# A Framework for Applying Reinforcement Learning to Deadlock Handling in Intralogistics

Marcel Müller, research fellow, Institute of Logistics and Material Handling Systems, Otto von Guericke University Magdeburg

*Lorena Silvana Reyes-Rubiano, research fellow, Chair of Data and Business Analytics, RWTH Aachen University, Germany* 

*Tobias Reggelin, research manager, Institute of Logistics and Material Handling Systems, Otto von Guericke University Magdeburg* 

*Hartmut Zadek, professor, Institute of Logistics and Material Handling Systems, Otto von Guericke University Magdeburg* 

**Summary.** Intralogistics systems, while complex, are crucial for a range of industries. One of their challenges is deadlock situations that can disrupt operations and decrease efficiency. This paper presents a four-stage framework for applying reinforcement learning algorithms to manage deadlocks in such systems. The stages include Problem Formulation, Model Selection, Algorithm Selection, and System Deployment. We carefully identify the problem, select an appropriate model to represent the system, choose a suitable reinforcement learning algorithm, and finally deploy the solution. Our approach provides a structured method to tackle deadlocks, improving system resilience and responsiveness. This comprehensive guide can serve researchers and practitioners alike, offering a new avenue for enhancing intralogistics performance. Future research can explore the framework's effectiveness and applicability across different systems.

Keywords. intralogistics planning, deadlock prevention, model and algorithm selection

# 1. Introduction

Intralogistics systems, responsible for the internal movement and organization of materials within a facility, play a pivotal role in ensuring efficiency and productivity in various industries. However, these complex systems face numerous challenges, with one of the most significant being the occurrence of deadlocks (Coffman, Elphick, and Shoshani 1971). Deadlocks can drastically impede the system's performance, manifesting as blocked resources and leading to suboptimal throughput and increased throughput time.

The severity and complexity of deadlock situations make them a critical aspect of logistics system design and operation. Yet, traditional planning processes often do not adequately consider deadlock handling in their design and operation. This oversight may stem from the inherent

complexity and dynamism of deadlocks, which often require a sophisticated understanding of system interactions and a capability for dynamic decision-making.

Recent advancements in artificial intelligence, particularly reinforcement learning (RL), present an exciting opportunity to address this challenge (Sutton and Barto 2018). RL, with its ability to learn from interactions and adapt to dynamic environments, could be a potential game-changer for deadlock handling in intralogistics systems.

# 2. Related Work

## 2.1. Strategic and Operational Intralogistics Planning

Process models for intralogistics planning are systematic processes for solving (decision-making) problems while considering subjective objectives in the logistics environment (Gudehus 2010). In this context, reinforcement learning can be a potent tool. This machine learning method enables agents to learn how to behave in an environment by performing certain actions and receiving rewards, thus learning a policy that maximizes the total reward over time (Esteso et al. 2022; van Heeswijk 2022; Yan et al. 2022). Logistics planning can be divided into strategic and operational logistics planning. Strategic logistics planning refers to long-term planning, while operational logistics planning is assigned to a short- to medium-term planning horizon.

*Strategic intralogistics planning* involves decisions that affect the long-term direction of a company's logistics operations. These could include the choice of automation technologies, warehouse layout and design, selection of material handling equipment, and overall process design. The main objective is to ensure a seamless, efficient, and flexible flow of materials within the facility. Reinforcement learning can be used in this context to optimize resource allocation, routing, scheduling, and other factors over a long-term horizon. For instance, reinforcement learning can be used to develop policies that minimize the occurrence of deadlocks in the long term, such as rules for allocating resources in a way that avoids potential deadlock scenarios (Esteso et al. 2022; van Heeswijk 2022; Yan et al. 2022). At the strategic level, deadlocks can be prevented by careful design and layout of the facility, choice of equipment, and integration of advanced technologies:

- Facility Design and Layout: The design and layout of a warehouse or production facility should aim to minimize the chances of deadlocks. This might involve designing sufficient space for movement, creating multiple paths to the same location, and arranging workstations or storage areas in a way that reduces conflicts (Pérez-Gosende, Mula, and Díaz-Madroñero 2021).
- Equipment Selection: The selection of material handling equipment can significantly influence the likelihood of deadlocks. For instance, choosing automated guided vehicles (AGVs) with advanced sensors and intelligent control systems can help in detecting and avoiding potential deadlocks.
- Integration of Advanced Technologies: Advanced technologies such as Warehouse Management Systems (WMS), Manufacturing Execution Systems (MES), or advanced planning and scheduling (APS) systems can be used to manage resources, tasks, and material flows. These systems can use sophisticated algorithms to prevent and handle deadlocks.

