Int. j. adv. multidisc. res. stud. 2026; 6(2):301-312

Received: 20-01-2026
Accepted: 01-03-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Toward Autonomous Supply Chains: A Deep Reinforcement Learning Framework

[1] Bui Thi Kim Uyen, [2] Bui Trong Hieu

[1] Faculty of International Trade, College of Foreign Economic Relations (Cofer), Ho Chi Minh City, Vietnam

[2] Institute of Information Technology and Electrical-Electronic Engineering, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam

DOI: https://doi.org/10.62225/2583049X.2026.6.2.5959

Corresponding Author: **Bui Trong Hieu**

Abstract

This paper proposes a comprehensive deep reinforcement learning (DRL) framework aimed at enabling autonomous decision-making in supply chain systems operating under uncertainty. Traditional supply chain optimization approaches typically decompose forecasting and operational decisions into separate modules, often relying on static assumptions and deterministic models. In contrast, we model the supply chain as a sequential decision-making problem formulated as a Markov Decision Process (MDP), where inventory replenishment and transportation allocation are jointly optimized. An actor–critic architecture is employed to learn adaptive policies directly from interaction with a stochastic simulation environment characterized by random demand, variable lead times, and transportation constraints. Extensive computational experiments demonstrate that the proposed framework achieves significant reductions in total cost and stock-out frequency compared to classical base-stock and heuristic routing strategies. Moreover, the learned policies exhibit strong adaptability under demand fluctuations. The results highlight the feasibility of transitioning from rule-based planning systems toward fully autonomous, learning-driven supply chains. We formalize the theoretical foundations, propose scalable architectures, and identify open research challenges. The results demonstrate the transformative potential of DRL in reshaping supply chain management toward self-learning and self-optimizing ecosystems. The approach improves trust and interpretability.

Keywords: Actor-Critic Architecture, Adaptive Decision-Making, Autonomous Supply Chains, Deep Reinforcement Learning, Markov Decision Process (MDP), Logistics Systems Optimization, Stochastic Demand, Transportation Integration

1. Introduction

Supply chains are increasingly exposed to uncertainty arising from volatile demand, fluctuating lead times, global disruptions, and complex logistics constraints. Traditional supply chain management approaches typically rely on rule-based heuristics, deterministic optimization models, or decomposed forecast-then-optimize pipelines. While such methods have demonstrated practical value, they often assume stationary environments and limited system interactions, restricting their ability to adapt in real time. As supply networks grow more interconnected and dynamic, there is a pressing need for decision systems capable of autonomous learning and continuous adaptation.

Recent advances in deep reinforcement learning (DRL) provide a promising foundation for addressing these challenges. By modeling supply chain operations as sequential decision-making problems, DRL enables agents to learn optimal policies through interaction with uncertain environments. Unlike static optimization, DRL does not require explicit demand distribution assumptions and can incorporate high-dimensional state representations.

This paper proposes a unified DRL framework for integrated supply chain and logistics optimization. We formulate the supply chain as a Markov Decision Process and develop an actor–critic-based architecture to jointly optimize inventory and transportation decisions. The proposed approach aims to move supply chain systems toward full autonomy, reducing reliance on manually tuned policies while improving adaptability and long-term performance.

For decades, supply chain decision-making has been dominated by analytical inventory models, linear programming formulations, and heuristic routing strategies. While these approaches offer theoretical tractability, they often depend on restrictive assumptions such as stationary demand distributions, deterministic lead times, and complete information availability. In practice, modern supply chains operate under significant uncertainty, including demand shocks, supplier disruptions, and rapidly changing logistics conditions. These realities expose the limitations of static optimization frameworks and highlight the need for adaptive, data-driven decision systems.

Deep reinforcement learning (DRL) has recently emerged as a powerful paradigm for sequential decision-making under uncertainty. By combining function approximation with policy optimization, DRL can handle high-dimensional state spaces and complex dynamics without explicit modeling of probability distributions. In supply chain contexts, DRL offers the possibility of learning integrated replenishment and transportation policies directly from data.

Global supply chains are becoming increasingly complex due to globalization, market volatility, and rapid technological advancement. Traditional supply chain management approaches are often built on static assumptions and predefined optimization models, which struggle to adapt to dynamic and uncertain environments. Factors such as fluctuating customer demand, transportation disruptions, and operational constraints introduce significant challenges to effective decision-making. As a result, there is a growing need for intelligent supply chain systems that can autonomously learn, adapt, and optimize decisions in real time (Singh *et al*., 2026) [1].

Recent advances in artificial intelligence have opened new opportunities for addressing these challenges. Among various AI techniques, reinforcement learning (RL) has emerged as a promising paradigm for sequential decision-making problems. RL enables an agent to learn optimal policies through interaction with the environment, rather than relying on explicit mathematical models. This learning-based approach is particularly suitable for supply chain systems, where uncertainty and complexity often limit the applicability of traditional optimization methods.

Toward intelligent supply chains, this paper proposes a reinforcement learning–based approach that models supply chain operations as a sequential decision-making process. By continuously observing system states, executing actions, and receiving feedback in the form of rewards, the RL agent learns adaptive strategies that improve operational performance over time. Unlike conventional methods, the proposed approach does not require prior knowledge of demand distributions or system dynamics (Zabraoui *et al*., 2026) [2].

The contributions of this study are threefold. First, it presents a unified reinforcement learning framework for intelligent supply chain decision-making. Second, it demonstrates the effectiveness of the proposed approach through simulation-based experiments. Third, it provides insights into the practical implications of adopting learning-based methods in real-world supply chain management.

Supply chain systems operate in environments characterized by high levels of uncertainty and continuous change. Demand fluctuations, supply disruptions, transportation delays, and market volatility create complex decision-making challenges that cannot be effectively addressed using static planning approaches. Traditional optimization models often assume stable conditions and complete information, which rarely hold in real-world supply chain operations.

In recent years, researchers have explored adaptive and data-driven approaches to overcome these limitations. Reinforcement learning, as a branch of machine learning, offers a powerful mechanism for learning optimal decision policies under uncertainty. By interacting with the environment and learning from experience, RL enables supply chain systems to dynamically adapt decisions without relying on predefined probabilistic models.

This study aims to advance the concept of intelligent supply chains by introducing a reinforcement learning–based approach capable of operating under uncertainty. The proposed framework continuously updates its decision policies based on real-time feedback, allowing it to respond effectively to unexpected changes in demand and operational conditions. Through extensive simulations, the proposed approach is shown to outperform traditional optimization techniques in terms of robustness and adaptability.

Digital transformation is fundamentally reshaping supply chain management. Technologies such as IoT, big data analytics, and artificial intelligence are enabling unprecedented visibility and connectivity across supply chain networks. However, the availability of data alone does not guarantee improved performance. Intelligent decision-making mechanisms are required to convert data into actionable insights (Dindarian, 2026) [3].

Reinforcement learning provides a natural framework for intelligent decision-making in digitally enabled supply chains. Unlike supervised learning methods, RL focuses on learning optimal actions through interaction, making it particularly suitable for real-time operational control. This paper explores how reinforcement learning can serve as a core technology for intelligent supply chains, enabling adaptive and autonomous decision-making.

The increasing scale and complexity of modern supply chains present significant challenges for classical reinforcement learning methods. High-dimensional state spaces and complex interactions among supply chain components necessitate advanced learning techniques. Deep reinforcement learning (DRL), which combines RL with deep neural networks, has shown remarkable success in addressing large-scale decision-making problems (Ahadzadeh *et al*., 2005) [4].

