***Original Research Article***

# Online Reinforcement Learning for Real-Time Monitoring and Control in Production Systems

---

## Abstract

Modern production systems are increasingly complex and variable, requiring adaptive and intelligent solutions for real-time monitoring and control. Traditional methods, such as linear models and static optimization, often fall short in addressing the dynamic, high-dimensional demands of industrial environments. Online reinforcement learning (RL) offers a compelling alternative by enabling systems to continuously learn and optimize decision-making through real-time interactions with their environment. This review explores the current advancements in online RL, focusing on its applications in predictive maintenance, dynamic scheduling, and process optimization. Key methodologies, including Deep RL, policy-based approaches, and hybrid frameworks, are examined for their ability to enhance scalability, adaptability, and efficiency in Industry 4.0 ecosystems.While online RL holds great promise, challenges such as computational demands, algorithmic stability, and limited real-world validation remain significant barriers to its widespread adoption. The lack of standardized benchmarks further hinders the evaluation and comparability of RL solutions across different industrial contexts. To address these gaps, this review identifies critical research directions, including the development of efficient algorithms, the integration of domain knowledge for improved stability, and the deployment of multi-agent RL systems for distributed manufacturing networks.By synthesizing recent advancements and identifying unresolved challenges, this paper underscores the transformative potential of online RL in creating intelligent, autonomous, and resilient production systems.

**Key words;**Online Reinforcement Learning, Real-Time Monitoring, Production Systems, Dynamic Scheduling, Predictive Maintenance, Process Optimization, Industry 4.0

# 1.Introduction

## 1.1 Background and Context

Modern production systems form the backbone of industries, facilitating large-scale manufacturing and supply chain operations[1]. These systems, however, are often subjected to dynamic environments characterized by fluctuating workloads, complex interdependencies, and high variability[2]. Maintaining efficiency, reducing downtime, and ensuring quality in these environments demands effective monitoring and control systems capable of functioning in real-time[3].

Traditional approaches, such as linear learning models and static optimization methods, have long been the standard tools for tasks like fault detection, scheduling, and resource allocation. Despite their popularity, these methods are inherently limited in adaptability and scalability, making them inadequate for handling the complexities of modern production systems [4].

In contrast, reinforcement learning (RL) has emerged as a powerful paradigm for dynamic decision-making. RL offers a data-driven and adaptive approach to optimization by enabling an agent to discover optimal actions through trial-and-error interactions with its environment[5]. What sets online RL apart from traditional offline RL is its ability to continuously learn and update policies in real time, making it particularly well-suited for production systems where operating conditions are constantly changing [6].

The application of online RL in real-time monitoring and control aligns closely with the principles of Industry 4.0, where systems are expected to self-optimize, adapt autonomously, and make decisions based on live data [7].For example, predictive maintenance, where equipment failures are anticipated and mitigated before they occur, can benefit significantly from online RL's ability to learn patterns and adjust decision-making dynamically. Similarly, tasks like real-time process optimization and resource scheduling can be revolutionized by its adaptive capabilities, leading to improved production efficiency and cost savings [8].

However, integrating online RL into industrial settings is not without challenges. Computational requirements for real-time decision-making, stability in learning algorithms, and the need for robust handling of noisy, high-dimensional data streams are just a few of the hurdles that researchers and practitioners face [9]. These challenges underscore the need for a comprehensive exploration of online RL's capabilities, limitations, and potential applications in production systems.

## 1.2 Problem Statement and Motivation

Achieving real-time adaptability in production systems is one of the most critical challenges faced by modern industries. Traditional approaches, such as linear models, have been effective for tasks like anomaly detection and fault diagnosis due to their simplicity and interpretability. However, these models rely on fixed assumptions and predefined patterns, rendering them unsuitable for managing the complexity and variability of non-linear production processes [10]. Similarly, static optimization and offline decision-making methods fail to account for rapidly changing real-time conditions, which require systems to continuously adjust and learn from dynamic environments.

Online reinforcement learning (RL) offers a promising alternative by combining real-time learning with adaptive decision-making. Unlike static systems, online RL enables continuous improvement through iterative feedback, allowing production processes to respond dynamically to changing scenarios. For instance, an online RL model can simultaneously detect anomalies and determine corrective actions to minimize downtime, making it a powerful tool for both proactive and reactive process management [11]. Online RL's ability to learn and optimize in real time makes it a vital solution for meeting the demands of modern industrial systems.

The motivation for this review arises from the growing need for adaptive, real-time solutions to address the complexities and variability of modern production systems. Traditional methods, constrained by their static nature and reliance on pre-defined patterns, struggle to meet the dynamic demands of today's industrial environments. Online reinforcement learning (RL) stands out as a transformative solution, enabling continuous adaptation, intelligent decision-making, and optimization in real time. By bridging gaps in understanding and addressing unresolved challenges, this review aims to provide researchers and practitioners with a critical resource for advancing the field. It offers a comprehensive analysis of state-of-the-art developments, key obstacles, and actionable pathways to harness the full potential of online RL in revolutionizing production systems.

### 1.3 Purpose and Scope

The primary purpose of this article is to critically examine the role of online reinforcement learning in advancing real-time monitoring and control within production systems. Specifically, this review aims to:

- Provide an overview of the foundational principles and technical frameworks underpinning online RL.
- Analyze recent research and applications of online RL in industrial contexts, with a focus on real-time decision-making.
- Identify the key challenges and limitations that hinder its widespread adoption in production systems.
- Highlight opportunities for further innovation, including integration with emerging technologies like IoT and edge computing.

This review is particularly relevant for researchers seeking to bridge the gap between traditional anomaly detection methods, such as linear models, and modern adaptive AI-driven approaches. It also provides practical insights for industries aiming to enhance operational efficiency through cutting-edge automation technologies.

### 1.4 Structure of the Article

To address these objectives, the article is structured as follows:

Section 2: Literature Review examines the evolution of monitoring and control methods in production systems, contrasting traditional approaches like linear models with the adaptive capabilities of online RL. This chapter also categorizes and evaluates existing research on online RL applications.

Section 3: Technical Frameworks and Concepts delves into the foundational principles of online RL, highlighting key algorithms, methodologies, and their suitability for real-time production systems.

Section 4: Applications and Challenges explores real-world implementations of online RL in areas such as predictive maintenance, dynamic scheduling, and process optimization. It also discusses the practical and technical challenges associated with deploying these systems in industrial settings.

Section 5: Conclusion summarizes the findings, identifies research gaps, and provides recommendations for future exploration of online RL in production systems

## 2. Literature Review

### 2.1 Traditional Approaches in Monitoring and Control

Traditional methods have long formed the foundation for monitoring and control in industrial production systems[12]. These approaches, including linear learning models, statistical methods, and rule-based systems, were designed for stable environments where processes were predictable and computational resources were

limited[13].Their simplicity and ease of implementation made them indispensable tools for early industrial automation, particularly for tasks such as fault detection, process monitoring, and anomaly identification[14].

Linear learning models are among the most widely adopted techniques in industrial monitoring and control. These models assume a linear relationship between input features, such as sensor data, and output variables, such as production quality or system efficiency[

15]. Linear regression, for example, is frequently applied to monitor process parameters like temperature, pressure, or flow rates, detecting deviations that may signal faults or inefficiencies [16],[17]. In steel manufacturing, regression-based models have been used to track rolling mill temperatures, with deviations from expected ranges triggering maintenance actions[18].Similarly, autoregressive models have been applied to time-series data, such as machine vibrations, to predict bearing failures in rotating machinery. These models are valued for their interpretability, computational efficiency, and ability to handle relatively stable production environments[19].

However, the reliance on linear assumptions presents a significant limitation when applied to modern production systems, which often exhibit complex, non-linear behaviors[20]. In chemical manufacturing, for example, interactions between temperature, pressure, and reactant concentrations create non-linear dependencies that linear models fail to capture[21]. Moreover, the growing adoption of IoT-enabled devices in manufacturing has introduced high-dimensional data streams, further straining the applicability of these models to real-time, dynamic environments [22].

Statistical techniques, such as principal component analysis (PCA) and statistical process control (SPC), are also prominent in traditional monitoring and control. PCA reduces the dimensionality of large sensor datasets by isolating key variables that contribute to process variations, while SPC uses control charts to monitor production metrics such as cycle time and yield[23],[24]. These methods are effective for detecting anomalies and monitoring trends based on historical data. For instance, in semiconductor manufacturing, SPC has been employed to maintain process stability by flagging deviations from acceptable quality ranges[25]. Similarly, PCA is widely used in continuous manufacturing to identify shifts in process variables that might indicate faults[26].

Despite their utility, statistical methods have inherent drawbacks. They rely on historical patterns and fixed thresholds, which limits their adaptability to evolving conditions[27]. For instance, in dynamic production environments where product specifications and operational requirements frequently change, such as in mass customization settings, statistical thresholds may become obsolete[28]. Additionally, these methods are not inherently designed for real-time operation, as they often rely on batch processing of historical data rather than live streaming data[29].

