International Journal of Production Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tprs20

Deep reinforcement learning in production systems: a systematic literature review

Marcel Panzer & Benedict Bender

To cite this article: Marcel Panzer & Benedict Bender (2021): Deep reinforcement learning in production systems: a systematic literature review, International Journal of Production Research, DOI: 10.1080/00207543.2021.1973138

To link to this article: https://doi.org/10.1080/00207543.2021.1973138

Published online: 17 Sep 2021.

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
https://doi.org/10.1080/00207543.2021.1973138

Taylor & Francis
Taylor & Francis Group

Deep reinforcement learning in production systems: a systematic literature review

Marcel Panzer and Benedict Bender

Chair of Business Informatics, Processes and Systems, University of Potsdam, Potsdam, Germany

ABSTRACT
Shortening product development cycles and fully customisable products pose major challenges for production systems. These not only have to cope with an increased product diversity but also enable high throughputs and provide a high adaptability and robustness to process variations and unforeseen incidents. To overcome these challenges, deep Reinforcement Learning (RL) has been increasingly applied for the optimisation of production systems. Unlike other machine learning methods, deep RL operates on recently collected sensor-data in direct interaction with its environment and enables real-time responses to system changes. Although deep RL is already being deployed in production systems, a systematic review of the results has not yet been established. The main contribution of this paper is to provide researchers and practitioners an overview of applications and to motivate further implementations and research of deep RL supported production systems. Findings reveal that deep RL is applied in a variety of production domains, contributing to data-driven and flexible processes. In most applications, conventional methods were outperformed and implementation efforts or dependence on human experience were reduced. Nevertheless, future research must focus more on transferring the findings to real-world systems to analyse safety aspects and demonstrate reliability under prevailing conditions.

ARTICLE HISTORY
Received 6 April 2021
Accepted 18 August 2021

KEYWORDS
Machine learning; reinforcement learning; production control; production planning; manufacturing processes; systematic literature review

1. Introduction

Nowadays, companies must cope with mass customisation and shortening development cycles that pose major challenges for smart production facilities. They must be capable to operate in highly uncertain market conditions and satisfy the increasingly challenging standards of product quality and sustainability in the shortest possible time. To meet these challenges, Germany launched the Industry 4.0 initiative in 2013 to support the development of flexible and adaptive production systems (Kagermann, Wahlster, and Helbig 2013). Although the initiative's potential and possible impact is huge, Xu, Xu, and Li (2018) indicate that many of today's Industry 4.0 implementations are not yet applying corresponding advanced techniques such as machine learning. This also becomes apparent in Liao et al. (2017), who states that while modelling, virtualisation, or big data techniques are increasingly in the focus of production research, machine learning is not. This impression has already been countered by Kang, Catal, and Tekinerdogan (2020), who highlighted the broad application landscape of machine learning in modern production and their ability to reach state-of-the-art performance. Going further into detail, our review specifically considers deep Reinforcement Learning (RL) as an online data-driven optimisation approach and highlights its beneficial properties for production systems.

The field of machine learning consists of (semi-) supervised, unsupervised, and reinforcement learning. Whereas supervised and unsupervised learning require a (pre-labelled) set of data, RL differs in particular by the learning in direct interaction with its environment. It learns by a trial-and-error principle without requiring any pre-collected data or prior (human) knowledge and has the ability to adapt flexibly to uncertain conditions (Sutton and Barto 2017). Considering these flexible and desired features in modern production, our paper aims to capture the current state-of-the-art of real or simulated deep RL applications in production systems. Besides, we seek to identify existing challenges and help to define future fields of research.

Already in 1998, Mahadevan and Theocharous (1998) demonstrated the potential of RL in production manufacturing and its superiority in inventory minimisation compared to a Kanban system. In recent years, since neural networks are emerging, neural network-based

CONTACT Marcel Panzer marcel.panzer@wi.uni-potsdam.de Chair of Business Informatics, Processes and Systems, University of Potsdam, Karl-Marx-Street 67, 14482 Potsdam, Germany

RL reached impressive success with Google DeepMind's AlphaGo (Silver et al. 2017), and is now increasingly being transferred to production systems. Based on recently collected sensor data, deep RL enables online data-driven decisions in real-time and supports a responsive reaction-driven and adaptive system design (Han and Yang 2020). It can increase production stability and robustness and reaches superior performances compared to state-of-the-art heuristics (as in Li et al. (2020)).

However, in production related reviews, deep RL has often been considered only in the context of other machine learning techniques as in Kang, Catal, and Tekinerdogan (2020) or Arinez et al. (2020) and is not mentioned in an industrial intelligence context in Peres et al. (2020), lacking in consolidation of the already obtained results. This is also apparent in other technology fields such as energy (Mishra et al. 2020), process industry (Lee, Shin, and Realff 2018), or tool condition monitoring (Serin et al. 2020).

In contrast, other disciplines have consolidated the obtained research findings of deep RL and highlighted its adaptive behaviour and the ability to generalise past experiences. This includes communications and networking (Luong et al. 2019), cyber-physical-systems (Liu et al. 2019), economic applications (Mosavi et al. 2020), internet of things (Lei et al. 2020), object grasping (Mohammed, Chung, and Chyi Chua 2020), power and energy systems (Cao et al. 2020), robotics (Khan et al. 2020), robotic manipulations tasks (Nguyen and La 2019), and dynamic task scheduling (Shyalika, Silva, and Karunananda 2020), which reflects the broad range of research and underlines the ongoing focus on implementing deep RL applications to significantly increase the adaptability and robustness of the respecting processes.

To the best of our knowledge, this is the first attempt to capture general applications of deep RL in production systems. We intend to provide a systematic overview of ongoing research to assist scholars in identifying deep RL research directions and potential future applications. The review also serves practitioners in considering possible deployment scenarios and motivate them to transfer research findings to real-world systems. For this purpose, we attempt to answer the following research questions.

- RQ1: What are deep RL applications in specific production system domains?
- RQ2: What are current implementation challenges of deep RL in production systems?
- RQ3: What future research needs to be conducted to address existing challenges of deep RL in production systems?

The Paper is structured as follows. Section 2 describes the basics of deep RL and gives an overview of essential algorithms. Section 3 defines the methodology and the conceptual framework that guides the literature review. Section 4 answers RQ1 based on the conducted review and provides the basis for Section 5, which analyses specific barriers and challenges (RQ2) and outlines fields for future research to address these (RQ3). Section 6 discusses the results and provides managerial insights and given limitations. Finally, a conclusion is given in Section 7.

2. Introduction to reinforcement learning

Reinforcement learning (RL) is a subcategory of machine learning and distinguishes itself from supervised and unsupervised learning in particular by the trial and error learning approach in direct interaction with its environment (Sutton and Barto 2017). It does not need supervision or a pre-defined labelled or unlabelled set of data and comes into consideration whenever challenges have to be met in dynamic environments that require a real-time and reaction-driven decision-making process. It is able to generalise its previously learned knowledge (Wang et al. 2020) and enables an online adaptation to changing environmental conditions by sequential decision-making (as in Palombarini and Martínez (2019)).

In RL, the agent learns a policy that outputs an action according to the received state as illustrated in Figure 1. To achieve this, conventional RL often employs a Q-table to map the policy, which requires discretisation of state and action spaces. The Q-table lists Q-values that quantify the action quality of performing an action in a given state, which are updated through ongoing training of the agent. In many cases, Q-learning outperformed several conventional approaches such as FIFO in flow production scheduling, which reduced makespan, whereby states were described by machine runtimes and buffer occupancy, and executable actions by unit movements (Lee and Kim 2021). Other successful examples are the superior performance compared to multiple scheduling approaches in an adaptive assembly process (Wang, Sarker, and Jian Li 2020) by choosing scheduling rules based on waiting queues, interval times,

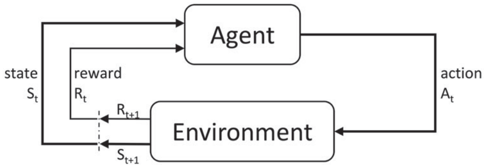


Figure 1. Agent–environment interaction; Sutton and Barto (2017).

remaining processing time, and processing status, or the condition-based maintenance control in Xanthopoulos et al. (2018) that reduced costs compared to a Kanban method by authorising maintenance actions based on finished goods, backorders, and facility deterioration states.

However, the required action and state discretisation impose the curse of dimensionality in high-dimensional problem spaces, which causes an exponentially increasing table size and leads to high iterative computational costs, low learning efficiencies, and degraded performances (Bellman 1957). To address this, and as proposed by Lee and Kim (2021) among others, deep RL attempts to solve this problem by combining the advantages of RL with those of deep learning. In deep RL, the policy is mapped by a neural network as a function approximator, which is capable of processing large amounts of unsorted and raw input data (Lange, Riedmiller, and Voigtlander 2012).

(Deep) RL can be further classified into model-free and model-based algorithms. Model-based algorithms such as the AlphaZero get or learn a model of the environment to predict next values or states (Silver et al. 2017). In contrast, model-free algorithms neither learn the dynamics of the environment nor a state-transition function (Sutton and Barto 2017). Model-free algorithms, as the major group in this review, can be further classified into policy-based, value-based, and hybrid algorithms. Policy-based algorithms such as a PPO provide a continuous action space and try to directly map a state to an action by building a representation of the actual behaviour policy (Sewak 2019). In contrast, value-based algorithms such as a DQN learn a value function for discrete action spaces to evaluate each of the potential actions (Watkins and Dayan 1992). Algorithms like the DDPG utilise a hybrid actor-critic structure which combines previous methods advantages (Lillicrap et al. 2016). Other possible modifications such as a prioritised experience replay, which takes particular account of important experiences during updates, can be integrated into the deep RL framework (Schaul et al. 2016).

Besides basic algorithmic settings, particular consideration is required for the choice of hyperparameters. The discount factor, which determines the relevance of short-term or distant future rewards, the learning rate, which determines the balance between learning speed and stability, and other algorithmic as well as neural network parameters in deep RL strongly affect the final performance. Specific considerations should also reflect the optimal design of the state/action space and reward design. Appropriate interference between these can lead to optimal system behaviour and help in the search for optimal control strategies (Sewak 2019). In particular, the reward function must be designed concerning the agent's objective and system dynamics and must be able to account for short- as well as long-term outcomes. For further algorithmic insights we would like to refer to Wang et al. (2020) or Naeem, Rizvi, and Coronato (2020) for an extended introduction and in-depth analysis of (deep) RL algorithms.

Initially limited to the Atari platform in Mnih et al. (2013), deep RL is being deployed in an increasing number of applications which benefit from its flexibility and online adaption capabilities. Potential applications such as smart scheduling benefit from the distributed multi-agent capabilities and collaborative properties, which could significantly increase robustness as proposed in Rossit, Tohmé, and Frutos (2019). It makes deep RL being a promising technique to improve the performance of modern production systems and enable the transition towards industry 4.0. However, unlike other algorithmic overviews or the general descriptions of machine intelligence applications in production, the intersection of deep RL in different production system domains was not specifically covered. To address this gap and highlight the benefits, an representative review of the intersection might assist to identify individual applications, challenges, and future fields of research.

3. Research methodology

This section outlines the basic literature review process of deep RL applications in production systems. To ensure a systematic and representative review, we follow Tranfield, Denyer, and Smart (2003) and Antônio Márcio Tavares, Felipe Scavarda and José Scavarda (2016) who provide guidelines for the content analysis. This enables a consolidation and evaluation of existing literature and provides the state-of-the-art in the focused domain at a given time. The consolidation shall assist researchers and others to identify research gaps and provides research incentives and managerial insights (Petticrew and Roberts 2006).

According to the guideline proposed by Antônio Márcio Tavares, Felipe Scavarda and José Scavarda (2016), the systematic literature review (SLR) can be organised into 8 (iterative) steps. These main steps are outlined sequentially in Figure 2 and will be considered in the subsequent review process.

3.1. *Review focus*

The formulation of the research questions and clarification of the problem, is outlined in Section 1. The composition of the review team consisted of the two authors who worked through each step separately and finally combined their work.

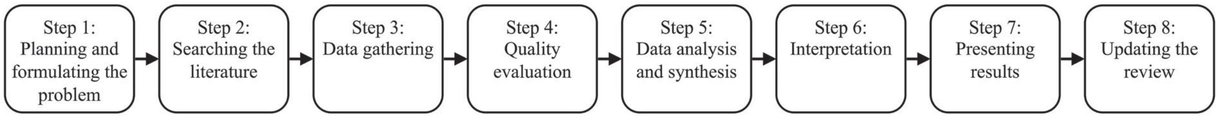

Step 1: Planning and formulating the problem
Step 2: Searching the literature
Step 3: Data gathering
Step 4: Quality evaluation
Step 5: Data analysis and synthesis
Step 6: Interpretation
Step 7: Presenting results
Step 8: Updating the review

Figure 2. Eight step approach to conduct an SLR.

Table 1. Taxonomy framework of the SLR.

Characteristic	Categories			
(1) Focus	Research outcomes	Research methods	Theories	Applications
(2) Goal	Integration	Criticism	Central issues	
(3) Perspective	Neutral representation		Espousal of position	
(4) Coverage	Exhaustive	Exhaustive and selective	Representative	Central/pivotal
(5) Organisation	Historical	Conceptual	Methodological	
(6) Audience	Specialised scholars	General scholars	Practitioners/politicians	General public

To define the scope of the problem and simplify the review process, the more in-depth planning relies on Brocke et al. (2009) and follows the associated taxonomy framework by Cooper (1988, Table 1). The gray highlighted cells represent the selection of underlying characteristics of this SLR and the associated goals and foci.

Following the taxonomy, this SLR focuses on presenting existing applications and achieved research results of deep RL in production systems (1). Its goal (2) is to present existing research in an integrative and synthesising manner while highlighting central future application and maturity issues. We try to maintain a neutral perspective (3) and provide a representative coverage of our focused content (4). The organisation of the review is conceptually designed (5). In particular, the application concept in the respective discipline shall be reflected rather than the historical or methodological organisation. Finally, we try to address a broad audience (6). We do not explain technical details in-depth, which benefits general scholars and practitioners, and at the same time, we try to give specialised scholars an overview of their quickly expanding research field. Altogether, we intend to clarify the relevance of deep RL in production systems and to provide stimuli for potential applications.

3.2. Literature search

For conducting the review, we initially defined the search terms and determined the underlying databases. The found literature is then filtered to obtain the final subset for the later in-depth analysis.

3.2.1. Phase 1 – database and iterative keyword selection

The search databases utilised in our review are the Web of Science (all fields), ScienceDirect (title, abstract or author-specified keywords), and IEEE Xplore (journals), similar to Lohmer and Lasch (2020) or other scholars.

To ensure a representative coverage of the research literature, we defined the keywords in an interactive process and had a rather broad focus, which comprised an algorithmic, a general, and a more specific domain. Within the iterative process, besides *production* and *manufacturing*, we incorporated *assembly*, *automation*, and *industry* as general keywords. To avoid missing any sub-discipline, additional subsets were incorporated into the search and included *quality control*, *maintenance*, and others as listed in Table 2. Because the term of *deep RL* is not always mentioned, we also linked RL with *artificial intelligence*, *deep learning*, and *machine learning*.

3.2.2. Phase 2 – defining inclusion and exclusion criteria

To systematically narrow the scope and ensure a high review quality, we defined several inclusion and exclusion criteria. For quality reasons, we only considered publications from peer-reviewed journals, proceedings, conference papers, and books (as in Light and Pillemer (1984) and Durach, Kembro, and Wieland (2017)). We excluded

Table 2. Defined keywords for the SLR.

