Online ISSN: 3104-9338
Print ISSN: 3104-932X

Dynamic Optimization of AI-Enabled Supply Chains under Demand and Disruption Uncertainty: A Distributional Forecasting and Model-Predictive Control Approach

Rufeng He*

School of Business Administration, Guizhou University of Finance and Economics, China

Received: June 24, 2026

Revised: June 25, 2026

Accepted: July 1, 2026

Published online: July 5, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 2, No. 8 (August 2026)

* Corresponding Author: Rufeng He (1415476577@qq.com)

Abstract. Supply chains are sequential decision systems in which procurement, production, inventory positioning, allocation and expediting decisions are repeatedly adjusted under demand volatility, lead-time uncertainty and disruption risk. Artificial intelligence has improved the quality of forecasts and early-warning signals, yet prediction alone does not determine an operational policy. This study develops an artificial-intelligence-enabled dynamic optimization model for supply chain management. The model treats inventory, pipeline orders, backlog, capacity, supplier reliability and carbon exposure as evolving state variables. A distributional learning layer updates demand, lead-time and disruption beliefs from operational data, while a rolling-horizon stochastic model-predictive controller selects feasible actions under cost, service, resilience and sustainability constraints. safe policy-improvement mechanism based on Constrained Proximal Policy Optimization (C-PPO) is further introduced to tune policy parameters in a digital twin without relaxing managerial constraints. A reproducible numerical experiment with 1,000 simulated demand-disruption scenarios shows that the proposed AI-enabled dynamic optimization model reduces total cost by 55.3% relative to a static base-stock benchmark, maintains a fill rate of 0.983, and lowers emissions relative to a conventional stochastic AI-MPC design. The results indicate that the managerial value of AI in supply chain management depends less on isolated forecasting accuracy than on the closed-loop conversion of predictive distributions into constrained sequential decisions.

Keywords: *Artificial Intelligence; Supply Chain Management; Dynamic Optimization; Stochastic Model Predictive Control; Reinforcement Learning; Disruption risk*

1. Introduction

The operational environment of modern supply chains has become increasingly dynamic and complex. Demand patterns frequently change due to various factors such as promotions, platform traffic fluctuations, macroeconomic shocks, and extreme events. Additionally, lead times can fluctuate significantly as logistics networks face congestion and constraints regarding supplier capacity. Firms are now expected to not only maintain high service performance but also control working capital and minimize carbon exposure simultaneously. These mounting pressures have made Artificial Intelligence (AI) an attractive solution in the realm of supply chain management. Techniques such as machine learning, probabilistic forecasting, and anomaly-detection models are capable of extracting valuable signals from a variety of systems including enterprise resource planning systems, warehouse management systems, transportation management systems, supplier records, and external market data. Nevertheless, the practical value of AI is not fully realized when a model merely predicts demand or classifies supplier risk. A supply chain manager must make critical decisions regarding how much to order, which supplier to use, how much inventory to hold, when to expedite shipments, and how to allocate scarce capacity across various customers effectively. Ultimately, optimizing these decisions requires a comprehensive understanding of both internal and external factors.

This study, therefore, thoroughly examines the integration of AI in supply chain management from the perspective of dynamic optimization. The central argument presented here is that AI should be seamlessly embedded within a sequential decision architecture rather than merely utilized as an isolated forecasting module. While a prediction-only system can enhance visibility across the supply chain, it fails to ensure that the resulting actions are feasible, cost-effective, resilient, or sustainable in practice. In contrast, a dynamic optimization system actively receives and processes updated probabilistic information, transforming it into well-informed decisions that explicitly account for state transitions, service requirements, budget restrictions, supplier constraints, and risk limits. This paradigm shift represents a more holistic approach to managing supply chain complexities and uncertainties effectively.

