* 58(2):628–646.
[133] Krüger J, Surdilovic D (2008) Robust Control of Force-Coupled Human–Robot-Interaction in Assembly Processes. *CIRP Annals* 57(1):41–44.
[134] Krüger J, Wang L, Verl A, Bauernhansl T, Carpanzano E, Makris S, Fleischer L, Reinhart G, Franke J, Pellegrinelli S (2017) Innovative Control of Assembly Systems and Lines. *CIRP Annals* 66(2):707–730.
[135] Kuhnle A, Kaiser JP, Theiß F, Stricker N, Lanza G (2021) Designing an Adaptive Production Control System Using Reinforcement Learning. *Journal of Intelligent Manufacturing* 32:855–876.
[136] Kumar AS, Khan MA, Thiraviam R, Sornakumar T (2006) Machining Parameters Optimization for Alumina Based Ceramic Cutting Tools Using Genetic Algorithm. *Machining Science and Technology* 10(4):471–489.
[137] Kuschan J, Krüger J (2021) Fatigue Recognition in Overhead Assembly Based on a Soft Robotic Exosuit for Worker Assistance. *CIRP Annals* 70(1):9–12.
[138] Laborie P, Rogerie J, Shaw P, Vilím P (2018) IBM ILOG CP Optimizer for Scheduling. *Constraints* 23(2):210–250.
[139] Lanza G, Ferdows K, Kara S, Mourtzis D, Schuh G, Váncza J, Wang L, Wiendahl HP (2019) Global Production Networks: Design and Operation. *CIRP Annals* 68 (2):823–841.
[140] Le Hesran C, Ladier AL, Botta-Genoulaz V, Laforest V (2019) Operations Scheduling for Waste Minimization: A Review. *Journal of Cleaner Production* 206:211–226.
[141] LeCun Y, Bengio Y, Hinton GE (2015) Deep Learning. *Nature* 521(7553):436–444.
[142] LeCun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L (1989) Handwritten Digit Recognition with a Back-Propagation Network. *Advances in Neural Information Processing Systems* 2:396–404.
[143] Lee AX, Lu H, Gupta A, Levine S, Abbeel P (2015) Learning Force-based Manipulation of Deformable Objects from Multiple Demonstrations. *IEEE International Conference on Robotics and Automation (ICRA)* : 177–184.
[144] Lee MA, Zhu Y, Srinivasan K, Shah P, Savarese S, Fei-Fei L, Garg A, Bohg J (2019) Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks. *International Conference on Robotics and Automation (ICRA)*, 8943–8950.
[145] Lei Y, Lin J, He Z, Zuo MJ (2013) A Review on Empirical Mode Decomposition in Fault Diagnosis of Rotating Machinery. *Mechanical Systems and Signal Processing* 35(1-2):108–126.
[146] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK (2020) Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. *Mechanical Systems and Signal Processing* 138:106587.
[147] Leo Kumar SP (2017) State of the Art-Intense Review on Artificial Intelligence Systems Application in Process Planning and Manufacturing. *Engineering Applications of Artificial Intelligence* 65:294–329.
[148] Levine S, Finn C, Darrell T, Abbeel P (2016) End-to-End Training of Deep Visuomotor Policies. *The Journal of Machine Learning Research* 17(1):1334–1373.
[149] Levine S, Pastor P, Krizhevsky A, Ibarz J, Quillen D (2018) Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection. *The International Journal of Robotics Research* 37(4-5):421–436.
[150] Li B, Hou B, Yu W, Lu X, Yang C (2017) Applications of Artificial Intelligence in Intelligent Manufacturing: A Review. *Frontiers of Information Technology & Electronic Engineering* 18:86–96.
[151] Li C, Zheng P, Yin Y, Wang B, Wang L (2023) Deep Reinforcement Learning in Smart Manufacturing: A Review and Prospects. *CIRP Journal of Manufacturing Science and Technology* 40:75–101.
[152] Li K, Zhang Z, Lin J, Sato R, Matsukuma H, Gao W (2023) Angle Measurement based on Second Harmonic Generation Using Artificial Neural Network. *Nanomanufacturing and Metrology* 6(1):28.
[153] Li T, Zhao Z, Sun C, Cheng L, Chen X, Yan R, Gao R (2021) WaveletKernelNet: An Interpretable Deep Neural Network for Industrial Intelligent Diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52(4):2302–2312.