Algorithmic keywords			General keywords		Specific keywords
Deep RL[a] OR RL[a] AND	Artificial intelligence OR Deep learning OR Machine learning	AND	Assembly OR Automation OR Industry OR Manufacturing OR Production	OR	Logistics OR Maintenance OR Process control OR Quality control OR Real − time control OR Tool control

[a] RL, Reinforcement learning.

workings papers, pre-prints and other non-peer reviewed publications. We also excluded publications that were not written in English and since significant successes of deep RL were especially observed with the publication from Mnih et al. (2013), we only included papers that were published after 2010.

A thematic definition of the inclusion and exclusion criteria is ensured by the defined research questions and taxonomy framework. Based on our target to identify industrial deep RL applications, we excluded papers that focus primarily on the development of methodologies, theories, or algorithms without transferring the results to a production use case. Review papers were used as appropriate to identify potential additional studies of relevance. Given the focus of our study, we reviewed papers that address the direct application of deep RL in real or simulated production environments and seek to leverage system performances. Only papers, that apply deep RL methods for policy approximation were considered. In contrast, papers dealing with conventional RL methods (i.e. a Q-table) were not reviewed.

3.2.3. Phase 3 – conducting the literature search

The literature search was conducted from December 2020 and a final extract was retrieved from the mentioned databases on February 10, 2021. A summary of the whole process is given in Figure 3 and starts with the aggregation of the articles found in the three databases. In total, 1255 papers were collected based on the defined keywords. Duplicates were removed and years filtered before applying in-depth thematic criteria.

According to Antônio Márcio Tavares, Felipe Scavarda and José Scavarda (2016), to ensure a high search quality, we examined the remaining 809 papers by their title, keywords, and abstract regarding the defined inclusion and exclusion criteria and the research questions. If possible, we already captured the applied algorithms, considered processes, and the application objective. In this step, many papers were excluded due to a missing production context or a non-deep RL implementation, which reduced the number to 141 papers. In the next step, we conducted a full-text review based on the same criteria. Besides capturing the first essential information for the later analysis, the full-text review provided the remaining 91 papers as a basis for the subsequent backward/forward search.

Following the review structure proposed by Webster and Watson (2002), the backward/forward search is an important extension to the previously conducted keyword-based search. Similarly, Greenhalgh and Peacock (2005) underlines the importance of this last literature search step to identify further interdisciplinary literature beyond the self-defined search scope. After this final search, we identified 29 additional papers in scope, resulting in a total set of 120 papers.

3.2.4. Phase 4 – data gathering

To conduct the subsequent literature analysis, we developed a concept matrix with regard to Antônio Márcio Tavares, Felipe Scavarda and José Scavarda (2016) and Webster and Watson (2002), that focused on the initial research questions. The categorisation and coding of the final data set was based on the production discipline, industry or process background including the specific application, optimisation objective, applied deep RL algorithm and neural network, benchmark results, and the application in a simulated and/or real environment.

3.3. Analysis of yearly and outlet related contributions

An initial analysis based on publication years allows conclusions to be drawn about the general research development. Figure 4(a) indicates a strong increase of deep RL publications in a production context since 2018. While 3 papers were published in 2017, there were already 8 in 2018, 27 in 2019, and 69 in 2020. In 2021, 9 papers were published in January / February up to the time of the database query. This indicates the growing relevance of deep RL in a production context and its rising attention within the research community.

One reason for this development could be due to Mnih et al. (2013) as described earlier, who laid a foundation for high performance deep RL in 2013. This also becomes evident in Figure 4(b) in which we compared deep and non-deep RL publications in the Web of Science database (with keywords from Table 2, non-reviewed). While in 2017 1 deep RL and 26 non-deep RL papers were

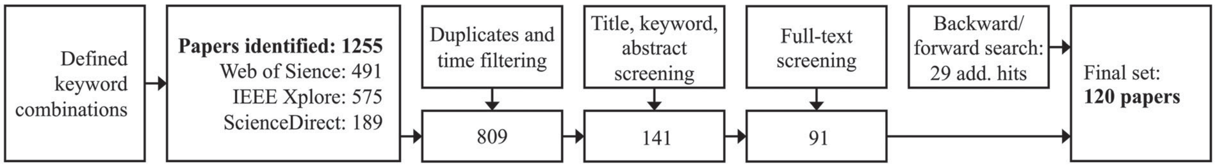
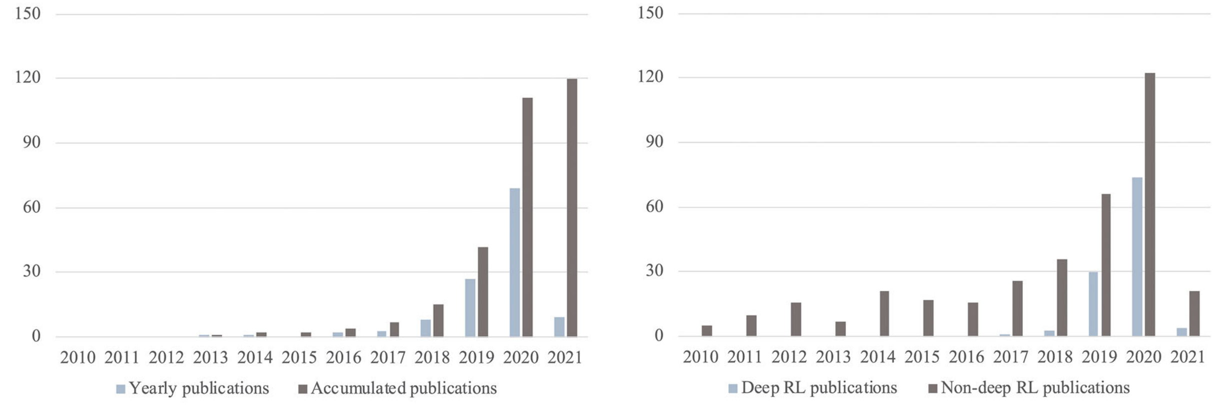

Defined keyword combinations
Papers identified: 1255
Web of Sience: 491
IEEE Xplore: 575
ScienceDirect: 189
Duplicates and time filtering
809
Title, keyword, abstract screening
141
Full-text screening
91
Backward/ forward search: 29 add. hits
Final set: 120 papers

Figure 3. Conducted review process.



(a) Yearly and accumulated deep RL publications (b) Yearly deep and non-deep RL publications

Figure 4. Analysis of yearly deep RL publications, 2021 includes Jan./Feb. (a) Yearly and accumulated deep RL publications (b) Yearly deep and non-deep RL publications.

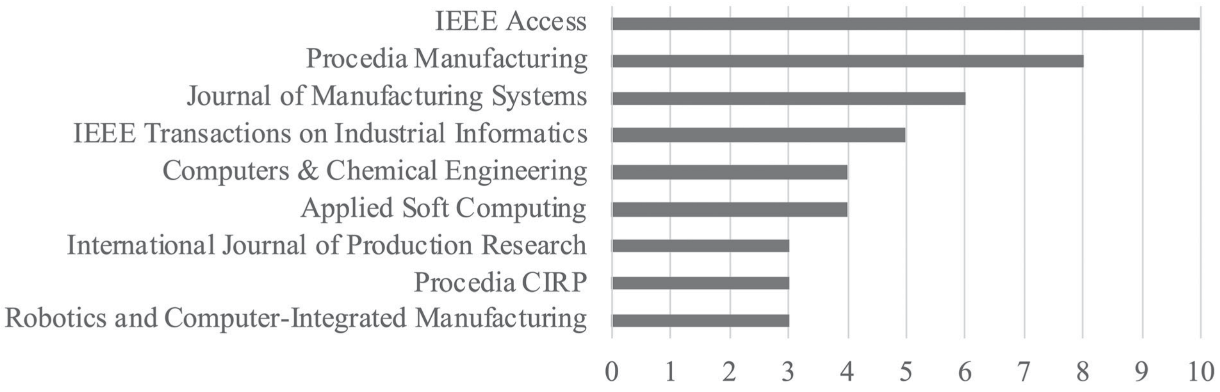


Figure 5. Number of publications per outlet; 2010–2021.

published, there were 74 deep and 122 non-deep publications in 2020. While this suggests a significant increase in both fields, it highlights the ongoing focus on neural network-based RL.

Figure 5 lists the most frequently cited outlets with more than three published papers from 2010 to 2021. Most papers were published in journals (92, 76%) followed by conference papers (14, 12%) and proceedings (14, 12%). In total, the papers were accessed from 54 journals, 16 conferences, and 4 proceedings. This not only indicates the high quality of the selected papers, but also reflects the broad application range of deep RL in various fields of production related systems.

4. Literature analysis

To address RQ1 we first outline existing domains of deep RL applications in production systems. Figure 6 contains the disciplines obtained after the final iterative review step and the respective number of publications.

Most of the reviewed papers were published in the field of process planning followed by scheduling, and assembly. The application landscape covers almost all relevant disciplines in a production system and confirms the ability of deep RL to address a variety of tasks. The further analysis is organised according to the structure indicated in Figure 6.

4.1. Process control

To circumvent a conventional model-based approach and an online adaption to continuous process modifications, Noel and Pandian (2014) initially developed a deep RL approach to control the liquid levels of multiple connected tanks. The controller minimised the target state difference and adjusted inlet flow rates between multiple tanks accordingly. Whereas conventional methods struggle to compensate for large changes in system parameters, the deep RL approach optimised control and simultaneously reduced process fluctuations and overshoot. Spielberg, Gopaluni, and Loewen (2017) and Spielberg et al. (2019) proposed a model-free controller design for single-/multiple-input and -output processes that was applied to various

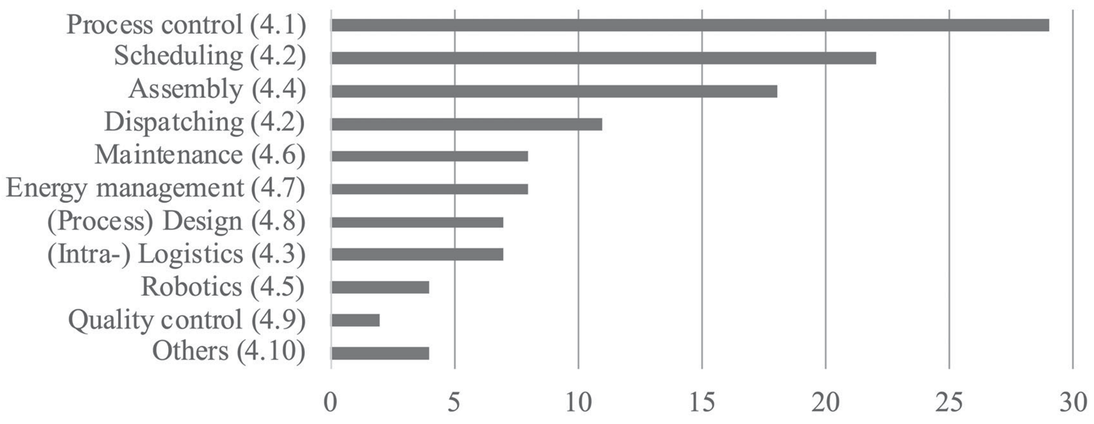


Figure 6. Number of publications allocated to the production disciplines.

application scenarios. The controller reduced maintenance efforts and computation costs and was capable of regulating the desired states and set-points. Similarly, deep RL approaches in chemical-mechanical polishing (Yu and Guo 2020) and microdroplet reactions (Zhou, Li, and Zare 2017), outperformed conventional methods in minimising process deviations and enabled an interactive and data-driven decision making and online process control which reduced temporal and monetary expenditures.

Deep RL outperformed 13 out of 16 conventional benchmarks and improved system performances. To reach such performance, a reward function is required, which transforms process targets into rewards, allowing to learn the optimal policy. The reward design can be based on different target variables, such as real-time profits (Powell, Machalek, and Quah 2020), cost-per-time function (Quah, Machalek, and Powell 2020), or similarity measures based on specified performance criteria (He et al. 2020). The individual goal-oriented design enables a broad application in further applications such as flotation processes to reduce non-dynamic drawbacks of model-based approaches (Jiang et al. 2018), in laser welding to increase process repeatabilities (Masinelli et al. (2020), and others), or in injection molding to broaden up narrow process windows of conventional methods in ultra-high precision processes (Guo et al. 2019). A detailed list of all process control applications and related publications can be found in Table 3. Besides, the table lists the applied algorithm and, if conducted, the performance result compared to conventional benchmarks. A significant portion of the papers conducted their testings in simulated environment and only 4 papers conducted real world testings. The implementation hurdles in the area of process control are large and require highest levels of process reliability, which prevents a rapid implementation for research purposes in real processes.

The majority of publications (79%) utilised policy-based or hybrid algorithms, which benefit from a continuous action space and do not require action discretisation. Thus, process parameters can be set smoothly and do not require a step-wise control approach. Beyond that, the motivation for applying deep RL is often an inaccurate mapping of conventional methods that cannot adequately cope with non-linearities (as in Lu et al. (2016)) or relies too much on error-prone expert knowledge (as in Mazgualdi et al. (2021)). With their adaptive and non-discretised action space, deep RL can thus avoid waste, especially in sensitive processes, and keep processes stable, which might be problematic with static or human-based process modelling (Andersen et al. 2019).

4.2. Production scheduling and dispatching

Already in 1995, Zhang and Dietterich described a neural network based job-shop scheduling approach which demonstrated superior performance and reduced costs for manual system design. Followed by other approaches such as Riedmiller and Riedmiller (1999) or Gabel and Riedmiller (2007), the advantage of deep RL in production planning and control was emphasised early on, but could not prevail, among other reasons, due to the lack of computational resources.

4.2.1. Production scheduling

The complexity of production scheduling is caused by high uncertainties regarding customised products, shutdowns, or similar. To cope with the complexities and to reduce human-based decisions, Lin et al. (2019) proposed a multi-class DQN approach that feeds local information to schedule job shops in semiconductor manufacturing. Based on the edge framework, the DQN demonstrated superior performance and reduced makespans, and average flow times. To reduce the high

Table 3. Summary of deep RL applications in process control.

	Application	Algorithm	Superiority	Source
1	Batch process	DDPG	Superior	Xu, Xie, and Shi (2020)
2	Brine injection process	Actor-critic	Superior	Andersen et al. (2019)
3	Liquid moulding process	DQN	Superior	Szarski and Chauhan (2021)
4	Chemical microdroplet reactions	Actor-critic	Superior	Zhou, Li, and Zare (2017)
5	Colour fading	Actor-critic	Superior	He et al. (2020)
6	Continuously stirred tank reactor	Actor-critic	–	Pandian and Noel (2018)
7	Continuously stirred tank reactor	DQN	Comparable	Powell, Machalek, and Quah (2020)
8	Continuously stirred tank reactor	Actor-critic	Comparable	Quah, Machalek, and Powell (2020)
9	Interacting tank liquid level control	Actor-critic	Superior	Noel and Pandian (2014)
10	Double dome draping	Actor-critic	–	Zimmerling, Poppe, and Kärger (2020)
11	General discrete-time processes	Actor-critic	–	Spielberg et al. (2019)
12	General discrete-time processes	Actor-critic	–	Spielberg, Gopaluni, and Loewen (2017)
13	Goethite iron-removal process	DDPG	–	Chen et al. (2020)
14	Hematite iron ore processing	DQN/PG	–	Li et al. (2020)
15	Laser welding	DQN	–	Günther et al. (2016)
16	Laser welding	Value based	–	Jin and Hongming Gao (2019)
17	Laser welding	Value based	–	Masinelli et al. (2020)
18	Laser welding	Actor-critic	Superior	Zou and Lan (2020)
19	Metal sheet deep drawing	ANN-PSO/PPO	–	Dornheim, Link, and Gumbsch (2020)
20	Non-lin. semi-batch polymerisation	DQN/DPG/REI	Comparable	Ma et al. (2019)
21	One-stage mineral grinding	Actor-Critic	–	Lu et al. (2016)
22	Optical lens manufacturing	PPO	Superior	Guo et al. (2019)
23	Propylene oxide batch polymerisation	DQN	Superior	Yoo et al. (2021)
24	Rot. chemical mechanical polishing	DDPG	Superior	Yu and Guo (2020)
25	Single-cell flotation process	DDPG	Superior	Jiang et al. (2018)
26	Single-cell flotation process	DDPG	–	Jiang et al. (2019)
27	Single-circuit ball mill grinding	DRO	Superior	Guo, Wang, and Zhang (2019)
28	Tempered glass manufacturing	Actor-critic	–	Mazgualdi et al. (2021)
29	Well surveillance	Actor-critic	Superior	Tewari, Liu, and Papageorgiou (2020)

setup and computational costs of conventional solutions in job-shop scheduling, Liu, Chang, and Tseng (2020) and Baer et al. (2019) adopted a self-learning multi-agent approach to meet local and global production objectives and to ensure an increased adaptation to prevent rescheduling cost. To train multiple agents Baer et al. (2019) employed a multi-stage learning strategy in which a single agent was trained locally first, while others applied chosen heuristics. Subsequently, all agents were trained individually and finally optimised together towards a global goal. Besides, Baer et al. (2020) demonstrated the agent's ability to adapt to new scenarios and proofed its scalability. The training of 700 scheduling topologies took only twice as long as the training of a single one. The deep RL policy learned basic task principles and modified its policy slightly concerning the new task specifics and thereby reduced re-configuration times and costs compared to conventional methods.