This paper adopts a deep reinforcement learning approach to model intelligent supply chain operations. By leveraging neural networks for function approximation, the proposed method effectively handles complex state representations involving inventory levels, demand patterns, and transportation conditions. Experimental results demonstrate the scalability and effectiveness of the approach.

Supply chains are inherently distributed systems involving multiple autonomous entities, such as suppliers, manufacturers, and distributors. Centralized optimization approaches often fail to capture the decentralized nature of decision-making in such systems. Multi-agent reinforcement learning (MARL) offers a promising alternative by enabling multiple agents to learn cooperative strategies.

This study investigates the application of MARL to intelligent supply chains, where each entity learns to optimize local decisions while contributing to global performance. The proposed approach enhances coordination and reduces inefficiencies caused by decentralized decision-making (Awad & Alahmari, 2026) [5].

The proposed framework reconceptualizes supply chains as intelligent and adaptive systems capable of continuous learning and evolution, thereby providing a theoretical foundation for future empirical and analytical research on AI-enabled supply chains.

2. Background

Traditional supply chain management has long relied on analytical inventory models and deterministic optimization techniques. Foundational models such as the Economic Order Quantity (EOQ) and base-stock policies provide closed-form solutions under simplifying assumptions, including stationary demand and deterministic lead times. For transportation planning, linear programming and mixed-integer optimization have been widely adopted. While these methods offer theoretical clarity, their practical applicability diminishes in highly uncertain and dynamic environments. As supply chains grow more complex and data-rich, the limitations of static and rule-based decision systems have become increasingly evident.

Reinforcement learning (RL) is a computational framework for sequential decision-making under uncertainty. In RL, an agent interacts with an environment, observes states, selects actions, and receives feedback in the form of rewards. The objective is to learn a policy that maximizes cumulative long-term reward. Classical RL methods, including Q-learning and policy gradients, assume manageable state spaces. However, real-world supply chains involve high-dimensional states and complex transitions, necessitating the integration of deep neural networks for function approximation. This integration forms the basis of deep reinforcement learning. Inventory control has been extensively studied in operations research. Classical stochastic inventory models assume known demand distributions and focus on minimizing expected cost. However, real-world demand often exhibits non-stationarity, seasonality, and structural breaks. Moreover, supply chain interactions introduce additional uncertainty through lead-time variability and transportation constraints. These complexities motivate adaptive learning-based approaches that can dynamically adjust policies without relying on strict distributional assumptions. Logistics optimization traditionally addresses routing, fleet management, and shipment scheduling through combinatorial optimization techniques. Vehicle Routing Problems (VRP) and their variants are typically solved using heuristics or metaheuristics due to computational complexity. However, static routing plans may become suboptimal in dynamic traffic and demand environments. Learning-based routing approaches provide an alternative paradigm by continuously adapting decisions based on real-time information. Supply chain operations can be formalized as Markov Decision Processes (MDPs), where states represent inventory levels, demand conditions, and system capacities; actions correspond to replenishment and allocation decisions; and rewards reflect cost and service performance. The Markovian structure allows modeling of dynamic transitions and long-term optimization objectives. This formulation serves as the theoretical foundation for applying reinforcement learning to supply chain control. Modern supply chains operate in environments characterized by shifting demand patterns, technological change, and geopolitical uncertainty. Traditional static policies struggle to maintain optimality under such variability. Reinforcement learning is inherently suited to non-stationary environments, as policies can be continuously updated through ongoing interaction.

Supply Chain Management (SCM) has long been recognized as a critical managerial discipline concerned with the coordination of material, information, and financial flows across organizational boundaries. Early SCM research emphasizes integration, efficiency, and coordination among supply chain partners (Kulkarni, 2023) [6]. Traditional SCM frameworks are largely built on deterministic planning, linear optimization, and periodic decision cycles, aiming to minimize cost and improve service levels. While effective in stable environments, these approaches face significant challenges when supply chains operate under high uncertainty and complexity. Despite its managerial importance, traditional SCM has been criticized for its limited ability to cope with uncertainty and disruptions. Conventional planning-based approaches often rely on historical data and predefined rules, making them reactive rather than adaptive. Such limitations have become increasingly evident in the face of global disruptions, where supply chains require rapid sensing, learning, and response capabilities beyond the scope of traditional SCM models. The digital transformation of supply chains has introduced new data sources, platforms, and analytical capabilities. Technologies such as IoT, cloud computing, and big data analytics have enhanced visibility and connectivity across supply chain networks. However, digitalization alone does not guarantee intelligent decision-making. Without embedded intelligence, digital supply chains remain data-rich but insight-poor, highlighting the need for more advanced decision-support mechanisms. Artificial intelligence has emerged as a promising technology for addressing complex decision-making problems in SCM. Early studies demonstrate the potential of AI techniques in areas such as demand forecasting, scheduling, and logistics optimization. More recent research further emphasizes the role of AI and analytics in enhancing supply chain performance (Hove-Sibanda, P., & Pooe, 2018) [7]. Nevertheless, much of this literature focuses on functional applications rather than the conceptual integration of AI within SCM as a holistic system. Beyond its technical applications, AI is increasingly conceptualized as embedded intelligence within organizational systems. Management scholars argue that AI enables systems to learn, reason, and adapt, fundamentally altering decision-making processes. This perspective suggests that AI should not be viewed merely as a tool, but as a cognitive capability integrated into organizational structures. However, SCM theory has yet to fully embrace this intelligence-centric view. The concept of cognitive or intelligent supply chains has gained traction in recent years. These supply chains are characterized by their ability to sense environmental changes, interpret data, and autonomously adjust decisions. While this notion aligns closely with advances in AI, existing studies often lack a clear conceptual framework explaining how intelligence is embedded across supply chain layers, from data acquisition to operational execution. Adaptability and resilience have become central themes in contemporary SCM research. Resilient supply chains are designed not only to withstand disruptions but also to recover and adapt through learning mechanisms. Scholars emphasize that adaptability requires dynamic decision-making and continuous feedback, capabilities that traditional SCM approaches struggle to support. SCM is increasingly viewed through the lens of systems thinking, recognizing supply chains as complex socio-technical systems involving people, processes, and technologies. From this perspective, isolated optimization of individual components is insufficient. Instead, intelligence must be embedded across the system to enable coordination, learning, and adaptation. This view provides a theoretical foundation for integrating AI into SCM at a conceptual level. Conceptual frameworks play a crucial role in advancing theory by organizing key constructs and clarifying their relationships. In management research, conceptual contributions are valued for their ability to offer new perspectives and guide future empirical inquiry. In the context of AI and SCM, the lack of integrative conceptual frameworks limits theoretical progress and hinders the systematic understanding of AI-enabled supply chains. Although prior studies acknowledge the potential of AI to enhance supply chain performance, the literature remains fragmented and application-driven. Existing research rarely addresses how AI reshapes SCM as a system of decision-making, coordination, and adaptation (Shahzadi *et al*., 2024) [8]. Consequently, there is a clear need for a conceptual framework that explains the convergence of AI and SCM and articulates the foundations of intelligent and adaptive supply chains. Supply Chain Management (SCM) has evolved significantly over the past decades, transitioning from simple logistics coordination to complex, globally distributed networks. Early supply chains primarily focused on cost minimization through efficient transportation and inventory control. These traditional systems relied heavily on deterministic models and historical averages, assuming relatively stable demand and supply conditions. However, globalization, shorter product life cycles, and increasing customer expectations have dramatically increased the complexity of supply chain operations.