[154] Liao S, Xue T, Jeong J, Webster S, Ehmann K, Cao J (2023) Hybrid Thermal Modeling of Additive Manufacturing Processes Using Physics-Informed Neural Networks for Temperature Prediction and Parameter Identification. *Computational Mechanics* 72:499–512.
[155] Lichtenthäler C, Peters A, Griffiths S, Kirsch A (2013) Social Navigation-Identifying Robot Navigation Patterns in a Path Crossing Scenario. *International Conference in Social Robotics*, 84–93.
[156] Lindenmeyer A, Webster S, Zaeh MF, Ehmann KF, Cao J (2021) Template-Bayesian Approach for the Evaluation of Melt Pool Shape and Dimension of a DED-Process from In-Situ X-Ray Images. *CIRP Annals* 70(1):183–186.
[157] Liu CL, Chang CC, Tseng CJ (2020) Actor-Critic Deep Reinforcement Learning for Solving Job Shop Scheduling Problems. *IEEE Access* 8:71752–71762.
[158] Liu S, Jiang H, Wu Z, Li X (2022) Data Synthesis Using Deep Feature Enhanced Generative Adversarial Networks for Rolling Bearing Imbalanced Fault Diagnosis. *Mechanical Systems and Signal Processing* 163:108139.
[159] Liu Z, Liu Q, Xu W, Wang L, Zhou Z (2022) Robot Learning Towards Smart Robotic Manufacturing: A Review. *Robotics and Computer-Integrated Manufacturing* 77:102360.
[160] Lowe DG (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60:91–110.
[161] Lu SY, ElMaraghy W, Schuh G, Wilhelm R (2007) A Scientific Foundation of Collaborative Engineering. *CIRP Annals* 56(2):605–634.
[162] Lu Y, Witherell P, Lopez F, Assouroko I (2016) Digital Solutions for Integrated and Collaborative Additive Manufacturing. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* 50084. V01BT02A033.
[163] Lu YJ, Tsao Y, Watanabe S (2021) A Study on Speech Enhancement based on Diffusion Probabilistic Model. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 659–666.
[164] Lundberg SM, Lee SI (2017) A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems* 30:4765–4774.
[165] Luo Y, Peng J, Ma J (2020) When Causal Inference Meets Deep Learning. *Nature Machine Intelligence* 2(8):426–427.
[166] Maier M, Rupenyan A, Bobst C, Wegener K (2020) Self-Optimizing Grinding Machines using Gaussian Process Models and Constrained Bayesian Optimization. *The International Journal of Advanced Manufacturing Technology* 108:539–552.
[167] Makris S, Dietrich F, Kellens K, Hu SJ (2023) Automated Assembly of Non-Rigid Objects. *CIRP Annals* 72(2):513–539.
[168] Malus A, Kozjek D (2020) Real-Time Order Dispatching for a Fleet of Autonomous Mobile Robots Using Multi-Agent Reinforcement Learning. *CIRP Annals* 69 (1):397–400.
[169] Mangold S, Steiner C, Friedmann M, Fleischer J (2022) Vision-Based Screw Head Detection for Automated Disassembly for Remanufacturing. *Procedia CIRP* 105:1–6.
[170] Markus A, Hatvany J (1987) Matching AI Tools to Engineering Requirements. *CIRP Annals* 36(1):311–315.
[171] McCarthy J (1978) *History of LISP*, Academic Press.
[172] McCulloch WS, Pitts W (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics* 5(4):115–133.
[173] Mehta M, Shao C (2022) Federated Learning-Based Semantic Segmentation for Pixel-Wise Defect Detection in Additive Manufacturing. *Journal of Manufacturing Systems* 64:197–210.
[174] Mikolov T, Chen K, Corrado G, Dean J (2013) *Efficient Estimation of Word Representations in Vector Space*, .
[175] Minsky M (1961) Steps toward Artificial Intelligence. *Proceedings of the IRE* 49 (1):8–30.
[176] Minsky ML, Papert SA (1988) *Perceptrons*, MIT Press.
[177] Mitchell T (1997) *Machine Learning*, McGraw Hill.
[178] Möhring HC, Wiederkehr P (2016) Intelligent Fixtures for High Performance Machining. *Procedia CIRP* 46:383–390.