In total, 89% of the benchmarked deep RL implementations increased the scheduling performance and reached lower total tardiness, higher profits, or other problem-specific objectives as indicated in Table 4. Zhou, Zhang, and Horn (2020) managed to minimise the completion time of all given tasks, for random incoming orders. Similarly, Wu et al. (2020) demonstrated that deep RL-based rescheduling can operate faster and more efficiently than heuristics. The deep RL approach reduced CPU times remarkably for the high volatile medical mask production in times of Covid-19. Besides mask production, deep RL demonstrated superior performances in batch processing which reduced tardiness for repair scheduling operations (Palombarini and Martinez 2018; Palombarini and Martínez 2019), in chemical scheduling to increase profitability and deal with fluctuating prices, shifting demands, and stoppages (Hubbs et al. 2020), and in paint job scheduling to minimise costs of colour changeovers within the automotive industry (Leng et al. 2020). Discipline-specific scheduling objectives were addressed by Lee, Cho, and Lee (2020), who increased sustainability and minimised tardiness in injection mold scheduling, or by Xie, Zhang, and Rose (2019) who reduced total throughput time and lateness in single-machine processes.

From an industry perspective, the semiconductor industry is one of the most competitive and capital-intensive. Interconnected machines must operate at full capacity, and production schedules need to be continuously optimised online (Kang, Catal, and Tekinerdogan 2020). Due to a large number of machines and process steps, the die attach and wire bonding process poses a major challenge that cannot optimally be solved by single heuristics. To cope with the complexities, Park et al. (2020) feed all relevant process information such as setup status continuously into the PPO neural network. It was able to outperform conventional heuristics such as shortest setup or processing time and reduced total makespan and computation times after training. A further increase in generalisation was reached by Park et al. (2021) by applying a graph neural network (GNN). The GNN learned the basic spatial structure of the problem in form of a graph that could be transferred to new problems and adapted its mapped policy. Thus, the GNN-PPO was not only able to adapt to novel job shop problems, but also outperformed algorithms that were configured scenario-specific.

Based on all reviewed papers in the field of production scheduling, 67% applied value-based algorithms. These assume a discrete action space, which must be determined beforehand. However, for scheduling-related problems, the action space can often be discretised according to possible transition actions such as *transfer* or *idle* (as in Shi et al. (2020)). It is noticeable that in comparison to process control, even fewer approaches have been adopted in a real environment. In scheduling and additionally in the subsequent dispatching and logistics section, a fast implementation of the scheduling policies in an established production environment would be complex and increase research efforts significantly.

4.2.2. Production dispatching

Personalised production has an enormous impact on the complexity of production control due to individual product configuration options. Depending on the customer requirements, the products must be dispatched to where they can be processed, under consideration of several technical and logistic constraints and optimisation variables (Waschneck et al. 2018a).

To meet the requirements in wafer fabrication dispatching, Altenmüller et al. (2020) implemented a single-agent DQN that processed 210 data points as a single state input (such as machine loading status or machine setup). This enabled the DQN to meet strict time constraints better than competitive heuristics (TC, FIFO) while reaching predefined work-in-progress (WIP) targets as a secondary goal. Stricker et al. (2018) and Kuhnle et al. (2020) proposed a single-agent adaptive production control system that maximised machine utilisation and reduced lead and throughput times compared to conventional methods that struggle partially known environments. Waschneck et al. (2018a) proposed a multi-agent system to meet flexible objectives within wafer processing and enable higher flexibilities with fewer delays. Similar to Waschneck et al. (2018b), the algorithms targeted plant-wide parameters to reduce the risk of a local operation optimisation. Besides, the simulations considered complex job shop specifics such as re-entrant flows, sequence-dependent setups, and varying processing times, reaching comparable performances against multiple heuristics. For general production dispatching, Marc-André and Fohlmeister (2020) introduced a multi-agent system with global performance objectives to avoid local optimisation tendencies. Although the agents received detailed local state information, they not only selected the fastest local dispatching actions but also improved the global logistics performance. Further, the distributed agents enabled real-time responses, a feature also emphasised by Kumar, Dimitrakopoulos, and Maulen (2020) for the short-term value stream adaptation in a copper mining complex. Based on the current mining process and component data, the single-agent framework allowed to deliver continuous updates regarding the latest plant status and increased the expected net present value by 6.5%. Considering capital constraints in production, Kanban or Conwip cards are often employed to limit WIP levels. As an alternative to those conventional pull production controls and to optimise local and global production indices in parallel, Tomé and Azevedo (2019) proposed a deep RL algorithm that balanced conflicting throughput and WIP level targets. Despite the trade-off between these, WIP levels were reduced by 43% compared to conventional methods through dynamic adjustments without affecting the total throughput.

A mixed-rule dispatching approach was proposed by Luo (2020) and Heger and Voß (2020) for general job shop systems to enable a dynamic dispatching adaptation to changing production conditions. Based on current state information, the algorithm determined which of the predefined rules (i.e. Heger and Voß: *SPT, EDD, FIFO, SIMSET*) should be activated in the current situation to reduce the mean and total tardiness. Table 4 briefly summarises the reviewed literature and contains the implemented algorithms of the respective papers and their performance results compared to conventional methods.

4.3. (Intra-) logistics

The review results for intralogistics are briefly summarised in Table 4. Beginning with Malus, Kozjek, and Vrabič (2020), an intralogistics-related dispatching solution was implemented to meet real-time requirements and handle a rapidly changing production by utilising autonomous mobile robots (AMRs). Based on the observations of the individual agents, they could negotiate with each other and virtually raised bids for orders. Similarly, Feldkamp, Bergmann, and Strassburger (2020) simulated

Table 4. Summary of deep RL applications in production scheduling, dispatching, and (intra-) logistics.

	Application	Algorithm	Superiority	Source
	Scheduling			
30	Chemical scheduling	A2C	Superior	Hubbs et al. (2020)
31	Cloud manufacturing	DQN	Superior	Dong et al. (2020)
32	Cloud manufacturing	PG	Superior	Zhu et al. (2020)
33	Dynamic scheduling	DQN	–	Zhou, Zhang, and Horn (2020)
34	Dynamic scheduling	DQN	Superior	Hu et al. (2020)
35	Flow shop scheduling	Reinforce	Superior	Wu et al. (2020)
36	Job-shop scheduling	DDPG	Superior	Liu, Chang, and Tseng (2020)
37	Job-shop scheduling	PPO	Superior	Park et al. (2021)
38	Job-shop scheduling	DQN	–	Baer et al. (2020)
39	Job-shop scheduling	–	–	Baer et al. (2019)
40	Job-shop scheduling	DQN	Superior	Zhou et al. (2021)
41	Job-shop scheduling	DDDQN	Superior	Han and Yang (2020)
42	Job-shop scheduling	(M)DQN	Superior	Lin et al. (2019)
43	Lot scheduling	PPO	Superior	Rummukainen and Nurminen (2019)
44	Mold scheduling	DQN	Superior	Lee, Cho, and Lee (2020)
45	Multichip production	DQN	Superior	Park et al. (2020)
46	Packaging line scheduling	DQN	Superior	Chen et al. (2019)
47	Paint job scheduling	Double DQN	Superior	Leng et al. (2020)
48	Parallel, re-entrant production	DQN	Comparable	Shi et al. (2020)
49	Rescheduling	DQN	Superior	Palombarini and Martinez (2018)
50	Rescheduling	DQN	Superior	Palombarini and Martínez (2019)
51	Single machine scheduling	DQN	Comparable	Xie, Zhang, and Rose (2019)
	Dispatching			
	Application	**Algorithm**	**Superiority**	**Source**
52	General job-shop	DQN	Comparable	Marc-André and Fohlmeister (2020)
53	General job-shop	double DQN	Superior	Luo (2020)
54	General job-shop	DQN	Comparable	Heger and Voß (2020)
55	General job-shop	Reinforce	Superior	Zheng, Gupta, and Serita (2020)
56	Mining materials flow	PG	Superior	Kumar, Dimitrakopoulos, and Maulen (2020)
57	Wafer fabrication	DQN	Comparable	Waschneck et al. (2018a)
58	Wafer fabrication	TRPO	Comparable	Kuhnle et al. (2020)
59	Wafer fabrication	DQN	Comparable	Waschneck et al. (2018b)
60	Wafer fabrication	DQN	Superior	Stricker et al. (2018)
61	Wafer fabrication	DQN	Superior	Altenmüller et al. (2020)
62	WIP bounding	DQN	Superior	Tomé and Azevedo (2019)
	(Intra-) logistics			
	Application	**Algorithm**	**Superiority**	**Source**
63	AGV scheduling	DQN	Superior	Feldkamp, Bergmann, and Strassburger (2020)
64	AGV scheduling	DQN	Superior	Hu et al. (2020)
65	AMR dispatching	TD3	Superior	Malus, Kozjek, and Vrabič (2020)
66	Item betching into trays	PPO	–	Hildebrand, Andersen, and Bøgh (2020)
67	QoS service composition model	duelingDQN	Superior	Liang et al. (2021)
68	Syringe filling process	doubleDQN	Superior	Xia et al. (2020)
69	Three-grid sortation system	DQN	–	Kim et al. (2020)

a self-regulating modular production system. Depending on current job information, station status, and others, the algorithm determined the optimal machine and reduced lead times compared to the benchmarked methods. In another approach, Hu et al. (2020) implemented a mixed rule dispatching approach that determines the dispatching rule (*FCFS, STD, EDD, LWT, NV*) for an automated guided vehicle (AGV) depending on its observed state which reduced the makespan and delay ratio by approximately 10% compared to the benchmarks.

Regarding conveyor systems, Kim et al. (2020) proposed a deep RL control to enable a faster product distribution for a 3-grid sorting system in which all of the 9 fields and corresponding inputs and outputs were controlled by respective agents. The pick and place of items from a conveyor belt into baskets was investigated by Hildebrand, Andersen, and Bøgh (2020). To reach a pre-defined weight, the trays should still be filled quickly to prevent dead-locks. Without an initial parameter tuning, which would have been necessary for conventional probability-based methods, the PPO reached a remarkable success rate of 48% after training. A further collaborative task completion of two robots for adaptive stacking was considered in Xia et al. (2020) which highlighted the flexible virtual commissioning abilities and demonstrated an above-human performance.

4.4. Assembly

A significant share of the reviewed assembly-related papers focused on the peg-in-hole task (56%). It comprises the insertion of a specific object into a hole under defined assembly conditions, utilising a robotic arm in most cases. To avoid large fluctuations in execution and to ensure a high level of safety, most papers utilised a post-processing force controller that processes the neural network outputs (as in Kim, Ahn, and Song (2020)).

The deep RL implementation was often motivated by disadvantages of conventional algorithms such as limited adaptability (Li et al. 2019), complex online optimisation processes (Inoue et al. 2017) or the need for re-programming in case of new tasks due to hand-engineered parameters (Luo et al. 2018).

Beginning with hole position uncertainties, Beltran-Hernandez et al. (2020) trained a transfer learning supported deep RL algorithm to fit a cuboid-shaped plug into a hole with 0.1 mm tolerance and reached a 100% success rate. Also, the insertion of electronic connectors (success rate: 65%), Lan connectors (60%), and USB connectors (80%) was investigated but reached lower success rates.

For contact-rich tasks (Kim, Ahn, and Song 2020; Lämmle et al. 2020; Beltran-Hernandez et al. 2020) proposed a imitation learning supported force-regulated approach consisting of hole approach, alignment, and insertion for the square-shaped peg assembly (tolerance: 0.1 mm). For smaller tolerances in high-precision assembly, Zhao et al. (2020) and Inoue et al. (2017) reached success rates of up to 86.7% and 100% with tolerances of 0.02 mm and 0.01/0.02 mm, respectively. Whereas (Zhao et al. 2020) thereby minimised the number of required interactions, Inoue et al. (2017) was able to significantly reduce online parameter adjustment efforts that are required by conventional methods. The insertion of the peg into a deformable hole with a smaller diameter was investigated by Luo et al. (2018) who utilised a force-torque controller for task completion.

For the double peg-in-hole task and a tolerance of 0.04mm for each peg, Xu et al. (2019) reached a success rate of 100%. In case of a changed start position, the success rate was reduced and required re-training. Not only stiff but also dangling pegs have been investigated by Hoppe et al. (2019), that required a contact-rich assembly. Through a combined global state space exploration and learning by demonstration strategy, the DDPG reached a 100% success rate.The learning by demonstration was also investigated by Wang et al. (2020), taking into account bigger arm and fine hand motions. Assuming different peg objects with a tolerance of 4.2 mm and a one-shot demonstration, a success rate of 67% was reached. The assembly of a circuit breaker housing was addressed by Li et al. (2019). Divided into free movement, movement under contact, and insertion phase, the two housings with four mounting spots were assembled with success rates of up to 88%.

Other deep RL applications included the vision-based insertion of a Misumi Model-E connector (success rate: up to 100%, Schoettler et al. (2020)), a long/short-term memory supported shoe-tongue assembly (up to 97%, Tsai et al. (2020)), and a space-force controller and force/torque information supported gear-set assembly (up to 100%, Luo et al. (2019)).

A multi-component assembly sequence planning approach to increase human–robot collaboration efficiencies was proposed by Yu, Huang, and Chang (2020). Assuming an adjustable desk as an example, the scheduling process was transformed into a chessboard-shaped planning structure that was able to complete planning significantly faster than conventional methods. Besides, to increase planning efficiencies while obtaining better generalisation, Zhao et al. (2019) combined a DQN with curriculum learning and parameter transfer techniques. Compared to a simple DQN, this increased the learning speed and adaptability to other environments.

Remarkably, assembly research conducted the highest number of real-world testings, and 15 out of 18 reviewed papers transferred results to reality under the prevailing conditions and on real hardware. Deep RL based assembly research benefits from low preconditions and well-scoped scenarios compared to the other domains, which reduces testing complexity and safety constraints.

The summary of applications in Table 5 differs from the previous ones due to the lack of compared algorithms. Only in 4 cases a benchmark was compared, which was outperformed by the deep RL algorithm in each case. Instead, the general task itself, as well as the specific use-case were referred for further classification.

