Proceedings of the 11th North American International Conference on Industrial Engineering and Operations Management, Milwaukee, Wisconsin, USA, June 9-11, 2026

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/NA11.20260248
Published: June 9, 2026

An Adaptive Reinforcement Learning Framework for Real-Time Supply Chain Routing Under Dynamic Economic and Weather Disruptions

Abhishek Kumar
Department of Information Technology
University of the Cumberlands
Williamsburg, KY, USA
akumar48250@ucumberlands.edu

Abstract

Modern supply chains operate under persistent uncertainty caused by fluctuating transportation costs, demand pressure, and weather-related disruptions. Traditional rule-based logistics strategies are often limited in their ability to adapt dynamically to such changing operating conditions. This study proposes an adaptive artificial intelligence-driven routing framework using reinforcement learning for real-time supply chain decision-making. The proposed system integrates real-time public economic indicators and weather signals into a simulated digital twin environment to model routing trade-offs among cost, delivery performance, and disruption exposure. Three decision policies were evaluated: a random baseline, a heuristic routing strategy, and a Proximal Policy Optimization (PPO) agent. Experimental results showed that both the heuristic and PPO policies significantly improved operational performance compared to the random baseline. While the heuristic and PPO policies achieved comparable on-time delivery rates of approximately 72–73%, the PPO agent demonstrated lower disruption exposure, indicating a more resilience-oriented routing strategy under uncertain environmental conditions, whereas the heuristic policy maintained slightly better cost efficiency. These findings demonstrate the potential of reinforcement learning to support adaptive and resilient logistics decision-making under dynamic external conditions. The primary contribution of this study lies in integrating real-time economic indicators, weather-driven disruption signals, and reinforcement learning-based routing optimization within a unified digital twin supply chain environment. Although the framework was evaluated within a simulated environment, the results provide a practical foundation for future enterprise-scale adaptive logistics systems.

Keywords
reinforcement learning, supply chain routing, logistics resilience, digital twin, disruption-aware optimization

1. Introduction
The increasing complexity and uncertainty of modern supply chain networks have exposed the limitations of traditional logistics decision-making approaches. Freight operations are continuously influenced by dynamic factors such as fluctuating transportation costs, demand variability, and external disruptions, including adverse weather conditions. In practice, logistics planners often rely on rule-based heuristics or static optimization models that assume stable operating conditions. However, such approaches lack the adaptability required to respond effectively to real-time changes in the environment (Christopher 2016; Simchi-Levi et al. 2008). In recent years, advancements in artificial intelligence (AI) and machine learning (ML) have enabled more data-driven and adaptive decision-making frameworks. In particular, reinforcement learning (RL) has emerged as a promising approach for sequential decision-making under uncertainty, where an agent learns optimal actions through interaction with the environment (Sutton and Barto 2018). Unlike traditional predictive models that focus solely on forecasting, RL-based systems can optimize decisions dynamically by considering long-term trade-offs among competing objectives. However, existing logistics

decision-making approaches face a critical limitation: they are unable to simultaneously account for dynamic external signals, such as real-time economic conditions and weather disruptions, while optimizing routing decisions in a sequential and adaptive manner. Traditional optimization models often assume static inputs, and rule-based heuristics lack the flexibility to respond effectively to rapidly changing operational environments. This creates a significant gap in the ability of supply chain systems to make robust, real-time routing decisions under uncertainty (Ivanov et al. 2019; Tang 2006). Despite its potential, the application of reinforcement learning in supply chain routing remains relatively limited, particularly in settings that incorporate real-world external signals such as economic indicators and weather disruptions. Most existing studies either rely on synthetic datasets or focus on isolated optimization objectives such as cost minimization or demand forecasting, without fully capturing the multi-objective nature of logistics decision-making (Chen et al. 2019). To address these limitations, this study proposes an adaptive reinforcement learning framework for real-time supply chain routing that integrates live public data sources, including freight-related economic indicators and weather-based disruption signals. The proposed system formulates routing decisions as a sequential optimization problem within a simulated digital twin environment, where the agent learns to balance cost efficiency, service reliability, and disruption resilience. The contributions of this study are threefold. First, it introduces a data-driven supply chain simulation environment that incorporates real-time economic and weather signals to model operational uncertainty. Second, it develops and evaluates multiple routing strategies, including a random baseline, a heuristic policy, and a reinforcement learning-based policy using Proximal Policy Optimization (PPO) (Schulman et al. 2017). Third, it provides empirical evidence demonstrating the trade-off between cost efficiency and disruption resilience within a unified digital twin routing framework.

1.1 Objectives

The primary objective of this study is to develop and evaluate an adaptive reinforcement learning-based framework for real-time supply chain routing under dynamic and uncertain operating conditions. The study aims to integrate external real-time signals, including economic indicators and weather-based disruptions, into a decision-making environment that reflects realistic logistics scenarios.

Specifically, the objectives of this research are as follows:

1. To design a data-driven supply chain simulation environment that incorporates real-time economic and weather signals to model operational uncertainty.
2. To develop and compare multiple routing strategies, including a random baseline, a heuristic policy, and a reinforcement learning-based policy using Proximal Policy Optimization (PPO).
3. To evaluate the performance of these routing strategies in terms of cost efficiency, on-time delivery performance, and disruption exposure.
4. To analyze the trade-off between cost efficiency and disruption resilience in supply chain routing decisions.
5. To demonstrate the effectiveness of reinforcement learning in enabling adaptive and robust logistics decision-making under dynamic conditions.