*Operational intralogistics planning*, on the other hand, focuses on short- to medium-term decisions that affect the daily operations within the facility. This includes scheduling of tasks, routing

of vehicles, inventory control, and management of resources. Here, reinforcement learning can be particularly useful for dynamic, real-time decision-making. For instance, a reinforcement learning agent could be used to monitor the logistics system in real time, identify potential deadlock situations as they arise, and take actions to prevent or resolve them. The agent could be trained to recognize patterns of resource usage that could lead to a deadlock and to take corrective actions, such as reallocating resources or rescheduling tasks, in response. At the operational level, strategies to handle, prevent, and avoid deadlocks can be built into the dayto-day decision-making processes:

- Scheduling: Tasks and resources should be scheduled in a way that minimizes conflicts and potential deadlocks. This might involve careful sequencing of tasks, ensuring that resources are not overcommitted, and leaving sufficient buffer time between tasks.
- Routing: Intralogistics operations often involve the movement of goods within a facility, and routing decisions can significantly influence the likelihood of deadlocks. Effective routing strategies can ensure smooth material flow, avoiding congestion and potential deadlocks.
- Real-Time Decision Making: In dynamic intralogistics environments, decisions often need to be made in real-time based on the current state of the system. Advanced decision-support systems can help to make these decisions quickly and accurately, helping to avoid potential deadlocks.

# 2.2. Planning Principles

The principles outlined below are foundational to effective intralogistics planning and management:

*Prioritizing Effectiveness over Efficiency (Fottner et al. 2022):* This principle underscores the necessity to confirm the system's capability to meet its predefined objectives (effectiveness) before embarking on optimizing resources (efficiency). When considering deadlocks, the logistics system must initially demonstrate its capability to operate without any deadlock scenarios. The selection of appropriate technology is pivotal in this regard. For instance, systems with inherent deadlock handling or avoidance mechanisms should be given preference (Coffman, Elphick, and Shoshani 1971). Once this level of effectiveness is ensured, we can shift our focus towards enhancing efficiency by streamlining resource allocation and curtailing the time needed for deadlock resolution.

Adherence to a Consistent Process and Flow Orientation (Fottner et al. 2022): This principle accentuates the need to maintain an uninterrupted, consistent operational flow within the logistics system. Deadlocks can significantly interrupt this flow, causing delays and inefficiencies. Hence, the formulation and implementation of deadlock prevention and avoidance strategies are paramount (Havender 1968). These strategies could encompass the design of processes and systems with a specific focus on deadlock avoidance, such as employing specific algorithms like Dijkstra's Banker's Algorithm. Moreover, maintaining a flow-oriented outlook could necessitate establishing a system for the timely detection and resolution of potential deadlocks.

*Robustness Against Short-term Deviations, Errors, and Disruptions (Fottner et al. 2022):* A robust logistics system is characterized by its resilience against disruptions, continuing to function effectively despite them. In the context of deadlocks, it implies the existence of strategies to manage deadlocks as and when they occur. This could entail developing contingency plans, deploying redundant systems, or incorporating swift recovery mechanisms to tackle deadlocks. It also necessitates the design of the system with a focus on deadlock

avoidance, such as evading circular wait conditions or ensuring resources are not strained beyond their capacity (Fottner et al. 2022).

*Principle of Decentralization:* This principle encourages the distribution of decision-making authority to various points in the system rather than concentrating it at a single point. This can enhance resilience, flexibility, and responsiveness, particularly in complex and dynamic environments (Le-Anh and M. de Koster 2006; Lombard et al. 2016). It is important to note that while decentralization can offer several benefits, it is not without challenges. For instance, it can lead to sub-optimal decisions if the subsystems do not have access to all necessary information or if they make decisions that benefit them individually but are detrimental to the system. Therefore, achieving a balance between centralization and decentralization is crucial and highly dependent on the specific context. Here we define resilience, flexibility, and responsiveness as:

- Resilience: A decentralized system is typically more robust in the face of disturbances or failures, as the system does not rely on a single point of control. If one component of the system fails, others can continue to operate independently.
- Flexibility: As each subsystem or component in a decentralized system can adjust its behavior independently based on local information, such systems can adapt more easily to changes in demand or other conditions.
- Responsiveness: Decentralized systems can react more quickly to changes, as they do not need to wait for instructions from a central authority. This is particularly valuable in environments where conditions change rapidly and unpredictably.