4.5. Robotics

To obtain a significantly smoothed motion planning, Scheiderer, Thun, and Meisen (2019) compensated disadvantages of existing RL planning approaches due to time discretisation. If the robot exceeded a certain trajectory mark, the observation of the next step was triggered, and a Bézier curve was generated that aligned smoothly with the previous one. Similarly, Li et al. (2020) investigated the smoothing of CNC trajectories to enable high-speed machining. Based on a high-speed $x-y$ motion platform, a real-time smoothing could be realised, which processed a pre-computed tool trajectory and smoothed

Table 5. Summary of deep RL applications in assembly and robotics.

	Assembly			
	Application	**Use-case**	**Algorithm**	**Source**
70	Sequence planning	Building block model	DQN	Watanabe and Inada (2020)
71	Sequence planning	Lift desk assembly	As AlphaGoZero	Yu, Huang, and Chang (2020)
72	Sequence planning	Seven parts assembly process	DQN	Zhao et al. (2019)
73	High precision insertion	Circuit breaker housing	DQN	Li et al. (2019)
74	High precision insertion	Gear set assembly	Model-based	Luo et al. (2019)
75	High precision insertion	Peg-in-hole	DQN	Inoue et al. (2017)
76	High precision insertion	Peg-in-hole	DQN, DDPG	Li et al. (2019)
77	High precision insertion	Peg-in-hole	SAC	Zhao et al. (2020)
78	Insertion task	Peg-in-hole (contact-rich, deform.)	AC	Luo et al. (2018)
79	Insertion task	Double peg-in-hole	DDPG	Xu et al. (2019)
80	Insertion task	Double peg-in-hole (contact-rich)	DDPG	Hoppe et al. (2019)
81	Insertion task	Peg-in-hole	DDPG	Wang et al. (2020)
82	Insertion task	Peg-in-hole	DDPG	Lämmle et al. (2020)
83	Insertion task	Peg-in-hole (cuboid)	SAC	Beltran-Hernandez et al. (2020)
84	Insertion task	Peg-in-hole (square)	DDPG	Kim, Ahn, and Song (2020)
85	Insertion task	Ring-insertion	SAC	Beltran-Hernandez et al. (2020)
86	Plug insertion tasks	Model-E connector	SAC, TD3	Schoettler et al. (2020)
87	Shoe tongue assembly	Soft fabric shoe tongues	DQN	Tsai et al. (2020)

	Robotics			
	Application	**Use-case**	**Algorithm**	**Source**
88	Intelligent gripping	Find optimal grasp position	PPO	Park et al. (2020)
89	Motion planning	Real-time CNC traj. smoothing	DQN, DDPG	Li et al. (2020)
90	Motion planning	Bézier curve trajectory smoothing	DDPG	Scheiderer, Thun, and Meisen (2019)
91	Visual control	Low-cost servo control	Sarsa, DQN	Miljković et al. (2013)

out the path, calculated tool velocities, and emitted servo commands. An early image-based control of servos by deep RL was proposed in Miljković et al. (2013). The robot processed the captured images as states which were processed by a SARSA or DQN and ejected as spatial camera velocities. Thus, high robustness and accuracy of the control process were reached despite calibration errors and sensor noises. Following the same structure as the assembly domain, Table 5 summarises the main review results and includes the general application and the specific use case due to the lack of benchmarks.

4.6. Maintenance

The interaction of several linked machines in a serial production line was considered by Huang, Chang, and Arinez (2020). Based on a large state space that contained buffer levels, operating inputs, and fault indicators for each machine, the algorithm made decisions about which individual machines needed to be turned off at a time for service. Conventional methods often rely on the static recommendations of machine manufacturers and do not take system dependencies into account. In comparison, deep RL reduced the average maintenance costs by approximately 20% compared to a run-to-failure strategy, 7% compared to an age-dependant, and 5% compared to an opportunistic maintenance strategy. The same interdependencies between multiple components with competing failure probabilities were considered by Zhu et al. (2020) to avoid static and ineffective maintenance limits of conventional methods in large-scale systems. In several scenarios, the deep RL algorithm was able to reduce maintenance cost in multi-component systems without requiring experience-based or predefined thresholds.

The issue of limited resources to perform maintenance due to insufficient monetary, technical, and human capital was considered by Liu, Chen, and Jiang (2020). Conventional methods only take the success of a single maintenance mission as a success factor, but neglect possible follow-up missions. Compared to benchmarks, deep RL thus demonstrated a 30% higher number of successful maintenance missions. Regarding, rotary machines fault diagnosis, Dai et al. (2020) and Ding et al. (2019) employed deep RL to detect faults from machine data at an early stage in real environments. Whereas Dai et al. focused on the detection of faulty components such as a cracked gear, Ding et al. focused on non-linear correlations between possible fault conditions by measuring raw sensor signals. Both times, errors could be detected at an early stage without the need for manual tuning efforts, expert experience, or pre-filtering of the data as required by conventional methods.

Despite conducting only three benchmarks, the deep RL algorithms demonstrated superior maintenance-specific performance in all of these. Additional maintenance related publications of deep RL in recent years are listed in Table 6.

4.7. Energy management

In modern production, not only the maximum process performance but also the energy consumption and environmental impact become more crucial. To meet the challenge of greener production, Leng et al. (2021) addressed the order acceptance in the energy- and resource-intensive PCB fabrication under the assumption of resource constraints and environmental metrics. Compared to conventional methods (FIFO, random forest), the deep RL algorithm was able to increase profits and minimise carbon consumption, while optimising lead time and cost. Considering a steel powder manufacturing process, Huang et al. (2019) proposed a model-free control design to optimise the energy consumption plan based on current energy costs and individual process components (i.e. atomizer, crusher). Compared to conventional methods, which often require a complex system model and neglect price fluctuations, the controller adjusted the production schedule to the electricity prices, which reduced energy costs by 24%. The same objective was addressed by Lu et al. (2020) for a lithium-ion battery assembly process which reduced electricity costs by 10%.

Anapproach to enable more energy-efficient and high reliable transmissions in low latent networks was proposed by Yang et al. (2020). Based on the channel status and other indicators, the algorithm selected radio frequency or visible light communication. It assigned an appropriate channel and performed the transmission power management. Thereby, energy efficiency, number of successful services, and latency were improved and a higher fulfillment of compulsory quality-of-service requirements was accomplished. Other applications included the single-machine energy optimisation (Bakakeu et al. 2018), blast furnace gas tank energy scheduling for steel industry (Zhang and Si 2020), and others as listed in Table 6.

A total of 4 benchmarks were carried out in the field of energy management, in which the deep RL algorithms again outperformed conventional ones. However, no real-world testing was conducted in the domain of energy management. Similar to previous categories, this would have entailed extraordinarily high expenses and would have caused a significantly increased implementation efforts at an early stage.

4.8. (Process) design

Beginning with integrated circuit design, Liao et al. (2020) addressed the global routing process, which became a major challenge due to increased transistors densities and multiple design constraints. To cope with the complexity, Liao et al. modelled the circuit as a grid graph from which information was fed into the DQN router and outperformed the conventional A∗ approach.

Oh et al. (2020) proposed a deep RL algorithm for the design and fine-tuning of notch filters, which are commonly used in servo systems to suppress resonances. In complex cases, however, the filters not only need to be deployed in large numbers but also fine-tuned manually based on expert knowledge. The proposed notch tuning automatism avoided these and optimised several

Table 6. Summary of deep RL applications in maintenance, energy management, and (process) design.

	Maintenance			
	Application	Use-case	Algorithm	Source
92	Condition-based maintenance	Minimise cost rates in multiple stages	Dou.DQN	Zhang and Si (2020)
93	Machine fault diagnosis	Gear root crack analysis	DQN	Dai et al. (2020)
94	Machine fault diagnosis	Rolling bearing fault	DQN	Ding et al. (2019)
95	Oportunistic maintenance	Minimise prod./ maint. interference	PPO	Kuhnle, Jakubik, and Lanza (2019)
96	Real-time prev. maintenance	Minimise long-run costs in serial prod.	Dou.DQN	Huang, Chang, and Arinez (2020)
97	Selective maintenace	Maximise maint. mission success	AC	Liu, Chen, and Jiang (2020)
98	Self-diagnosis and self-repair	Optimise self-repair in prod. lines	DQN	Epureanu, Aydin Nassehi, and Koren (2020)
99	Sensor-driven maintenance	Calc. remaining useful (turbofan)	DQN	Skordilis and Moghaddass (2020)
	Energy management			
	Application	Use-case	Algorithm	Source
100	Energy system balancing	Tank level scheduling	DQN	Zhang et al. (2020)
101	Multi-agent energy optimisation	CPS energy coordination	AC	Bakakeu et al. (2020)
102	Network resource management	Energy efficient RF/VLC network	DQN	Yang et al. (2020)
103	PCB order acceptance	Real-time order acceptance decisions	DQN	Leng et al. (2021)
104	Production-energy schedule opt.	Single machine optimisation	DQN	Bakakeu et al. (2018)
105	Production-energy schedule opt.	Lithium-ion battery assembly	DDPG	Lu et al. (2020)
106	Production-energy schedule opt.	Steel powder manufacturing process	AC	Huang et al. (2019)
107	Sustainable joint energy control	Two machine, one buffer system	DQN	Hu et al. (2019)

(*continued*)

Table 6. Continued.

	(Process) Design			
	Application	Use-case	Algorithm	Source
108	Clamping position optimisation	Milling machine	SAC	Samsonov et al. (2020)
109	Computer-aided process planning	Constrained machining	AC	Wu et al. (2021)
110	Integrated circuit design	Global IC routing	DQN	Liao et al. (2020)
111	Notch filter design	Industrial servo systems	DDPG	Oh et al. (2020)
112	Rectangular item placement	2D Strip Packing	DQN	Zhu and Xiang Dong Li (2020)
113	SaaS remote training	Heavy plate rolling	SAC	Scheiderer et al. (2020)
114	Tool path design	Geometric impeller optimisation	DDPG	Zhou et al. (2020)

notch filters simultaneously and successfully stabilised a belt-drive servo system. Zhou et al. (2020) addressed the machining optimisation of centrifugal impellers for a five-axis flank milling processing. By considering aerodynamic and machining parameters, an optimised path planning for the machine tool was developed, which reduces development time and cost.

Among the publications listed in Table 6, other design approaches included 2D-strip packing to improve space utilisation in Zhu and Xiang Dong Li (2020) which reduced average gaps by 20% compared to several benchmark algorithms, or the design of a SaaS architecture in Scheiderer et al. (2020) which significantly reduced optimisation times in heavy-plate rolling compared to manual tuning.

4.9. Quality control

The field of quality control is affected by the increased product diversification and must adapt accordingly to carry out necessary component inspections. To support the workforce in quality related tasks, cobots can contribute to more stable processes. Brito et al. (2020) addressed the collaborative cooperation to combine the accuracy of the robot with the flexibility of the workforce. In case of an unforeseen inspection incident, the workforce taught the robot its new path, which was learned and reproduced by the DDPG. Unlike other methods that require an interruption of the production process, the DDPG enabled an online adaptation and significantly increased productivity and reduced stoppages.

Another approach for real-time quality monitoring of additive manufacturing processes was proposed by Wasmer et al. (2019). Conventional methods often rely on temperature data or high-resolution images, which have difficulties in reflecting the processes below the surface. To provide further process information, the implemented algorithm took acoustic emissions as an input for the process analysis and could thereby derive a pore concentration based quality categorisation with an accuracy of up to 82% in real testings.

4.10. Further applications

Further categories with single publications are listed in Table 7. These include specific topics such as building an agent swapping framework to allow learning in a non-real-time environment and execution in a real-time environment (Schmidt, Schellroth, and Riedel 2020) or the deep RL based selection of optimal prediction models in the semiconductor manufacturing domain to cope with demand fluctuations and avoid shortages and overstock (Chien, Lin, and Lin 2020).

5. Implementation challenges and research agenda

In the previous section, the broad application base and benefits associated with the deployment of the deep RL algorithms were highlighted. Nevertheless, there are some challenges and hurdles that must be overcome that

Table 7. Summary of deep RL applications in quality control and further applications.

	Quality control			
	Application	Use-case	Algorithm	Source
115	In-situ quality moniitoring	Subsurface dynamics analysis	sim. AlphaGO	Wasmer et al. (2019)
116	Quality inspection	Path teaching and adaption	AC	Brito et al. (2020)

	Further applications			
	Application	Use-case	Algorithm	Source
117	Novel PLC learning/acting arch.	Real-time framework	–	Schmidt, Schellroth, and Riedel (2020)
118	Select opt. demand forecast model	Semiconductor components	DQN	Chien, Lin, and Lin (2020)
119	Multi-task policy generalisation	Mfg. system with various tasks	DQN	Wang et al. (2019)
120	Investigation of malicious behaviours	Function-/performance attacks	DQN	Liu et al. (2021)

Table 8. Summary of the key findings from the review analysis.

Production domain	#Publications	Most frequent algorithm	Most frequent neural netw.	Simulation-only share	Superiority (#benchmarks)
Process control	29	AC	FFNN	86%	81% (16)
Scheduling	22	DQN	FFNN	100%	89% (19)
Dispatching	11	DQN	FFNN	100%	55% (11)
(Intra-) Logistics	7	DQN	FFNN	86%	100% (5)
Assembly	18	DQN/DDPG	FFNN	17%	100% (4)
Robotics	4	DQN/DDPG	FFNN	75%	– (0)
Maintenance	8	DQN	FFNN	75%	100% (3)
Energy Management	8	DQN	FFNN	100%	100% (4)
(Process) Design	7	(S)AC	FFNN	71%	100% (4)
Quality Control	2	AC/AlphaGo	FFNN	0%	– (0)
Others	4	DQN	FFNN	75%	100% (2)

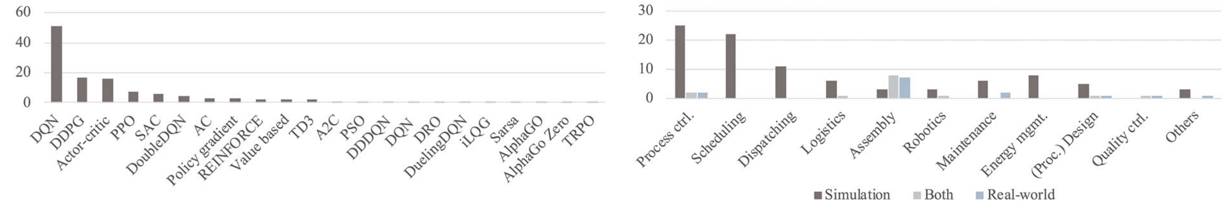
prevent an extensive deployment (RQ2) and need to be addressed in future research (RQ3).

5.1. Implementation challenges and research gaps

The key insights of the review analysis are summarised in Table 8. The table is aligned in its sequence with the previous chapter and comprises the most frequently applied algorithms and neural networks as well as the simulation-only and superiority share (related to the conducted benchmarks).

Table 8 highlights some of the challenges we identified during the literature review. We categorised those into the following 4 major and subsequent minor application challenges and research gaps.