The paper presents a novel AI-enabled Dynamic Optimization Model (AI-DOM) specifically designed for multi-period supply chain planning. This advanced model is firmly grounded in a combination of dynamic programming, stochastic inventory control, model predictive control, and safe reinforcement learning techniques. Its primary contribution lies not in the development of a singular forecasting algorithm, but in the creation of an integrated decision-making framework that effectively links distributional AI estimates to constrained rolling-horizon optimization processes. This comprehensive framework is particularly well-suited for diverse environments such as manufacturing, retail, and platform-based distribution, where both demand uncertainty and supplier disruptions play critical roles in shaping overall operational performance and efficiency.

The contribution of the paper is threefold.

First, it formalizes the supply chain as a controlled stochastic system with observable and latent state variables.

Second, it proposes a hybrid AI-DOM algorithm that combines distributional forecasting, stochastic model predictive control and safe policy improvement in a digital twin.

Third, it reports a reproducible simulation that compares the proposed approach with static inventory control, deterministic rolling-horizon optimization and stochastic AI-MPC.

The analysis provides both methodological value for researchers and implementation guidance for practitioners.

2. Literature Review

AI applications in supply chain management have evolved significantly over the years, transitioning from early expert systems and simple rule-based decision support mechanisms to more advanced methodologies such as machine learning, deep learning, and reinforcement learning. Min [1] identified both early theoretical frameworks and practical operational applications of AI in the field of supply chain management. Subsequently, later studies have placed a greater emphasis on the critical role that large-scale operational data plays in informing logistics, production, and distribution decisions within this domain [2]. In contemporary practice, AI technologies are now widely utilized for various purposes including demand forecasting, customer segmentation, predictive maintenance, routing optimization, supplier-risk monitoring, and abnormal-event detection. This evolution highlights the transformative potential of AI in enhancing efficiency and responsiveness in supply chain processes.

A second stream of research addresses the critical issues of supply chain risk and resilience. Tang [5] provided a comprehensive summary of various risk-management perspectives applicable to supply chain systems, while Christopher and Peck [20] placed significant emphasis on the essential structural foundations that contribute to resilience. More recent studies focusing on epidemic outbreaks and the associated ripple effects have revealed that disruption risk can propagate through complex supplier networks, inventories, and transportation links [3-4]. These findings collectively demonstrate that static planning assumptions are becoming increasingly inadequate, particularly as demand and supply shocks interact dynamically over time, highlighting the need for more adaptive strategies in supply chain management.

A third research stream focuses on dynamic optimization, inventory control and model predictive control. Classical inventory theory provides base-stock, order-up-to and multi-echelon control principles [6-7]. Dynamic programming gives a rigorous foundation for sequential decisions [9-10], but exact dynamic programming becomes computationally burdensome in high-dimensional supply networks. Model predictive control offers a practical alternative because it solves a finite-horizon problem at each decision epoch and implements only the first action before re-optimizing with new data [11]. Robust and stochastic optimization further improve the treatment of parameter uncertainty and tail risk [12-13].

Reinforcement learning has recently gained significant attention for its potential applications in inventory management and supply chain decision-making processes. Gijsbrechts et al. [14] demonstrate that deep reinforcement learning can achieve impressive performance in selected inventory-related challenges, while Rolf et al. [15] provide a comprehensive review of various reinforcement-learning applications within the realm of supply chain management. However, it is important to note that unconstrained learning policies may pose difficulties when deployed in real-world managerial settings, where critical factors such as service-level contracts, procurement limits, carbon budgets, and governance requirements impose binding constraints. The present study aims to address this notable gap by effectively combining safe learning methods with explicit optimization constraints, thereby enhancing the applicability of reinforcement learning in practice.

3. Problem Description

Consider a focal firm that plans replenishment, production or allocation decisions over periods $t = 0,1,\dots,T$. The firm observes inventories, pipeline orders, backlogs, supplier information and market signals. At each period, it chooses a control action that determines new orders, production quantities, backup-supplier usage, transshipment and expediting decisions. Demand, lead time, supplier reliability and external shocks are uncertain. The objective is to minimize expected operating cost while satisfying service, resilience and sustainability constraints.