Rule-based systems, another cornerstone of traditional control methods, rely on predefined “if-then” rules to monitor and manage production processes[30]. These systems are particularly valued for their simplicity, interpretability, and ease of deployment, making them effective in static and well-defined environments[31]. For instance, in food processing industries, rule-based systems might monitor oven temperatures and trigger alarms if values exceed a predefined threshold[32]. Similarly, in automotive assembly lines, rule-based systems are used to reject defective parts based on sensor readings [33].

However, as production systems become more complex and dynamic, the rigidity of rule-based systems becomes a major drawback [34]. Developing rules to account for every possible scenario in a dynamic environment is not feasible, and unforeseen events can cause these systems to fail. For example, during unexpected equipment wear or fluctuating supply chain demands, rule-based systems may struggle to adapt and continue flagging anomalies without offering actionable solutions. This rigidity often leads to unnecessary downtime and increased operational inefficiencies [35].

While these traditional methods have been instrumental in advancing industrial automation, they share common limitations that hinder their effectiveness in modern production systems[36]. Linear learning models, statistical

techniques, and rule-based systems all rely on fixed assumptions and lack the flexibility to adapt to real-time changes[37]. Furthermore, their scalability is limited in the face of high-dimensional data and complex system dynamics[38]. These limitations have driven the search for more adaptive and intelligent solutions capable of handling the non-linear, high-dimensional, and dynamic nature of modern production environments.

In some cases, traditional methods have evolved to remain relevant in specific scenarios. For instance, hybrid systems combining rule-based logic with machine learning algorithms have shown promise in leveraging the interpretability of traditional methods while adding adaptability through AI [39]. Similarly, statistical methods like SPC are often integrated into broader frameworks that include predictive analytics for anomaly detection[40]. These advancements indicate that while traditional methods are no longer standalone solutions for modern production challenges, they can still play a complementary role when integrated with more advanced approaches.

To address the growing demands of real-time, adaptive monitoring and control, reinforcement learning (RL) has emerged as a powerful alternative[41]. Unlike traditional methods, RL offers dynamic, data-driven solutions that continuously learn and optimize in response to changing [42]. The evolution of RL from offline optimization to real-time online learning marks a significant shift in how industrial systems are managed, paving the way for intelligent and autonomous production processes.

## 2.2 Evolution of Reinforcement Learning in Manufacturing

The growing complexity of modern production systems has driven a transition from static optimization approaches to more dynamic and intelligent solutions. Reinforcement learning (RL) has emerged as a promising framework to meet these demands by enabling systems to learn optimal strategies through interactions with their environment[43]. Over time, the application of RL in manufacturing has evolved significantly, starting with offline methods that addressed static problems and progressing toward online RL, which offers real-time adaptability and decision-making[44],[45].

In its early stages, RL was primarily applied in manufacturing through offline methods. These approaches trained models on historical data to derive optimal policies, which were then deployed without further updates during operation. Offline RL proved effective for solving static optimization problems, such as job-shop scheduling, inventory management, and resource allocation[46]. For instance, RL has been used to optimize production schedules in multi-stage systems, minimizing production time and maximizing resource utilization and also was employed to sequence production tasks, reducing bottlenecks and improving overall throughput[47]. These methods often relied on value-based algorithms, such as Q-learning, which helped agents associate states with actions that maximized cumulative rewards[48]. However, offline RL's reliance on fixed policies limited its ability to adapt to dynamic environments, making it less effective in real-time production systems where conditions like demand variability, machine performance, or raw material availability frequently shift [49].

The emergence of online reinforcement learning addressed the limitations of offline approaches by enabling systems to continuously update their policies through real-time interactions with the environment. Unlike offline RL, which remains static after deployment, online RL allows agents to balance exploration of new strategies with exploitation of learned policies to optimize performance[50]. This adaptability makes online RL particularly suited for dynamic and uncertain production environments.

One of the key strengths of online RL is its capability to integrate real-time data into the decision-makingprocess [51]. In predictive maintenance, for example, online RL can monitor equipment health by analyzing continuous sensor data, predict potential failures, and adjust maintenance schedules dynamically to minimize downtime and costs.[52]highlights an application where RL-based systems proactively reschedule tasks to balance maintenance needs with operational priorities, improving overall system efficiency. Similarly,

in dynamic scheduling, online RL can allocate resources adaptively, ensuring optimal utilization even as constraints or production demands shift in real time[53].

The evolution of online RL has been further accelerated by several technological innovations. One significant advancement is the integration of RL with deep learning, known as Deep Reinforcement Learning (Deep RL). By leveraging neural networks, Deep RL algorithms, such as Deep Q-Networks (DQNs), can handle high-dimensional data from IoT sensors, enabling RL agents to make decisions in complex environments[54]. Another breakthrough is the development of policy-based methods like Proximal Policy Optimization (PPO) and Actor-Critic algorithms, which are particularly well-suited for continuous control tasks[55]. These methods enable smoother policy updates, reducing the risk of instability during real-time learning.

The transition from offline to online RL marks a paradigm shift in the way industrial processes are optimized and controlled. While offline RL laid the foundation for understanding how reinforcement learning could be applied to manufacturing, online RL has unlocked its full potential by making systems adaptive and responsive to the demands of real-time environments. This evolution has positioned RL as a key enabler of intelligent, autonomous production systems, capable of navigating the complexities of modern industrial operations[56].

## 2.3 Current Trends in Online RL Research

In recent years, research on online reinforcement learning (RL) has expanded rapidly, driven by the need for adaptive, data-driven solutions in dynamic industrial environments. Online RL's unique capability to learn and adjust policies in real time has made it a critical tool for solving complex challenges in manufacturing. Current research trends focus on two primary areas: practical applications of online RL in production systems and ongoing advancements in RL methodologies designed to enhance its performance in real-time scenarios. The practical applications of online RL span multiple facets of production systems, with significant progress made in predictive maintenance, dynamic scheduling, and real-time process optimization.

One of the most impactful applications of online RL is in predictive maintenance, where the focus is on anticipating equipment failures and minimizing unplanned downtime[57]. Unlike traditional approaches that rely on predefined maintenance schedules or statistical models, online RL adapts to real-time sensor data to identify when maintenance is truly needed. By continuously monitoring and analyzing live operational data, the system reduces unnecessary maintenance interventions while preventing catastrophic failures, ensuring minimal disruption to production workflows.

Another critical application is dynamic scheduling and resource allocation, which addresses the challenge of managing production tasks in environments with fluctuating workloads or resource constraints[58]. Online RL provides a flexible solution by continuously updating its scheduling policies based on real-time feedback. For instance, [59]applied RL to task scheduling in semiconductor manufacturing, were real-time adjustments significantly reduced cycle times and enhanced resource utilization. These advancements highlight the ability of RL to maintain operational efficiency even in highly variable and time-sensitive production settings.

In process optimization and control, online RL has been utilized to fine-tune production parameters to maximize output quality while minimizing waste and energy consumption[60]. This adaptability is especially valuable in industries where even small deviations can result in substantial financial or environmental consequences.Alongside its diverse applications, online RL research has seen significant progress in the development of methodologies that improve its scalability, efficiency, and stability in industrial environments[61].

Recent advancements in online reinforcement learning (RL) have been built upon foundational innovations such as Deep Reinforcement Learning (Deep RL) and policy-based methods. These approaches, already noted for their ability to process high-dimensional data and optimize continuous control tasks, continue to play a pivotal role in enhancing real-time decision-making. Deep RL, with techniques like Deep Q-Networks (DQNs), and

policy-based methods such as Proximal Policy Optimization (PPO) and Actor-Critic algorithms, remain essential for managing complex production systems, particularly in IoT-enabled environments [62].

A more recent trend is the development of hybrid RL frameworks, which combine the predictive capabilities of model-based RL with the adaptability of model-free approaches which leverage simulations to anticipate outcomes while retaining the flexibility to adapt in real-time, striking a balance between computational efficiency and responsiveness[63]. For instance, ahybrid RL application in supply chain management that dynamically optimizes inventory and logistics based on real-time fluctuations in demand and transportation constraints[64]. Such frameworks are particularly valuable in scenarios where uncertainty and complexity require fast and reliable decision-making.

Another emerging area is multi-agent reinforcement learning (MARL), where multiple agents collaborate or compete to optimize system-wide performance[65]. This approach has shown promise in complex industrial networks, such as collaborative robotics and distributed production systems. By enabling agents to share information and coordinate actions, MARL offers a pathway to achieving greater efficiency and resilience in interconnected manufacturing ecosystems.

# 3. Technical Frameworks and Concepts in Online RL

Reinforcement Learning (RL) has long been recognized as a potent framework for dynamic decision-making, allowing agents to learn optimal policies through trial-and-error interactions with their environments [66].

## 3.1 Foundations of Online Reinforcement Learning

The fundamental principles of RL include the interaction between an agent and its environment, where the agent navigates through a defined state space by taking actions that result in state transitions governed by probabilistic dynamics [67] A reward signal provides feedback, guiding the agent toward an optimal policy that maximizes cumulative rewards over time. Online RL builds on these principles by integrating new data as it becomes available, refining the agent's policy continuously [68] This capability makes online RL particularly suitable for applications requiring adaptability and responsiveness, such as real-time monitoring and control in production systems [69].