- ***Algorithm selection:*** After identification of a potential implementation, the question arises which algorithm and parameters should be used for the planned scenario. Although these have a significant impact on the resulting performance, there are no or only a few guidelines that can assist during the selection and parameter optimisation process. As mentioned by Rummukainen and Nurminen (2019), Yoo et al. (2021), and others, this selection is a central issue that can worsen the resulting performance and hinder the full development of deep RL capabilities. Table 8 and Figure 7(a) demonstrate the reliance on standard algorithms that may result from missing guidelines. A majority of the reviewed papers implemented a DQN, although possible improvements like the doubleDQN can significantly improve performances (van Hasselt, Guez, and Silver 2016). As one of the few examples, Li et al. (2019) thus improved success rates by 13% to 94% utilising a DDPG for a robot assembly process compared to a DQN. Similarly, a significantly improved performance was reached through learning rate and batch size modifications in Baer et al. (2020).
- ***Further modifications:*** A majority of the papers utilised a common (convolutional) neural network (85%). Only a small fraction utilised LSTM (6%) or recurrent neural networks (1%), which can better reflect long-term experiences. Besides, only four papers compared more than one network and six papers more than one RL algorithm, leading to a gap in benchmarks and performance correlations. Other extensions such as a prioritised experience replay were only applied occasionally although they might increase production performance significantly, similar to the testings within the Atari environment (Schaul et al. 2016).
- ***Transfer of results:***Another major challenge is the transfer of the simulation results to real-world scenarios. Overall, 76% of the papers have validated the proposed solution within simulations. Only 24% of the papers conducted real-world testings, half of which considered either a purely real-world scenario or both. The percentage in assembly was particularly high, as 83% of tests were conducted in real or simulation-based environments. Assembly benefits in particular from confined and segregated environments, which limit hurdles and mitigate risks. In contrast, no real testing was carried out in production scheduling and dispatching as indicated in Figure 7(b). Particularly high safety and reliability-related entry hurdles must be met, besides high system implementation efforts, that prevent large-scale and rapid testings in those fields. Besides, simulations only obtain a simplified representation of the real problem. Due to the considerable differences in complexity between the simulation and real applications, a reduced performance of the approaches after the real-world transfer is to be expected. In particular, the implementation into large real-world systems is rather challenging and has to cope with many unconsidered parameters and a non-preprocessed set of data.



(a) Number of implemented algorithms

(b) Number of conducted testings

Figure 7. Quantitative analysis of applied algorithms and testing environments.

- ***Local optimization:*** In addition to the aforementioned challenges, there is the risk that the algorithms only perform local optimisation (Marc-André and Fohlmeister 2020). As discussed by Rossit, Tohmé, and Frutos (2019), smart scheduling, in particular, is composed of decentralised structures in which multiple deep RL agents can interact and perform their tasks in the defined task domain. This might result in potential small-scale control loops that are optimised intrinsically and exploit local information, but neglect larger interdependencies. Besides, a non-optimal problem-solving strategy may arise from a lack of exploration of the state and action space, resulting in the selection of non-optimal actions and a non-existent optimality guarantee (Spielberg et al. 2019; Guo et al. 2019).
  In addition to the challenges mentioned above, others arise from exponentially growing action and state spaces in complex production systems that require high computational efforts or multitasking scenarios that can not be managed by a single agents (Wang et al. 2019; Beltran-Hernandez et al. 2020). Besides, non-smooth execution due to jumps in the policy decision can result in the inability to execute optimal actions and negatively impact process qualities (Noel and Pandian 2014). Last, differences in training performance (Ma et al. 2019) or vibration during the training of complex tasks (Shi et al. 2020), can lead to less repeatable processes and lower predictability, resulting in low reliability and raising safety concerns.

5.2. Future research agenda

Although the hurdles and challenges described above do not yet enable a full-scale adoption of deep RL in production systems, further efforts can assist in accelerating the process towards industrial maturity. The bullet points below address the outlined challenges and provide research colleagues and practitioners incentives for future research.

- ***Standardised implementation approach:*** Future deep RL based production research can incorporate more insightful benchmarks by considering advanced algorithms, modifications, and parameter tuning within the same simulations. Similar to van Hasselt, Guez, and Silver (2016) in the Atari environment, this could yield a significant increase in performance without causing high adaption efforts. To assist future research, the benchmarks could additionally serve as a basis to derive further guidance for optimisation and control problems with similar state and action spaces circumventing expert advice needed for a fast system adoption and applicability.
  The generation of prototype evaluations can also benefit from the definition of model environments, similar to the Atari environment. Frameworks such as the *SimRLFab* for production dispatching (Kuhnle 2020) can be integrated quickly and enable algorithm benchmarks without requiring large implementation efforts.
- ***Accelerated simulation to real-world transfer:*** To enable a faster integration to real production environments, the respective system requirements must be satisfied. This primarily involves the consideration of safety-relevant parameters to avoid critical actions and threats. In this context, a constraint-driven approach in non-deep RL was proposed by Ge et al. (2019), in which permitted actions were limited through preliminary filtering, or by Xiong and Diao (2021) who proposed a safety-based evaluation of policy robustness. Further studies should approximate the simulations and frameworks to real-world conditions even more, which includes consideration of hard real-time requirements, significant parameters, uncertainties, and indeterminacies. Thus, by establishing a digital twin that copies reality, hardware-in-the-loop environments (HiL), and separate training and testing sequences, the gap between research and practical testings is narrowed and the transfer of results and validation can be accelerated and performed with less risk. The HiL approach would enable a real-time use of machine data and also address the data quality issue. In this context, data pre-processing is essential and may be integrated in the simulation, but can

only be matched to reality with great efforts. The same applies to the state-action-reward design, which must process the changed or even additional input variables and cope with unknown process variables. The algorithms could be thoroughly investigated under real conditions in the hybrid HiL environment, parameters optimised and the real system dynamics between input and output variables analysed. Especially, domains that require large-scale implementations like production scheduling, might benefit from such a step-by-step HiL approach that anticipates transfer issues and identifies unknown disturbances at an early stage.

- ***Generalisability:*** The ability of the agents to adapt more effectively to changing production conditions should be considered to further optimise their learning stability and robustness. Even though this has already been considered by learning general behaviours instead of specific policies in Baer et al. (2020), it was also observed that small deviations of the starting conditions led to performance reductions (as in Beltran-Hernandez et al. (2020)). Future research should therefore focus on methods that enable agents to adapt to different scenarios as quickly as possible. This not only includes a particularly fast re-training under changed conditions, but also an accelerated transfer of the adapted policy to the real agent. Such a swift transfer could be facilitated by applying a permanently trained agent within the digital twin and a subsequent policy transfer. Another approach to increase generalisability and performance under changing conditions could be addressed by implementations that go beyond the use of isolated deep RL solutions. Combining deep RL with classical approaches such as scenario analysis, combined rule decisions, or task decomposition could help circumvent common drawbacks such as low sample-efficiencies and reduce error-proneness.
- ***Handling production complexity:*** If the network receives too many state inputs and has to decide on a large number of possible actions, this increases problem complexity and significantly complicates optimal decision making. Thus, to keep large-scale production problems manageable, they must be reduced in their dimension and problem complexity to circumvent the curse of dimensionality. For this purpose, the complexity of whole production systems could be decomposed by decentralised structures and allocated to multiple agents. Having been trained to optimise specific parameters, these individual agents can be deployed situation-dependent. Through the associated orchestration and complexity break-down, a significantly improved scalability might be reached, since no individual agent has to cope with the entire complexity and the exponentially growing state and action space in large-scale applications. Local and global optimisation loops could run in parallel and minimise the risk of a local optimisation.

  Although Wang et al. (2019) already demonstrated such an ability of deep RL to optimise multiple objectives utilising generalised policies, further research should elaborate on multitasking and leverage the generalisability of deep RL algorithms.

  Besides, research should focus on transfer learning to enable agents to learn and perform complex tasks faster and better. Thus, in multi-agent systems, single agents could benefit from the experience gained by others and cope better with unfamiliar situations. The development of such swarm intelligence could better exploit local and global information and enable a flexible response and adaptation of the production system to unforeseen incidents.
- ***Coordinated optimisation:*** In distributed production systems, local optimisation of individual agents must be opposed by adjusting input variables, reward functions, and training strategy. Agents must receive essential global and local information and should be evaluated on individual as well as multi-agent performance criteria. This can include maximising the utilisation of machines in the local agent environment while minimising the total cycle time of the overall multi-agent process. Further research could scale this sensitivity towards multiple objectives which might be accomplished by staged training sequences in which individual agents first find optimal local solutions and subsequently target global objectives in a multi-agent training phase (as in Baer et al. (2019)).

  Besides, the exploration strategy of a single agent must be determined by appropriate parameters to avoid an intrinsic local optimisation. This can be remedied by specific tuning and should be considered more in-depth in deep RL controlled multi-agent production systems.

6. Discussion

Today's production systems must cope with increasingly sophisticated customer requirements, shorter product and development cycles, and short-term fluctuations in demand. One approach to address these challenges in production is deep RL, which differs from other machine learning methods primarily through its online adaptability and real-time processing of sensor data. Although other technical domains have already emphasised the benefits of deep RL, a focused review in production systems has yet to be conducted. Our purpose was to provide a systematic literature review of current deep RL applications in production systems and to outline challenges and fields of future research to address these. Based on a taxonomy framework, 120 retrieved papers from three databases were reviewed and classified according to their manufacturing discipline, industry background, specific application, optimisation objective, applied deep RL algorithm, and neural network, heuristic benchmark results, and its application in a simulated and/or real environment.

An application of deep RL was found in a wide range of production engineering disciplines. Although a large portion of the applications were implemented in simulations (76%), the superiority of deep RL driven production optimisation was evident. In more than 85% of the total comparisons, deep RL algorithms outperformed the corresponding benchmarks and increased problem-specific performances.

6.1. *Managerial implications*

Future factories will be increasingly interconnected, products and processes will become more complex, and development cycles will be more accelerated. To cope with these, companies should challenge current practices and consider alternatives to minimise process risks and to fully exploit algorithmic performances and organisational capabilities. To give a first introduction, this literature review presents a variety of possible applications of deep RL in production systems and helps managers to identify potential internal use cases. As a reference, the surveyed papers can provide valuable guidance for own deep RL implementation approaches and assist in the further selection of algorithms and parameters.

In contrast to static methods that can react to changing conditions only to a limited extent, deep RL algorithms were able to increase productions robustness and adaptability. In most applications, it proofed its practical relevance and not only improved technical parameters, but in some applications increased cash flow and reduced (online) conversion costs. Through deep RL, companies can limit the dependency on increasingly scarce human capital and leverage data-driven operations proactively to reduce cost-intensive manual and expert-based processes.

6.2. *Limitations*

Although the work is based on a taxonomy and methodology framework, we would like to emphasise the existing limitations of our review. We conducted the literature search based on three selected databases and an iterative keyword search, in which we tried to determine essential domains, but may miss some that would have yielded relevant supplementary results. To compensate for this bias, we conducted a forward and backward search to aggregate correlated publications. To satisfy our claim of providing a representative review and to provide a broad foundation, we also included proceedings and conference papers, which may cause bias compared to other reviews. However, by ensuring peer review we sought to reduce this bias and to meet all quality requirements. Besides, a limitation arises from the definition of a restricted review scope. Publications from enterprise research or other domains that may have interfaces to production were not specifically considered. Specific reviews can provide insights for the application of deep RL in these production related environments, which we recommend and encourage.

7. Conclusion

It became evident that deep RL is widely used from process control to maintenance and other domains, outperforming conventional algorithms in most cases, demonstrating its ability to adapt to a variety of scenarios and deal with existing production uncertainties (RQ1). This not only reduced lead times and WIP levels, reached high accuracies in assembly, or developed robust scheduling policies, but also mitigated current drawbacks of conventional methods such as limited adaptation capabilities, cost intensive re-optimisations, or high dependencies on human-based decisions.

Nevertheless, some challenges still prevent widespread adoption in production systems (RQ2). Besides missing hands-on guidelines and limited use of the available algorithm base, only a few deep RL applications have been evaluated in reality and optimised in-depth, making further validation mandatory. In future research (RQ3), the simulations need to be further refined to incorporate additional uncertainties, reduce current transfer barriers, and enable real-world applications. Additional optimisation alternatives such as more powerful deep RL algorithms that are currently less utilised, extensive elaboration on increased generalisability, alternative training strategies, and reduction of production task complexities can be further considered to realise more optimal performances.

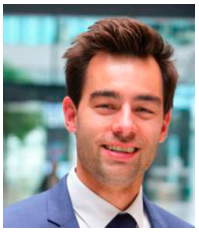
The challenge remains of defining a thorough approach that will assist scholars and practitioners through the application and optimisation process, providing guidelines for deployment, and accelerating the implementation in potential use-cases. Further research efforts on collaborative and hierarchical multi-agent architectures, as well as the use of fleet intelligence, can further strengthen the application of deep RL in production systems and make it a widely applicable and robust edge and global optimisation method.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Marcel Panzer, M.Sc. (∗1994) studied mechanical engineering at the Karlsruhe Institute of Technology and has been working as a research assistant at the University of Potsdam, Chair of Busines Informatics, esp. Process and Systems since 2020. His research is focused on AI-based production planning and control.



Benedict Bender, Dr. (∗1989) studied business informatics at the University of Potsdam, the Humboldt University of Berlin as well as the University of St. Gallen. His research interests include Industry 4.0 and aspects of IT security and privacy. Furthermore, he deals with digital platforms and business ecosystems.

ORCID

Marcel Panzer http://orcid.org/0000-0003-4099-0179

References

Altenmüller, Thomas, Tillmann Stüker, Bernd Waschneck, Andreas Kuhnle, and Gisela Lanza. 2020. "Reinforcement Learning for An Intelligent and Autonomous Production Control of Complex Job-Shops Under Time Constraints." *Production Engineering* 14 (3): 319–328. doi:10.1007/s11740-020-00967-8.

Andersen, Rasmus E., Steffen Madsen, Alexander B. K. Barlo, Sebastian B. Johansen, Morten Nør, Rasmus S. Andersen, and Simon Bøgh. 2019. "Self-Learning Processes in Smart Factories: Deep Reinforcement Learning for Process Control of Robot Brine Injection." *Procedia Manufacturing* 38: 171–177. doi:10.1016/j.promfg.2020.01.023.

Antônio Márcio Tavares, Thomé, Luiz Felipe Scavarda, and Annibal José Scavarda. 2016. "Conducting Systematic Literature Review in Operations Management." *Production Planning & Control* 27 (5): 408–420. doi:10.1080/09537287.2015.1129464.

Arinez, Jorge F., Qing Chang, Robert X. Gao, Chengying Xu, and Jianjing Zhang. 2020. "Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook." *Journal of Manufacturing Science and Engineering* 142 (11): 110804. doi:10.1115/1.4047855.

Baer, Schirin, Jupiter Bakakeu, Richard Meyes, and Tobias Meisen. 2019, September. "Multi-Agent Reinforcement Learning for Job Shop Scheduling in Flexible Manufacturing Systems." 2019 Second International Conference on Artificial Intelligence for Industries (AI4I). Laguna Hills, CA: IEEE, pp. 22–25.

Baer, Schirin, Danielle Turner, Punit Mohanty, Vladimir Samsonov, Romuald Bakakeu, and Tobias Meisen. 2020. "Multi Agent Deep Q-Network Approach for Online Job Shop Scheduling in Flexible Manufacturing." 2020 International Conference on Manufacturing System and Multiple Machines, Tokyo, Japan, pp. 1–9.

Bakakeu, Jupiter, Schirin Baer, Jochen Bauer, Hans-Henning Klos, Jörn Peschke, Adrian Fehrle, Werner Eberlein, et al. 2018. "An Artificial Intelligence Approach for Online Optimization of Flexible Manufacturing Systems." *Applied Mechanics and Materials* 882: 96–108. doi:10.4028/www.scientific.net/AMM.882.96.

Bakakeu, Jupiter, Dominik Kisskalt, Joerg Franke, Shirin Baer, Hans-Henning Klos, and Joern Peschke. 2020, August 30–September 2. "Multi-Agent Reinforcement Learning for the Energy Optimization of Cyber-Physical Production Systems." 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada. IEEE, pp. 2–8.

Bellman, Richard. 1957. *Dynamic Programming*. 1 vols. Princeton: Princeton University Press.

Beltran-Hernandez, Cristian C., Damien Petit, Ixchel G. Ramirez-Alpizar, and Kensuke Harada. 2020. "Variable Compliance Control for Robotic Peg-in-Hole Assembly: A Deep-Reinforcement-Learning Approach." *Applied Sciences* 10 (19): 6923. doi:10.3390/app10196923.