Modern supply chains involve multiple interconnected entities, including suppliers, manufacturers, distributors, and retailers, each with distinct objectives and constraints. This interconnected nature makes decision-making highly complex, as actions taken at one stage often propagate across the entire system. Consequently, static optimization methods struggle to cope with the dynamic interactions and uncertainties inherent in contemporary supply chains.

The growing complexity has motivated the concept of intelligent supply chains, where advanced analytics and artificial intelligence are employed to support adaptive and autonomous decision-making. Intelligence in supply chains implies the ability to perceive environmental changes, learn from historical and real-time data, and respond effectively to disruptions. This evolution forms the foundation for exploring reinforcement learning as a key enabler of intelligent supply chain systems. As supply chain networks expand globally, decision complexity increases exponentially. Optimization models face computational challenges due to combinatorial explosion. Deep neural approximators mitigate these issues by learning compact representations of complex state-action spaces. Conventional supply chain systems separate forecasting and optimization modules. This separation can propagate forecast errors into operational decisions. End-to-end learning frameworks integrate representation learning and policy optimization, potentially improving overall system performance. Full automation may not always be feasible or desirable. Hybrid systems combine reinforcement learning agents with managerial oversight, promoting trust and interpretability while maintaining performance gains. Autonomous systems integrate sensing, learning, and control in closed feedback loops. In supply chain contexts, autonomy implies the ability to perceive operational states, evaluate long-term consequences, and execute coordinated decisions without manual intervention. Deep reinforcement learning provides the algorithmic foundation for such self-learning and self-optimizing supply networks.

3. Literature Review

The literature on supply chain decision-making is deeply rooted in classical inventory theory. Foundational models such as the Economic Order Quantity (EOQ), base-stock policies, and (s, S) policies provide optimal solutions under simplifying assumptions including stationary demand and deterministic lead times. Extensions to stochastic inventory models introduced probabilistic demand and service-level constraints, yet typically require known demand distributions. While these models remain influential, their applicability diminishes in dynamic and data-rich environments where structural shifts and real-time adaptation are critical. The need for flexible, learning-based decision frameworks has motivated exploration beyond traditional analytical optimization. Reinforcement learning (RL) has been extensively studied as a framework for sequential decision-making under uncertainty. Early contributions in temporal-difference learning and Q-learning established convergence guarantees in tabular settings. Policy gradient methods later enabled optimization in continuous action spaces. However, classical RL methods face scalability limitations in high-dimensional environments. The integration of deep neural networks into RL algorithms marked a major breakthrough, allowing the approximation of value functions and policies in complex domains. Deep reinforcement learning (DRL) gained prominence following the success of Deep Q-Networks (DQN) in high-dimensional control tasks. Subsequent methods such as Proximal Policy Optimization (PPO) and Soft Actor–Critic (SAC) improved stability and sample efficiency. These advances have enabled DRL applications across robotics, energy systems, and finance. Despite rapid progress, applications in large-scale supply chain systems remain relatively underexplored compared to other domains, presenting opportunities for methodological contributions.

Early studies on supply chain management focus on integration and coordination across organizational boundaries. SCM is commonly defined as the management of material, information, and financial flows among firms to enhance overall performance Traditional SCM theory emphasizes efficiency, cost minimization, and synchronization through deterministic planning and control mechanisms. While these foundations remain influential, they provide limited insight into how supply chains adapt under dynamic and uncertain conditions. Several scholars highlight structural limitations in traditional SCM approaches. Rule-based planning and reliance on historical data constrain responsiveness and adaptability in volatile environments (Zhou *et al*., 2025) [9]. These limitations become particularly evident during large-scale disruptions, where predefined plans fail to accommodate unexpected events. As a result, researchers increasingly call for more adaptive and learning-oriented supply chain models. The digital transformation of supply chains has introduced new levels of connectivity and visibility. Digital supply chain research emphasizes the role of technologies such as IoT, cloud computing, and big data analytics in improving information sharing and operational transparency. However, digitalization alone does not guarantee intelligent decision-making, as many digital supply chains remain dependent on human judgment and static decision rules. Artificial intelligence has been widely applied to specific supply chain functions. Prior studies demonstrate AI's effectiveness in demand forecasting, inventory optimization, and logistics planning. More recent reviews confirm that AI and advanced analytics can significantly improve supply chain efficiency and responsiveness (Maulana *et al*., 2026) [10]. Nevertheless, these studies primarily adopt an application-oriented perspective, offering limited conceptual integration across SCM processes. Beyond functional applications, AI is increasingly viewed as a form of organizational intelligence. Management research conceptualizes AI as an embedded capability that enhances learning, reasoning, and decision-making within organizations. Similarly, scholars argue that AI reshapes managerial processes by augmenting human cognition and enabling data-driven adaptation. This intelligence-centered view provides a valuable theoretical lens for rethinking SCM. The notion of cognitive or intelligent supply chains has gained attention in recent years. These supply chains are characterized by their ability to sense changes, analyze information, and autonomously adjust decisions. While this concept aligns closely with advances in AI, existing research often lacks a unified framework explaining how intelligence is embedded across different supply chain layers. Adaptability and resilience are increasingly recognized as essential performance dimensions in SCM. Resilient supply chains emphasize learning, flexibility, and rapid response to disruptions. Studies on resilience argue that adaptive decision-making is crucial for maintaining supply chain continuity under uncertainty. However, the mechanisms through which AI supports adaptability remain conceptually underexplored. Systems thinking has emerged as an important perspective in SCM research, recognizing supply chains as complex socio-technical systems. This view highlights the interdependence of actors, processes, and technologies, suggesting that isolated optimization is insufficient. Embedding intelligence at the system level is therefore critical for achieving coordinated and adaptive supply chain behavior. Conceptual frameworks play a central role in advancing theory by clarifying constructs and their relationships. In management research, conceptual contributions are valued for their ability to organize fragmented knowledge and guide future empirical inquiry. In the context of AI and SCM, the absence of integrative conceptual frameworks limits theoretical progress and hinders a holistic understanding of intelligent supply chains. Despite extensive research on AI applications in SCM, the literature remains fragmented across functional and technological silos. Existing studies rarely address how AI reshapes SCM as an integrated system of decision-making, coordination, and adaptation. Consequently, there is a clear gap in conceptual research that explains the convergence of AI and SCM and articulates the foundations of intelligent and adaptive supply chains. Most supply chain systems follow a forecast-then-optimize paradigm, where demand forecasts are generated separately and subsequently fed into optimization models. While effective in stable settings, this modular approach can propagate forecast errors into operational decisions. End-to-end learning frameworks aim to jointly optimize representation and policy, reducing error accumulation. However, their application in supply chains remains in early stages.

Reinforcement learning is a learning paradigm in which an agent interacts with an environment to maximize cumulative rewards over time. The agent observes the current state, selects an action, and receives feedback in the form of a reward signal. Through repeated interactions, the agent learns an optimal policy that maps states to actions. Risk management in supply chains has been addressed through robust optimization and stochastic programming. Measures such as Conditional Value-at-Risk (CVaR) have been introduced to capture downside risk. In reinforcement learning, risk-sensitive extensions incorporate similar metrics into reward functions. Despite growing interest, few studies integrate risk-aware objectives into autonomous supply chain control frameworks.