*Principle of Standardization:* Standardizing procedures and components can improve efficiency, predictability, and interoperability in an intralogistics system. It aims to reduce variability, improve predictability, and enhance efficiency. This principle can be applied to various aspects of intralogistics such as standardized containers, standardized methods of operation, standardized routes, etc. (R. de Koster, Le-Duc, and Roodbergen 2007; Klug 2018; Pohl, Meller, and Gue 2011). We define efficiency, interoperability, and predictability as follows:

- Efficiency: Standardization reduces the need for custom solutions, thereby improving the efficiency of operations by reducing complexity, errors, and rework.
- Interoperability: Standardization enables different systems or components within the logistics chain to work together seamlessly because they adhere to the same set of standards.
- Predictability: Standardized processes are more predictable, which can improve planning and scheduling.

*Principle of Integration:* This principle refers to the concept of creating a cohesive, interconnected system where all parts work together seamlessly for a common goal. This involves coordinating and synchronizing different activities and processes to ensure smooth operations and minimize inefficiencies (Goetschalckx 2002; Ivanov, Dolgui, and Sokolov 2019). We define the Streamlining Operations, Information Sharing. It is important to note that while integration offers many benefits, it also comes with challenges. These can include the complexities of coordinating different processes, potential difficulties in implementing changes across an integrated system, and the need for effective communication and information sharing mechanisms. The benefits, however, often outweigh the challenges, making integration a key principle in intralogistics planning. We define the Streamlining Operations, Information Sharing Operations, Information Sharing operations, Information Sharing and information sharing, and Improved Visibility as follows:

• Streamlining Operations: Integration can help streamline operations by eliminating unnecessary redundancies and ensuring different parts of the system are in alignment. This can lead to improved productivity and efficiency.

- Information Sharing: An integrated system facilitates better communication and information sharing across different parts of the logistics chain, leading to more informed decision making.
- Improved Visibility: Integration can provide a holistic view of the entire logistics process, making it easier to identify bottlenecks, inefficiencies, and opportunities for improvement.

Principle of Automation: Intralogistics planning can benefit from the use of automated systems and technologies, which can enhance efficiency, accuracy, and consistency (Vis 2006).

*Principle of Sustainability:* This principle emphasizes the importance of considering the environmental, social, and economic impacts of intralogistics operations, and seeking to minimize negative impacts (Dekker, Bloemhof, and Mallidis 2012). Here we provide a description of sustainable dimensions:

- Environmental Sustainability: Intralogistics systems can be designed to reduce energy consumption, minimize waste, and lessen the carbon footprint. This can be achieved through optimizing routes, using energy-efficient equipment, and implementing recycling initiatives.
- Social Sustainability: This involves ensuring fair labor practices, maintaining safe working conditions, and contributing positively to the local community.
- Economic Sustainability: This principle stresses the need for intralogistics operations to be economically viable. This can involve improving efficiency, reducing costs, and ensuring the longevity of the business.

*Principle of Resilience:* The Principle of Resilience in intralogistics planning refers to the ability of the system to adapt and recover quickly from disruptions and changes, maintaining high levels of performance under a range of conditions. Resilience involves aspects like flexibility, robustness, redundancy, and the capacity for rapid recovery. While sustainability and resilience are both crucial principles in intralogistics planning, tensions can arise between them. For instance, sustainability often involves streamlining operations, reducing waste, and minimizing redundancy, which can potentially decrease a system's resilience. On the other hand, building resilience often involves maintaining a certain level of redundancy and flexibility, which could lead to higher costs and resource usage, potentially impacting the sustainability goals. It is important to balance sustainability and resilience in intralogistics planning. While the two principles can sometimes conflict, they can also be mutually supportive in many cases. For instance, a more sustainable logistics system may be more resilient to disruptions related to environmental regulations or resource scarcity. Similarly, a resilient logistics system can better withstand disruptions and thus ensure long-term sustainability. Thus, the key is to find the right balance that maximizes both sustainability and resilience (Tukamuhabwa et al. 2015).