According to [70]. Reinforcement Learning (RL) is a decision-making framework where an agent learns to make sequential decisions by interacting with an environment with the goal to maximize cumulative rewards over time, guided by several core elements:

- **States** $(S)$: The state $s_t$ at time $t$ represents the agent's perception of the environment. In production systems, states may encompass machine conditions, inventory levels, or system throughput.
- **Actions** $(A)$: Actions $a_t$ are decisions made by the agent to influence the environment. For example, $a_t$ in manufacturing could be setting machine parameters or scheduling maintenance.
- **Rewards** $(R)$: The scalar signal $r_t$ provides feedback on the agent's actions, helping it evaluate success. In real-time systems, rewards might represent production efficiency, cost savings, or fault avoidance.
- **Policies** $(\pi(a \mid s))$: The policy $\pi(a \mid s)$ defines the probability of taking action $a$ in state $s$. Policies can be deterministic or stochastic, depending on the application.
- **Value Functions**: These include the state value function $V(s)$, representing the expected cumulative reward starting from state $s$, and the state-action value function $Q(s, a)$, which evaluates the expected reward for taking action $a$ in state $s$.

Online RL agents operate by navigating through state spaces $(S)$and making decisions from a set of actions $(A)$. These actions transition the environment into new states based on probabilistic dynamics ($P(s_{t+1} \mid s_t, a_t)$), and the agent receives a reward signal ($R_t$) to guide learning, as illustrated in Figure 1. The agent's goal is to optimize its policy ($\pi(a \mid s)$), which dictates the probability of selecting actions in specific states, to maximize the expected cumulative reward ($\sum_{t=0}^{\infty} \gamma^t R_t$) (Padakandla 2021).

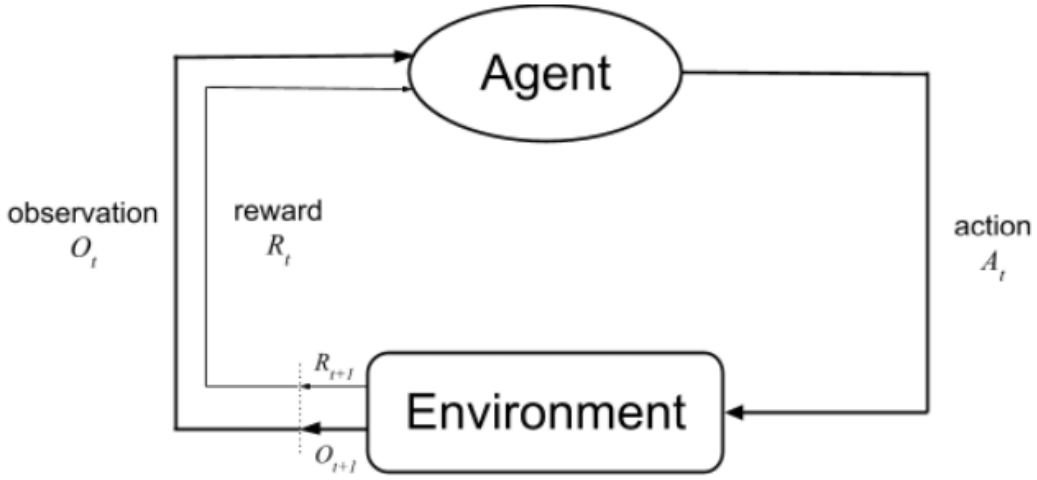



Figure 1: RL cycle (Agent-Environment Interaction), showing observation (state), actions, rewards, policy updates, and the feedback loop [71].

In real-time production systems, this framework offers several advantages. For instance, real-time adaptation enables RL agents to dynamically adjust policies as operating conditions evolve, such as sudden machine malfunctions or shifts in production schedules [72]. Continuous learning ensures that RL systems remain effective even in the face of long-term changes, such as wear and tear in equipment [72]. These features make online RL indispensable in scenarios where responsiveness and adaptability are essential, such as the complex environments of Industry 4.0 [73].

## 3.2 Frameworks for Real-Time Monitoring and Control

The architecture of an online RL system significantly impacts its effectiveness in real-time settings. Broadly, RL algorithms are categorized into value-based, policy-based, and hybrid methods, with further distinction between model-based and model-free approaches [74].

**Value-BasedMethods**

According to [75], value-based methods, such as Q-learning and Deep Q-Networks (DQN), focus on estimating the state-action value function $Q(s, a)$. Q-learning iteratively updates this function using the Bellman equation:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)],$$

where:

- $r$ is the immediate reward,
- $s'$ is the next state,
- $a'$ is the next action,

- $\alpha$ is the learning rate,
- $\gamma$ is the discount factor for future rewards [76].

DQN extends Q-learning to handle high-dimensional state spaces by employing deep neural networks (DNNs) to approximate the $Q(s,a)$ function [77]. In production systems such as semiconductor fabrication plants, DQN has been used to optimize tasks such as dynamic resource allocation and robotic assembly, where traditional methods fail due to the complexity of the environment [78].

**Policy-BasedMethods**

According [79] policy-based methods optimize the policy $\pi(a \mid s)$ directly, which determines the agent's behavior, with the objective to maximize the expected cumulative reward $J(\pi)$:

$$J(\pi) = \mathbb{E}_{\pi}[\sum_{t=0}^{\infty} \gamma^t r_t].$$

making them more suitable for environments with continuous action spaces. Algorithms such as Proximal Policy Optimization (PPO) and Trust Region Policy Optimization (TRPO) adjust policies incrementally to ensure stable learning [80] and it's a used algorithm that balances exploration and exploitation while maintaining stability during training. It updates policies using a clipped objective function:

$$L^{CLIP}(\theta) = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)],$$

where:

- $r_t(\theta)$ is the probability ratio between new and old policies,
- $\hat{A}_t$ is the advantage estimate at time $t$,
- $\epsilon$ controls the magnitude of updates.

Policy-based methods are particularly suitable for applications requiring continuous control, such as robotic manipulation or process optimization in manufacturing [81].

**HybridMethodsand Actor-Critic Approach**

Actor-Critic methods combine the strengths of value-based and policy-based algorithms [82]. The**Actor** adjusts the policy based on feedback from the Critic, which evaluates the policy using a value function [83]. These methods are particularly effective in scenarios requiring long-term planning and real-time adaptability, such as predictive maintenance and fault mitigation in industrial systems [73].

# 4. Applications and Challenges in Real-Time Production Systems

The adoption of online RL in real-time production systems is transforming traditional industrial processes, enabling greater adaptability and efficiency in dynamic environments. This chapter examines the key areas where online RL is applied, considers the challenges that hinder its implementation, and explores emerging opportunities that can drive further advancements. By integrating specific techniques and real-world examples, this chapter highlights the tangible impact of online RL while addressing the complexities of its deployment.

## 4.1 Applications in Real-Time Monitoring and Control

Online RL has demonstrated significant potential across several areas of production systems, including predictive maintenance, dynamic resource allocation, and process optimization (Panzer and Bender 2022). These applications leverage the continuous learning capabilities of online RL to address the challenges posed by variability, complexity, and uncertainty in industrial settings [72].

#### 4.1.1 Predictive Maintenance

Predictive maintenance, a cornerstone of modern industrial operations, seeks to preempt equipment failures by using advanced monitoring techniques [84]. Online RL enhances traditional approaches by dynamically learning and adapting maintenance strategies based on real-time data [85].

For instance, Rolls-Royce integrates RL with its Intelligent Borescope System to analyze engine blades by mapping irregularities and inconsistencies, thereby reducing the time to carry out specific inspections by 75% and is projected to save up to £100m (approximately $128m) in inspection costs over five years [86]. Techniques like Deep Q-Networks (DQN) are used to model state-action values, where states represent machine conditions, and actions correspond to maintenance decisions [87]. By employing experience replay buffers, the system ensures efficient learning from historical and real-time data [88].

Another example is Siemens Senseye Predictive Maintenance solution [89], where online RL models use Actor-Critic frameworks to schedule predictive maintenance for critical machinery. The Actor suggests optimal maintenance schedules, while the Critic evaluates the long-term impact of these schedules on production continuity [90], this dual approach balances immediate needs with broader operational goals, resulting in reduced downtime and optimized resource allocation.

#### 4.1.2 Dynamic Resource Allocation

Dynamic resource allocation is critical in production systems, particularly in scenarios involving fluctuating demands and constrained resources. Online RL enables efficient distribution of resources such as materials, labor, and machinery by continuously optimizing decision-making based on real-time inputs.

Amazon Fulfillment Centers utilize RL-powered robotic systems to manage inventory retrieval and placement [91], using RL agents trained with Proximal Policy Optimization **(PPO)** optimize task allocation for thousands of autonomous robots [92]. Each robot operates as an individual agent, dynamically adjusting its actions based on proximity to inventory, current task loads, and real-time traffic within the warehouse [91]. The PPO algorithm ensures stability during training by using clipped policy updates, which prevent sudden shifts in decision-making that could disrupt operations.

In automotive manufacturing, **Toyota** employs RL-based scheduling systems that use Deep Deterministic Policy Gradient **(DDPG)** to optimize production workflows [93]. DDPG, a hybrid approach combining value-based and policy-based methods, enables continuous control over scheduling variables such as machine availability, production line sequencing, and worker assignments [94]. By incorporating delayed reward signals, the system ensures that short-term decisions do not compromise long-term throughput.