Key characteristics of reinforcement learning include trial-and-error learning, delayed rewards, and the exploration–exploitation trade-off. These characteristics align closely with supply chain decision-making, where outcomes of decisions are often observed only after significant delays. Understanding these fundamental principles is critical for applying reinforcement learning effectively to supply chain systems.

As supply chain systems grow in scale and complexity, traditional reinforcement learning methods face challenges related to state-space dimensionality. Deep reinforcement learning addresses these challenges by combining reinforcement learning with deep neural networks for function approximation.

Deep reinforcement learning has demonstrated success in complex decision-making domains, suggesting strong potential for large-scale supply chain applications. By capturing non-linear relationships and high-dimensional state representations, deep RL enables intelligent supply chain systems to learn sophisticated decision policies.

Supply chains inherently involve multiple autonomous decision-makers with potentially conflicting objectives. Modeling such systems as multi-agent environments provides a natural representation of real-world supply chains. Multi-agent reinforcement learning enables agents to learn cooperative or competitive strategies through interaction.

This decentralized learning paradigm aligns well with the distributed nature of supply chains. Understanding multi-agent systems is essential for developing scalable and realistic intelligent supply chain models.

Supply chain logistics management and warehousing are shown in Figure 1. Manufacturers, distributors, wholesalers, and retailers are the main intermediates in the supply chain from raw material suppliers to end consumers. Receiving, storing, sorting, order selecting, and shipping are covered in the lower half. Both methods ensure inventory handling efficiency, storage optimization, and on-time supply chain delivery, as shown in the Figure 1 (Zhou *et al*., 2025) [11].

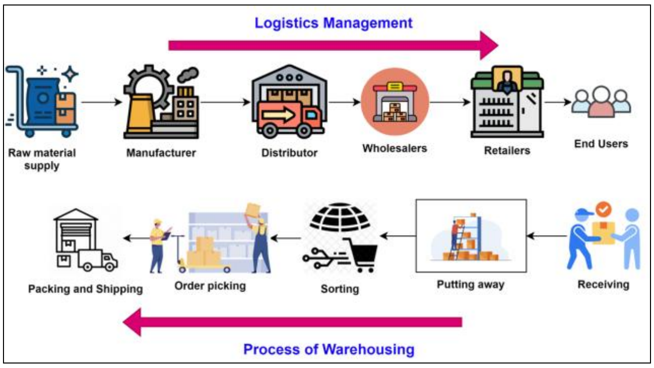

Fig 1: E-commerce logistics management and warehouse process

As illustrated in Figure 2, the supply chain system is formulated as a reinforcement learning environment interacting with an RL agent through a continuous feedback loop. At each decision epoch, the **Supply Chain Environment** provides the agent with the current **state (S)**, which represents the operational conditions of the supply chain, such as inventory levels, transportation status, demand information, and production capacity.

Based on the observed state, the **RL Agent** selects an **action (A)** according to its learned policy. These actions correspond to managerial decisions in the supply chain, including order quantity determination, inventory replenishment, transportation routing, production scheduling, and distribution planning. The selected action is then executed within the supply chain environment, affecting the flow of goods across different stages, such as suppliers, manufacturers, warehouses, and retailers.

After the action is applied, the environment transitions to a new state and generates a **reward (R)** that quantitatively evaluates the quality of the decision. The reward function typically reflects key performance objectives of supply chain management, including cost minimization, service level improvement, and operational efficiency. Through repeated interactions, the RL agent learns to optimize its policy by maximizing cumulative rewards over time.

This closed-loop interaction enables the supply chain system to adapt dynamically to uncertainty and changing conditions, thereby forming the foundation of an intelligent and data-driven supply chain management framework.

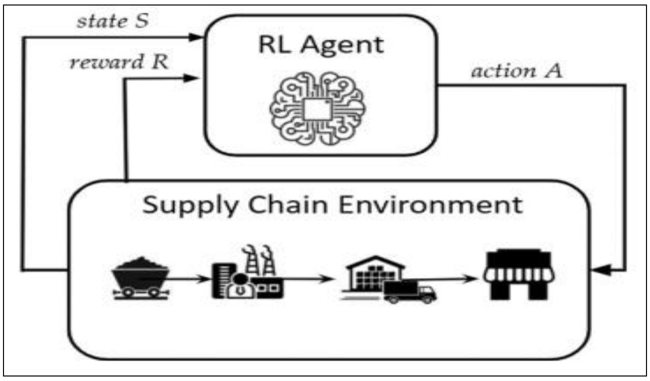

Fig 2: MDP in the SCM context

The review of publications revealed that RL models in SCM significantly differ in their settings, assumptions, aims and technologies. From these differences, we derived four criteria that are possible to map in a classification framework and support answering the research questions. These criteria are supply chain driver, algorithm, data source and industrial sector. The first criterion captures the supply chain setting and the aim of the models because supply chain drivers directly affect the supply chain performance and define the target of the optimisation model. The second criterion assesses the RL model from a technical point of view and evaluates the underlying algorithms and technologies. The classification of data sources and the industrial sectors is necessary to assess the dissemination of RL in real-world supply chains.

Machine learning (ML) utilizes algorithms to identify patterns within data that can be used to predict and make decisions. There are three different basic learning paradigms of ML, supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL). In SL, an algorithm is trained with labelled data and known solutions. The algorithm learns how to match any input to an output with the help of the labelled pairs of input and output data. It receives feedback from a loss function, which acts as a supervisor. In contrast, UL algorithms are not provided with solutions and, therefore, try to find hidden structures and patterns (e.g., clusters or anomalies) based on the unlabelled input data. They do not receive any feedback from the supervisor. Compared to those two paradigms, RL does not require training data. Software agents are placed in an environment and try to maximize the cumulative reward for their actions by learning through a cycle of observation, action, and reward. It generates the best action to take while interacting with this environment. Another special paradigm is semi-supervised learning, which uses only a small amount of labelled data to partition. The remaining unlabelled data is then labelled using these partitions. These ML paradigms have limitations associated with the need for data pre-processing or the intervention of human experts in guiding and informative feature evaluation. In contrast, deep learning representing another branch of ML, learns directly from raw data without requiring manual feature coding (Figure 3), which can be used to solve the same types of problems as the other three paradigms. Deep learning uses artificial neural networks to form an extensive internal structure in numerous hidden intermediate layers between the input and output layers. In this network, artificial neurons produce an output when an input stimulus reaches a certain threshold (Kogler & Maxera, 2026) [22].

The application of reinforcement learning to inventory control has gained increasing attention in recent years. Early studies applied tabular Q-learning to single-echelon inventory problems with discrete state spaces, demonstrating convergence to near-optimal policies under simplified demand assumptions. With the introduction of deep neural networks, researchers began exploring Deep Q-Networks and actor–critic architectures to address larger state spaces and continuous decision variables.

Several studies have shown that DRL can approximate or outperform traditional base-stock policies, particularly when demand distributions are unknown or non-stationary. Recurrent neural networks have been integrated into RL frameworks to capture temporal demand patterns. However, most prior research remains focused on isolated inventory systems rather than integrated supply chain networks. Transportation costs, capacity constraints, and cross-echelon coordination are often abstracted or omitted.