[179] Möhring HC, Wiederkehr P, Erkorkmaz K, Kakinuma Y (2020) Self-Optimizing Machining Systems. *CIRP Annals* 69(2):740–763.
[180] Molina LM, Teti R, Alvir EM (2023) Quality, Efficiency and Sustainability Improvement in Machining Processes Using Artificial Intelligence. *Procedia CIRP* 118:501–506.
[181] Monostori L, Csaji BC, Egri P, Kis KB, Váncza J, Ochs J, Jung S, König N, Pieske S, Wein S, Schmitt R (2021) Automated Stem Cell Production by Bio-Inspired Control. *CIRP Journal of Manufacturing Science and Technology* 33:369–379.
[182] Monostori L, Viharos ZJ (2001) Hybrid, AI-and Simulation-Supported Optimisation of Process Chains and Production Plants. *CIRP Annals* 50(1):353–356.
[183] Monostori L, Váncza J, Kumara SRT (2006) Agent-Based Systems for Manufacturing. *CIRP Annals* 55(2):697–720.
[184] Moor J (2006) The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine* 27(4):87.
[185] Morgan JN, Sonquist JA (1963) Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American statistical association* 58(302):415–434.
[186] Mourtzis D, Angelopoulos J, Panopoulos N (2022) A Literature Review of the Challenges and Opportunities of the Transition from Industry 4.0 to Society 5.0. *Energies* 15(17):6276.
[187] Naderi B, Ruiz R, Roshanaei V (2023) Mixed-Integer Programming vs. Constraint Programming for Shop Scheduling Problems: New Results and Outlook. *INFORMS Journal on Computing* 35(4):817–843.
[188] Narvekar S, Peng B, Leonetti M, Sinapov J, Taylor ME, Stone P (2020) Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey. *The Journal of Machine Learning Research* 21(1):7382–7431.
[189] Nassehi A, Essink W, Barclay J (2015) Evolutionary Algorithms for Generation and Optimization of Tool Paths. *CIRP Annals* 64(1):455–458.
[190] National Academies of Sciences, Engineering, and Medicine. *Foundational Research Gaps and Future Directions for Digital Twins*, The National Academies Press. Washington, DC. https://doi.org/10.17226/26894. Accessed 15 March 2024.
[191] Newton I (1833) *Philosophiae Naturalis Principia Mathematica*, G. Brookman.
[192] Niraula NB, Whyatt D, Kao A (2018) A Novel Approach to Part Name Discovery in Noisy Text. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 170–176.
[193] Nonaka Y, Erdős G, Kis T, Nakano T, Váncza J (2012) Scheduling with Alternative Routings in CNC Workshops. *CIRP Annals* 61(1):449–454.
[194] O'Hare GMP, Chisholm AWJ (1990) Distributed Artificial Intelligence: An Invaluable Technique for the Development of Intelligent Manufacturing Systems. *CIRP Annals* 39(1):485–488.
[195] OpenAI. *Introducing ChatGPT*, https://openai.com/blog/chatgpt. Accessed 9 April 2024.
[196] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A, Schulman J (2022) Training Language Models to Follow Instructions with Human Feedback. *Advances in Neural Information Processing Systems* 35:27730–27744.
[197] Panzer M, Bender B (2022) Deep Reinforcement Learning in Production Systems: A Systematic Literature Review. *International Journal of Production Research* 60 (13):4316–4341.
[198] Paris PC (1961) A Rational Analytic Theory of Fatigue. *Trends Engineering* 13:9–14.
[199] Parker-Holder J, Nguyen V, Roberts SJ (2020) Provably Efficient Online Hyperparameter Optimization with Population-Based Bandits. *Advances in Neural Information Processing Systems* 33:17200–17211.
[200] Patel SB, Lam K (2023) ChatGPT: The Future of Discharge Summaries? *The Lancet Digital Health* 5(3):e107–e108.
[201] Pfrommer S, Halm M, Posa M (2021) Contactnets: Learning Discontinuous Contact Dynamics with Smooth, Implicit Representations. *Conference on Robot Learning* : 2279–2291.
[202] Pochet Y, Wolsey LA (2006) *Production Planning By Mixed Integer Programming*, SpringerBerlin.