Beltran-Hernandez, C. C., D. Petit, I. G. Ramirez-Alpizar, T. Nishi, S. Kikuchi, T. Matsubara, and K. Harada. 2020. "Learning Force Control for Contact-Rich Manipulation Tasks With Rigid Position-Controlled Robots." *IEEE Robotics and Automation Letters* 5 (4): 5709–5716. doi:10.1109/LRA.2020.3010739.

Brito, Thadeu, Jonas Queiroz, Luis Piardi, Lucas A. Fernandes, José Lima, and Paulo Leitão. 2020. "A Machine Learning Approach for Collaborative Robot Smart Manufacturing Inspection for Quality Control Systems." *Procedia Manufacturing* 51: 11–18. doi:10.1016/j.promfg.2020.10.003.

Cao, Di, Weihao Hu, Junbo Zhao, Guozhou Zhang, Bin Zhang, Zhou Liu, Zhe Chen, and Frede Blaabjerg. 2020. "Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review." *Journal of Modern Power Systems and Clean Energy* 8 (6): 1029–1042. doi:10.35833/MPCE.2020.000552.

Chen, Ning, Shuhan Luo, Jiayang Dai, Biao Luo, and Weihua Gui. 2020. "Optimal Control of Iron-Removal Systems Based on Off-Policy Reinforcement Learning." *IEEE Access* 8: 149730–149740. doi:10.1109/ACCESS.2020.3015801.

Chen, Baotong, Jiafu Wan, Lan Yanting, Muhammad Imran, Di Li, and Nadra Guizani. 2019. "Improving Cognitive Ability of Edge Intelligent IIoT Through Machine Learning." *IEEE Network* 33 (5): 61–67. doi:10.1109/MNET.001.1800505.

Chien, Chen-Fu, Yun-Siang Lin, and Sheng-Kai Lin. 2020. "Deep Reinforcement Learning for Selecting Demand Forecast Models to Empower Industry 3.5 and An Empirical Study for a Semiconductor Component Distributor." *International Journal of Production Research* 58 (9): 2784–2804. doi:10.1080/00207543.2020.1733125.

Cooper, Harris M. 1988. "Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews." *Knowledge in Society* 1: 104–126.

Dai, Wenxin, Zhenling Mo, Chong Luo, Jing Jiang, Heng Zhang, and Qiang Miao. 2020. "Fault Diagnosis of Rotating Machinery Based on Deep Reinforcement Learning and Reciprocal of Smoothness Index." *IEEE Sensors Journal* 20 (15): 8307–8315. doi:10.1109/JSEN.2020.2970747.

Ding, Yu, Liang Ma, Jian Ma, Mingliang Suo, Laifa Tao, Yujie Cheng, and Chen Lu. 2019. "Intelligent Fault Diagnosis for Rotating Machinery Using Deep Q-Network Based Health State Classification: A Deep Reinforcement Learning Approach." *Advanced Engineering Informatics* 42: 100977. doi:10.1016/j.aei.2019.100977.

Dong, Tingting, Fei Xue, Chuangbai Xiao, and Juntao Li. 2020. "Task Scheduling Based on Deep Reinforcement Learning in a Cloud Manufacturing Environment." *Concurrency and Computation: Practice and Experience* 32 (11): e5654. doi:10.1002/cpe.5654.

Dornheim, Johannes, Norbert Link, and Peter Gumbsch. 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. doi:10.1007/s12555-019-0120-7.

Durach, Christian F., Joakim Kembro, and Andreas Wieland. 2017. "A New Paradigm for Systematic Literature Reviews in Supply Chain Management." *Journal of Supply Chain Management* 53 (4): 67–85. doi:10.1111/jscm.12145.

Epureanu, Bogdan I., Xingyu Li, Aydin Nassehi, and Yoram Koren. 2020. "Self-Repair of Smart Manufacturing Systems by Deep Reinforcement Learning." *CIRP Annals* 69 (1): 421–424. doi:10.1016/j.cirp.2020.04.008.

Feldkamp, Niclas, Soeren Bergmann, and Steffen Strassburger. 2020. "Simulation-Based Deep Reinforcement Learning for Modular Production Systems." Proceedings of the 2020 Winter Simulation Conference, 1596–1607. IEEE.

Gabel, Thomas, and Martin Riedmiller. 2007. "Adaptive Reactive Job-Shop Scheduling with Reinforcement Learning Agents." *International Journal of Information Technology and Intelligent Computing* 24 (4): 14–18.

Ge, Yangyang, Fei Zhu, Xinghong Ling, and Quan Liu. 2019. "Safe Q-Learning Method Based on Constrained Markov Decision Processes." *IEEE Access*7: 165007–165017. doi:10.1109/ACCESS.2019.2952651.

Greenhalgh, Trisha, and Richard Peacock. 2005. "Effectiveness and Efficiency of Search Methods in Systematic Reviews of Complex Evidence: Audit of Primary Sources." *BMJ (Clinical Research Ed.)* 331 (7524): 1064–1065. doi:10.1136/bmj.38636.593461.68.

Günther, Johannes, Patrick M. Pilarski, Gerhard Helfrich, Hao Shen, and Klaus Diepold. 2016. "Intelligent Laser Welding Through Representation, Prediction, and Control Learning: An Architecture with Deep Neural Networks and Reinforcement Learning." *Mechatronics* 34: 1–11. doi:10.1016/j.mechatronics.2015.09.004.

Guo, Li, Huan Wang, and Jun Zhang. 2019, August 10–12. "Data-Driven Grinding Control Using Reinforcement Learning." 2019 IEEE 21st International Conference on High Performance Computing and Communications, Zhangjiajie, China. IEEE.

Guo, Fei, Xiaowei Zhou, Jiahuan Liu, Yun Zhang, Dequn Li, and Huamin Zhou. 2019. "A Reinforcement Learning Decision Model for Online Process Parameters Optimization From Offline Data in Injection Molding." *Applied Soft Computing* 85: 105828. doi:10.1016/j.asoc.2019.105828.

Han, Bao-An, and Jian-Jun Yang. 2020. "Research on Adaptive Job Shop Scheduling Problems Based on Dueling Double DQN." *IEEE Access* 8: 186474–186495. doi:10.1109/ACCESS.2020.3029868.

He, Zhenglei, Kim-Phuc Tran, Sebastien Thomassey, Xianyi Zeng, Jie Xu, and Changhai Yi. 2020. "A Deep Reinforcement Learning Based Multi-Criteria Decision Support System for Optimizing Textile Chemical Process." *Computers in Industry* 125: 103373. doi:10.1016/j.compind.2020.103373.

Heger, Jens, and Thomas Voß. 2020. "Dynamically Changing Sequencing Rules Wirth Reinforcement Learning in a Job Shop System with Stochastic Influences." Proceedings of the 2020 Winter Simulation Conference 1608–1618. Orlando, FL: IEEE.

Hildebrand, Max, Rasmus S. Andersen, and Simon Bøgh. 2020. "Deep Reinforcement Learning for Robot Batching Optimization and Flow Control." *Procedia Manufacturing* 51: 1462–1468. doi:10.1016/j.promfg.2020.10.203.

Hoppe, Sabrina, Zhongyu Lou, Daniel Hennes, and Marc Toussaint. 2019. "Planning Approximate Exploration Trajectories for Model-Free Reinforcement Learning in Contact-Rich Manipulation." *IEEE Robotics and Automation Letters* 4 (4): 4042–4047. doi:10.1109/LRA.2019.2928212.

Hu, Hao, Xiaoliang Jia, Qixuan He, Shifeng Fu, and Kuo Liu. 2020. "Deep Reinforcement Learning Based AGVs Real-Time Scheduling with Mixed Rule for Flexible Shop Floor in Industry 4.0." *Computers & Industrial Engineering* 149: 106749. doi.10.1016/j.cie.2020.106749.

Hu, Liang, Zhenyu Liu, Weifei Hu, Yueyang Wang, Jianrong Tan, and Fei Wu. 2020. "Petri-Net-Based Dynamic Scheduling of Flexible Manufacturing System Via Deep Reinforcement Learning with Graph Convolutional Network." *Journal of Manufacturing Systems* 55: 1–14. doi:10.1016/j.jmsy.2020.02.004.

Hu, Wenqing, Zeyi Sun, Yunchao Zhang, and Yu Li. 2019. "Joint Manufacturing and Onsite Microgrid System Control Using Markov Decision Process and Neural Network Integrated Reinforcement Learning." *Procedia Manufacturing* 39: 1242–1249. doi:10.1016/j.promfg.2020.01.345.

Huang, Jing, Qing Chang, and Jorge Arinez. 2020. "Deep Reinforcement Learning Based Preventive Maintenance Policy for Serial Production Lines." *Expert Systems with Applications* 160: 113701. doi:10.1016/j.eswa.2020.113701.

Huang, Xuefei, Seung Ho Hong, Mengmeng Yu, Yuemin Ding, and Junhui Jiang. 2019. "Demand Response Management for Industrial Facilities: A Deep Reinforcement Learning Approach." *IEEE Access* 7: 82194–82205. doi:10.1109/ACCESS.2019.2924030.

Hubbs, Christian D., Can Li, Nikolaos V. Sahinidis, Ignacio E. Grossmann, and John M. Wassick. 2020. "A Deep Reinforcement Learning Approach for Chemical Production Scheduling." *Computers & Chemical Engineering* 141: 106982. doi.10.1016/j.compchemeng.2020.106982.

Inoue, Tadanobu, Giovanni De Magistris, Asim Munawar, Tsuyoshi Yokoya, and Ryuki Tachibana. 2017. "Deep Reinforcement Learning for High Precision Assembly Tasks." In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC, September 24–28.

Jiang, Yi, Jialu Fan, Tianyou Chai, and Frank L. Lewis. 2019. "Dual-Rate Operational Optimal Control for Flotation Industrial Process With Unknown Operational Model."

IEEE Transactions on Industrial Electronics 66 (6): 4587–4599. doi:10.1109/TIE.2018.2856198.

Jiang, Yi, Jialu Fan, Tianyou Chai, Jinna Li, and Frank L. Lewis. 2018. "Data-Driven Flotation Industrial Process Operational Optimal Control Based on Reinforcement Learning." *IEEE Transactions on Industrial Informatics* 14 (5): 1974–1989. doi:10.1109/TII.2017.2761852.

Jin, Zeshi, and Haichao Li Hongming Gao. 2019. "An Intelligent Weld Control Strategy Based on Reinforcement Learning Approach." *The International Journal of Advanced Manufacturing Technology*100 (9-12): 2163–2175. doi:10.1007/s00170-018-2864-2.

Kagermann, Henning, Wolfgang Wahlster, and Johannes Helbig. 2013. *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0 – Securing the Future of German Manufacturing Industry*. Acatech – National Academy of Science and Engineering.

Kang, Ziqiu, Cagatay Catal, and Bedir Tekinerdogan. 2020. "Machine Learning Applications in Production Lines: A Systematic Literature Review." *Computers & Industrial Engineering* 149: 106773. doi:10.1016/j.cie.2020.106773.

Khan, Al-Masrur, Rashed Jaowad Khan, Abul Tooshil, Niloy Sikder, M. A. Parvez Mahmud, Abbas Z. Kouzani, and Abdullah-Al Nahid. 2020. "A Systematic Review on Reinforcement Learning-Based Robotics Within the Last Decade." *IEEE Access* 8: 176598–176623. doi.10.1109/ACCESS.2020.3027152.

Kim, Young-Loul, Kuk-Hyun Ahn, and Jae-Bok Song. 2020. "Reinforcement Learning Based on Movement Primitives For Contact Tasks." *Robotics and Computer-Integrated Manufacturing* 62: 101863. doi:10.1016/j.rcim.2019.101863.

Kim, Ju-Bong, Ho-Bin Choi, Gyu-Young Hwang, Kwihoon Kim, Yong-Geun Hong, and Youn-Hee Han. 2020. "Sortation Control Using Multi-Agent Deep Reinforcement Learning in N-Grid Sortation System." *Sensors* 20 (12): 3401. doi.10.3390/s20123401.

Kuhnle, Andreas. 2020. *SimRLFab: Simulation and Reinforcement Learning Framework for Production Planning and Control of Complex Job Shop Manufacturing Systems*. GitHub (accessed 20 March 2021. https://github.com/AndreasKuhnle/SimRLFab).

Kuhnle, Andreas, Johannes Jakubik, and Gisela Lanza. 2019. "Reinforcement Learning for Opportunistic Maintenance Optimization." *Production Engineering* 13 (1): 33–41. doi:10.1007/s11740-018-0855-7.

Kuhnle, Andreas, Jan-Philipp Kaiser, Felix Theiß, Nicole Stricker, and Gisela Lanza. 2020. "Designing An Adaptive Production Control System Using Reinforcement Learning." *Journal of Intelligent Manufacturing* 32: 855–876. doi:10.1007/s10845-020-01612-y.

Kumar, Ashish, Roussos Dimitrakopoulos, and Marco Maulen. 2020. "Adaptive Self-Learning Mechanisms for Updating Short-Term Production Decisions in An Industrial Mining Complex." *Journal of Intelligent Manufacturing* 31 (7): 1795–1811. doi:10.1007/s10845-020-01562-5.

Lämmle, Arik, Thomas König, Mohamed El-Shamouty, and Marco F. Huber. 2020. "Skill-Based Programming of Force-controlled Assembly Tasks Using Deep Reinforcement Learning." *Procedia CIRP*93: 1061–1066. doi:10.1016/j.procir.2020.04.153.

Lange, Sascha, Martin Riedmiller, and Arne Voigtlander. 2012. "Autonomous Reinforcement Learning on Raw Visual Input Data in A Real World Application." In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, Australia, Juni 10–15.

Lee, Seunghoon, Yongju Cho, and Young Hoon Lee. 2020. "Injection Mold Production Sustainable Scheduling Using Deep Reinforcement Learning." *Sustainability* 12 (20): 8718. doi:10.3390/su12208718.

Lee, Jun-Ho, and Hyun-Jung Kim. 2021. "Reinforcement Learning for Robotic Flow Shop Scheduling with Processing Time Variations." *International Journal of Production Research,* 1–23. doi:10.1080/00207543.2021.1887533.

Lee, Jay H., Joohyun Shin, and Matthew J. Realff. 2018. "Machine Learning: Overview of the Recent Progresses and Implications for the Process Systems Engineering Field." *Computers & Chemical Engineering* 114: 111–121. doi:10.1016/j.compchemeng.2017.10.008.

Lei, Lei, Yue Tan, Kan Zheng, Shiwen Liu, Kuan Zhang, and Xuemin Shen. 2020. "Deep Reinforcement Learning for Autonomous Internet of Things: Model, Applications and Challenges." *IEEE Communications Surveys & Tutorials* 22 (3): 1722–1760. doi:10.1109/COMST.2020.2988367.

Leng, Jinling, Chun Jin, Alexander Vogl, and Huiyu Liu. 2020. "Deep Reinforcement Learning for A Color-Batching Resequencing Problem." *Journal of Manufacturing Systems* 56: 175–187. doi:10.1016/j.jmsy.2020.06.001.

Leng, Jiewu, Guolei Ruan, Yuan Song, Qiang Liu, Yingbin Fu, Kai Ding, and Xin Chen. 2021. "A Loosely-Coupled Deep Reinforcement Learning Approach for Order Acceptance Decision of Mass-Individualized Printed Circuit Board Manufacturing in Industry 4.0." *Journal of Cleaner Production* 280: 124405. doi:10.1016/j.jclepro.2020.124405.

Li, Jinna, Jinliang Ding, Tianyou Chai, and Frank L. Lewis. 2020. "Nonzero-Sum Game Reinforcement Learning for Performance Optimization in Large-Scale Industrial Processes." *IEEE Transactions on Cybernetics* 50 (9): 4132–4145. doi:10.1109/TCYB.2019.2950262.