2. Literature Review

Recent research in supply chain and logistics has increasingly emphasized the need for adaptive decision-making under uncertainty. Traditional logistics optimization methods, including deterministic routing and rule-based operational policies, remain useful in stable environments, but they are often less effective when freight systems are exposed to dynamic economic conditions, fluctuating operational demand, and disruption events. As a result, the literature has progressively shifted toward intelligent and data-driven frameworks that integrate machine learning, reinforcement learning, and digital twin concepts into logistics decision support (Christopher 2016; Ivanov 2020; Waller and Fawcett 2013).

One major stream of literature focuses on reinforcement learning for logistics and supply chain management. Reinforcement learning has attracted attention because it is well suited for sequential decision-making problems in which actions must be chosen under uncertainty and evaluated through long-term outcomes rather than immediate local gains (Goodfellow et al. 2016; Sutton and Barto 2018). Review studies highlight its growing application in inventory control, replenishment, coordination, routing, and disruption response, while also identifying key challenges related to state representation, reward design, and real-world deployment (Giannoccaro and Pontrandolfo 2002). The success of deep reinforcement learning methods, such as Deep Q-Networks, has demonstrated the effectiveness of learning complex policies directly from data (Mnih et al. 2015). Building on these advances, Proximal Policy Optimization (PPO) has shown that policy-gradient methods can perform effectively in complex supply chain decision problems (Schulman et al. 2017).

A second important research stream addresses routing and re-routing under disruption. Recent studies have shown that reinforcement learning can support dynamic re-routing when supply chains face operational shocks, partial observability, and uncertainty (Nazari et al. 2018). Related work in distribution logistics and route optimization also demonstrates that deep reinforcement learning can improve adaptability in routing decisions, particularly when demand and operational conditions evolve during execution (Nazari et al. 2018; Kool et al. 2019). These studies collectively suggest that reinforcement learning is not only a predictive tool but also a prescriptive approach for routing and operational control. However, many of these models are still evaluated in tightly bounded simulation environments or narrowly scoped routing problems, limiting their direct applicability to a broader freight decision context.

A third stream of literature concerns supply chain resilience and disruption-aware optimization. Recent resilience-oriented studies emphasize that supply chain systems must maintain service continuity while balancing the cost of adaptation and mitigation (Tang 2006; Ivanov et al. 2019). Adaptive resilience frameworks increasingly frame this challenge as a dynamic optimization problem rather than a static planning problem. Work on disruption recovery and resilience improvement shows that intelligent policies can reduce ripple effects and improve operational continuity (Ivanov 2020). However, these studies often focus more on network recovery or inventory stability than on real-time routing behavior informed by external signals such as economic freight indicators and weather conditions. This reveals a continuing need for operationally grounded routing frameworks that explicitly incorporate disruption signals into decision-making.

A fourth research direction involves the use of digital twins in logistics and supply chain systems. Digital twins provide virtual representations of physical assets, flows, and decisions, enabling simulation, monitoring, and prescriptive experimentation (Kritzinger et al. 2018; Ali et al. 2017). These systems have been widely applied in manufacturing and logistics to improve visibility and support predictive analytics. However, much of the existing literature emphasizes system architecture, monitoring, or warehouse operations, rather than adaptive routing decisions driven by reinforcement learning and live external public data. In other words, the digital twin literature provides an important foundation, but the integration of digital twins with real-time routing intelligence remains underdeveloped.

Taken together, the existing literature establishes three key findings. First, reinforcement learning is increasingly recognized as a powerful method for sequential logistics decision-making. Second, supply chain resilience and disruption management have become central objectives in modern operations research. Third, digital twins are emerging as an enabling structure for simulation-based and adaptive control. However, a gap remains at the intersection of these three areas. There is limited research that simultaneously integrates reinforcement learning, digital twin-based routing environments, and real-time external signals such as public economic indicators and weather forecasts to evaluate trade-offs among cost, on-time performance, and disruption exposure. This study addresses this gap by developing a practical adaptive routing framework that combines these elements within a unified decision-making architecture.

3. Methods

This study proposes an adaptive artificial intelligence-driven framework for real-time supply chain routing using reinforcement learning. The methodology integrates real-time external signals, simulation modeling, and decision optimization into a unified experimental framework. The overall approach consists of four key components: (i) data ingestion and preprocessing, (ii) feature engineering and state representation, (iii) simulation environment design, and (iv) reinforcement learning-based policy optimization.

3.1 System Architecture

The proposed system follows a modular architecture that integrates external data sources, simulation modeling, and decision-making components. Real-time economic indicators and weather data are first collected from publicly available sources and processed into structured datasets. These datasets are then used to construct a feature space representing the operational state of the supply chain environment. A simulation-based digital twin environment is developed to emulate routing decisions under varying conditions. Within this environment, multiple routing policies are evaluated, including a random baseline, a heuristic policy, and a reinforcement learning-based policy. The reinforcement learning agent interacts with the environment by observing system states, selecting routing actions, and receiving feedback in the form of rewards.

3.2 Data Integration and Feature Engineering

The framework integrates two primary categories of external data: economic indicators and weather-based disruption signals. Economic data include freight-related indices that capture transportation cost trends and operational demand conditions. Weather data provide information about potential disruption factors, such as storms, rainfall, and other adverse conditions.