[203] Qiao B, Zhu J, Wei Z (1999) Learning Force Control for Position Controlled Robotic Manipulator. *CIRP Annals* 48(1):1–4.
[204] Raatz A, Blankemeyer S, Recker T, Pischke D, Nyhuis P (2020) Task Scheduling Method for HRC Workplaces Based on Capabilities and Execution Time Assumptions for Robots. *CIRP Annals* 69(1):13–16.
[205] Radford A, Narasimhan K, Salimans T, Sutskever I (2018) *Improving Language Understanding by Generative Pre-Training*, OpenAI.
[206] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ (2020) Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21(140):1–67.
[207] Rajpathak DG, Motta E, Zdrahal Z, Roy R (2006) A Generic Library of Problem Solving Methods for Scheduling Applications. *IEEE Transactions on Knowledge and Data Engineering* 18(6):815–828.
[208] Ramos L (2015) Semantic Web for Manufacturing, Trends and Open Issues: Toward a State of the Art. *Computers & Industrial Engineering* 90:444–460.
[209] Rao K, Harris C, Irpan A, Levine S, Ibarz J, Khansari M (2020) Rl-cyclegan: Reinforcement Learning Aware Simulation-to-Real. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11157–11166.
[210] Renken V, Albinger S, Goch G, Neef A, Emmelmann C (2017) Development of an Adaptive, Self-Learning Control Concept for an Additive Manufacturing Process. *CIRP Journal of Manufacturing Science and Technology* 19:57–61.
[211] Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI*, 234–241.
[212] Rosenblatt F (1960) Perceptron Simulation Experiments. *Proceedings of the IRE* 48(3):301–309.
[213] Rowe WB, Yan L, Inasaki I, Malkin S (1994) Applications of Artificial Intelligence in Grinding. *CIRP Annals* 43(2):521–531.
[214] Rumelhart DE, Hinton GE, Williams RJ (1986) Learning Representations by Back-Propagating Errors. *Nature* 323(6088):533–536.
[215] Russell M, Wang P (2023) Maximizing Model Generalization for Manufacturing with Self-Supervised Learning and Federated Learning. *Journal of Manufacturing Systems* 71:274–285.
[216] Russell S, Norvig P (1995) *Artificial Intelligence — A Modern Approach*, Pearson.
[217] Russell S, Wefald E (1991) *Do the Right Thing-Studies in Limited Rationality*, MIT Press.
[218] Sagan C, Druyan A (1993) *Shadows of Forgotten Ancestors: A search For Who We Are*, Random House.
[219] Sagan C, Druyan A (2011) *Pale Blue Dot: A Vision of the Human Future in Space*, Random House.
[220] Sánchez M, Cruz-Duarte JM, carlos Ortíz-Bayliss J, Ceballos H, Terashima-Marin H, Amaya I (2020) A Systematic Review of Hyper-Heuristics on Combinatorial Optimization Problems. *IEEE Access* 8:128068–128095.
[221] Sanfilippo EM, Belkadi F, Bernard A (2019) Ontology-Based Knowledge Representation for Additive Manufacturing. *Computers in Industry* 109:182–194.
[222] Schleich B, Anwer N, Mathieu L, Wartzack S (2017) Shaping the Digital Twin for Design and Production Engineering. *CIRP Annals* 66(1):141–144.
[223] Scholz-Reiter B, Weimer D, Thamer H (2012) Automated Surface Inspection of Cold-Formed Micro-Parts. *CIRP Annals* 61(1):531–534.
[224] Schwartz R, Dodge J, Smith NA, Etzioni O (2020) Green AI. *Communications of the ACM* 63(12):54–63.
[225] Semeraro F, Griffiths A, Cangelosi A (2023) Human–Robot Collaboration and Machine Learning: A Systematic Review of Recent Research. *Robotics and Computer-Integrated Manufacturing* 79:102432.
[226] Serrano-Ruiz JC, Mula J, Poler R (2021) Smart Manufacturing Scheduling: A Literature Review. *Journal of Manufacturing Systems* 61(1):265–287.
[227] Shao S, Wang P, Yan R (2019) Generative Adversarial Networks for Data Augmentation in Machine Fault Diagnosis. *Computers in Industry* 106:85–93.
[228] Sharp M, Ak R, Hedberg Jr T (2018) A Survey of the Advancing Use and Development of Machine Learning in Smart Manufacturing. *Journal of Manufacturing Systems* 48:170–179.