Logistics planning principles are intrinsically linked to strategies aimed at managing, preventing, and avoiding deadlocks. By prioritizing effectiveness, adhering to a consistent process and flow orientation, and ensuring robustness against disruptions, logistics planners can architect systems that demonstrate resilience to deadlocks and possess the capability to recover rapidly in their occurrence. In the pursuit of boosting intralogistics planning through reinforcement learning, several research gaps have been identified. Foremost among these is the application of reinforcement learning to real-world logistics systems, which present a high degree of complexity, uncertainty, and non-linearity. The current success of reinforcement learning in simplified or simulated scenarios does not readily translate to these complex systems, making the learning and decision-making process challenging. Scalability also emerges as a significant issue, particularly for deep-learning-based reinforcement learning methods. As the logistics system expands, the state-action space to be explored by the reinforcement learning agent can become prohibitively large, rendering the learning process inefficient. Moreover, these reinforcement learning methods often lack interpretability, functioning as "black boxes" and providing limited insight into the decision-making process (Yan et al. 2022). This opaqueness is particularly problematic in logistics settings where understanding the rationale behind decisions is crucial. Reinforcement learning methods also typically operate in isolation from existing logistics planning methods, indicating a research deficit in the integration of reinforcement learning with traditional logistics planning methods. Lastly, the focus of reinforcement learning on immediate reward often overlooks the long-term consequences of actions, especially in the context of deadlock prevention. Addressing these research gaps, such as enhancing the capability of reinforcement learning methods to manage real-world logistics complexity, improving scalability and interpretability, and ensuring consideration of long-term consequences, provides promising directions for future research.

## 2.3. Application of Reinforcement Learning in Logistics

Table 1 shows related examples of applying reinforcement learning for logistics systems and gives a brief overview of typical applications.

Authors	Objective	Methodology	Real-world application	Remarks
(Li et al., 2018)	Improve task selection by autonomous material handling vehicles	Deep reinforcement learning	Material handling and warehousing	This paper presents a deep reinforcement learning methodology for task selection by autonomous material handling vehicles that performs well under the given conditions.
(Estanjini et al., 2011)	Improve forklift dispatching and reduce costs	Stochastic learning method to generate dispatching recommendations	Warehousing	This paper presents a successful deployment of an inexpensive mobile wireless sensor network in a commercial warehouse served by a fleet of forklifts to improve forklift dispatching and reduce costs associated with loading/unloading delivery trucks

(Kamoshida & Kazama, 2017)	AGV route planning method for an AGV picking system	Deep reinforcement learning	The proposed method was evaluated using an AGV picking system simulator.	Deep reinforcement learning was used to develop an AGV route planning policy
(Andersen et al., 2019)	Increases sample efficiency and enables algorithms with low sample efficiency to function better in real- world environments	A model-based reinforcement learning algorithm	Warehouse simulated environment	This paper presents a model-based reinforcement learning algorithm called The Dreaming Variational Autoencoder v2 (DVAE-2) that increases sample efficiency and enables algorithms with low sample efficiency to function better in real-world environments.
(Cals et al., 2021)	Minimize the number of tardy orders	Deep Reinforcement Learning with a heuristic algorithm	Warehousing	This paper presents a Deep Reinforcement Learning (DRL) approach for deciding how and when orders should be batched and picked in a warehouse to minimize the number of tardy orders that performs well under the given conditions.
(Hu et al., 2020)	Optimal transport time and equipment for each task	Adaptive deep reinforcement learning	Flexible shop floor	A new Deep Q-Network (DQN) method is devised to establish the optimal combined rule policy. This policy guides the selection of appropriate dispatching rules and Automated Guided Vehicles (AGVs) for diverse scheduling scenarios.

(Rimélé et Minimiz al., 2021) average time	te the Partially e cycle observable Markov decision process and using a deep Q- learning agent	Warehousing	This paper presents a method for learning a storage policy for a warehouse with a Robotic Mobile Fulfillment System (RMFS) using Deep Q- learning that performs well under the given conditions.
--	---	-------------	---

Table 1. Examples of applying reinforcement learning for logistics systems.

# 3. Framework

#### 3.1. Overview

The framework is based on the previous phases in logistics planning. There are many different approaches in logistics planning, but most of them are essentially similar. As an example, we use the planning phases of Gudehus (2010) for general logistics planning and the approach of Fottner et al. (2022) for an intralogistics-specific approach. Table 2 compares the two planning methodologies and adds the necessary extensions in a third column for a sufficient consideration to deal with deadlocks.

Logistics system (Gudehus 2010)	Intralogistics (Fottner et al. 2022)	Necessary extension for deadlock handling with RL
Target planning	Setting targets/ constraints	Resilience as a target or significant influence on other target parameters
	Requirement analysis	Deadlock consideration
System planning	Rough concepts/ planning Variants	Preventive deadlock design
Detailed planning	System design	(Simulation) Modeling Reinforcement learning integration
Tender	Evaluation/ decision	
System construction	Realization	Deadlock handling solution deployment
		Monitoring & continuous Learning

Table 2. Comparison of selected procedure models for logistics system (Gudehus2010) and intralogistics system (Fottner et al. 2022) with necessary extensionfor deadlock handling with reinforcement learning (RL).