#### 4.1.3 Process OptimizationThrough Reinforcement Learning and Digital Twins

Process optimization focuses on improving throughput, reducing costs, and ensuring consistent product quality in production systems. Online RL enhances this domain by enabling systems to adapt to changing conditions and optimize parameters dynamically. BASF serves as a leading example of how reinforcement learning (RL) and digital twin technologies can transform industrial processes across sectors.

**The Issue: Inefficiency in Complex Systems**

Traditional methods in chemical manufacturing and other production systems often fail to adapt to dynamic conditions, relying on pre-defined parameters that are inefficient for real-time adjustments. These limitations

result in wasted energy, inconsistent product quality, and excessive exploration costs, particularly in high-stakes environments like chemical production. Additionally, the use of harmful solvents and resource-intensive processes exacerbates environmental and operational challenges, highlighting the need for smarter, more adaptive solutions.

**The Solution: RL-Powered Digital Twins and Data-Driven Tools**

BASF addresses these challenges using RL agents embedded in digital twin environments. In chemical manufacturing, digital twins act as virtual replicas of physical processes, enabling RL agents to simulate various scenarios and predict optimal adjustments to parameters such as temperature, pressure, and reactant flow rates [95].By integrating Model-Based RL, BASF minimizes costly trial-and-error experimentation while maximizing yields and reducing energy consumption [96].

To extend its innovation pipeline, BASF partnered with Imperial College London to launch Solve, a spinout company focused on high-value chemical manufacturing. Solve combines machine learning with flow chemistry to create large datasets from precisely controlled experiments, enabling rapid predictions of optimal production methods. This approach reduces waste, reliance on harmful solvents, and costs while improving scalability and sustainability [97].

BASF also applies digital twin technology beyond chemicals through its involvement in the European Union's SPHERE project. SPHERE leverages digital twins to enhance energy efficiency and lifecycle management in buildings. SPHERE relies on simulation and predictive modelingcore principles of RL systems. BASF's subsidiary, Master Builders Solutions, contributes its expertise in Building Information Modeling (BIM), providing adaptive planning and management tools across the construction lifecycle [98].

**How It Helped: Achieving Adaptive, Efficient Optimization**

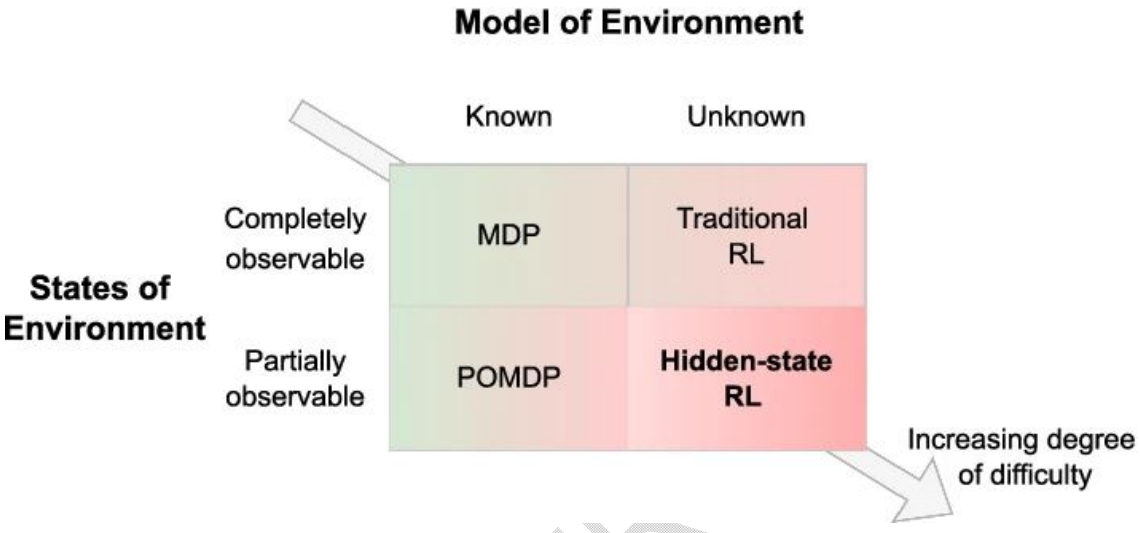
BASF's integration of RL, digital twins, and machine learning has significantly improved process efficiency and sustainability:

- Chemical Manufacturing: RL-driven digital twins enable real-time optimization, ensuring maximum yields, reduced exploration costs, and energy efficiency. Solve's machine learning models further enhance chemical production by improving scalability and reducing the environmental impact of harmful solvents.
- Construction: Through SPHERE, BASF has demonstrated the versatility of digital twins in optimizing energy use and lifecycle costs, extending adaptive technologies to large-scale building projects.
- By transitioning from static systems to data-driven optimization, BASF exemplifies how RL and digital twin technologies can address the challenges of real-time monitoring and control in production systems. These efforts underline BASF's leadership in advancing Industry 4.0 through intelligent, adaptive solutions.

Similarly, Tesla's Gigafactory is known for its reliance on advanced automation and machine learning to optimize battery production. Reinforcement learning (RL) has been highlighted in research studies as a potential tool for optimizing operations in such high-tech manufacturing environments. For instance, RL agents trained with Soft Actor-Critic (SAC) algorithms could adjust robotic arm movements to enhance precision and speed, ensuring stable performance even when environmental variables, such as assembly line speeds, fluctuate [99],[100]. While specific public documentation of Tesla's direct application of RL in Gigafactories is limited, Tesla's extensive use of robotic systems and its pursuit of end-to-end AI-driven solutions for manufacturing underscore the potential for such innovations. Tesla's commitment to AI and automation, as showcased in its AI Day presentations and the development of its humanoid robot [101], further reflects its broader vision for integrating intelligent learning systems across its operations.

## 4.2 Key Challenges and Research Gaps

Despite its success, the implementation of online RL in production systems is not without challenges. These challenges stem from the computational demands of real-time operations, the quality, availability of data, state of environment, and the need to balance stability with adaptability [102]. Real-time decision-making demands that RL algorithms process vast amounts of data within tight latency constraints [103]. High-dimensional state-action spaces exacerbate computational burdens, leading to delays that can compromise system performance[104]. Figure 2 illustrates the increasing degree of difficulty base on the state of the environment.




Increasing Degree of Difficulty for RL agent in Predictive Maintenance [105].

**Techniques for Mitigation**:

1. **Edge Computing**: Offloading computations to edge devices reduce latency by processing data closer to its source, for example, in smart factories, RL agents deployed on edge devices can process sensor data locally, ensuring real-time responsiveness [106].

2. **Parallelized Training**: Distributed RL frameworks, such as Ape-X, enable parallel training of RL agents, significantly reducing computation times while maintaining accuracy [107].

### 4.2.2 Data-Related Issues

The quality and availability of data significantly impacts the performance of RL agents, with common issues including noisy sensor readings, incomplete state observations, and continuous data streams that overwhelm traditional RL frameworks [108],[103].

**Techniques for Mitigation**:

1. Partially Observable Markov Decision Processes **(POMDPs)**: POMDPs extend traditional RL models by incorporating probabilistic reasoning to handle missing or uncertain data [109].

2. Noise-Filtering Techniques: Tools like Kalman filters and Gaussian noise reduction smooth out fluctuations in sensor data, ensuring more reliable state representations [110].

### 4.2.3 Stability vs. Adaptability

Balancing stability and adaptability are critical in dynamic environments. While exploration is necessary for discovering new strategies, excessive exploration can lead to performance degradation, especially in high-stakes applications [111].

**Techniques for Mitigation**:

1. **Reward Shaping**: Modifying reward structures to emphasize long-term goals helps guide exploration without compromising stability [112].

2. **Adaptive Exploration Strategies**: Techniques like epsilon-greedy decay gradually reduce exploration over time, allowing the agent to converge on optimal policies [113].

### 4.3 Emerging Opportunities

Emerging technologies and interdisciplinary approaches are paving the way for significant advancements in online reinforcement learning (RL), addressing current limitations and expanding its applicability. One of such developments is the integration of the Internet of Things (IoT) and edge computing, which has revolutionized the ability of RL agents to operate in real time. IoT-enabled systems provide high-frequency, rich data streams that capture the nuances of dynamic industrial environments, while edge computing reduces latency by processing data closer to the source. For instance, GE's Digital Twin Technology leverages IoT sensors and edge devices to create real-time virtual models of industrial assets, allowing RL agents to simulate operations and optimize parameters dynamically [114]. This combination enables more precise decision-making and accelerates the implementation of adaptive strategies in production systems.

Another breakthrough in online RL is the adoption of Multi-Agent Reinforcement Learning (MARL), where multiple agents collaborate to achieve system-wide optimization. Unlike single-agent systems, MARL distributes tasks among agents, allowing them to work in parallel while coordinating their actions to optimize resource allocation and operational efficiency. A compelling application of MARL is seen at DHL, where fleets of autonomous vehicles in logistics hubs are managed collaboratively using this approach. By employing MARL, DHL ensures seamless task allocation, real-time route optimization, and efficient resource management, significantly improving throughput and reducing delays [115].