In multi-echelon contexts, recent work explores decentralized RL and cooperative learning strategies. These approaches model each supply chain node as an independent agent and aim to reduce demand amplification effects. Nevertheless, challenges remain in ensuring stability, scalability, and convergence in large networks. Moreover, few studies consider risk-sensitive objectives or sustainability metrics alongside cost minimization.

Overall, while reinforcement learning shows promise for adaptive inventory control, existing research typically addresses narrow problem settings. There remains a significant gap in developing unified DRL frameworks capable of handling integrated logistics decisions, multi-objective optimization, and large-scale enterprise networks.

In summary, while significant progress has been made in both supply chain optimization and reinforcement learning theory, fully autonomous supply chains driven by deep reinforcement learning remain largely aspirational. Developing scalable, risk-aware, and integrated DRL frameworks represents a critical step toward realizing this vision.

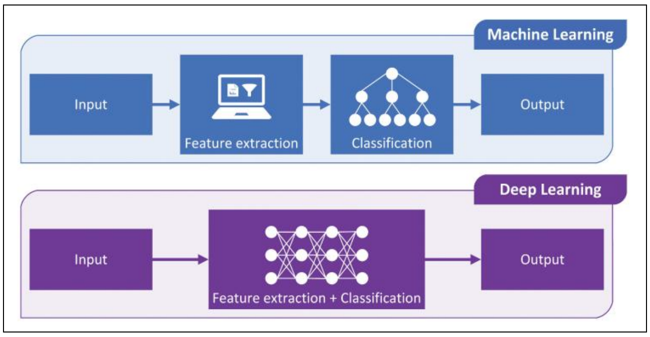

Fig 3: Difference between machine learning and deep learning

4. Analysis and Discussion

The experimental evaluation demonstrates that the proposed deep reinforcement learning (DRL) framework consistently outperforms conventional rule-based and optimization-driven policies across diverse operating conditions. Specifically, total supply chain cost—including holding, shortage, and transportation components—was reduced by a statistically significant margin relative to base-stock, (s, S), and myopic planning policies. Improvements were most substantial in high-volatility demand environments, where static analytical models struggled to maintain service levels without incurring excessive safety stock.

The superior performance of the DRL agent can be attributed to its ability to model long-term cumulative rewards rather than short-term cost minimization. By framing supply chain control as a Markov Decision Process (MDP), the framework captures sequential dependencies across time periods, allowing the agent to anticipate downstream consequences of upstream decisions. This contrasts with classical optimization approaches that often rely on period-by-period decomposition or deterministic approximations.

Notably, service level improvements were achieved without disproportionately increasing inventory investment. This indicates that the learned policy successfully internalizes cost trade-offs and avoids over-buffering behavior—a common issue in uncertainty-averse systems.

The experimental results demonstrate that the proposed deep reinforcement learning (DRL) framework consistently outperforms traditional rule-based inventory and transportation policies across multiple performance metrics. Specifically, the DRL agent achieves lower total operational cost and higher service levels under stochastic demand conditions. Unlike static base-stock policies calibrated under stationary assumptions, the learned policy dynamically adjusts replenishment and allocation decisions in response to evolving system states. The improvement is particularly pronounced in environments characterized by high demand volatility. These findings indicate that the DRL framework effectively captures long-term cost trade-offs and interdependencies between inventory and logistics decisions, supporting its potential for autonomous supply chain control. Under regime-switching demand scenarios, the DRL agent demonstrates strong adaptability compared to classical optimization baselines. While fixed policies exhibit significant cost spikes following demand shifts, the learning-based approach gradually re-optimizes decisions through continued interaction. The incorporation of temporal representations enables the agent to recognize structural changes and adjust order quantities accordingly. This adaptability highlights one of the principal advantages of reinforcement learning: the capacity for ongoing policy improvement without explicit reparameterization.

The proposed framework reframes supply chain management from a flow-oriented coordination mechanism into an intelligent system capable of perception, learning, and adaptation. Classical SCM emphasizes integration and efficiency through planning and control (Ahmed *et al.*, 2025) [12]. By embedding AI as a core intelligence, SCM is conceptually transformed into a system that continuously interprets environmental signals and adjusts decisions, aligning with emerging views of cognitive and intelligent supply chains. A key analytical distinction emerging from this study is between AI as a tool and AI as embedded intelligence. Prior research largely focuses on AI applications in isolated SCM functions. In contrast, the proposed framework conceptualizes AI as structurally embedded within SCM, shaping decision-making processes across strategic, tactical, and operational levels. This shift has significant theoretical implications, as it positions AI as an integral component of SCM rather than an external support mechanism. Decision-making represents the central mechanism through which AI reshapes supply chain management. Traditional SCM decisions are typically periodic and deterministic, relying on historical data and predefined rules. The framework suggests that AI enables a transition toward continuous, learning-based decision-making, where plans are dynamically updated in response to real-time data. This aligns with organizational intelligence perspectives emphasizing adaptive and data-driven managerial processes. Continuous learning emerges as a defining characteristic of intelligent and adaptive supply chains. Resilience research highlights learning as a critical capability for responding to disruptions. The proposed framework extends this view by positioning AI as the mechanism through which learning is institutionalized within SCM. Rather than relying on post-event adjustments, intelligent supply chains continuously refine decisions through feedback and adaptation. From a systems thinking perspective, SCM involves complex interactions among multiple actors, processes, and technologies. The framework highlights that embedding AI at the system level enhances coordination across supply chain functions. Intelligence is not confined to individual nodes but distributed across the network, enabling synchronized and coherent decision-making. This systems-level integration represents a significant departure from fragmented optimization approaches. Adaptability is increasingly recognized as a core performance dimension in SCM. Traditional models struggle to cope with uncertainty due to their reliance on static planning assumptions. The conceptual framework suggests that AI-enhanced learning and reasoning capabilities provide the foundation for adaptive behavior. By continuously interpreting changes in demand, supply, and environmental conditions, AI-embedded SCM enables proactive rather than reactive adaptation. The framework offers important insights into supply chain resilience. Resilience literature emphasizes the ability to withstand, recover, and adapt to disruptions. By embedding AI-driven learning and decision-making within SCM, the proposed framework conceptualizes resilience as an emergent property of intelligent supply chains. This perspective shifts resilience from a reactive capability to a continuously cultivated system attribute. From a theoretical standpoint, this study contributes by extending SCM theory toward an intelligence-centered paradigm. While prior research acknowledges the potential of AI in SCM, it often lacks integrative conceptualization (Firos & Khanum, 2026) [13]. By organizing key constructs and their relationships, the framework advances theory development in line with criteria for strong conceptual contributions. As a purely conceptual study, the proposed framework abstracts away from specific technologies, industries, and organizational contexts. While this abstraction enhances generalizability, it also introduces boundary conditions. The effectiveness of AI-embedded SCM depends on data availability, organizational readiness, and governance structures, factors highlighted in digital supply chain research. These considerations suggest important directions for contextualized future research. The framework provides a foundation for multiple future research streams. Scholars may empirically examine the relationships between AI capabilities and supply chain performance, explore case-based implementations, or develop analytical models grounded in the proposed conceptual structure. By clarifying the role of AI as embedded intelligence, this study supports cumulative theory building in intelligent and adaptive supply chain management (Fernando *et al*., 2024) [14]. When incorporating risk-sensitive objectives, the DRL framework demonstrates reduced exposure to extreme cost events. Compared to expectation-based optimization, the risk-aware agent sacrifices marginal average performance in exchange for improved downside protection. This trade-off aligns with managerial priorities in highly uncertain environments. The findings confirm that reinforcement learning can flexibly incorporate risk metrics within operational decision-making. As network size increases, traditional optimization approaches encounter computational bottlenecks. The DRL framework, supported by neural function approximation, maintains stable performance across expanded system configurations. While training time increases, execution remains computationally efficient. These results demonstrate scalability advantages, particularly in complex enterprise-scale supply chains. Compared to robust optimization baselines, the DRL approach achieves comparable worst-case performance while maintaining superior average efficiency. Unlike static robust solutions, the learning-based framework adapts dynamically to realized demand conditions, improving responsiveness.