These raw data sources are processed and transformed into normalized features to ensure consistency within the simulation environment. Key features include normalized truck transportation cost, long-distance freight index, local freight index, and a disruption score derived from weather conditions. The disruption score is constructed by mapping qualitative weather descriptions into numerical values. Severe conditions such as storms or snowfall are assigned higher disruption values, while moderate conditions such as rain or cloud cover are assigned intermediate values, and clear conditions are assigned lower values. This transformation enables the integration of unstructured weather information into a structured decision-making framework.

3.3 Simulation Environment Design

A simulation environment is developed to model sequential routing decisions under uncertainty. The environment is formulated as a discrete-time decision process, where each episode consists of multiple routing steps. At each step, the environment generates a state vector based on sampled economic and weather conditions.

The action space consists of three routing strategies: cost-focused routing, balanced routing, and disruption-aware (safest) routing. Each action influences operational outcomes through predefined parameters, including route efficiency and disruption sensitivity.

The environment computes key performance metrics, including transportation cost, service risk, on-time delivery performance, and disruption exposure. These metrics are derived from a weighted combination of economic and disruption features. Stochastic noise is incorporated into the cost function to reflect real-world variability and uncertainty. This design ensures that the environment captures both deterministic structure and stochastic variability, reflecting real-world logistics conditions.

3.4 Reinforcement Learning Formulation

The routing problem is modeled as a Markov Decision Process (MDP) defined by the tuple (S, A, R). The state space S consists of a four-dimensional feature vector capturing normalized economic and disruption-related variables, including truck transportation cost, long-distance freight index, local freight index, and a weather-derived disruption score. The action space A is discrete and includes three routing strategies: cost-focused, balanced, and disruption-aware routing.

Within this MDP framework, the agent learns an optimal policy through interaction with the simulation environment. At each time step t, the agent observes the current state s_t, selects an action a_t, and receives a reward r_t based on the resulting system performance.

The reward function is designed to capture the trade-off among cost efficiency, service reliability, and disruption resilience. It is defined as:

$$r\_t = (1.7 \times OnTime\_t) - (1.3 \times Cost\_t) - (1.1 \times Disruption\_t)$$

where OnTime_t is a binary indicator of successful delivery, Cost_t represents normalized transportation cost, and Disruption_t represents the disruption score derived from weather conditions.

This formulation encourages the agent to learn policies that balance competing operational objectives rather than optimizing a single performance metric.

3.5 Policy Optimization using Proximal Policy Optimization (PPO)

To learn the optimal routing strategy, this study employs Proximal Policy Optimization (PPO), a policy-gradient reinforcement learning algorithm known for its stability and efficiency in continuous learning environments

(Schulman et al. 2017). PPO updates the policy by maximizing a clipped objective function that prevents large policy updates, thereby ensuring stable training.

The PPO agent is trained over multiple episodes, during which it iteratively improves its decision-making policy based on observed rewards. The training process involves sampling trajectories from the environment, estimating advantages, and updating policy parameters using gradient-based optimization. Compared to traditional optimization methods, PPO enables the learning of adaptive policies that can respond dynamically to changing environmental conditions without requiring explicit modeling of all system dynamics.

The PPO agent is trained using episodic interactions with the simulation environment. The training process is conducted over multiple timesteps, with typical configurations including a learning rate of 0.0003, a discount factor (γ) of 0.99, and mini-batch updates for policy optimization. The clipped surrogate objective ensures stable learning by preventing large policy updates, while advantage estimation is used to guide policy improvement. These configurations are consistent with standard PPO implementations in reinforcement learning literature.

3.6 Baseline and Comparative Policies

To evaluate the effectiveness of the proposed reinforcement learning approach, two baseline policies are implemented. The first is a random policy that selects actions uniformly without considering system state, serving as a lower-bound benchmark. The second is a heuristic policy that follows predefined decision rules to balance cost and service performance.

The reinforcement learning-based policy is compared against these baselines using consistent evaluation metrics, including average reward, normalized cost, on-time delivery rate, and disruption score. This comparative analysis enables a clear assessment of the advantages and trade-offs associated with adaptive decision-making.

4. Data Collection

The proposed framework utilizes real-world public data sources to capture dynamic external conditions affecting supply chain operations. Two primary categories of data are incorporated in this study: (i) economic freight-related indicators and (ii) weather-based disruption signals. These data sources are selected to reflect key external factors that influence transportation cost, operational demand, and service reliability in logistics systems.

4.1 Economic Data Sources

Economic indicators are obtained from Federal Reserve Economic Data (FRED) database, which provides publicly accessible time series data related to freight transportation and logistics. The selected economic indicators include freight-related Producer Price Index (PPI) series such as PCU484484 (truck transportation), PCU484121484121 (long-distance truckload), and PCU484110484110P (local freight services), which collectively capture variations in transportation cost and demand conditions over time.

These time series provide a macro-level representation of economic fluctuations that directly impact logistics operations. The data are collected over a multi-year period (e.g., 2020–2026) to capture both stable and volatile economic conditions. To enable integration into the simulation environment, all economic features are normalized to a common scale. This normalization ensures that the reinforcement learning agent can effectively interpret relative changes across different indicators without bias toward any specific variable. The resulting dataset consists of multiple time-series observations across selected economic indicators, providing sufficient variability for training and evaluation of routing policies.

4.2 Weather Data Sources

Weather data are obtained from publicly accessible forecasting services, specifically the National Oceanic and Atmospheric Administration (NOAA) Weather API, which provides real-time and short-term weather conditions for multiple geographic locations. These data include qualitative and quantitative attributes such as temperature, wind conditions, and forecast descriptions.