Li, Fengming, Qi Jiang, Wei Quan, Shibo Cai, Rui Song, and Yibin Li. 2019. "Manipulation Skill Acquisition for Robotic Assembly Based on Multi-Modal Information Description." *IEEE Access* 8: 6282–6294. doi:10.1109/ACCESS.2019.2934174.

Li, Fengming, Qi Jiang, Sisi Zhang, Meng Wei, and Rui Song. 2019. "Robot Skill Acquisition in Assembly Process Using Deep Reinforcement Learning." *Neurocomputing* 345: 92–102. doi:10.1016/j.neucom.2019.01.087.

Li, Bingran, Hui Zhang, Peiqing Ye, and Jinsong Wang. 2020. "Trajectory Smoothing Method Using Reinforcement Learning for Computer Numerical Control Machine Tools." *Robotics and Computer-Integrated Manufacturing* 61: 101847. doi:10.1016/j.rcim.2019.101847.

Liang, Huagang, Xiaoqian Wen, Yongkui Liu, Haifeng Zhang, Lin Zhang, and Lihui Wang. 2021. "Logistics-Involved QoS-Aware Service Composition in Cloud Manufacturing With Deep Reinforcement Learning." *Robotics and Computer-Integrated Manufacturing* 67: 101991. doi:10.1016/j.rcim.2020.101991.

Liao, Yongxin, Fernando Deschamps, Eduardo de Freitas Rocha Loures, and Luiz Felipe Pierin Ramos. 2017. "Past, Present and Future of Industry 4.0 – A Systematic Literature Review and Research Agenda Proposal." *International Journal of Production Research* 55 (12): 3609–3629. doi:10.1080/00207543.2017.1308576.

Liao, Haiguang, Wentai Zhang, Xuliang Dong, Barnabas Poczos, Kenji Shimada, and Levent Burak Kara. 2020. "A Deep Reinforcement Learning Approach for Global Routing." *Journal of Mechanical Design*142 (6): 061701. doi:10.1115/1.4045044.

Light, Richard J., and David B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Cambridge: Harvard University Press.

Lillicrap, Timothy P., Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2016. "Continuous Control with Deep Reinforcement Learning." In *Proceedings of 4th International Conference on Learning Representations* ArXiv: 1509.02971.

Lin, Chun-Cheng, Der-Jiunn Deng, Yen-Ling Chih, and Hsin-Ting Chiu. 2019. "Smart Manufacturing Scheduling With Edge Computing Using Multiclass Deep Q Network." *IEEE Transactions on Industrial Informatics* 15 (7): 4276–4284. doi:10.1109/TII.2019.2908210.

Liu, Chien-Liang, Chuan-Chin Chang, and Chun-Jan Tseng. 2020. "Actor-Critic Deep Reinforcement Learning for Solving Job Shop Scheduling Problems." *IEEE Access*8: 71752–71762. doi:10.1109/ACCESS.2020.2987820.

Liu, Yu, Yiming Chen, and Tao Jiang. 2020. "Dynamic Selective Maintenance Optimization for Multi-State Systems Over a Finite Horizon: A Deep Reinforcement Learning Approach." *European Journal of Operational Research* 283 (1): 166–181. doi:10.1016/j.ejor.2019.10.049.

Liu, Xing, Hansong Xu, Weixian Liao, and Wei Yu. 2019. "Reinforcement Learning for Cyber-Physical Systems." In *2019 IEEE International Conference on Industrial Internet (ICII)*, Orlando, FL, USA, November 11–12.

Liu, Xing, Wei Yu, Fan Liang, David Griffith, and Nada Golmie. 2021. "On Deep Reinforcement Learning Security for Industrial Internet of Things." *Computer Communications* 168: 20–32. doi:10.1016/j.comcom.2020.12.013.

Lohmer, Jacob, and Rainer Lasch. 2020. "Production Planning and Scheduling in Multi-Factory Production Networks: A Systematic Literature Review." *International Journal of Production Research*59 (7): 2028–2054. doi:10.1080/00207543.2020.1797207.

Lu, Xinglong, Bahare Kiumarsi, Tianyou Chai, and Frank L. Lewis. 2016. "Data-Driven Optimal Control of Operational Indices for A Class of Industrial Processes." *IET Control Theory & Applications* 10 (12): 1348–1356. doi:10.1049/iet-cta.2015.0798.

Lu, Renzhi, Yi-Chang Li, Yuting Li, Junhui Jiang, and Yuemin Ding. 2020. "Multi-Agent Deep Reinforcement Learning Based Demand Response for Discrete Manufacturing Systems Energy Management." *Applied Energy* 276: 115473. doi:10.1016/j.apenergy.2020.115473.

Luo, Shu. 2020. "Dynamic Scheduling for Flexible Job Shop with New Job Insertions by Deep Reinforcement Learning." *Applied Soft Computing* 91: 106208. doi:10.1016/j.asoc.2020.106208.

Luo, Jianlan, Eugen Solowjow, Chengtao Wen, Juan Aparicio Ojea, and Alice M. Agogino. 2018. "Deep Reinforcement Learning for Robotic Assembly of Mixed Deformable and Rigid Objects." In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, October 1–5.

Luo, Jianlan, Eugen Solowjow, Chengtao Wen, Juan Aparicio Ojea, Alice M. Agogino, Aviv Tamar, and Pieter Abbeel. 2019. "Reinforcement Learning on Variable Impedance Controller for High-Precision Robotic Assembly." In *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, May 20–24.

Luong, Nguyen Cong, Dinh Thai Hoang, Shimin Gong, Dusit Niyato, Ping Wang, Ying-Chang Liang, and Dong In Kim. 2019. "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey." *IEEE Communications Surveys Tutorials* 21 (4): 3133–3174. doi:10.1109/COMST.2019.2916583.

Ma, Yan, Wenbo Zhu, Michael G. Benton, and José Romagnoli. 2019. "Continuous Control of A Polymerization System with Deep Reinforcement Learning." *Journal of Process Control* 75: 40–47. doi:10.1016/j.jprocont.2018.11.004.

Mahadevan, Sridhar, and Georgios Theocharous. 1998. "Optimizing Production Manufacturing Using Reinforcement Learning." In *Proceedings of the Eleventh International FLAIRS Conference*, 372–377.

Malus, Andreja, Dominik Kozjek, and Rok Vrabič. 2020. "Real-Time Order Dispatching for A Fleet of Autonomous Mobile Robots Using Multi-Agent Reinforcement Learning." *CIRP Annals* 69 (1): 397–400. doi:10.1016/j.cirp.2020.04.001.

Marc-André, Dittrich, and Silas Fohlmeister. 2020. "Cooperative Multi-Agent System for Production Control Using Reinforcement Learning." *CIRP Annals* 69 (1): 389–392. doi:10.1016/j.cirp.2020.04.005.

Masinelli, Giulio, Tri Le-Quang, Silvio Zanoli, Kilian Wasmer, and Sergey A. Shevchik. 2020. "Adaptive Laser Welding Control: A Reinforcement Learning Approach." *IEEE Access* 8: 103803–103814. doi:10.1109/ACCESS.2020.2998052.

Mazgualdi, Choumicha El, Tawfik Masrour, Ibtissam El Hassani, and Abdelmoula Khdoudi. 2021. "A Deep Reinforcement Learning (DRL) Decision Model for Heating Process Parameters Identification in Automotive Glass Manufacturing." In *Artificial Intelligence and Industrial Applications*, Vol. 1193, 77–87. Cham: Springer International Publishing.

Miljković, Zoran, Marko Mitić, Mihailo Lazarević, and Bojan Babić. 2013. "Neural Network Reinforcement Learning for Visual Control of Robot Manipulators." *Expert Systems with Applications*40 (5): 1721–1736. doi:10.1016/j.eswa.2012.09.010.

Mishra, Manohar, Janmenjoy Nayak, Bighnaraj Naik, and Ajith Abraham. 2020. "Deep Learning in Electrical Utility Industry: A Comprehensive Review of A Decade of Research." *Engineering Applications of Artificial Intelligence* 96: 104000. doi:10.1016/j.engappai.2020.104000.

Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller.. 2013. "Playing Atari with Deep Reinforcement Learning." ArXiv:1312.5602.

Mohammed, Marwan Qaid, K. L. Chung, and Shing Chyi Chua. 2020. "Review of Deep Reinforcement Learning-Based Object Grasping: Techniques, Open Challenges, and Recommendations." *IEEE Access*8: 178450–178481. doi:10.1109/ACCESS.2020.3027923.

Mosavi, Amirhosein, Yaser Faghan, Pedram Ghamisi, Puhong Duan, Sina Faizollahzadeh Ardabili, Ely Salwana, and

Shahab S. Band. 2020. "Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics." *Mathematics* 8 (10): 1640. doi:10.3390/math8101640.

Naeem, Muddasar, Syed T. H. Rizvi, and Antonio Coronato. 2020. "A Gentle Introduction to Reinforcement Learning and Its Application in Different Fields." *IEEE Access* 8: 209320–209344. doi:10.1109/ACCESS.2020.3038605.

Nguyen, Hai, and Hung La. 2019. "Review of Deep Reinforcement Learning for Robot Manipulation." In *2019 Third IEEE International Conference on Robotic Computing (IRC)*, Naples, Italy, February 25–27.

Noel, Mathew Mithra, and B. Jaganatha Pandian. 2014. "Control of a Nonlinear Liquid Level System Using a New Artificial Neural Network Based Reinforcement Learning Approach." *Applied Soft Computing* 23: 444–451. doi:10.1016/j.asoc.2014.06.037.

Oh, Tae-Ho, Ji-Seok Han, Young-Seok Kim, Dae-Young Yang, Sang-Hoon Lee, and Dong-Il 'Dan' Cho. 2020. "Deep RL Based Notch Filter Design Method for Complex Industrial Servo Systems." *International Journal of Control, Automation and Systems* 18 (12): 2983–2992. doi:10.1007/s12555-020-0153-y.

Palombarini, Jorge A., and Ernesto C. Martinez. 2018. "Automatic Generation of Rescheduling Knowledge in Socio-Technical Manufacturing Systems using Deep Reinforcement Learning." In *2018 IEEE Biennial Congress of Argentina (ARGENCON)*, San Miguel de Tucumán, Argentina, June 6–8.

Palombarini, Jorge A., and Ernesto C. Martínez. 2019. "Closed-loop Rescheduling Using Deep Reinforcement Learning." *IFAC-PapersOnLine* 52 (1): 231–236. doi:10.1016/j.ifacol.2019.06.067.

Pandian, B. Jaganatha, and Mathew M. Noel. 2018. "Tracking Control of a Continuous Stirred Tank Reactor Using Direct and Tuned Reinforcement Learning Based Controllers." *Chemical Product and Process Modeling* 13 (3): 1–10. doi:10.1515/cppm-2017-0040.

Park, Junyoung, Jaehyeong Chun, Sang Hun Kim, Youngkook Kim, and Jinkyoo Park. 2021. "Learning to Schedule Job-Shop Problems: Representation and Policy Learning Using Graph Neural Network and Reinforcement Learning." *International Journal of Production Research* 59 (11): 3360–3377. doi:10.1080/00207543.2020.1870013.

Park, In-Beom, Jaeseok Huh, Joongkyun Kim, and Jonghun Park. 2020. "A Reinforcement Learning Approach to Robust Scheduling of Semiconductor Manufacturing Facilities." *IEEE Transactions on Automation Science and Engineering* 17 (3): 1420–1431. doi:10.1109/TASE.2019.2956762.

Park, JoungMin, SangYoon Lee, JaeWoon Lee, and Jumyung Um. 2020. "GadgetArm–Automatic Grasp Generation and Manipulation of 4-DOF Robot Arm for Arbitrary Objects Through Reinforcement Learning." *Sensors* 20 (21): 6183. doi:10.3390/s20216183.

Peres, Ricardo Silva, Xiaodong Jia, Jay Lee, Keyi Sun, Armando Walter Colombo, and Jose Barata. 2020. "Industrial Artificial Intelligence in Industry 4.0 – Systematic Review, Challenges and Outlook." *IEEE Access* 8: 220121–220139. doi:10.1109/ACCESS.2020.3042874.

Petticrew, Mark, and Helen Roberts. 2006. *Systematic Reviews in the Social Sciences*. Oxford: Blackwell Publishing Ltd.

Powell, By Kody M., Derek Machalek, and Titus Quah. 2020. "Real-Time Optimization Using Reinforcement Learning." *Computers & Chemical Engineering* 143: 107077. doi:10.1016/j.compchemeng.2020.107077.

Quah, Titus, Derek Machalek, and Kody M. Powell. 2020. "Comparing Reinforcement Learning Methods for Real-Time Optimization of a Chemical Process." *Processes* 8 (11): 1497. doi:10.3390/pr8111497.

Riedmiller, Simone, and Martin Riedmiller. 1999. "A Neural Reinforcement Learning Approach to Learn Local Dispatching Policies in Production Scheduling." *Proceedings of the 16th International Joint Conference on Artificial Intelligence* 2: 764–769.

Rossit, Daniel Alejandro, Fernando Tohmé, and Mariano Frutos. 2019. "Industry 4.0: Smart Scheduling." *International Journal of Production Research* 57 (12): 3802–3813. doi:10.1080/00207543.2018.1504248.

Rummukainen, Hannu, and Jukka K. Nurminen. 2019. "Practical Reinforcement Learning -Experiences in Lot Scheduling Application." *IFAC-PapersOnLine* 52 (13): 1415–1420. doi:10.1016/j.ifacol.2019.11.397.

Samsonov, Vladimir, Chrismarie Enslin, Hans-Georg Köpken, Schirin Baer, and Daniel Lütticke. 2020. "Using Reinforcement Learning for Optimization of a Workpiece Clamping Position in a Machine Tool." In *Proceedings of the 22nd International Conference on Enterprise Information Systems* 506–514. doi:10.5220/0009354105060514.

Schaul, Tom, John Quan, Ioannis Antonoglou, and David Silver. 2016. "Prioritized Experience Replay." In *International Conference on Learning Representations*, San Juan, Puerto Rico, May.

Scheiderer, Christian, Timo Thun, Christian Idzik, Andrés Felipe Posada-Moreno, Alexander Krämer, Johannes Lohmar, Gerhard Hirt, and Tobias Meisen. 2020. "Simulation-as-a-Service for Reinforcement Learning Applications by Example of Heavy Plate Rolling Processes." *Procedia Manufacturing* 51: 897–903. doi:10.1016/j.promfg.2020.10.126.

Scheiderer, Christian, Timo Thun, and Tobias Meisen. 2019. "Bézier Curve Based Continuous and Smooth Motion Planning for Self-Learning Industrial Robots." *Procedia Manufacturing* 38: 423–430. doi:10.1016/j.promfg.2020.01.054.

Schmidt, Alexander, Florian Schellroth, and Oliver Riedel. 2020. "Control Architecture for Embedding Reinforcement Learning Frameworks on Industrial Control Hardware." In *Proceedings of the 3rd International Conference on Applications of Intelligent Systems*, Las Palmas de Gran Canaria Spain, January 7–12. doi:10.1145/3378184.3378198.

Schoettler, Gerrit, Ashvin Nair, Jianlan Luo, Shikhar Bahl, Juan Aparicio Ojea, Eugen Solowjow, and Sergey Levine. 2020. "Deep Reinforcement Learning for Industrial Insertion Tasks with Visual Inputs and Natural Rewards." In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, Nevada, USA, October 25-29. doi:10.1109/IROS45743.2020.9341714.

Serin, Gokberk, Batihan Sener, Murat Ozbayoglu, and Hakki Ozgur Unver. 2020. "Review of Tool Condition Monitoring in Machining and Opportunities for Deep Learning." *The International Journal of Advanced Manufacturing Technology* 109 (3-4): 953–974. doi:10.1007/s00170-020-05449-w.