[229] Shortliffe EH. MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection (Doctoral dissertation, Stanford University).
[230] Simonyan K, Vedaldi A, Zisserman A (2014) Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. In: *Proceedings of the International Conference on Learning Representations*, 1–8.
[231] Smith SF, Becker MA (1997) An Ontology for Constructing Scheduling Systems. *Working Notes of 1997 AAAI Symposium on Ontological Engineering*, 120–127.
[232] Smolensky P (1987) Connectionist AI, Symbolic AI, and the Brain. *Artificial Intelligence Review* 1(2):95–109.
[233] Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems* 25:2951–2959.
[234] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research* 15(1):1929–1958.
[235] Stability AI (2022) *Stable Diffusion Launch Announcement*, https://stability.ai/blog/stable-diffusion-announcement. Accessed 15 March 2023.
[236] Stark R, Kind S, Neumeyer S (2017) Innovations in Digital Modelling for Next Generation Manufacturing System Design. *CIRP Annals* 66(1):169–172.
[237] Stricker N, Kuhnle A, Hofmann C, Deininger P (2021) Self Adjusting Multi-Objective Scheduling Based on Monte Carlo Tree Search for Matrix Production Assembly Systems. *CIRP Annals* 70(1):381–384.
[238] Su J, Adams S, Beling PA (2020) *Counterfactual Multi-Agent Reinforcement Learning with Graph Convolution Communication*, .
[239] Sutton R, Barto A (2018) *Reinforcement Learning: An Introduction*, MIT Press.
[240] Suvdaa B, Ahn J, Ko J (2012) Steel Surface Defects Detection and Classification using SIFT and Voting Strategy. *International Journal of Software Engineering and Its Applications* 6(2):161–166.
[241] Svegliato J, Sharma P, Zilberstein S (2020) A Model-Free Approach to Meta-Level Control of Anytime Algorithms. *IEEE International Conference on Robotics and Automation (ICRA)*, 11436–11442.
[242] Tan H (2017) A Brief History and Technical Review of the Expert System Research. *IOP Conference Series* 242(1):012111.
[243] Terkaj W, Qi Q, Urgo M, Scott PJ, Jiang X (2021) Multi-Scale Modelling of Manufacturing Systems Using Ontologies and Delta-Lenses. *CIRP Annals* 70 (1):361–364.
[244] Teti R, Kumara SRT (1997) Intelligent Computing Methods for Manufacturing Systems. *CIRP Annals* 46(2):629–652.
[245] Thiede S, Turetskyy A, Loellhoeffel T, Kwade A, Kara S, Herrmann C (2020) Machine Learning Approach for Systematic Analysis of Energy Efficiency Potentials in Manufacturing Processes: A Case of Battery Production. *CIRP Annals* 69 (1):21–24.
[246] Tian Q, Guo S, Guo Y (2020) A Physics-Driven Deep Learning Model for Process-Porosity Causal Relationship and Porosity Prediction with Interpretability in Laser Metal Deposition. *CIRP Annals* 69(1):205–208.
[247] Tian Y, Xu J, Li Y, Luo J, Sueda S, Li H, Willis KD, Matusik W (2022) Assemble Them All: Physics-Based Planning for Generalizable Assembly by Disassembly. *ACM Transactions on Graphics (TOG)* 41(6):1–11.
[248] Tolio T, Ceglarek D, ElMaraghy HA, Fischer A, Hu SJ, Laperrière L, Newman ST, Váncza J (2010) SPECIES—Co-Evolution of Products, Processes and Production Systems. *CIRP Annals* 59(2):672–693.
[249] Tomiyama T, Gu P, Jin Y, Lutters D, Kind C, Kimura F (2009) Design Methodologies: Industrial and Educational Applications. *CIRP Annals* 58(2):543–565.
[250] Tsutsumi D, Gyulai D, Kovács A, Tipary B, Ueno Y, Nonaka Y, Monostori L (2018) Towards Joint Optimization of Product Design, Process Planning and Production Planning in Multi-Product Assembly. *CIRP Annals* 67(1):441–446.
[251] Turing AM (2009) *Computing Machinery and Intelligence*, Springer.
[252] Valmeekam K, Marquez M, Sreedharan S, Kambhampati S (2023) *On the Planning Abilities of Large Language Models-A Critical Investigation*, .