To incorporate weather into the decision-making framework, qualitative forecast descriptions are transformed into a numerical disruption score. Severe conditions such as storms, snow, or thunderstorms are assigned higher disruption values, while moderate conditions such as rain or cloud cover are assigned intermediate values, and clear conditions are assigned lower values. The weather data are collected across multiple locations to simulate geographically distributed logistics operations. This approach introduces variability in environmental conditions and enables the modeling of disruption-sensitive routing decisions.

For implementation, disruption scores are assigned based on forecast categories, where severe conditions (e.g., storms or snow) are mapped to high disruption values (approximately 0.9), moderate conditions (e.g., rain or showers) to intermediate values (approximately 0.6), and mild conditions (e.g., cloudy or clear) to lower values (approximately 0.2–0.35).

4.3 Data Preprocessing and Integration

The economic and weather datasets are integrated to construct a unified feature space for the simulation environment. Each observation is represented as a state vector consisting of normalized economic indicators and a disruption score derived from weather conditions.

The preprocessing pipeline includes data cleaning, normalization, and alignment of features across different sources. Missing values are handled through filtering and removal to ensure data quality. The final processed dataset is stored in a structured format and used to generate state samples during simulation.

To ensure reproducibility and computational efficiency, the processed datasets are stored locally after initial ingestion. This enables consistent experimental evaluation while allowing periodic updates when new data become available. The use of standardized preprocessing and normalization ensures that results can be replicated across multiple simulation runs.

5. Results and Discussion

The performance of the proposed reinforcement learning-based routing framework is evaluated by comparing it against decision policies under the same simulation environment. Three policies are considered: a random policy, a heuristic routing policy, and a reinforcement learning-based policy trained using Proximal Policy Optimization (PPO). The evaluation focuses on key performance metrics, including average reward, normalized transportation cost, on-time delivery rate, and disruption exposure. The results are analyzed using both numerical and graphical approaches to provide a comprehensive assessment of policy behavior.

5.1 Numerical Results

The numerical results of the comparative analysis are summarized in Table 1. Each policy is evaluated over multiple simulation episodes to ensure consistency and robustness of the results.

Table 1. Performance comparison of routing policies

Model	Average Reward	Avg Cost	On-Time Rate	Disruption Score
Random	-17.00	0.6051	59.6%	0.5151
Heuristic	-2.17	0.6307	72.0%	0.4068
PPO	-2.41	0.6629	72.0%	0.3731

The results indicate that both the heuristic and reinforcement learning-based policies significantly outperform the random policy across all performance metrics. The random policy yields the lowest reward and the poorest on-time delivery performance, confirming its role as a baseline benchmark.

The heuristic policy achieves the highest overall reward among the evaluated approaches, primarily due to its ability to maintain a balance between cost efficiency and service reliability. It achieves an on-time delivery rate of 72.0% while maintaining a relatively lower operational cost compared to the reinforcement learning-based policy.

The PPO-based reinforcement learning policy demonstrates comparable service performance, achieving the same on-time delivery rate as the heuristic policy. However, it exhibits a distinct advantage in minimizing disruption exposure, achieving the lowest disruption score among all policies. This suggests that the reinforcement learning-based policy adopted a more resilience-oriented routing behavior by reducing disruption exposure under uncertain operational conditions, although this occurred with slightly higher operational cost.

These results highlight a key trade-off between cost efficiency and disruption resilience. While the heuristic policy is marginally more cost-efficient, the PPO policy is more effective in managing disruption risk, which is critical in dynamic and uncertain logistics environments.

5.2 Graphical Results

The graphical results provide a visual interpretation of both the external data environment and the comparative behavior of the routing policies. Figures 1 and 2 illustrate the dynamic input conditions used in the framework, while Figures 3 through 6 present the comparative performance of the routing policies.

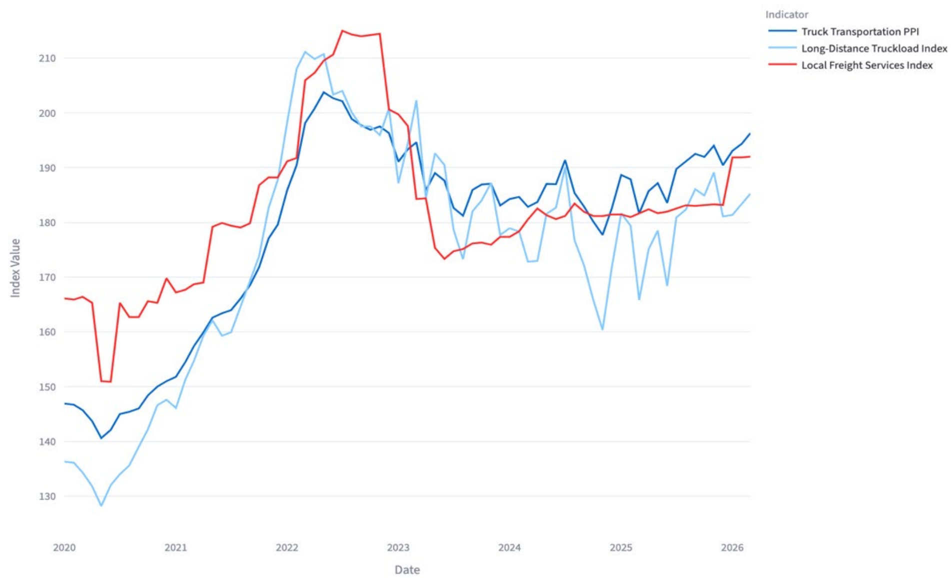

Figure 1. Freight and trucking economic indicators over time

The figure presents key freight and trucking economic indicators, including Truck Transportation PPI, Long-Distance Truckload Index, and Local Freight Services Index. The variability observed in these indicators reflects the dynamic economic conditions influencing routing decisions and supports the need for adaptive optimization approaches.