The model is designed for a multi-product and multi-node supply chain, but it is presented in a compact vector form. Let x_t denote the state vector, u_t the control vector, ω_t the random disturbance, and b_t the belief state generated by the AI layer. The state vector contains physical and informational variables:

$$x_t = (I_t, P_t, B_t, K_t, R_t, E_t)$$

Where I_t is on-hand inventory, P_t denotes pipeline orders, B_t is backlog, K_t is available capacity, R_t captures supplier reliability and disruption exposure, and E_t is cumulative or remaining carbon budget. The control vector is defined as:

$$u_t = (q_t, y_t, a_t, z_t, e_t)$$

Where q_t is the regular replenishment quantity, y_t is the production quantity, a_t is the allocation decision, z_t is the backup-supplier or transshipment decision, and e_t is the expediting decision.

4. Dynamic Optimization Model

4.1. State Transition

The supply chain evolves according to a controlled stochastic transition:

$$x_{t+1} = f(x_t, u_t, \omega_t; \theta_t)$$

Where θ_t represents latent environmental parameters, including demand regime, lead-time distribution and supplier-risk state. A representative inventory-balance equation for product p at node j is:

$$I_{j,p,t+1} = I_{j,p,t} + R_{j,p,t}(q_{t-L}) + Y_{j,p,t}(y_t) + Z_{j,p,t}(z_t) - D_{j,p,t} - S_{j,p,t}$$

Where $R_{j,p,t}(q_{t-L})$ denotes replenishment arriving after lead time L, $Y_{j,p,t}(y_t)$ is production output, $Z_{j,p,t}(z_t)$ is inbound transshipment or backup supply, $D_{j,p,t}$ is realized demand, and $S_{j,p,t}$ is the quantity shipped or allocated. Pipeline orders and backlog follow their corresponding conservation equations.

4.2. Objective Function

The dynamic optimization problem minimizes expected discounted cost:

$$\min_{\pi \in \Pi} \mathbb{E}_\pi \left[\sum_{t=0}^{T} \gamma^t \, C(x_t, u_t, \omega_t) + \gamma^{T+1} V_f(x_{T+1}) \right]$$

Where π is a feasible decision policy, γ is a discount factor and $V_f(\cdot)$ is a terminal value function. The period cost is decomposed as:

$$C_t = C_t^{ord} + C_t^{prd} + C_t^{hold} + C_t^{back} + C_t^{lost} + C_t^{exp} + C_t^{risk} + C_t^{carb}$$

This decomposition enables the model to generate clear and insightful managerial explanations. Consequently, a decision can be attributed to various cost drivers, such as holding costs, service penalties, disruption risks, or carbon exposure, rather than simply presenting it as an opaque algorithmic output that lacks transparency. This clarity helps stakeholders better understand the reasoning behind each decision.

4.3. Constraints

The feasible action set $\mathcal{U}(x_t)$ is defined by operational, risk and sustainability constraints. The main constraints are:

$$0 \le q_t \le \bar{q}_t, \quad 0 \le y_t \le \bar{K}_t, \quad 0 \le e_t \le \bar{e}_t$$

$$Pr(FR_t \ge \alpha \mid b_t) \ge \beta$$

$$\mathbb{E}\left[CO_{2,t}(u_t) \mid b_t\right] \le \bar{E}_t$$

Where FR_t is the fill rate, α is the service threshold, β is the confidence level, $CO_{2,t}$ is emissions generated by operational actions, and $\bar{E}_t$ is the carbon limit. The probability constraint in (8) is especially important because inventory decisions are often evaluated by service reliability rather than by average cost alone.

4.4. Bellman Representation

The dynamic structure can be written as a Bellman recursion:

$$V_t(x_t, b_t) = \min_{u_t \in \mathcal{U}(x_t)} \mathbb{E}[C(x_t, u_t, \omega_t) + \gamma V_{t+1}(x_{t+1}, b_{t+1}) \mid x_t, b_t, u_t]$$

The belief state b_t is updated after new observations are received. Exact solution of (10) is generally intractable for large networks; nevertheless, it clarifies why the problem should be treated as sequential and why a purely myopic forecast-based policy may be suboptimal.