(Rimélé et Minimiz al., 2021) average time	te the Partially e cycle observable Markov decision process and using a deep Q- learning agent	Warehousing	This paper presents a method for learning a storage policy for a warehouse with a Robotic Mobile Fulfillment System (RMFS) using Deep Q- learning that performs well under the given conditions.
--	---	-------------	---

Table 1. Examples of applying reinforcement learning for logistics systems.

# 3. Framework

#### 3.1. Overview

The framework is based on the previous phases in logistics planning. There are many different approaches in logistics planning, but most of them are essentially similar. As an example, we use the planning phases of Gudehus (2010) for general logistics planning and the approach of Fottner et al. (2022) for an intralogistics-specific approach. Table 2 compares the two planning methodologies and adds the necessary extensions in a third column for a sufficient consideration to deal with deadlocks.

Logistics system (Gudehus 2010)	Intralogistics (Fottner et al. 2022)	Necessary extension for deadlock handling with RL
Target planning	Setting targets/ constraints	Resilience as a target or significant influence on other target parameters
	Requirement analysis	Deadlock consideration
System planning	Rough concepts/ planning Variants	Preventive deadlock design
Detailed planning	System design	(Simulation) Modeling Reinforcement learning integration
Tender	Evaluation/ decision	
System construction	Realization	Deadlock handling solution deployment
		Monitoring & continuous Learning

Table 2. Comparison of selected procedure models for logistics system (Gudehus2010) and intralogistics system (Fottner et al. 2022) with necessary extensionfor deadlock handling with reinforcement learning (RL).

The avoidance of deadlock and the resulting robustness of the system becomes a central target in our framework. It is recognized not only as a standalone key figure but also represented by possible significant influence on other target parameters, as for example throughput or lead time. This recognition means that even if deadlocks are not directly considered as a target parameter by the logistics planner, their significant impact on these key figures necessitates their consideration in the planning process.

During the requirement analysis phase, we integrate considerations for potential deadlocks. This stage could include examining past data to identify conditions that have previously resulted in deadlocks, as well as predicting potential deadlock scenarios based on the proposed system design and operation. Additionally, this phase aims to estimate the degree of disruptions that could occur in the system, helping determine the level of effort needed for effective deadlock avoidance.

In the system planning stage, our focus shifts towards a preventive deadlock design. This approach underscores the importance of designing systems with the primary goal of reducing the risk of deadlocks as much as possible. This could be achieved by employing strategies such as optimizing the layout or routing, incorporating buffers for added resource allocation flexibility, and adhering to consistent process and flow orientation. This way, the system planning follows standard principles while taking extra measures for potential deadlock scenarios.

During the detailed planning stage, we actively incorporate simulation modeling to facilitate the application of reinforcement learning. We create a simulation model that accurately represents the dynamics and interactions between agents and resources within the system. This model serves as a controlled environment, ideal for training the reinforcement learning algorithm. Alternatively, when a system's dynamics can be accurately captured mathematically, a mathematical model can be used instead of a simulation. This alternative is particularly applicable for simpler systems or when comprehensive, high-quality data is available.

The integration of reinforcement learning algorithms occurs during the system operation phase. The selected algorithm, once trained on the simulation model, is introduced into the actual system, continually adjusting its strategy based on the system's performance and effectively managing deadlock situations.

Following the deployment of the deadlock handling solution, the system moves into a phase of continuous monitoring and learning. Performance is constantly tracked to identify changes or trends that might impact deadlock situations. The reinforcement learning algorithm, in turn, utilizes this ongoing data to continually learn and adapt its strategy, improving its deadlock handling capabilities over time.

We propose a framework, which is based on four main steps for applying reinforcement learning for deadlock handling in logistics. Figure 1 shows the framework as it accompanies each phase of logistics systems planning according to Gudehus. The framework starts from the system planning stage by preparing the use of reinforcement learning and deriving conclusion from the earlier target planning.



Figure 1: Framework for applying reinforcement learning for deadlock handling in logistics.

## **3.2. Problem Formulation**

The first stage of our framework is the "Reinforcement Learning (RL) Problem Formulation". In this stage, we focus on understanding the specific problem at hand within the context of the intralogistics system. The problems that we may encounter in an intralogistics setting can be diverse, including but not limited to, issues related to scheduling, path planning, and inventory management.