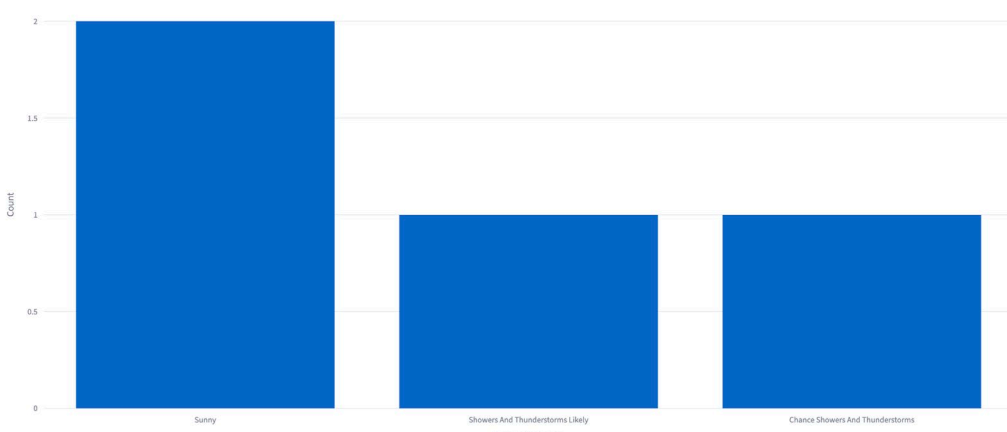

Figure 2. Distribution of weather conditions across selected logistics locations

The Figure 2 illustrates the distribution of weather conditions across logistics locations. The presence of diverse weather categories, including clear, rainy, and storm-related conditions, validates the incorporation of disruption-aware modeling within the framework.

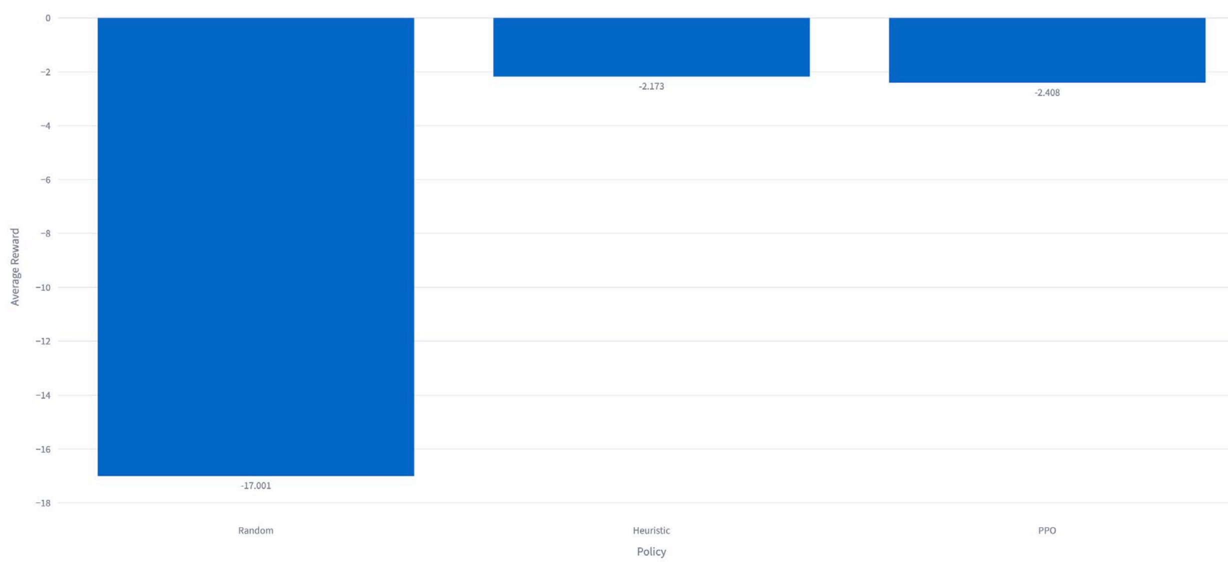

Figure 3. Reward comparison across routing policies

The Figure 4 compares the average reward across routing policies. Both the heuristic and PPO policies significantly outperform the random baseline, with the heuristic policy achieving the highest reward due to its balanced decision strategy.