To manage complexity in hierarchical systems, Hierarchical Reinforcement Learning (HRL) has gained traction. HRL structures decision-making by dividing tasks into layers, with high-level policies directing overarching strategies and low-level policies handling specific, granular actions. This approach is particularly beneficial for large-scale industrial systems where tasks must be coordinated across multiple levels. [116] highlights how HRL has been used in scenarios such as multi-stage manufacturing processes, where it helps streamline decision-making by delegating specific actions to appropriate sub-policies, thereby reducing computational overhead while maintaining system-wide coherence.

Hybrid methods that combine RL with traditional optimization techniques have also emerged as robust solutions for complex industrial challenges. By leveraging the strengths of RL in adaptability and learning with the computational efficiency of traditional methods, hybrid models offer superior performance. For example, Hyundai Steel integrates RL with linear programming models to optimize production schedules. This approach not only accelerates convergence during training but also minimizes waste, leading to enhanced sustainability and cost savings [117]. Such methods demonstrate the versatility of RL when paired with established optimization frameworks, opening new avenues for addressing dynamic industrial requirements.

Collectively, these innovations underscore the transformative potential of online RL when augmented with cutting-edge technologies and interdisciplinary approaches. From IoT-enabled digital twins and MARL in logistics to hierarchical structures and hybrid optimization models, these advancements are redefining the scope and effectiveness of RL in modern industrial systems.

## 5. Conclusion and Future Directionss

### 5.1 Summary of Key Findings

Online reinforcement learning (RL) has emerged as a transformative technology in revolutionizing real-time monitoring and control within production systems. By continuously learning from real-time feedback, online RL

enables systems to dynamically adapt to changing conditions, addressing challenges that traditional methods such as linear models and rule-based systems fail to resolve. Its strengths lie in adaptability, scalability, and the ability to process high-dimensional data, making it essential for modern production environments where complexity and variability are the norm.

Technological advancements such as Deep RL, policy-based methods like Proximal Policy Optimization (PPO), and hybrid frameworks have significantly enhanced RL's capabilities. These methodologies, discussed in earlier sections, enable RL systems to handle diverse industrial tasks, from dynamic scheduling and predictive maintenance to process optimization. For instance, in the automotive sector, online RL has been used to optimize assembly line operations, ensuring consistent production flow despite equipment variability. Similarly, in chemical manufacturing, RL-driven process control has improved yield while minimizing energy consumption. These successes highlight online RL's alignment with the goals of Industry 4.0, paving the way for adaptive, efficient, and intelligent production ecosystems.

## 5.2 Critical Gaps and Challenges

Despite its promise, several key challenges hinder the full realization of online RL in real-world production systems.

One major challenge is the lack of real-world validation. While most studies rely on simulated environments, these settings often fail to account for the unpredictability and complexity of real-world manufacturing. Factors such as noisy data, incomplete information, and unanticipated disruptions are underexplored [118]. Testing RL systems under live industrial conditions is essential to ensure robustness and reliability.

Another significant issue is computational demand. Real-time decision-making in high-dimensional environments, such as IoT-enabled factories, requires substantial processing power and memory. For example, training Deep RL models often involves extensive exploration, which can slow down critical operations, particularly in small-scale facilities with limited resources [119]. Addressing this challenge is vital for making online RL accessible to a broader range of industries.

Algorithmic stability is another critical concern. In live industrial settings, prolonged exploration or unstable policy updates can lead to erratic performance, which is unacceptable in high-stakes scenarios like robotic assembly lines or energy grids. Instabilities not only compromise operational efficiency but can also result in costly disruptions. Ensuring stable convergence of RL algorithms remains a pressing research need [120].

Finally, the absence of standardized benchmarks complicates progress in the field. Current research often prioritizes varying objectives, such as energy efficiency or throughput optimization, making it difficult to compare results across studies. Developing universal benchmarks tailored to diverse industrial needs will be crucial for fostering collaboration and accelerating innovation [121].

## 5.2 Future Research and Directions.

To overcome these challenges, future research must focus on the following key areas:

- **Real-World Deployment and Collaboration**: Collaboration between academia and industry is essential for deploying RL systems in live production environments. Real-world testing will uncover challenges not apparent in simulations, such as data quality issues and operational constraints, allowing algorithms to be refined for robustness and scalability. For example, partnerships with automotive or electronics manufacturers could help test RL-based scheduling systems in complex assembly lines.

- **Enhancing Computational Efficiency**: Developing lightweight RL models and integrating techniques such as model compression, edge computing, and federated learning will reduce computational demands, making RL systems feasible for small-scale factories and resource-constrained environments. These advancements can also enhance RL's applicability in decentralized settings, such as multi-factory networks.

• **Improving Algorithmic Stability**: Ensuring stable and reliable performance during real-time learning should be a priority. Integrating domain knowledge into RL frameworks can reduce the reliance on exploration, enabling faster convergence. Additionally, hybrid approaches that combine model-based simulations with model-free adaptability can balance efficiency and stability, especially in critical applications like predictive maintenance and process control.

• **Standardizing Benchmarks**: Establishing universal benchmarks for evaluating RL systems across industries will facilitate meaningful comparisons and accelerate progress. These benchmarks should account for diverse industrial goals, such as throughput, energy efficiency, sustainability, and cost-effectiveness, ensuring that RL solutions align with real-world objectives.

• **Exploring Multi-Agent RL**: Multi-agent reinforcement learning (MARL) offers exciting opportunities for collaborative decision-making in distributed production systems, such as interconnected factories or multi-robot assembly lines. Future research should explore how MARL can enhance resource allocation, improve system-wide efficiency, and enable seamless coordination across industrial networks.

• **Advancing Adaptive Learning in Manufacturing**: Adaptive learning systems, including reinforcement learning, hold significant promise for transforming manufacturing processes by enabling real-time optimization and dynamic decision-making. A dedicated exploration of how adaptive learning can streamline production, reduce waste, and enhance scalability is critical. This concept will be expanded in future work, including in "**Improving Manufacturing Processes Through Adaptive Learning**," which will focus on how adaptive systems can address key challenges in modern production environments.

### 5.4 Final Remarks

Online RL represents a paradigm shift in industrial automation, offering dynamic adaptability, scalability, and intelligent decision-making capabilities. Its success in predictive maintenance, process optimization, and scheduling demonstrates its potential to transform production systems into resilient, efficient, and autonomous ecosystems. However, achieving this vision will require addressing key challenges, including computational demands, stability issues, and the need for real-world validation.

As industries embrace the principles of Industry 4.0, online RL is poised to play a pivotal role in creating sustainable and adaptive manufacturing ecosystems. To unlock its full potential, interdisciplinary collaboration between researchers, practitioners, and industry stakeholders will be essential. By addressing existing gaps, refining algorithms, and exploring innovative applications, online RL can pave the way for the next generation of intelligent, real-time production systems, fundamentally reshaping the future of industrial operations.



# Reference

1. Guo, D., Li, M., Lyu, Z., Kang, K., Wu, W., Zhong, R.Y. and Huang, G.Q. (2021). Synchroperation in industry 4.0 manufacturing. *International journal of production economics*, 238, p.108171.
2. Durugbo, C.M. and Al-Balushi, Z. (2022). Supply chain management in times of crisis: a systematic review. *Management Review Quarterly*, [online] 73. doi:https://doi.org/10.1007/s11301-022-00272-x.
3. Li, C., Zheng, P., Yin, Y., Wang, B. and Wang, L. (2023). Deep reinforcement learning in smart manufacturing: A review and prospects. *CIRP Journal of Manufacturing Science and Technology*, 40, pp.75–101.