The problem formulation process begins with a thorough understanding of the logistics system, focusing on the nature and scope of the problem. We investigate the types of resources, agents, and processes involved in the system, as well as the objectives and constraints that the system operates under. Understanding these elements allows us to define the problem more accurately and tailor our approach for the best possible solution.

We also delve into historical data, looking for patterns and trends that may contribute to deadlock situations. By analyzing this data, we can gain insights into the conditions under which deadlocks typically occur and strategize our approach accordingly.

In this stage, we also determine the nature of the problem in terms of being a single-agent or multi-agent problem, and whether the system is fully observable or partially observable. These considerations are critical as they influence the implementation of the learning environment and RL algorithm in the next stages.

Next, we define the states, actions, and rewards for the RL problem. The "state" represents the current condition of the system, the "action" is the decision made by the agent, and the "reward" is the feedback that the agent receives after taking an action. These definitions form the foundation of the RL problem and guide the learning process of the RL algorithm.

## 3.3. Model Selection

The second stage, "Model Selection", necessitates the determination of an appropriate model to accurately represent the logistics system and the problem identified. Owing to the dynamic and stochastic characteristics of logistics systems prone to deadlocks, simulation models generally serve as the preferred choice, though mathematical models may be employed under certain circumstances.

The initial phase of this stage involves the selection of a suitable machine learning framework for reinforcement learning. Options for this might range from well-established frameworks such as Ray RLlib or Stable Baselines to a custom implementation developed to meet specific requirements. The choice of the framework is a significant decision, given its impact on the scalability and development of the solution.

Subsequently, we proceed to establish the learning environment for the reinforcement learning algorithm. This environment functions as the interface bridging the reinforcement learning algorithm and the simulation model. The creation process might involve integration with simulation software like Plant Simulation or AnyLogic or the development of a custom environment directly in a programming language such as Python. The learning environment encompasses the dynamics of the logistics system and integrates the definitions of states, actions, and rewards that were established during the problem formulation stage.

## 3.4. Algorithm Selection

The third stage of the framework is "Algorithm Selection", which involves determining and setting up the reinforcement learning algorithm that is best suited to address the problem defined in the learning environment. The choice of the algorithm is crucial, as different algorithms have varied strengths and weaknesses, making them more suited to some problems than others. For instance, algorithms such as Deep Q-Network (DQN) or Proximal Policy Optimization (PPO) might be considered, among others, depending on the specific characteristics of the problem.

The first sub-step in this stage is hyperparameter selection. Hyperparameters are parameters whose values are set prior to the commencement of the learning process and significantly impact the training of the reinforcement learning algorithm. These might include learning rate, discount factor, or the number of episodes, among others. The choice of hyperparameters typically involves a balance between exploration and exploitation and directly affects the speed and effectiveness of learning.

Following hyperparameter selection, the training process is initiated. This involves the reinforcement learning algorithm interacting with the learning environment over a series of episodes, each time learning from the reward feedback and improving its decision-making policy.

After a set of training episodes, the performance of the algorithm is evaluated and validated. This step involves analyzing the mean episode reward, a key performance metric that we aim to maximize. The evaluation and validation process helps ensure that the RL algorithm is learning effectively and that it can generalize well to new situations.

The hyperparameter selection, training, and evaluation & validation steps form a cycle that is repeated until a satisfactory combination of hyperparameters is found that allows the algorithm to solve the problem effectively. This iterative process allows for continuous refinement and improvement of the RL solution.

### **3.5. System Deployment**

The final stage of our framework is "System Deployment". This stage involves putting the trained RL algorithm into operation within the actual logistics system.

The first sub-step in this stage is Implementation. Here, the RL algorithm, which has been trained and validated in the simulated environment, is implemented into the real-world logistics system. The algorithm begins to make decisions in real-time, applying the learned policy to handle deadlock situations.

Following the implementation, the next sub-step is Monitoring. Given the dynamic nature of logistics systems, it is crucial to continuously monitor the performance of the RL algorithm once it has been deployed. Key performance indicators, such as throughput or throughput time, are tracked to ensure that the system is operating as expected and that the RL algorithm is effectively managing deadlocks.

The final sub-step is Continuous Learning. In an ever-evolving logistics system, new deadlock situations may arise that the RL algorithm has not encountered during training. Continuous learning allows the RL algorithm to learn from these new experiences and adapt its policy accordingly. This process involves periodically retraining the RL algorithm with updated data from the logistics system, allowing it to refine and improve its policy over time. Continuous learning ensures that the RL algorithm remains effective in the face of changing conditions within the logistics system.