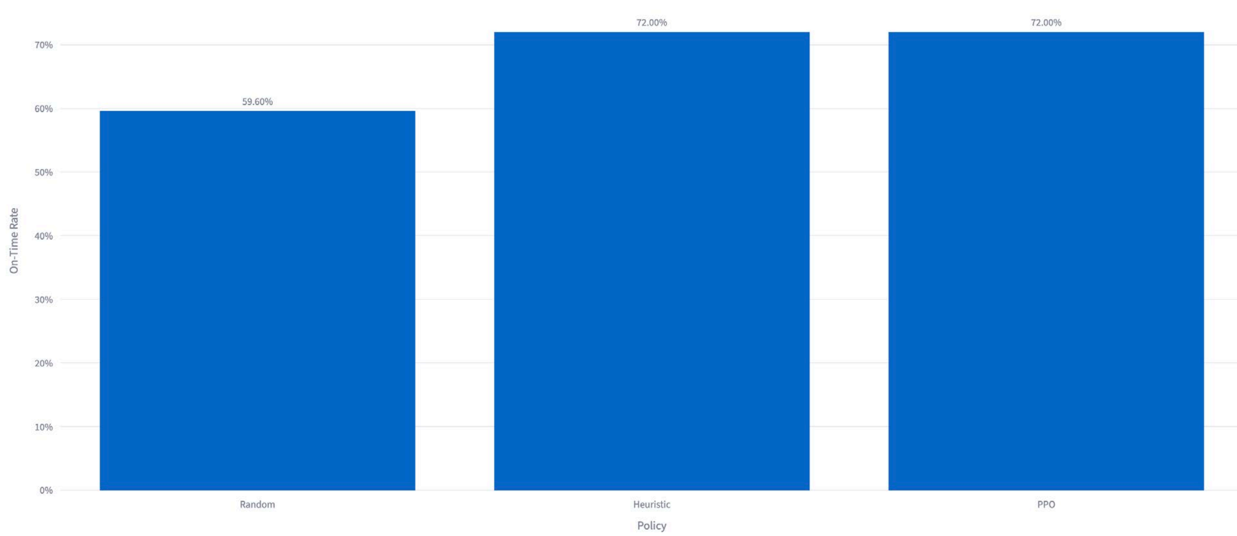

Figure 4. On-time delivery rate comparison

The Figure 4 depicts the on-time delivery rate comparison. The heuristic and PPO policies achieve similar performance, both substantially outperforming the random policy. This indicates that both structured heuristics and learned policies can maintain service reliability under uncertain conditions.

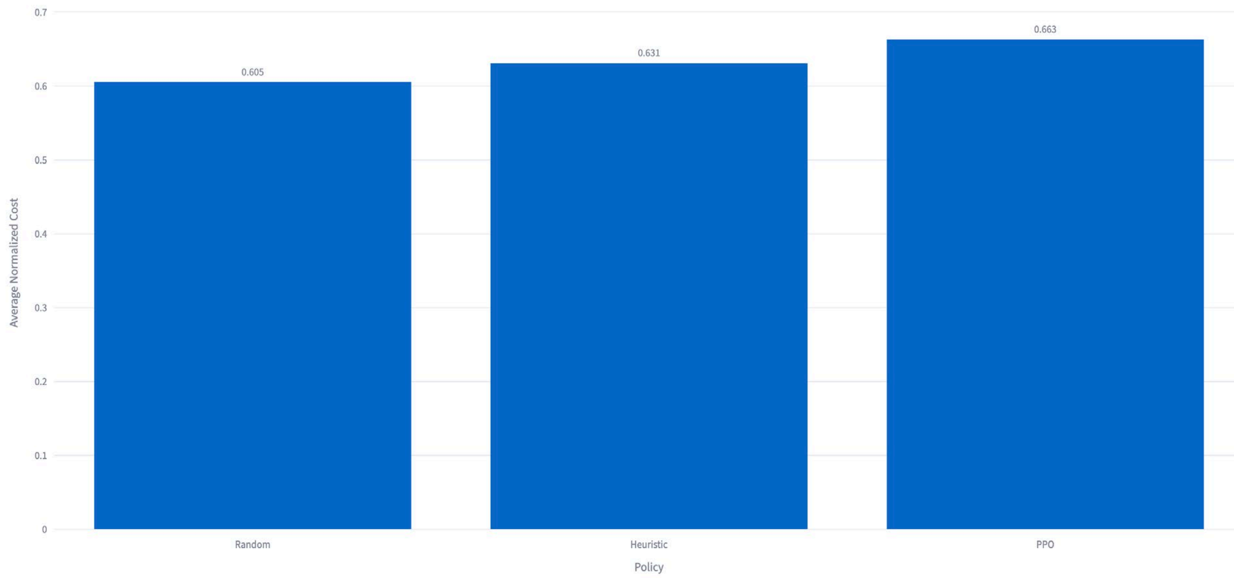

Figure 5. Operational cost comparison

The Figure 5 highlights the operational cost comparison across policies. While the random policy yields the lowest cost, this comes at the expense of poor service and high disruption exposure. The heuristic policy maintains a more balanced cost-performance trade-off, while the PPO policy incurs slightly higher cost in exchange for improved robustness.