4. Zope, K., Singh, K., Nistala, S.H., Basak, A., Rathore, P. and Runkana, V. (2019). Anomaly Detection and Diagnosis In Manufacturing Systems. *Annual Conference of the PHM Society*, 11(1). doi:https://doi.org/10.36001/phmconf.2019.v11i1.815.
5. Schreiber, T., Netsch, C., Baranski, M. and Mueller, D. (2021). Monitoring datadriven reinforcement learning controller training: A comparative study of different training strategies for a realworld energy system. *Energy and Buildings*, 239, p.110856.
6. Venkataswamy, V., Grigsby, J., Grimshaw, A. and Qi, Y. (2022). Launchpad: Learning to Schedule Using Offline and Online RL Methods. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.2212.00639.
7. Qin, Z. (2023). Adaptive Control Towards SelfOrganizing Manufacturing Network.
8. Okuyelu, O. and Adaji, O. (2024). AIDriven Realtime Quality Monitoring and Process Optimization for Enhanced Manufacturing Performance. *Journal of Advances in Mathematics and Computer Science*, 39(4), pp.81–89.
9. Srinivasan, A. (2023). Reinforcement Learning: Advancements, Limitations, and Real-world Applications. *Indian Scientific Journal Of Research In Engineering And Management*, 07(08). doi:https://doi.org/10.55041/ijsrem25118.
10. Pattyam, Sandeep Pushyamitra (2020). AI in Data Science for Predictive Analytics: Techniques for Model Development, Validation, and Deployment. *Journal of Science & Technology*, 1(1), pp.511–552.
11. Kumar, K.A., Sharma, A., Patel, N. and Singh, V. (2022). Leveraging Reinforcement Learning and Predictive Analytics for Continuous Improvement in Smart Manufacturing. *International Journal of AI and ML*, 3(9).
12. Severson, K., Chaiwatanodom, P. and Braatz, R.D. (2016). Perspectives on process monitoring of industrial systems. *Annual Reviews in Control*, 42, pp.190–200. doi:https://doi.org/10.1016/j.arcontrol.2016.09.001.
13. Gunasegaram, D., Barnard, A., Matthews, M., Jared, B., Andreaco, A., Bartsch, K. and Murphy, A. (2024). Machine learningassistedinsitu adaptive strategies for the control of defects and anomalies in metal additive manufacturing. *Additive Manufacturing*, p.104013.
14. Liu, Q., Hagenmeyer, V. and Keller, H.B. (2021). A review of rule learningbased intrusion detection systems and their prospects in smart grids. *IEEE Access*, 9, pp.57542–57564.
15. Yuan, X., Wang, Y., Yang, C., Ge, Z., Song, Z. and Gui, W. (2017). Weighted Linear Dynamic System for Feature Representation and Soft Sensor Application in Nonlinear Dynamic Industrial Processes. *IEEE Transactions on Industrial Electronics*, 65(2), pp.1508–1517. doi:https://doi.org/10.1109/tie.2017.2733443.
16. Tian, C. and Horne, R.N. (2015). Applying Machine Learning Techniques to Interpret Flow Rate, Pressure and Temperature Data From Permanent Downhole Gauges. doi:https://doi.org/10.2118/174034-ms.
17. Gbagba, S., Maccioni, L. and Concli, F. (2023). Advances in machine learning techniques used in fatigue life prediction of welded structures. *Applied Sciences*, 14(1), p.398.
18. Jung, C. (2019). Datadriven optimization of hot rolling processes.
19. Ding, P., Jia, M. and Yan, X. (2021). Stationary subspaces-vector autoregressive with exogenous terms methodology for degradation trend estimation of rolling and slewing bearings. *Mechanical Systems and Signal Processing*, [online] 150, p.107293. doi:https://doi.org/10.1016/j.ymssp.2020.107293.
20. Lin, J. and Naim, M.M. (2019). Why do nonlinearities matter? The repercussions of linear assumptions on the dynamic behaviour of assemble-to-order systems. *International Journal of Production Research*, 57(20), pp.6424–6451. doi:https://doi.org/10.1080/00207543.2019.1566669.
21. Alexander, R., Campani, G., Dinh, S. and Lima, F.V. (2020). Challenges and Opportunities on Nonlinear State Estimation of Chemical and Biochemical Processes. *Processes*, 8(11), p.1462. doi:https://doi.org/10.3390/pr8111462.
22. Bzai, J., Alam, F., Dhafer, A., Bojović, M., Altowaijri, S.M., Niazi, I.K. and Mehmood, R. (2022). Machine Learning-Enabled Internet of Things (IoT): Data, Applications, and Industry Perspective. *Electronics*, [online] 11(17), p.2676. doi:https://doi.org/10.3390/electronics11172676.

23. Xiao, B., Li, Y., Sun, B., Yang, C., Huang, K. and Zhu, H. (2021). Decentralized PCA modeling based on relevance and redundancy variable selection and its application to largescale dynamic process monitoring. *Process Safety and Environmental Protection*, 151, pp.85–100.
24. Hernandez, G. (2024). Nonparametric advances and improvements on Phase I process monitoring schemes for statistical process control.
25. Chang, S.Y., ShibanTiku and Luu, L. (2022). Advanced Process Monitoring through Fault Detection and Classification for Robust Statistical Process Control of Tantalum Nitride Reactive Sputtering. doi:https://doi.org/10.1109/issm55802.2022.10027148.
26. Park, Y.-J., Fan, S.-K.S. and Hsu, C.-Y. (2020). A Review on Fault Detection and Process Diagnostics in Industrial Processes. *Processes*, 8(9), p.1123. doi:https://doi.org/10.3390/pr8091123.
27. Woodall, W. (2023). *Recent Critiques of Statistical Process Monitoring Approaches*.
28. Ayvaz, S. and Alpay, K. (2021). Predictive Maintenance System for Production Lines in Manufacturing: A Machine Learning Approach Using IoT Data in Real-Time. *Expert Systems with Applications*, 173, p.114598. doi:https://doi.org/10.1016/j.eswa.2021.114598.
29. Colosimo, B.M., L. Allison Jones-Farmer, Megahed, F.M., Kamran Paynabar, Ranjan, C. and Woodall, W.H. (2024). Statistical Process Monitoring from Industry 2.0 to Industry 4.0: Insights into Research and Practice. *Technometrics*, pp.1–35. doi:https://doi.org/10.1080/00401706.2024.2327341.
30. Dimitris Mourtzis and Angelopoulos, J. (2024). Artificial intelligence for human–cyber-physical production systems. *Elsevier eBooks*, pp.343–378. doi:https://doi.org/10.1016/b978-0-443-13924-6.00012-0.
31. Mailer, D. (2023). Dynamic deployment of fault detection modelsa use case of the asset administration shell.
32. Gilles Trystram (2022). Automatic control of industrial food processes. *Elsevier eBooks*, pp.351–390. doi:https://doi.org/10.1016/b978-0-323-91158-0.00008-9.
33. Hossain, M.N., Rahman, M.M. and Ramasamy, D. (2024). Artificial IntelligenceDriven Vehicle Fault Diagnosis to Revolutionize Automotive Maintenance: A Review. *CMESComputer Modeling in Engineering & Sciences*, 141(2).
34. Musiał, W. and Witek, J. (2023). PROPOSAL FOR AN EXPERT SYSTEM TO AID DECISIONMAKING IN THE DESIGN AND MANAGEMENT OF FLEXIBLE MANUFACTURING SYSTEMS. *Scientific Papers of Silesian University of Technology. Organization & Management/ZeszytyNaukowePolitechnikiSlaskiej. SeriaOrganizacjiiZarzadzanie*, (186).
35. Grumbach, L. (2023). Flexible workflowsaconstraintandcasebased approach.
36. Liu, C., Tian, W. and Kan, C. (2022). When AI meets additive manufacturing: Challenges and emerging opportunities for humancentered products development. *Journal of Manufacturing Systems*, 64, pp.648–656.
37. Jimeno-Morenilla, A., Azariadis, P., Molina-Carmona, R., Kyratzi, S. and Moulianitis, V. (2021). Technology enablers for the implementation of Industry 4.0 to traditional manufacturing sectors: A review. *Computers in Industry*, 125, p.103390. doi:https://doi.org/10.1016/j.compind.2020.103390.
38. Thudumu, S., Branch, P., Jin, J. and Singh, J. (Jack) (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, 7(1). doi:https://doi.org/10.1186/s40537-020-00320-x.
39. Waqar, A. (2024). Intelligent decision support systems in construction engineering: An artificial intelligence and machine learning approaches. *Expert Systems with Applications*, [online] 249, p.123503. doi:https://doi.org/10.1016/j.eswa.2024.123503.
40. Zhao, L.-T., Yang, T., Yan, R. and Zhao, H.-B. (2022). Anomaly detection of the blast furnace smelting process using an improved multivariate statistical process control model. *Process Safety and Environmental Protection*, 166, pp.617–627. doi:https://doi.org/10.1016/j.psep.2022.08.035.
41. Massaoudi, M., Haitham Abu-Rub and Ghrayeb, A. (2023). Navigating the Landscape of Deep Reinforcement Learning for Power System Stability Control: A Review. *IEEE access*, 11, pp.134298–134317. doi:https://doi.org/10.1109/access.2023.3337118.
42. Aravind Kumar Kalusivalingam, Sharma, A., Patel, N. and Singh, V. (2022). Leveraging Reinforcement Learning and Predictive Analytics for Continuous Improvement in Smart

Manufacturing. *International Journal of AI and ML*, [online] 3(9). Available at: https://cognitivecomputingjournal.com/index.php/IJAIML-V1/article/view/71 [Accessed 7 Dec. 2024].