# 4. Discussion and Integration

Implementing the proposed framework in real-world intralogistics settings brings its own set of challenges and considerations. The framework offers a structured methodology, yet its practical application must be attuned to the specific context and conditions of each logistics system. For example, the chosen RL algorithm and model need to correspond to the characteristics of the logistics system and the complexity of the deadlock problem. Furthermore, practical implementation may entail overcoming potential obstacles such as the availability of training data for the RL algorithm, computational resources, or the integration of the RL solution with the existing system architecture.

The efficacy of the framework is not easily evaluated through standard performance metrics alone. While throughput, throughput time, and the impact of deadlocks (waiting time or detours) provide insight into the system's performance, they do not necessarily reflect the success of the framework in enabling effective deadlock handling through RL. The evaluation should, therefore, be multi-faceted, including not only these performance metrics but also the improvement in system robustness, the reduction in manual interventions for deadlock resolution, and the adaptability of the RL solution to changes in the system dynamics. It is also worth noting that these evaluation criteria should be aligned with the system's objectives and constraints as defined in the problem formulation stage.

While the framework is designed to be applicable to a wide range of intralogistics systems and deadlock scenarios, its actual adaptability and flexibility are dependent on the specific system requirements and constraints. The framework provides the structure and steps for applying RL for deadlock handling, but the specifics of each stage, such as the problem type in the RL problem formulation stage and the choice of model in the model selection stage, need to be adapted based on the system characteristics. The continuous learning capability of the framework allows

for ongoing adaptation of the RL solution to changes in the system dynamics, making the framework potentially robust across diverse intralogistics systems. However, the extent of this flexibility and adaptability needs to be explored further in practical applications.

# 5. Conclusion and Future Work

This paper has presented a comprehensive framework for applying reinforcement learning (RL) to handle deadlocks in intralogistics systems. The proposed framework extends conventional logistics planning by integrating RL considerations into each stage of the process. It provides a structured approach that guides practitioners through problem formulation, model selection, algorithm selection, and system deployment, with the aim of creating robust, resilient, and efficient logistics systems.

Our approach introduces RL as a tool for dynamic deadlock handling, offering potential improvements in system performance and resilience. However, the application of this framework in real-world scenarios needs to be explored further. While we provide a roadmap for the implementation of RL in logistics systems, the specifics of each step are dependent on the individual system's characteristics and requirements, necessitating a degree of adaptation and flexibility.

Future work could involve the application and evaluation of this framework across various practical cases, thereby providing a more comprehensive understanding of its effectiveness and adaptability. Such studies would not only provide validation for our approach but also shed light on potential challenges and obstacles in its implementation. Additionally, research could be conducted to refine the evaluation criteria for the success of the framework, ensuring a multi-faceted assessment that captures all relevant aspects of system performance.

Further enhancements to the framework could include more advanced RL techniques or the integration of other AI methods to increase the robustness and accuracy of the deadlock handling solution. As the field of RL continues to evolve, so too will the opportunities for its application in intralogistics systems, and our framework provides a starting point for this exciting journey.

# References

- Coffman, E. G., M. Elphick, and A. Shoshani. 1971. "System Deadlocks." ACM Comput. Surv. 3 (2): 67–78. https://doi.org/10.1145/356586.356588.
- Dekker, Rommert, Jacqueline Bloemhof, and Ioannis Mallidis. 2012. "Operations Research for Green Logistics – an Overview of Aspects, Issues, Contributions and Challenges." *European Journal of Operational Research* 219 (3): 671–79. https://doi.org/10.1016/j.ejor .2011.11.010.
- Esteso, Ana, David Peidro, Josefa Mula, and Manuel Díaz-Madroñero. 2022. "Reinforcement Learning Applied to Production Planning and Control." *International Journal of Production Research*, 1–18. https://doi.org/10.1080/00207543.2022.2104180.
- Fottner, Johannes, Stefan Galka, Sebastian Habenicht, Eva Marianne Klenk, Ingolf Meinhardt, and Thorsten Schmidt. 2022. *Planung von innerbetrieblichen Transportsystemen: Fahrzeugsysteme.* Intralogistik. Berlin, Heidelberg: Springer Vieweg.