[253] Van Brussel H, Wyns J, Valckenaers P, Bongaerts L, Peeters P (1998) Reference Architecture for Holonic Manufacturing Systems. *PROSA. Computers in Industry* 37(3):255–274.

[254] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is All You Need. *Advances in Neural Information Processing Systems* 30:5998–6008.

[255] Vlassis NN, Sun W (2023) Denoising Diffusion Algorithm for Inverse Design of Microstructures with Fine-Tuned Nonlinear Material Properties. *Computer Methods in Applied Mechanics and Engineering* 413:116126.

[256] Vongbunyong S, Kara S, Pagnucco M (2013) Application of Cognitive Robotics in Disassembly of Products. *CIRP Annals* 62(1):31–34.

[257] Vrabic R, Erkoyuncu JA, Farsi M, Ariansyah D (2021) An Intelligent Agent-Based Architecture for Resilient Digital Twins in Manufacturing. *CIRP Annals* 70 (1):349–352.

[258] Wabersich KP, Hewing L, Carron A, Zeilinger MN (2021) Probabilistic Model Predictive Safety Certification for Learning-Based Control. *IEEE Transactions on Automatic Control* 67(1):176–188.

*JournalofManufacturingSystems*57:298–310.

[260] Wang L (2019) From Intelligence Science to Intelligent Manufacturing. *Engineering* 5(4):615–618.

[261] Wang L, Gao R, Váncza J, Krüger J, Wang XV, Makris S, Chryssolouris G (2019) Symbiotic Human-Robot Collaborative Assembly. *CIRP Annals* 68(2):701–726.

[262] Wang L, Liu S, Cooper C, Wang XV, Gao R (2021) Function Block-Based Human-Robot Collaborative Assembly Driven by Brainwaves. *CIRP Annals* 70(1):5–8.

[263] Wang P, Gao R (2015) Adaptive Resampling-Based Particle Filtering for Tool Life Prediction. *Journal of Manufacturing Systems* 37:528–534.

[264] Wang P, Gao R, Yan R (2017) A Deep Learning-based Approach to Material Removal Rate Prediction in Polishing. *CIRP Annals* 66(1):429–432.

[265] Wang P, Karigiannis J, Gao R (2024) Ontology-Integrated Tuning of Large Language Model for Intelligent Maintenance. *CIRP Annals* 73(1):1–4.

[266] Wang P, Liu H, Wang L, Gao R (2018) Deep Learning-Based Human Motion Recognition for Predictive Context-Aware Human-Robot Collaboration. *CIRP Annals* 67(1):17–20.

[267] Wang P, Yan R, Gao R (2018) Multi-Mode Particle Filter for Bearing Remaining Life Prediction. *International Manufacturing Science and Engineering Conference* 51371:V003T02A031.

[268] Wang X, Williams RE, Sealy MP, Rao PK, Guo Y (2018) Stochastic Modeling and Analysis of Spindle Power During Hard Milling with a Focus on Tool Wear. *Journal of Manufacturing Science and Engineering* 140(11):111011.

[269] Wang YF (2020) Adaptive Job Shop Scheduling Strategy based on Weighted Q-Learning Algorithm. *Journal of Intelligent Manufacturing* 31(2):417–432.

[270] Wegener K, Damm O, Harst S, Ihlenfeldt S, Monostori L, Teti R, Wertheim R, Byrne G (2023) Biologicalisation in Manufacturing−Current State and future Trends. *CIRP Annals* 72(2):781–807.

[271] Weimer D, Scholz-Reiter B, Shpitalni M (2016) Design of Deep Convolutional Neural Network Architectures for Automated Feature Extraction in Industrial Inspection. *CIRP Annals* 65(1):417–420.

[272] Wen B, Lian W, Bekris K, Schaal S (2022) Catgrasp: Learning Category-Level Task-Relevant Grasping in Clutter from Simulation. *International Conference on Robotics and Automation (ICRA)* : 6401–6408.

[273] White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J, Schmidt DC (2023) *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*, .

[274] Wikipedia. *Artificial Intelligence in Fiction*, https://en.wikipedia.org/wiki/Artificial_intelligence_in_fiction Accessed 15 May 2021.

[275] Wu Y, Yan W, Kurutach T, Pinto L, Abbeel P (2020) Learning to Manipulate Deformable Objects Without Demonstrations. *Robotics: Science and Systems* 1:11 .

[276] Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY (2020) A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* 32(1):4–24.