43. Panzer, M. and Bender, B. (2022). Deep reinforcement learning in production systems: a systematic literature review. *International Journal of Production Research*, 60(13), pp.4316–4341.
44. Prudencio, R.F., Maximo, M.R. and Colombini, E.L. (2023). A survey on offline reinforcement learning: Taxonomy, review, and open problems. *IEEE Transactions on Neural Networks and Learning Systems*.
45. Wang, Z., Zhang, K., Chen, G., Zhang, J., Wang, W., Wang, H., Zhang, L., Yan, X. and Yao, J. (2023). Evolutionaryassisted reinforcement learning for reservoir realtime production optimization under uncertainty. *Petroleum Science*, 20(1), pp.261–276.
46. Liu, R., Rajesh Piplani and Toro, C. (2023). A deep multi-agent reinforcement learning approach to solve dynamic job shop scheduling problem. *Computers & Operations Research*, 159, pp.106294–106294. doi:https://doi.org/10.1016/j.cor.2023.106294.
47. Li, C. and Chang, Q. (2022). Hybrid feedback and reinforcement learning-based control of machine cycle time for a multi-stage production system. *Journal of Manufacturing Systems*, 65, pp.351–361. doi:https://doi.org/10.1016/j.jmsy.2022.09.020.
48. Wang, X., Wang, S., Liang, X., Zhao, D., Huang, J., Xu, X., Dai, B. and Miao, Q. (2022). Deep Reinforcement Learning: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, pp.1–15. doi:https://doi.org/10.1109/tnnls.2022.3207346.
49. Nuria Nievas, Pagès-Bernaus, A., FrancescBonada, Echeverria, L. and Domingo, X. (2024). Reinforcement Learning for Autonomous Process Control in Industry 4.0: Advantages and Challenges. *Applied Artificial Intelligence*, 38(1). doi:https://doi.org/10.1080/08839514.2024.2383101.
50. Ladosz, P., Weng, L., Kim, M. and Oh, H. (2022). Exploration in deep reinforcement learning: A survey. *Information Fusion*. doi:https://doi.org/10.1016/j.inffus.2022.03.003.
51. Powell, B.K.M., Machalek, D. and Quah, T. (2020). Real-time optimization using reinforcement learning. *Computers & Chemical Engineering*, 143, p.107077. doi:https://doi.org/10.1016/j.compchemeng.2020.107077.
52. Liu, R. (2022). Deep reinforcement learning-based dynamic scheduling. *DR-NTU (Nanyang Technological University)*. doi:https://doi.org/10.32657/10356/158353.
53. Ikonen, T.J., Heljanko, K. and Harjunkoski, I. (2020). Reinforcement learning of adaptive online rescheduling timing and computing time allocation. *Computers & Chemical Engineering*, 141, p.106994. doi:https://doi.org/10.1016/j.compchemeng.2020.106994.
54. Chen, W., Qiu, X., Cai, T., Dai, H.-N., Zheng, Z. and Zhang, Y. (2021). Deep Reinforcement Learning for Internet of Things: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, pp.1–1. doi:https://doi.org/10.1109/comst.2021.3073036.
55. Gautam, M. (2023). Deep Reinforcement Learning for Resilient Power and Energy Systems: Progress, Prospects, and Future Avenues. *Electricity*, 4(4), pp.336–380. doi:https://doi.org/10.3390/electricity4040020.
56. RaihaTallat, Ammar Hawbani, Wang, X., Al□Dubai, A., Zhao, L., Liu, Z., Min, G., Zomaya, A.Y. and Saeed HamoodAlsamhi (2023). Navigating Industry 5.0: A Survey of Key Enabling Technologies, Trends, Challenges, and Opportunities. *IEEE Communications Surveys and Tutorials*, pp.1–1. doi:https://doi.org/10.1109/comst.2023.3329472.
57. Ong, K.S.H., Wang, W., Niyato, D. and Friedrichs, T. (2022). Deep-Reinforcement-Learning-Based Predictive Maintenance Model for Effective Resource Management in Industrial IoT. *IEEE Internet of Things Journal*, 9(7), pp.5173–5188. doi:https://doi.org/10.1109/jiot.2021.3109955.
58. Zhou, T., Tang, D., Zhu, H. and Zhang, Z. (2021). Multi-agent reinforcement learning for online scheduling in smart factories. *Robotics and Computer-integrated Manufacturing*, 72, pp.102202–102202. doi:https://doi.org/10.1016/j.rcim.2021.102202.
59. Lee, Y.H. and Lee, S. (2022). Deep reinforcement learning based scheduling within production plan in semiconductor fabrication. *Expert Systems with Applications*, 191, p.116222. doi:https://doi.org/10.1016/j.eswa.2021.116222.

60. Lee, M.-F.R. (2023). A Review on Intelligent Control Theory and Applications in Process Optimization and Smart Manufacturing. *Processes*, [online] 11(11), p.3171. doi:https://doi.org/10.3390/pr11113171.
61. Gu, S., Yang, L., Du, Y., Chen, G., Walter, F., Wang, J. and Knoll, A. (2024). A Review of Safe Reinforcement Learning: Methods, Theories, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, [online] 46(12), pp.11216–11235. doi:https://doi.org/10.1109/tpami.2024.3457538.
62. Wu, G., Zhang, D., Miao, Z., Bao, W. and Cao, J. (2024). How to Design Reinforcement Learning Methods for the Edge: An Integrated Approach toward Intelligent Decision Making. *Electronics*, 13(7), pp.1281–1281. doi:https://doi.org/10.3390/electronics13071281.
63. Abdulsamad, H. and Peters, J. (2023). Model-Based Reinforcement Learning via Stochastic Hybrid Models. *IEEE Open Journal of Control Systems*, 2, pp.155–170. doi:https://doi.org/10.1109/ojcsys.2023.3277308.
64. Siva, V. (2021). Artificial Intelligence for Real-Time Logistics and Transportation Optimization in Retail Supply Chains: Techniques, Models, and Applications. *Journal of Machine Learning for Healthcare Decision Support*, 1(1), pp.88–126.
65. Lan, X., Qiao, Y. and Lee, B. (2021). Towards Pick and Place Multi Robot Coordination Using Multi-agent Deep Reinforcement Learning. doi:https://doi.org/10.1109/icara51699.2021.9376433.

66. Li, S. E., 2023. *Reinforcement learning for sequential decision and optimal control*. Springer
67. Hao, J., Yang, T., Tang, H., Bai, C., Liu, J., Meng, Z., Liu, P. and Wang, Z., 2023. Exploration in deep reinforcement learning: From single-agent to multiagent domain. *IEEE Transactions on Neural Networks and Learning Systems*.
68. Bello, H. O., Ige, A. B. and Ameyaw, M. N., 2024. Adaptive machine learning models: concepts for real-time financial fraud prevention in dynamic environments. *World Journal of Advanced Engineering Technology and Sciences*,12 (02), 021-034.
69. Nian, R., Liu, J. and Huang, B., 2020. A review on reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering*,139, 106886.
70. Ghasemi, M., Moosavi, A. H., Sorkhoh, I., Agrawal, A., Alzhouri, F. and Ebrahimi, D., 2024. An introduction to reinforcement learning: Fundamental concepts and practical applications. *arXiv preprint arXiv:2408.07712*.
71. Zobernig, V., Saldanha, R. A., He, J., van der Sar, E., van Doorn, J., Hua, J.-C., Mason, L. R., Czechowski, A., Indjic, D. and Kosmala, T., 2022. RangL: A Reinforcement Learning Competition Platform. *arXiv preprint arXiv:2208.00003*.

72. Kalusivalingam, A. K., Sharma, A., Patel, N. and Singh, V., 2022. Optimizing Autonomous Factory Operations Using Reinforcement Learning and Deep Neural Networks. *International Journal of AI and ML*,3 (9).
73. RUIZ RODRIGUEZ, M. L., 2024. Maintenance optimization in Industry 4.0: A Deep Reinforcement Learning Approach to Sustainable Policy Development.
74. Ginzburg-Ganz, E., Segev, I., Balabanov, A., Segev, E., Kaully Naveh, S., Machlev, R., Belikov, J., Katzir, L., Keren, S. and Levron, Y., 2024. Reinforcement learning model-based and model-free paradigms for optimal control problems in power systems: Comprehensive review and future directions. *Energies*,17 (21), 5307.
75. Ding, Z., Huang, Y., Yuan, H. and Dong, H., 2020. Introduction to reinforcement learning. *Deep reinforcement learning: fundamentals, research and applications*, 47-123.
76. Kim, J. and Yang, I., 2020. Hamilton-Jacobi-Bellman equations for Q-learning in continuous time, *Learning for Dynamics and Control* (pp. 739-748): PMLR.
77. Mantilla Calderón, L. C., 2021. Deep Q-learning.
78. Sakr, A. H., Aboelhassan, A., Yacout, S. and Bassetto, S., 2023. Simulation and deep reinforcement learning for adaptive dispatching in semiconductor manufacturing systems. *Journal of Intelligent Manufacturing*,34 (3), 1311-1324.
79. Ghosh, D., C Machado, M. and Le Roux, N., 2020. An operator view of policy gradient methods. *Advances in Neural Information Processing Systems*,33, 3397-3406.
80. Peng, Y., Chen, G., Zhang, M. and Xue, B., 2024. Proximal evolutionary strategy: improving deep reinforcement learning through evolutionary policy optimization. *Memetic Computing*,16 (3), 445-466.
81. Arents, J. and Greitans, M., 2022. Smart industrial robot control trends, challenges and opportunities within manufacturing. *Applied Sciences*,12 (2), 937.