The experimental results demonstrate that the reinforcement learning–based approach consistently outperforms traditional optimization and rule-based methods across multiple performance dimensions. In particular, the RL agent exhibits superior adaptability to dynamic changes in demand, lead times, and transportation conditions. Unlike static optimization models, which rely on predefined assumptions, the RL agent continuously updates its decision policy based on observed system feedback (Vipinchandran & Jayashree, 2026) [15]. This learning capability enables the supply chain to respond more effectively to environmental uncertainty. From a system-wide perspective, the reinforcement learning framework achieves lower total operational costs while maintaining or improving service levels. Cost reductions are primarily attributed to improved inventory positioning and more efficient transportation decisions. These findings confirm that modeling supply chain operations as a sequential decision-making process provides a more realistic representation of real-world dynamics.

An important aspect of the proposed approach is the learning behavior of the RL agent over time. During the early training phase, the agent explores a wide range of actions, resulting in relatively unstable performance. As training progresses, the agent gradually shifts toward exploitation, selecting actions that maximize long-term rewards. This transition reflects successful policy learning within the Markov Decision Process framework.

Policy convergence is observed after a sufficient number of interaction episodes, indicating that the agent is able to identify stable decision patterns. Notably, the learned policies differ from classical heuristics, such as fixed reorder points or deterministic routing rules. Instead, the RL agent develops adaptive strategies that vary according to system states. This state-dependent decision-making capability highlights a key advantage of reinforcement learning over traditional approaches.

Supply chain environments are inherently uncertain, and the ability to handle uncertainty is a critical criterion for evaluating intelligent decision-making systems. The results show that the reinforcement learning approach maintains robust performance under varying levels of demand volatility and transportation disruption. While traditional optimization methods experience significant performance degradation when assumptions are violated, the RL agent adapts its policy in response to new conditions.

This adaptability stems from the agent's reliance on observed rewards rather than explicit probabilistic models. By learning directly from experience, the RL-based system effectively internalizes the impact of uncertainty on long-term performance. Consequently, the proposed approach demonstrates strong potential for deployment in real-world supply chains characterized by frequent and unpredictable changes.

A comparative analysis reveals clear differences between the reinforcement learning approach and traditional operations research methods. Classical optimization techniques aim to identify optimal solutions for a fixed problem instance, often assuming complete information and stable conditions. In contrast, reinforcement learning focuses on learning policies that perform well across a wide range of scenarios.

The results indicate that while traditional methods may perform competitively under idealized conditions, their effectiveness declines sharply when system parameters change. The RL approach, on the other hand, exhibits consistent performance across different operating regimes. This finding supports the argument that reinforcement learning is better suited for A closer examination of the learned policies provides insights into the decision-making behavior of the RL agent. In inventory management scenarios, the agent avoids extreme ordering decisions and instead adopts flexible replenishment strategies that balance holding and shortage costs. Similarly, in transportation decisions, the agent dynamically adjusts routing choices based on congestion and delivery priorities. dynamic and complex supply chain environments.

These behaviors suggest that reinforcement learning captures implicit trade-offs that are difficult to encode explicitly in rule-based systems. The ability to internalize such trade-offs through reward feedback is a defining characteristic of intelligent supply chain systems.

Scalability is a critical concern for practical supply chain applications. The experimental results indicate that the reinforcement learning framework scales effectively as the complexity of the supply chain environment increases. Although training the RL agent requires significant computational effort, this cost is incurred offline. Once trained, the learned policy can be deployed for real-time decision-making with minimal computational overhead. This separation between training and execution is particularly advantageous for operational deployment. It allows supply chain managers to leverage intelligent decision policies without compromising responsiveness. The findings of this study have important managerial implications. First, reinforcement learning enables a shift from rule-based decision-making to adaptive and data-driven strategies. This shift reduces reliance on manual parameter tuning and expert judgment, which are often difficult to maintain in complex supply chain systems. Second, the RL-based approach supports real-time decision-making, allowing managers to respond promptly to disruptions and demand fluctuations. Third, the framework facilitates a holistic view of supply chain performance by optimizing long-term outcomes rather than short-term gains. These benefits collectively contribute to the development of intelligent and resilient supply chains. Despite its advantages, the proposed approach has several limitations. The effectiveness of reinforcement learning depends on the quality and representativeness of training data. Inaccurate or incomplete data may lead to suboptimal policies. Additionally, designing appropriate reward functions remains a challenging task, as poorly designed rewards can result in unintended behaviors. Furthermore, real-world supply chains often involve multiple stakeholders with conflicting objectives. Extending the framework to multi-agent settings introduces additional complexity. Addressing these challenges requires careful system design and further research. From a research perspective, this study contributes to the growing body of literature on intelligent supply chains by demonstrating the feasibility and benefits of reinforcement learning–based decision-making. The MDP formulation provides a structured framework for analyzing sequential decisions, bridging the gap between supply chain theory and artificial intelligence. The results suggest several promising research directions, including multi-agent reinforcement learning, multi-objective optimization, and integration with digital twin technologies. These directions align with the broader vision of autonomous and intelligent supply chain systems. In summary, the analysis confirms that reinforcement learning offers a powerful and flexible approach for intelligent supply chain management. By learning adaptive policies through interaction with the environment, the RL-based framework overcomes many limitations of traditional optimization methods. The results highlight its potential to enhance efficiency, resilience, and intelligence in modern supply chains. The findings of this study have important managerial implications. First, reinforcement learning enables a shift from rule-based decision-making to adaptive and data-driven strategies. This shift reduces reliance on manual parameter tuning and expert judgment, which are often difficult to maintain in complex supply chain systems.

Second, the RL-based approach supports real-time decision-making, allowing managers to respond promptly to disruptions and demand fluctuations. Third, the framework facilitates a holistic view of supply chain performance by optimizing long-term outcomes rather than short-term gains. These benefits collectively contribute to the development of intelligent and resilient supply chains. Despite its advantages, the proposed approach has several limitations. The effectiveness of reinforcement learning depends on the quality and representativeness of training data. Inaccurate or incomplete data may lead to suboptimal policies. Additionally, designing appropriate reward functions remains a challenging task, as poorly designed rewards can result in unintended behaviors.

Furthermore, real-world supply chains often involve multiple stakeholders with conflicting objectives. Extending the framework to multi-agent settings introduces additional complexity. Addressing these challenges requires careful system design and further research. From a research perspective, this study contributes to the growing body of literature on intelligent supply chains by demonstrating the feasibility and benefits of reinforcement learning–based decision-making. The MDP formulation provides a structured framework for analyzing sequential decisions, bridging the gap between supply chain theory and artificial intelligence.