[277] Xie X, Bennett J, Saha S, Lu Y, Cao J, Liu WK, Gan Z (2021) Mechanistic Data-Driven Prediction of As-Built Mechanical Properties In Metal Additive Manufacturing. *npj Computational Materials* 7(1):1–12.

[278] Xu X, Lu Y, Vogel-Heuser B, Wang L (2021) Industry 4.0 and Industry 5.0-Inception, Conception and Perception. *Journal of Manufacturing Systems* 61:530–535.

[279] Yan R, Gao R (2007) Approximate Entropy as a Diagnostic Tool for Machine Health Monitoring. *Mechanical Systems and Signal Processing* 21(2):824–839.

[280] Yan R, Liu Y, Gao R (2012) Permutation entropy: A Nonlinear Statistical Measure for Status Characterization of Rotary Machines. *Mechanical Systems and Signal Processing* 29:474–484.

[281] Yildiz E, Brinker T, Renaudo E, Hollenstein JJ, Haller-Seeber S, Piater JH, Wörgötter F (2020) A Visual Intelligence Scheme for Hard Drive Disassembly in Automated Recycling Routines. *International Conference on Robotics, Computer Vision, and Intelligent Systems*, 17–27.

[282] Yin YH, Nee AY, Ong SK, Zhu JY, Gu PH, Chen LJ (2015) Automating Design with Intelligent Human−Machine Integration. *CIRP Annals* 64(2):655–677.

[283] Zadeh LA (1965) Fuzzy sets. *Information and control* 8(3):338–353.

[284] Zakka K, Zeng A, Lee J, Song S (2020) Form2fit: Learning Shape Priors for Generalizable Assembly from Disassembly. *IEEE International Conference on Robotics and Automation (ICRA)*, 9404–9410.

[285] Zeiler MD, Fergus R (2014) Visualizing and Understanding Convolutional Networks. *European Conference on Computer Vision* : 818–833.

[286] Zhang J, Cooper C, Gao R (2022) Federated Learning for Privacy-Preserving Collaboration in Smart Manufacturing. *Global Conference on Sustainable Manufacturing*, 845–853.

[287] Zhang J, Liu C, Gao R (2022) Physics-Guided Gaussian Process for HVAC System Performance Prognosis. *Mechanical Systems and Signal Processing* 179:109336.

[288] Zhang J, Liu H, Chang Q, Wang L, Gao R (2020) Recurrent Neural Network for Motion Trajectory Prediction in Human-Robot Collaborative Assembly. *CIRP Annals* 69(1):9–12.

[289] Zhang L, Yu C, Wong TN (2021) A Graph-Based Constraint Programming Approach for the Integrated Process Planning and Scheduling Problem. *Computers & Operations Research* 131:105282.

[290] Zhang Q, Lin Y (2023) Integrating Multi-Agent Reinforcement Learning and 3D A* Search for Facility Layout Problem Considering Connector-Assembly. *Journal of Intelligent Manufacturing* : 1–26.

[291] Zhang T, McCarthy Z, Jow O, Lee D, Chen X, Goldberg K, Abbeel P (2018) Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. *IEEE International Conference on Robotics and Automation (ICRA)* : 5628–5635.

[292] Zhang W, Li X, Ma H, Luo Z, Li X2 (2021) Federated Learning for Machinery Fault Diagnosis with Dynamic Validation and Self-Supervision. *Knowledge-Based Systems* 213:106679.

[293] Zhang Y, Xiong R, He H, Pecht MG (2018) Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Transactions on Vehicular Technology* 67(7):5695–5705.

[294] Zhao TZ, Kumar V, Levine S, Finn C (2023) *Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware*, .

[295] Zheng C, Du Y, Sun T, Eynard B, Zhang Y, Li J, Zhang X (2023) Multi-Agent Collaborative Conceptual Design Method for Robotic Manufacturing Systems in Small- and Mid-Sized Enterprises. *Computers & Industrial Engineering* 183:109541.

[296] Zobeiry N, Humfeld KD (2021) A Physics-Informed Machine Learning Approach for Solving Heat Transfer Equation in Advanced Manufacturing and Engineering Applications. *Engineering Applications of Artificial Intelligence* 101:104232.

[297] Zweben M, Fox MS (1994) *Intelligent Scheduling*, Morgan Kaufmann.