Sewak, Mohit. 2019. "Policy-Based Reinforcement Learning Approaches: Stochastic Policy Gradient and the REINFORCE Algorithm." In *Deep Reinforcement Learning*, 127–140. Singapore: Springer Singapore.

Sewak, Mohit. 2019. *Deep Reinforcement Learning: Frontiers of Artificial Intelligence*. 1st ed. Singapore: Springer Singapore.

Shi, Daming, Wenhui Fan, Yingying Xiao, Tingyu Lin, and Chi Xing. 2020. "Intelligent Scheduling of Discrete Automated Production Line Via Deep Reinforcement Learning." *International Journal of Production Research* 58 (11): 3362–3380. doi:10.1080/00207543.2020.1717008.

Shyalika, Chathurangi, Thushari Silva, and Asoka Karunananda. 2020. "Reinforcement Learning in Dynamic Task Scheduling: A Review." *SN Computer Science* 1 (6): 306. doi:10.1007/s42979-020-00326-5.

Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, and Marc Lanctot, et al. 2017. "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm." arXiv:1712.01815 [cs].

Skordilis, Erotokritos, and Ramin Moghaddass. 2020. "A Deep Reinforcement Learning Approach for Real-Time Sensor-Driven Decision Making and Predictive Analytics." *Computers & Industrial Engineering* 147: 106–600. doi:10.1016/j.cie.2020.106600.

Spielberg, S., R. B. Gopaluni, and P. D. Loewen. 2017. "Deep Reinforcement Learning Approaches for Process Control." In *2017 6th International Symposium on Advanced Control of Industrial Processes*, Taipei, Taiwan, May 28–31.

Spielberg, Steven, Aditya Tulsyan, Nathan P. Lawrence, Philip D. Loewen, and R. Bhushan Gopaluni. 2019. "Toward Self-Driving Processes: A Deep Reinforcement Learning Approach to Control." *AIChE Journal* 65 (10): 16–689. doi:10.1002/aic.16689.

Stricker, Nicole, Andreas Kuhnle, Roland Sturm, and Simon Friess. 2018. "Reinforcement Learning for Adaptive Order Dispatching in The Semiconductor Industry." *CIRP Annals* 67 (1): 511–514. doi:10.1016/j.cirp.2018.04.041.

Sutton, Richard S., and Andrew G. Barto. 2017. *Reinforcement Learning: An Introduction*. 2nd ed. Adaptive computation and machine learning series. Cambridge: The MIT Press.

Szarski, Martin, and Sunita Chauhan. 2021. "Composite Temperature Profile and Tooling Optimization Via Deep Reinforcement Learning." *Composites Part A: Applied Science and Manufacturing* 142: 106–235. doi:10.1016/j.compositesa.2020.106235.

Tewari, Ashutosh, Kuang-Hung Liu, and Dimitri Papageorgiou. 2020. "Information-Theoretic Sensor Planning for Large-Scale Production Surveillance Via Deep Reinforcement Learning." *Computers & Chemical Engineering* 141: 106–988. doi:10.1016/j.compchemeng.2020.106988.

Tomé, Silva, and Américo Azevedo. 2019. "Production Flow Control Through the Use of Reinforcement Learning." *Procedia Manufacturing* 38: 194–202. doi:10.1016/j.promfg.2020.01.026.

Tranfield, David, David Denyer, and Palminder Smart. 2003. "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review." *British Journal of Management* 14 (3): 207–222. doi:10.1111/1467-8551.00375.

Tsai, Yu-Ting, Chien-Hui Lee, Tao-Ying Liu, Tien-Jan Chang, Chun-Sheng Wang, S. J. Pawar, Pei-Hsing Huang, and Jin-H. Huang. 2020. "Utilization of A Reinforcement Learning Algorithm for the Accurate Alignment of a Robotic Arm in a Complete Soft Fabric Shoe Tongues Automation Process." *Journal of Manufacturing Systems* 56: 501–513. doi:10.1016/j.jmsy.2020.07.001

van Hasselt, Hado, Arthur Guez, and David Silver. 2016. "Deep Reinforcement Learning with Double Q-learning." In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* 2094–2100. ArXiv: 1509.06461. doi:10.5555/3016100.3016191.

vom Brocke, J., A. Simons, Björn Niehaves, K. Riemer, Ralf Plattfaut, and A. Cleven. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process." Proceedings of the 17th European Conference on Information Systems (ECIS). Verona: Università di Verona, pp. 2206–2217.

Wang, Kaixin, Bingyi Kang, Jie Shao, and Jiashi Feng. 2020. "Improving Generalization in Reinforcement Learning with Mixture Regularization." In *34th Conference on Neural Information Processing Systems*, Vancouver, Canada, December 6–12.

Wang, Jun Ping, You Kang Shi, Wen Sheng Zhang, Ian Thomas, and Shi Hui Duan. 2019. "Multitask Policy Adversarial Learning for Human-Level Control With Large State Spaces." *IEEE Transactions on Industrial Informatics* 15 (4): 2395–2404. doi:10.1109/TII.2018.2881266.

Wang, Hao-nan, Ning Liu, Yi-yun Zhang, Da-wei Feng, Feng Huang, Dong-sheng Li, and Yi-ming Zhang. 2020. "Deep Reinforcement Learning: A Survey." *Frontiers of Information Technology & Electronic Engineering* 21 (12): 1726–1744. doi:10.1631/FITEE.1900533.

Wang, Haoxiang, Bhaba R. Sarker, Jing Li, and Jian Li. 2020. "Adaptive Scheduling for Assembly Job Shop with Uncertain Assembly Times Based on Dual Q-Learning." *International Journal of Production Research,* 1–17. doi:10.1080/00207543.2020.1794075.

Wang, Fei, Xingqun Zhou, Jianhui Wang, Xing Zhang, Zhenquan He, and Bo Song. 2020. "Joining Force of Human Muscular Task Planning With Robot Robust and Delicate Manipulation for Programming by Demonstration." *IEEE/ASME Transactions on Mechatronics* 25 (5): 2574–2584. doi:10.1109/TMECH.2020.2997799.

Waschneck, Bernd, André Reichstaller, Lenz Belzner, Thomas Altenmüller, Thomas Bauernhansl, Alexander Knapp, and Andreas Kyek. 2018b. "Optimization of Global Production Scheduling with Deep Reinforcement Learning." *Procedia CIRP* 72: 1264–1269. doi:10.1016/j.procir.2018.03.212.

Waschneck, Bernd, Andre Reichstaller, Lenz Belzner, Thomas Altenmuller, Thomas Bauernhansl, Alexander Knapp, and Andreas Kyk. 2018a. "Deep Reinforcement Learning for Semiconductor Production Scheduling." In *2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, Saratoga Springs, NY, USA, April 30 – May 3.

Wasmer, Kilian, Tri Le-Quang, Bastian Meylan, and Sergey Shevchik. 2019. "In Situ Quality Monitoring in AM Using Acoustic Emission: A Reinforcement Learning Approach." *Journal of Materials Engineering and Performance* 28 (2): 666–672. doi:10.1007/s11665-018-3690-2.

Watanabe, Keijiro, and Shuhei Inada. 2020. "Search Algorithm of the Assembly Sequence of Products by Using Past Learning Results." *International Journal of Production Economics* 226: 107–615. doi:10.1016/j.ijpe.2020.107615.

Watkins, Christopher J. C. H., and Peter Dayan. 1992. "Q-learning." *Machine Learning* 8 (3-4): 279–292. doi:10.1007/BF00992698.

Webster, Jane, and Richard T. Watson. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review." *MIS Quarterly* 26 (2): 13–23. Publisher: Management Information Systems Research Center, University of Minnesota.

Wu, Wenbo, Zhengdong Huang, Jiani Zeng, and Kuan Fan. 2021. "A Fast Decision-Making Method for Process Planning with Dynamic Machining Resources Via Deep Reinforcement Learning." *Journal of Manufacturing Systems* 58: 392–411. doi:10.1016/j.jmsy.2020.12.015.

Wu, Chen-Xin, Min-Hui Liao, Mumtaz Karatas, Sheng-Yong Chen, and Yu-Jun Zheng. 2020. "Real-Time Neural Network Scheduling of Emergency Medical Mask Production During COVID-19." *Applied Soft Computing* 97: 106–790. doi:10.1016/j.asoc.2020.106790.

Xanthopoulos, A. S., Athanasios Kiatipis, D. E. Koulouriotis, and Sepp Stieger. 2018. "Reinforcement Learning-Based and Parametric Production-Maintenance Control Policies for a Deteriorating Manufacturing System." *IEEE Access* 6: 576–588. doi:10.1080/00207543.2021.1887533.

Xia, Kaishu, Christopher Sacco, Max Kirkpatrick, Clint Saidy, Lam Nguyen, Anil Kircaliali, and Ramy Harik. 2020. "A Digital Twin to Train Deep Reinforcement Learning Agent for Smart Manufacturing Plants: Environment, Interfaces and Intelligence." *Journal of Manufacturing Systems* 58: 210–230. doi:10.1016/j.jmsy.2020.06.012.

Xie, Shufang, Tao Zhang, and Oliver Rose. 2019. "Online Single Machine Scheduling Based on Simulation and Reinforcement Learning." In *18. ASIM Fachtagung Simulation in Production und Logistik*, Chemnitz, September 19–20.

Xiong, Hao, and Xiumin Diao. 2021. "Safety Robustness of Reinforcement Learning Policies: A View From Robust Control." *Neurocomputing* 422: 12–21. doi:10.1016/j.neucom.2020.09.055.

Xu, Jing, Zhimin Hou, Wei Wang, Bohao Xu, Kuangen Zhang, and Ken Chen. 2019. "Feedback Deep Deterministic Policy Gradient With Fuzzy Reward for Robotic Multiple Peg-in-Hole Assembly Tasks." *IEEE Transactions on Industrial Informatics* 15 (3): 1658–1667. doi:10.1109/TII.2018.2868859.

Xu, Xinghai, Huimin Xie, and Jia Shi. 2020. "Iterative Learning Control (ILC) Guided Reinforcement Learning Control (RLC) Scheme for Batch Processes." In *2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)*, Liuzhou, Juni 19–21.

Xu, Li Da, Eric L. Xu, and Ling Li. 2018. "Industry 4.0: State of The Art and Future Trends." *International Journal of Production Research* 56 (8): 2941–2962. doi:10.1080/00207543.2018.1444806.

Yang, Helin, Arokiaswami Alphones, Wen-De Zhong, Chen Chen, and Xianzhong Xie. 2020. "Learning-Based Energy-Efficient Resource Management by Heterogeneous RF/VLC for Ultra-Reliable Low-Latency Industrial IoT Networks." *IEEE Transactions on Industrial Informatics* 16 (8): 5565–5576. doi:10.1109/TII.2019.2933867.

Yoo, Haeun, Boeun Kim, Jong Woo Kim, and Jay H. Lee. 2021. "Reinforcement Learning Based Optimal Control of Batch Processes Using Monte-Carlo Deep Deterministic Policy Gradient with Phase Segmentation." *Computers & Chemical Engineering* 144: 107–133. doi:10.1016/j.compchemeng.2020.107133

Yu, Jianbo, and Peng Guo. 2020. "Run-to-Run Control of Chemical Mechanical Polishing Process Based on Deep Reinforcement Learning." *IEEE Transactions on Semiconductor Manufacturing* 33 (3): 454–465. doi:10.1109/TSM.2020.3002896.

Yu, Tian, Jing Huang, and Qing Chang. 2020. "Mastering the Working Sequence in Human–Robot Collaborative Assembly Based on Reinforcement Learning." *IEEE Access* 8: 163868–163877. doi:10.1109/ACCESS.2020.3021904.

Zhang, Wei, and Thomas G. Dietterich. 1995. " A Reinforcement Learning Approach to Job-Shop Scheduling." Proceedings of the 14th International Joint Conference on Artificial Intelligence, Vol. 2, 1114–1120.

Zhang, Nailong, and Wujun Si. 2020. "Deep Reinforcement Learning for Condition-Based Maintenance Planning of Multi-Component Systems Under Dependent Competing Risks." *Reliability Engineering & System Safety* 203: 107–094. doi:10.1016/j.ress.2020.107094.

Zhang, Tai, Fan Zhou, Jun Zhao, and Wei Wang. 2020. "Deep Reinforcement Learning for Secondary Energy Scheduling in Steel Industry." In *2020 2nd International Conference on Industrial Artificial Intelligence (IAI)*, Shenyang, China, July 24–26.

Zhao, Minghui, Xian Guo, Xuebo Zhang, Yongchun Fang, and Yongsheng Ou. 2019. "ASPW-DRL: Assembly Sequence Planning for Workpieces Via a Deep Reinforcement Learning Approach." *Assembly Automation* 40 (1): 65–75. doi:10.1108/AA-11-2018-0211.

Zhao, Xin, Huan Zhao, Pengfei Chen, and Han Ding. 2020. "Model Accelerated Reinforcement Learning for High Precision Robotic Assembly." *International Journal of Intelligent Robotics and Applications*4 (2): 202–216. doi:10.1007/s41315-020-00138-z.

Zheng, Shuai, Chetan Gupta, and Susumu Serita. 2020. "Manufacturing Dispatching Using Reinforcement and Transfer Learning." *Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 655–671. doi:10.1007/978-3-030-46133-1_39.

Zhou, Zhenpeng, Xiaocheng Li, and Richard N. Zare. 2017. "Optimizing Chemical Reactions with Deep Reinforcement Learning." *ACS Central Science* 3 (12): 1337–1344. doi:10.1021/acscentsci.7b00492.

Zhou, Tong, Dunbing Tang, Haihua Zhu, and Liping Wang. 2021. "Reinforcement Learning With Composite Rewards for Production Scheduling in a Smart Factory." *IEEE Access* 9: 752–766. doi:10.1109/ACCESS.2020.3046784.

Zhou, Yu, Tong Xing, Yue Song, Yajing Li, Xiefeng Zhu, Guo Li, and Shuiting Ding. 2020. "Digital-Twin-Driven Geometric Optimization of Centrifugal Impeller with Free-Form Blades for Five-Axis Flank Milling." *Journal of Manufacturing Systems* 58: 22–35. doi:10.1016/j.jmsy.2020.06.019.

Zhou, Longfei, Lin Zhang, and Berthold K. P. Horn. 2020. "Deep Reinforcement Learning-Based Dynamic Scheduling in Smart Manufacturing." *Procedia CIRP* 93: 383–388. doi:10.1016/j.procir.2020.05.163.

Zhu, Kai, Naihua Ji, and Xiang Dong Li. 2020. “Hybrid Heuristic Algorithm Based On Improved Rules Reinforcement Learning for 2D Strip Packing Problem.” *IEEE Access* 8: 226784–226796. doi:10.1109/ACCESS.2020.3045905.

Zhu, Huayu, Mengrong Li, Yong Tang, and Yanfei Sun. 2020. “A Deep-Reinforcement-Learning-Based Optimization Approach for Real-Time Scheduling in Cloud Manufacturing.” *IEEE Access* 8: 9987–9997. doi:10.1109/ACCESS.2020.2964955.

Zimmerling, Clemens, Christian Poppe, and Luise Kärger. 2020. “Estimating Optimum Process Parameters in Textile Draping of Variable Part Geometries – A Reinforcement Learning Approach.” *Procedia Manufacturing* 47: 847–854. doi:10.1016/j.promfg.2020.04.263.

Zou, Yanbiao, and Rui Lan. 2020. “An End-to-End Calibration Method for Welding Robot Laser Vision Systems With Deep Reinforcement Learning.” *IEEE Transactions on Instrumentation and Measurement*69 (7): 4270–4280. doi:10.1109/TIM.2019.2942533.