82. Liu, Y.-t., Yang, J.-m., Chen, L., Guo, T. and Jiang, Y., 2020. Overview of reinforcement learning based on value and policy, *2020 Chinese Control And Decision Conference (CCDC)* (pp. 598-603): IEEE.
83. Barto, A. G., Sutton, R. S. and Anderson, C. W., 2020. Looking back on the actor–critic architecture. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*,51 (1), 40-50.
84. Kamgba, R., 2024. Development of Predictive Maintenance Technologies for Critical Industrial Systems Using AI and IoT. *J Data Analytic Eng Decision Making*,1 (2), 01-09.
85. Xiong, R., Cao, J. and Yu, Q., 2018. Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle. *Applied energy*,211, 538-548.
86. Roke, 2021. *Developing Rolls-Royce's Intelligent Borescope With AI* [online]. Roke. Available from: https://www.roke.co.uk/news-discover-the-latest-news-and-press-releases-roke/developing-rolls-royce-s-intelligent-borescope-with-ai [Accessed 02 December].
87. Salmani, M., 2023. *Application of Reinforcement Learning for Condition-based Maintenance of Multi-Unit Systems*. Concordia University.
88. Zhang, K., Wang, Z., Chen, G., Zhang, L., Yang, Y., Yao, C., Wang, J. and Yao, J., 2022. Training effective deep reinforcement learning agents for real-time life-cycle production optimization. *Journal of Petroleum Science and Engineering*,208, 109766.
89. Siemens, 2024. Generative artificial intelligence takes Siemens' predictive maintenance solution to the next level: Siemens.
90. Chen, Y., Liu, Z., Zhang, Y., Wu, Y., Chen, X. and Zhao, L., 2020. Deep reinforcement learning-based dynamic resource management for mobile edge computing in industrial internet of things. *IEEE Transactions on Industrial Informatics*,17 (7), 4925-4934.
91. About-Amazon, 2024. *Amazon's Newest Fufillment Centre Powered By AI and Robotics* [online]. Available from: https://www.aboutamazon.com/news/operations/amazon-fulfillment-center-robotics-ai [Accessed 02 December].
92. Yadav, U., Bondre, S. V. and Thakre, B., 2024. Deep Reinforcement Learning in Robotics and Autonomous Systems. *Deep Reinforcement Learning and Its Industrial Use Cases: AI for Real□World Applications*, 207-238.
93. Biswas, A., Acquarone, M., Wang, H., Miretti, F., Misul, D. A. and Emadi, A., 2024. Safe Reinforcement Learning for Energy Management of Electrified Vehicle with Novel Physics-Informed Exploration Strategy. *IEEE Transactions on Transportation Electrification*.
94. Ogunfowora, O. and Najjaran, H., 2023. Reinforcement and deep reinforcement learning-based solutions for machine maintenance planning, scheduling policies, and optimization. *Journal of Manufacturing Systems*,70, 244-263.
95. Thorpe, P., 2022. A Dynamically Reconciled Digital Twin for Operations Optimization and Decision Support, *Abu Dhabi International Petroleum Exhibition and Conference* (pp. D021S070R004): SPE.
96. Tai, X. Y., Zhang, H., Niu, Z., Christie, S. D. and Xuan, J., 2020. The future of sustainable chemistry and process: Convergence of artificial intelligence, data and hardware. *Energy and AI*,2, 100036.
97. Soci.org. (2024). Imperial and BASF spinout company uses AI to improve chemical manufacturing. [online] Available at: https://www.soci.org/news/2024/7/basf-and-imperial-college-spinout [Accessed 24 Dec. 2024].
98. Basf.com. (2019). BASF partners with European Union's BIM project 'SPHERE' for sustainable construction. [online] Available at: https://www.basf.com/global/en/media/news-releases/2019/05/p-19-215 [Accessed 24 Dec. 2024].
99. Belharouak, I., Nanda, J., Self, E. C., Hawley, W. B., Wood III, D. D., Du, Z., Li, J. and Graves, R. L., 2020. *Operation, manufacturing, and supply chain of lithium-ion batteries for electric vehicles*. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).
100. Shianifar, J., Schukat, M. and Mason, K., 2024. Optimizing Deep Reinforcement Learning for Adaptive Robotic Arm Control. *arXiv preprint arXiv:2407.02503*.
101. Ali Ahmad Malik, Masood, T. and Brem, A. (2023). Intelligent humanoids in manufacturing to address worker shortage and skill gaps: Case of Tesla Optimus. arXiv (Cornell University). doi:https://doi.org/10.48550/arxiv.2304.04949.
102. Powell, K. M., Machalek, D. and Quah, T., 2020. Real-time optimization using reinforcement learning. *Computers & Chemical Engineering*,143, 107077.

103. Lei, L., Tan, Y., Zheng, K., Liu, S., Zhang, K. and Shen, X., 2020. Deep reinforcement learning for autonomous internet of things: Model, applications and challenges. *IEEE Communications Surveys & Tutorials*,22 (3), 1722-1760.
104. Hu, J., Ye, Y., Tang, Y. and Strbac, G., 2023. Towards risk-aware real-time security constrained economic dispatch: A tailored deep reinforcement learning approach. *IEEE Transactions on Power Systems*,39 (2), 3972-3986.
105. Siraskar, R., Kumar, S., Patil, S., Bongale, A. and Kotecha, K., 2023. Reinforcement learning for predictive maintenance: A systematic technical review. *Artificial Intelligence Review*,56 (11), 12885-12947.
106. Li, C., Zheng, P., Yin, Y., Wang, B. and Wang, L., 2023. Deep reinforcement learning in smart manufacturing: A review and prospects. *CIRP Journal of Manufacturing Science and Technology*,40, 75-101.
107. Liu, Z., Xu, X., Qiao, P. and Li, D., 2024. Acceleration for Deep Reinforcement Learning using Parallel and Distributed Computing: A Survey. *ACM Computing Surveys*.
108. Fu, J., Kumar, A., Nachum, O., Tucker, G. and Levine, S., 2020. D4rl: Datasets for deep data-driven reinforcement learning. *arXiv preprint arXiv:2004.07219*.
109. Haklidir, M. and Temeltaş, H., 2021. Guided soft actor critic: A guided deep reinforcement learning approach for partially observable Markov decision processes. *IEEE Access*,9, 159672-159683
110. Kumari, N., Kulkarni, R., Ahmed, M. R. and Kumar, N., 2021. Use of kalman filter and its variants in state estimation: A review. *Artificial Intelligence for a Sustainable Industry 4.0*, 213-230
111. Li, H., Zhang, Q. and Zhao, D., 2019. Deep reinforcement learning-based automatic exploration for navigation in unknown environment. *IEEE transactions on neural networks and learning systems*,31 (6), 2064-2076.
112. Devidze, R., Kamalaruban, P. and Singla, A., 2022. Exploration-guided reward shaping for reinforcement learning under sparse rewards. *Advances in Neural Information Processing Systems*,35, 5829-5842.
113. Thadikamalla, S., Joshi, P., Raman, B., Manipriya, S. and Sanodiya, R. K., 2024. Optimizing Traffic Signal Control with Deep Reinforcement Learning: Exploring Decay Rate Tuning for Enhanced Exploration-Exploitation Trade-off, *2024 11th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 64-71): IEEE.
114. Singh, M., Srivastava, R., Fuenmayor, E., Kuts, V., Qiao, Y., Murray, N. and Devine, D., 2022. Applications of digital twin across industries: A review. *Applied Sciences*,12 (11), 5727.
115. Singh, J., 2023. Autonomous Vehicle Swarm Robotics: Real-Time Coordination Using AI for Urban Traffic and Fleet Management. *Journal of AI-Assisted Scientific Discovery*,3 (2), 1-44.
116. Löfwenberg, N., 2023. A hierarchical neural network approach to learning sensor planning and control
117. Song, Y., Ha, H., Lee, W., Lee, K.-Y. and Kim, J., 2023. Data□Driven Approach Using Supervised Learning for Predicting Endpoint Temperature of Molten Steel in the Electric Arc Furnace. *steel research international*,94 (10), 2300143.
118. Tang, C., Abbatematteo, B., Hu, J., Chandra, R., Martín-Martín, R. and Stone, P. (2024). Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes. *Annual Review of Control Robotics and Autonomous Systems*. doi:https://doi.org/10.1146/annurev-control-030323-022510.
119. Alwarafy, A., Abdallah, M., Ciftler, B.S., Al-Fuqaha, A. and Hamdi, M. (2022). The Frontiers of Deep Reinforcement Learning for Resource Management in Future Wireless HetNets: Techniques, Challenges, and Research Directions. *IEEE Open Journal of the Communications Society*, 3, pp.322–365. doi:https://doi.org/10.1109/ojcoms.2022.3153226.
120. Ginzburg-Ganz, E., Itay Segev, Balabanov, A., EliorSegev, Sivan KaullyNaveh, Ram Machlev, Juri Belikov, Liran Katzir, Keren, S. and Levron, Y. (2024). Reinforcement Learning Model-Based and Model-Free Paradigms for Optimal Control Problems in Power Systems: Comprehensive Review and Future Directions. *Energies*, 17(21), pp.5307–5307. doi:https://doi.org/10.3390/en17215307.
121. Rakholia, R., Suárez-Cetrulo, A.L., Singh, M. and Carbajo, R.S. (2024). Advancing Manufacturing Through Artificial Intelligence: Current Landscape, Perspectives, Best Practices, Challenges and Future Direction. *IEEE Access*, pp.1–1. doi:https://doi.org/10.1109/access.2024.3458830.