- Goetschalckx, Marc. 2002. "Strategic Network Planning." In *Supply Chain Management and Advanced Planning: Concepts, Models, Software and Case Studies*, ed. by Hartmut Stadtler and Christoph Kilger. 2<sup>nd</sup> ed., 105–21. Berlin, Heidelberg: Springer.
- Gudehus, Timm. 2010. *Logistik: Grundlagen Strategien Anwendungen.* 4. Aufl. Berlin, Heidelberg: Springer.
- Havender, J. W. 1968. "Avoiding Deadlock in Multitasking Systems." *IBM Syst. J.* 7 (2): 74–84. https://doi.org/10.1147/sj.72.0074.
- Ivanov, Dmitry, Alexandre Dolgui, and Boris Sokolov. 2019. "The Impact of Digital Technology and Industry 4.0 on the Ripple Effect and Supply Chain Risk Analytics." *International Journal of Production Research* 57 (3): 829–46. https://doi.org/10.1080/00207543 .2018.1488086.
- Klug, Florian. 2018. "Lean Logistics." In *Logistikmanagement in der Automobilindustrie*, ed. by Florian Klug, 287–326. Berlin, Heidelberg: Springer.
- Koster, René de, Tho Le-Duc, and Kees Jan Roodbergen. 2007. "Design and Control of Warehouse Order Picking: A Literature Review." *European Journal of Operational Research* 182 (2): 481–501. https://doi.org/10.1016/j.ejor.2006.07.009.
- Le-Anh, Tuan, and M.B.M. de Koster. 2006. "A Review of Design and Control of Automated Guided Vehicle Systems." *European Journal of Operational Research* 171 (1): 1–23. https://doi.org/10.1016/j.ejor.2005.01.036.
- Lombard, Alexandre, Florent Perronnet, Abdeljalil Abbas-Turki, and Abdellah El Moudni. 2016. "Decentralized Management of Intersections of Automated Guided Vehicles." *IFAC-PapersOnLine* 49 (12): 497–502. https://doi.org/10.1016/j.ifacol.2016.07.669.
- Pérez-Gosende, Pablo, Josefa Mula, and Manuel Díaz-Madroñero. 2021. "Facility Layout Planning. An Extended Literature Review." International Journal of Production Research 59 (12): 3777–3816. https://doi.org/10.1080/00207543.2021.1897176.
- Pohl, Letitia M., Russell D. Meller, and Kevin R. Gue. 2011. "Turnover-Based Storage in Non-Traditional Unit-Load Warehouse Designs." *IIE Transactions* 43 (10): 703–20. https://doi.org/10.1080/0740817X.2010.549098.
- Sutton, Richard S., and Andrew Barto. 2018. *Reinforcement Learning: An Introduction.* 2<sup>nd</sup> ed., Adaptive computation and machine learning. Cambridge, Massachusetts, London, England: The MIT Press.
- Tukamuhabwa, Benjamin R., Mark Stevenson, Jerry Busby, and Marta Zorzini. 2015. "Supply Chain Resilience: Definition, Review and Theoretical Foundations for Further Study." *International Journal of Production Research* 53 (18): 5592–5623. https://doi.org /10.1080/00207543.2015.1037934.
- van Heeswijk, W. J. A. 2022. "Strategic Bidding in Freight Transport Using Deep Reinforcement Learning." Ann Oper Res. https://doi.org/10.1007/s10479-022-04572-z.
- Vis, Iris F.A. 2006. "Survey of Research in the Design and Control of Automated Guided Vehicle Systems." *European Journal of Operational Research* 170 (3): 677–709. https://doi.org /10.1016/j.ejor.2004.09.020.

 Yan, Yimo, Andy H.F. Chow, Chin Pang Ho, Yong-Hong Kuo, Qihao Wu, and Chengshuo Ying.
2022. "Reinforcement Learning for Logistics and Supply Chain Management: Methodologies, State of the Art, and Future Opportunities." *Transportation Research Part E: Logistics and Transportation Review* 162:102712. https://doi.org/10.1016/j.tre .2022.102712.

This paper could be cited as:

Müller, Marcel, Lorena Silvana Reyes-Rubiano, Tobias Reggelin, and Hartmut Zadek. 2023. "A Framework for Applying Reinforcement Learning to Deadlock Handling in Intralogistics." In 2023 International Scientific Symposium on Logistics: Conference Volume, ed. by Thorsten Schmidt, Kai Furmans, Bernd Hellingrath, René de Koster, Anne Lange, and Hartmut Zadek, 107–121. Bremen: Bundesvereinigung Logistik. https://nbn-resolving.org/urn:nbn :de:bsz:14-qucosa2-855832 https://doi.org/10.25366/2023.127