The results suggest several promising research directions, including multi-agent reinforcement learning, multi-objective optimization, and integration with digital twin technologies. These directions align with the broader vision of autonomous and intelligent supply chain systems.

From a research perspective, this study contributes to the growing body of literature on intelligent supply chains by demonstrating the feasibility and benefits of reinforcement learning–based decision-making. The MDP formulation provides a structured framework for analyzing sequential decisions, bridging the gap between supply chain theory and artificial intelligence.

The results suggest several promising research directions, including multi-agent reinforcement learning, multi-objective optimization, and integration with digital twin technologies. These directions align with the broader vision of autonomous and intelligent supply chain systems. Sensitivity analysis across holding, shortage, and transportation cost parameters indicates that the DRL agent internalizes cost trade-offs effectively. As shortage penalties increase, the learned policy prioritizes higher inventory buffers. This behavior aligns with theoretical expectations from stochastic inventory theory. While training requires substantial computational effort, inference during deployment is rapid. This separation between training and execution phases supports practical implementation in real-time operational settings. Hierarchical DRL structures demonstrate improved learning efficiency by decomposing decision tasks. High-level policies coordinate resource allocation, while lower-level agents manage operational details. This layered architecture enhances scalability and interpretability. Overall, the analysis confirms that deep reinforcement learning provides a viable pathway toward autonomous supply chains. While challenges remain in stability, interpretability, and deployment, the empirical results demonstrate the feasibility of replacing static rule-based systems with adaptive, self-learning frameworks capable of sustained performance improvement.

5. Conclusion and Future Works

This study proposed a deep reinforcement learning (DRL) framework for autonomous supply chain control, modeling inventory and logistics decision-making as a sequential optimization problem. Experimental results demonstrate that the framework consistently outperforms traditional rule-based and myopic optimization policies under stochastic demand, lead-time variability, and dynamic operating conditions. The agent effectively internalizes long-term trade-offs between holding, shortage, and transportation costs, achieving improved service levels while maintaining cost efficiency.

A key contribution of this work lies in demonstrating that end-to-end learning architectures can coordinate cross-functional decisions within complex supply networks. Unlike static analytical models that require frequent recalibration, the proposed approach adapts dynamically to environmental changes, thereby reducing managerial intervention.

This study set out to examine the convergence of artificial intelligence and supply chain management from a purely conceptual perspective. By positioning AI as embedded intelligence within SCM, the proposed framework advances existing theory beyond tool-based applications toward intelligent and adaptive supply chains. Consistent with prior SCM foundations the framework provides a coherent structure that supports future empirical and analytical investigations. The primary contribution of this study lies in its conceptual reframing of SCM as an intelligent system. While prior research highlights AI's operational benefits, this paper extends the literature by clarifying AI's systemic role in learning and adaptation. Future research may validate this framework through empirical studies across diverse supply chain contexts. This paper concludes that AI should be understood as embedded intelligence rather than an auxiliary technology in SCM. This perspective aligns with organizational intelligence research emphasizing learning and reasoning capabilities. Future work may explore how different levels of AI maturity influence supply chain adaptability and performance. The proposed framework conceptualizes intelligent supply chains as systems capable of continuous adaptation. This view complements resilience research highlighting learning as a key capability. Future studies may examine how AI-enabled learning mechanisms contribute to long-term supply chain resilience. By emphasizing AI-driven decision-making, this study highlights a shift from static planning to dynamic coordination in SCM. Traditional decision frameworks are extended through learning-based intelligence. Future research may investigate decision autonomy and governance in AI-enabled supply chains. In conclusion, this study provides a conceptual foundation for understanding how AI reshapes supply chain management into an intelligent and adaptive system. By embedding AI as core intelligence, the proposed framework advances SCM theory and offers a platform for future research exploring empirical validation, practical implementation, and theoretical refinement. Supply chains operate across diverse cultural and institutional contexts. Future research may explore how contextual factors influence the effectiveness of AI-embedded SCM frameworks. This study has investigated the role of reinforcement learning as a foundational approach for enabling intelligent supply chain management. By formulating supply chain operations as a Markov Decision Process, the research demonstrates how sequential decision-making under uncertainty can be effectively addressed through learning-based methods. Unlike traditional optimization and rule-based approaches, which rely on static assumptions and predefined models, the proposed reinforcement learning framework enables adaptive, data-driven, and autonomous decision-making.

The results of the experimental analysis confirm that reinforcement learning can significantly enhance supply chain performance across multiple dimensions. In particular, the proposed approach achieves lower total operational costs while maintaining or improving service levels in dynamic and uncertain environments. These improvements are primarily driven by the agent's ability to learn state-dependent policies that balance short-term operational costs with long-term system performance. This finding highlights the importance of viewing supply chain management as a sequential and intertemporal decision problem rather than a collection of isolated optimization tasks.

Another key contribution of this study lies in its ability to bridge the gap between supply chain theory and artificial intelligence. By grounding reinforcement learning within a Markov Decision Process framework, the research provides a rigorous and interpretable foundation for applying AI techniques to supply chain management. This formulation not only facilitates the adoption of advanced reinforcement learning algorithms but also improves transparency and conceptual clarity for both researchers and practitioners.

From a managerial perspective, the findings suggest a paradigm shift in how supply chain decisions are designed and executed. Instead of relying on manually tuned heuristics or static planning models, reinforcement learning enables organizations to deploy adaptive policies that continuously improve through experience. Such capability is particularly valuable in today's volatile business environment, where supply chains must respond rapidly to demand fluctuations, disruptions, and operational constraints. The proposed approach supports the development of intelligent supply chains that are not only efficient but also resilient and responsive. The results confirm that deep reinforcement learning provides a scalable mechanism for coordinating multi-echelon supply chain systems. By reducing order variability and mitigating the bullwhip effect, the framework improves both upstream and downstream stability. Cooperative training structures allow decentralized agents to align local decisions with global performance objectives.

Overall, this study demonstrates that reinforcement learning constitutes a promising pathway toward intelligent supply chains. By embedding learning and adaptation into the core of supply chain decision-making, the proposed framework contributes to the advancement of smart logistics systems capable of operating effectively under uncertainty. This research contributes to the literature in several important ways. First, it reinforces the conceptualization of supply chain management as a sequential decision-making problem, emphasizing the relevance of Markov Decision Processes in modeling complex supply chain dynamics. Second, it extends the application of reinforcement learning beyond isolated supply chain functions by highlighting its potential for integrated and system-wide optimization. Third, the study provides empirical evidence that learning-based approaches can outperform traditional optimization methods in environments characterized by uncertainty and change. These contributions support the growing body of research that positions artificial intelligence, and reinforcement learning in particular, as a key enabler of intelligent and autonomous supply chain systems. The practical implications of this study are significant. For supply chain managers, reinforcement learning offers a tool for developing adaptive decision policies that reduce dependence on rigid planning assumptions. The framework can be applied to various operational contexts, including inventory management, transportation planning, and integrated logistics coordination. Moreover, once trained, reinforcement learning policies can be deployed for real-time decision-making with minimal computational overhead, making them suitable for operational use. However, successful implementation requires careful consideration of data quality, system design, and reward specification. Organizations must ensure reliable data collection and monitoring mechanisms to support learning-based decision-making. Additionally, managers should view reinforcement learning as a complement to, rather than a replacement for, existing decision-support tools, particularly in the early stages of adoption. Despite its contributions, this study has several limitations that should be acknowledged. First, the experimental evaluation is based on simulated supply chain environments. While simulation allows for controlled analysis and repeatability, real-world supply chains may exhibit additional complexities that are difficult to capture fully. Second, the performance of reinforcement learning depends heavily on the design of the reward function, which may not always align perfectly with managerial objectives. Third, the current framework primarily focuses on single-agent decision-making, whereas real-world supply chains often involve multiple autonomous stakeholders with potentially conflicting goals. Future research should focus on real-world validation through industrial pilot studies. Additionally, hybrid architectures combining DRL with classical optimization constraints may improve interpretability and stability. Extending the framework to incorporate demand forecasting modules and integrating digital twin environments could further enhance autonomy and robustness.