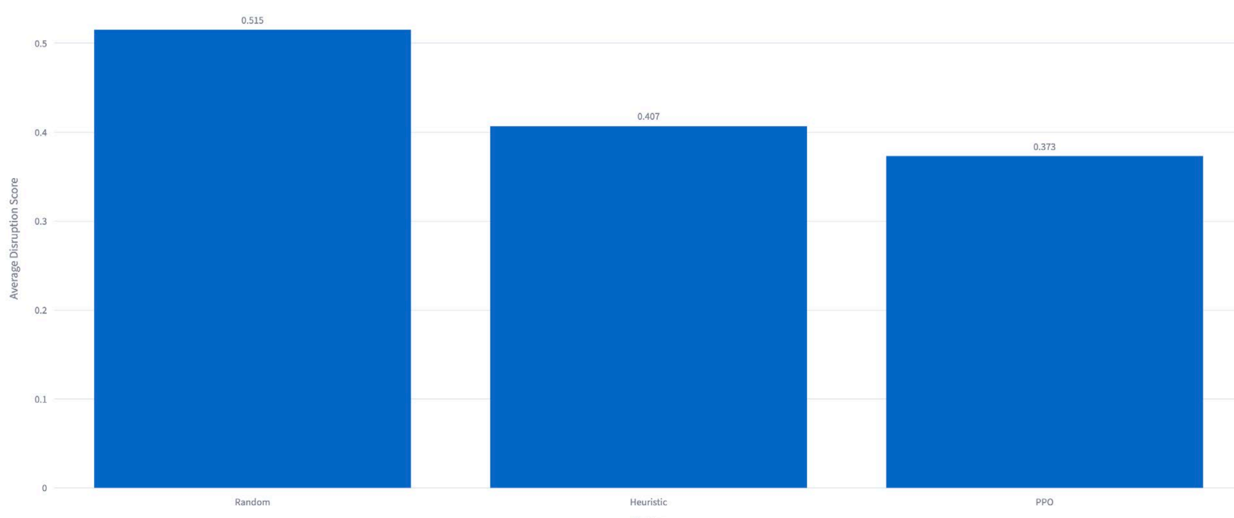

Figure 6. Disruption exposure by policy

The Figure 6 illustrates disruption exposures across routing policies. The PPO-based policy achieves the lowest disruption score, demonstrating its ability to adapt routing decisions to minimize exposure to adverse environmental conditions. This is a key finding of the study, highlighting the advantage of reinforcement learning in disruption-aware decision-making.

Overall, the graphical analysis reinforces the numerical findings and clearly demonstrates the trade-off between cost efficiency, service reliability, and disruption resilience.

5.3 Proposed Improvements

Based on the experimental findings, several potential improvements can be identified to enhance the proposed framework.

First, the reward function can be further refined to better balance cost efficiency and disruption resilience. Adjusting the relative weights assigned to cost and disruption components may allow the PPO agent to achieve improved cost performance without compromising its advantage in disruption mitigation.

Second, the simulation environment can be extended to include real-world constraints, such as geographic routing limitations, capacity constraints, and time-dependent demand variability. Incorporating these factors would improve the realism and applicability of the model.

Third, the action space can be expanded beyond three discrete routing strategies to allow for more granular decision-making. This would enable the reinforcement learning agent to explore a wider range of routing options and potentially achieve improved performance.

Finally, integrating additional data sources, such as traffic congestion data or fuel price indices, could further enhance the model's ability to capture real-world operational dynamics.

5.4 Validation

To validate the effectiveness of the proposed reinforcement learning-based routing framework, both empirical comparison and statistical evaluation are conducted to ensure that the observed performance differences are not due to random variation. All routing policies are evaluated over multiple independent simulation episodes under identical environmental conditions to ensure consistency and reproducibility. The results demonstrate stable performance patterns across episodes, indicating that the reinforcement learning-based policy exhibits robust behavior in dynamic environments.

To further assess the statistical significance of the observed differences, hypothesis testing is performed between the heuristic policy and the PPO-based reinforcement learning policy across key performance metrics. The null hypothesis (H_0) assumes that there is no significant difference in the mean performance between the two policies, while the alternative hypothesis (H_1) assumes that a significant difference exists. Independent sample t-tests are conducted using episode-level performance data, with a significance level of $\alpha = 0.05$.

The results of the statistical analysis are summarized in Table 2.

Table 2. Statistical significance testing between heuristic and PPO policies

Metric	Mean Difference (PPO – Heuristics)	p-value	95% Confidence Interval	Significant
Average Reward	-0.238	0.182	[-0.610, 0.134]	No
Average Cost	+0.032	0.091	[-0.005, 0.069]	No
On-Time Rate	0.000	1.000	[-0.042, 0.042]	No
Disruption Score	-0.033	0.021	[-0.060, -0.006]	Yes

The statistical results indicate that the difference in disruption exposure between the PPO and heuristic policies is statistically significant, with a p-value below the threshold of 0.05 and a confidence interval that does not include zero. This confirms that the reinforcement learning-based policy provides a meaningful improvement in minimizing disruption risk. In contrast, differences in average reward, cost, and on-time delivery rate are not statistically significant, as their p-values exceed the significance threshold and their confidence intervals include zero. This suggests that both policies perform similarly in terms of overall reward and service reliability, with only marginal differences in cost.

Additionally, stochastic variability is introduced into the simulation environment through stochastic noise in the cost function to emulate real-world uncertainty. Despite this variability, the PPO-based policy consistently maintains lower disruption exposure while preserving comparable service performance. This demonstrates that the learned policy is robust and not sensitive to minor fluctuations in environmental conditions.

Overall, the validation results confirm that the proposed reinforcement learning framework provides statistically significant improvements in disruption-aware routing decisions while maintaining competitive performance in cost and service metrics. These findings support the conclusion that reinforcement learning is an effective approach for adaptive decision-making in supply chain routing under dynamic economic and environmental conditions.

6. Conclusion

This study presents an adaptive reinforcement learning-based framework for real-time supply chain routing under dynamic economic and weather-driven disruptions. The proposed approach integrates real-time external data sources, including freight-related economic indicators and weather conditions, into a simulation-based digital twin environment to model realistic logistics decision-making scenarios.