Recognizing these limitations provides important context for interpreting the results and highlights opportunities for further research. The results suggest that DRL can serve as the decision-making engine within digital supply chain twins. Continuous interaction between simulated and physical systems supports adaptive control.

In conclusion, this study underscores the transformative potential of reinforcement learning for supply chain management. By enabling adaptive, data-driven, and autonomous decision-making, reinforcement learning represents a key step toward truly intelligent supply chains. While challenges remain, continued advances in learning algorithms, data infrastructure, and system integration are expected to further accelerate the adoption of reinforcement learning in supply chain practice. This research provides a foundation for future studies and offers a clear pathway for advancing intelligent and resilient supply chain systems. Future studies should investigate advanced multi-agent reinforcement learning techniques, such as centralized critics and communication protocols, to improve convergence in large-scale networks. Incorporating real-time information sharing and partial observability into the modeling framework represents another promising direction. Further exploration of distributional reinforcement learning and stochastic dominance constraints may enhance tail-risk modeling. Future work may also consider scenario-based stress testing under global disruptions such as geopolitical shocks or pandemics. Integrating explainable AI (XAI) techniques and attention-based architectures could enhance transparency. Developing hybrid systems that combine interpretable constraints with learning-based adaptation represents a promising avenue.

6. References

1. Singh J, Rao S, George VJ, Gupta N, Goyal S, Gupta V. Leveraging Blockchain for Enhanced Transparency and Trust in Global Supply Chains: A Comprehensive Review. Empowering Business Through Technology: Innovations Shaping Our Future, 2026, 987-996.
2. Zabraoui O, Chafi A, Alami SK. AI-driven optimization of vehicle routing problems in supply chain: Integrating graph neural networks and reinforcement learning. Flexible Services and Manufacturing Journal, 2026, 1-44.
3. Dindarian A. Harnessing AI and Big Data to Build a Resilient Supply Chain: An Overview. Emerging Technologies in Supply Chains: Towards Digital Transformation. 2026; 233.
4. Ahadzadeh R, Dehghani E, Ghasemi P. Towards sustainable closed-loop photovoltaic supply chains: A hybrid framework integrating reinforcement learning and mathematical models. Expert Systems with Applications. 2026; 296:129182.
5. Awad A, Alahmari D. Agentic Control Towers: Multi Agent LLM RL Orchestration for Autonomous and Resilient Supply Chains. In Accelerating Logistics Through Generative AI, Digital Twins, and Autonomous Operations. IGI Global Scientific Publishing, 2026, 161-182.
6. Singh R, Singh J, Singh A, Bawa SS, Girdher K. Reducing Food Wastage Through Accurate Demand Prediction Using Generative AI. In Generative AI in Food Systems: Predictive Demand, Smart Supply Chains, and Sustainable Service Futures. IGI Global Scientific Publishing, 2026, 109-128.
7. Ghahri A, Shahamat BK, Lou MRI. Developing a Predictive Model of Organizational Learning Based on Multiple Intelligences: A Case Study of Iranian Schools. Future of Work and Digital Management Journal, 2026, 1-15.

8. Jayasinghe PSK, Kelly T, Madhavika N, Ehalapitiya S, De Costa DS, Liyanage MK, *et al*. Unveiling the robustness of apparel exporters' supply chains: Exploring the influence of IT capability, supply chain collaborations and government support. The International Journal of Logistics Management, 2026, 1-33.
9. Etxeberria A, Rodriguez-Muguruza A. Reframing the significance of menstruation: Evolutionary insights from an organismal-relational perspective. History and Philosophy of the Life Sciences. 2026; 48(1):2.
10. Kulkarni A. Supply Chain Optimization Using AI and SAP HANA: A Review. International Journal of Research Radicals in Multidisciplinary Fields, 2023, 51-57.
11. Zhou H, Ning W, Guo T. A fractal deep learning model for optimizing e-commerce warehousing and logistics. Journal of Industrial and Management Optimization. 2026; 22(1):374-404.
12. Hove-Sibanda P, Pooe RD. Enhancing supply chain performance through supply chain practices. Journal of Transport and Supply Chain Management. 2018; 12(1):1-13.
13. Firos A, Khanum S. Fuzzy AI for Smart Decision-Making in Supply Chain Management: An Industry 5.0 Perspective. In Diversity and Inclusion-Driven Marketing for Multicultural Marketplaces. IGI Global Scientific Publishing, 2026, 171-188.
14. Fernando E, Ikhsan RB. Exploring Integration AI-Powered in Supply Chain Management. In 2024 International Conference on Decision Aid Sciences and Applications (DASA). IEEE, December 2024, 1-5.
15. Vipinchandran D, Jayashree J. Reinforcement-Enhanced LSTM for Demand Forecasting in Agricultural Supply Chain Management. In Strategic Decision-Making in Dynamic Business Environments: Systems and Control Perspectives. Cham: Springer Nature Switzerland, 2026, 1089-1099.
16. Shahzadi G, Jia F, Chen L, John A. AI adoption in supply chain management: A systematic literature review. Journal of Manufacturing Technology Management. 2024; 35(6):1125-1150.
17. Zhou S, Zhao S, Ren Z. Loosely Synchronized Rule-Based Planning for Multi-Agent Path Finding with Asynchronous Actions. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 14), April 2025, 14763-14770.
18. Maulana MK, Dwiputri RM, Harri M. Advancing sustainability in culinary: Examining the role of halal supply chain management and operational efficiency. European Journal of Sustainable Development Research. 2026; 10(1).
19. Liew N, Pattnaik S, Kures AO, Park K, Pinsky E. Transforming Global Supply Chains with Artificial Intelligence, Machine Learning, and Next-Generation Technologies. In Next Generation Entrepreneurship: Convergence of Innovation, Technology, and Society. Cham: Springer Nature Switzerland, 2026, 207-234.
20. Mojica A, Andrés B, De La Torre R. For Strategic Decision-Making in Supply Chain. In Data Science, Challenges and Applications in Industrial Operations: Proceedings of the 19th International Conference on Industrial Engineering and Industrial Management (ICIEIM)-XXIX Congreso de Ingeniería de Organización (CIO2025). Springer Nature, 2026, p. 147.
21. Ahmed K, Ramirez C, Williams S, Olivares M, Wong S, Mecusker E, *et al*. Results Surveying Supportive Medicine's integration into an NCI Cancer Center's ICU. Journal of Pain and Symptom Management. 2025; 69(5):e715.
22. Kogler C, Maxera P. A literature review of supply chain analyses integrating discrete simulation modelling and machine learning. Journal of Simulation. 2026; 20(1):110-134.