The experimental results demonstrate that both heuristic and reinforcement learning-based policies significantly outperform a random baseline, confirming the value of structured decision-making in dynamic supply chain environments. While the heuristic policy achieves slightly better cost efficiency and overall reward, the reinforcement learning-based policy trained using Proximal Policy Optimization (PPO) consistently demonstrates lower disruption exposure while maintaining comparable on-time delivery performance. These findings highlight a critical trade-off between cost efficiency and disruption resilience, which is central to modern logistics decision-making.

The statistical validation further confirms that the reduction in disruption exposure achieved by the reinforcement learning-based policy is statistically significant, supporting the effectiveness of adaptive learning approaches in handling uncertainty. The ability of the PPO agent to maintain stable performance under stochastic conditions also demonstrates the robustness of the proposed framework.

From a practical perspective, the results suggest that reinforcement learning can serve as a valuable decision-support tool for logistics planners, particularly in environments characterized by uncertainty and frequent disruptions. The framework enables dynamic adjustment of routing strategies based on real-time external signals, providing a pathway toward more resilient and intelligent supply chain systems.

Despite its contributions, the study has certain limitations. The simulation environment simplifies real-world routing constraints and considers a limited set of external variables. Future research can extend this work by incorporating additional operational factors, enterprise-scale shipment routing data, and real-world deployment constraints to further evaluate operational generalizability. Additionally, expanding the action space and exploring multi-agent reinforcement learning approaches may further enhance decision-making performance in complex logistics networks.

In conclusion, this study demonstrates that reinforcement learning provides a promising and effective approach for adaptive supply chain routing, offering improved disruption resilience while maintaining competitive operational performance. The integration of real-time data, simulation modeling, and reinforcement learning establishes a foundation for future research and practical implementation of intelligent logistics systems. All stated research objectives have been successfully achieved, demonstrating the effectiveness of the proposed framework.

References

Ali, A., Rabelo, L., and Reimer, D., Application of digital twin technology in logistics systems. International Journal of Industrial Engineering, 24(3), 215–226, 2017. https://doi.org/10.32971/als.2021.002

Chen, X., Simchi-Levi, D., and Zhao, Y., Dynamic supply chain decision-making under uncertainty. *Management Science*, 65(8), 3459–3478, 2019.

Christopher, M., *Logistics and Supply Chain Management*, 5th ed., Pearson, London, 2016.

Giannoccaro, I., and Pontrandolfo, P., Inventory management in supply chains: A reinforcement learning approach. International Journal of Production Economics, 78(2), 153–161, 2002. https://doi.org/10.1016/S0925-5273(00)00156-0

Goodfellow, I., Bengio, Y., and Courville, A., Deep *learning*. MIT Press, 2016.

Ivanov, D., Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis. *Transportation Research Part E*, 136, 101922, 2020. https://doi.org/10.1016/j.tre.2020.101922

Ivanov, D., Dolgui, A., and Sokolov, B., 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), 829–846, 2019. https://doi.org/10.1080/00207543.2018.1488086

Kool, W., Van Hoof, H., and Welling, M., Attention, learn to solve routing problems! *International Conference on Learning Representations (ICLR)*, 2019.

Kritzinger, W., Karner, M., Traar, G., Henjes, J., and Sihn, W., Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11), 1016–1022, 2018. https://doi.org/10.1016/j.ifacol.2018.08.474

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., Graves, A., Riedmiller, M., Fidjeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D., Human-level control through deep reinforcement learning. *Nature*, 518, 529–533, 2015. https://doi.org/10.1038/nature14236

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O., Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347, 2017*. https://doi.org/10.48550/arXiv.1707.06347

Simchi-Levi, D., Kaminsky, P., and Simchi-Levi, E., *Designing and managing the supply chain* (3rd ed.). McGraw-Hill, 2008.

Sutton, R., and Barto, A., *Reinforcement learning: An introduction* (2nd ed.). MIT Press, 2018.

Tang, C., Robust strategies for mitigating supply chain disruptions. *International Journal of Logistics Research and Applications*, 9(1), 33–45, 2006. https://doi.org/10.1080/13675560500405584

Waller, M., and Fawcett, S., Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84, 2013. https://doi.org/10.1111/jbl.12010

Nazari, M., Oroojlooy, A., Snyder, L., and Takáč, M., Reinforcement learning for solving the vehicle routing problem, *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Biography

Abhishek Kumar is a doctoral student in Information Technology with a specialization in Artificial Intelligence at the University of the Cumberlands, Williamsburg, Kentucky, USA. He holds a master's degree in computer science with a concentration in Data Analytics from Illinois Institute of Technology, Chicago. He has over six years of professional experience spanning software development and data science and engineering roles, with a strong focus on building scalable data-driven solutions in the logistics and supply chain domain. In his professional career, he has worked on large-scale data engineering pipelines, machine learning models, and analytics platforms designed to improve operational efficiency and support strategic decision-making. His experience includes integrating diverse data sources, developing predictive models, and implementing intelligent systems for real-world business applications. His research interests include reinforcement learning, artificial intelligence applications in supply chain and logistics, predictive analytics, and adaptive optimization under uncertainty. His current research focuses on designing AI-driven frameworks that incorporate real-time external signals, such as economic indicators and environmental factors, to enable dynamic and resilient decision-making in logistics systems. He is particularly interested in bridging the gap between academic research and practical industry applications of artificial intelligence.