5. AI-Enabled Distributional Updating

The AI layer estimates predictive distributions rather than point forecasts. Let $\mathcal{D}_{0:t}$ denote all information available up to period t, including internal transaction data, supplier records, logistics status and external market variables. The AI module produces:

$$p(D_{t:t+H}, L_{t:t+H}, \rho_{t:t+H} \mid \mathcal{D}_{0:t})$$

Where D denotes demand, L denotes lead time and ρ denotes disruption probability over the rolling horizon H. The distribution may be obtained through ensemble learning, quantile regression, probabilistic gradient boosting, Bayesian neural networks or temporal deep-learning models. The exact forecasting method is less important than calibration, stability and interpretability because the optimization layer depends on tail behavior and constraint satisfaction.

Belief updating is expressed as:

$$b_{t+1} = \Psi(b_t, \mathcal{D}_{t+1}; \phi)$$

Where $\Psi(\cdot)$ is the updating operator and ϕ represents estimated learning parameters. In an implementation, b_t may include demand quantiles, supplier-risk probabilities, lead-time distributions and anomaly indicators. These quantities enter the rolling optimization problem as scenarios, chance constraints or uncertainty sets.

6. AI-DOM Solution Method

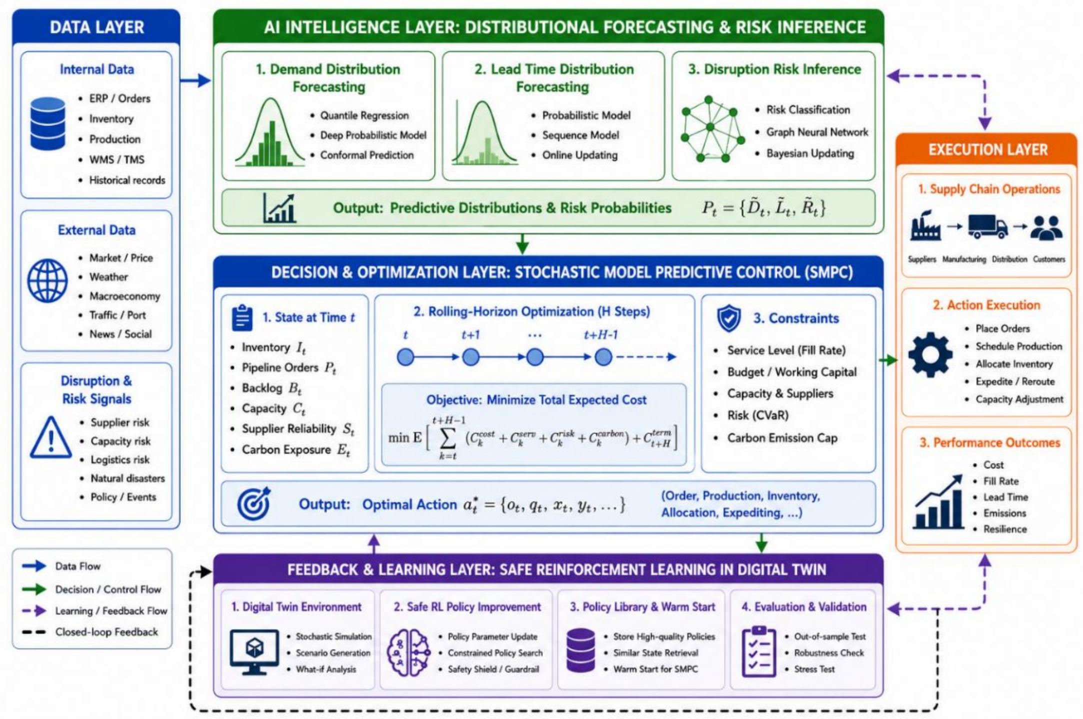

Figure 1. Closed-loop architecture of the proposed AI-DOM framework.

Note. The AI-DOM framework forms a closed-loop decision system. Operational and external data are first collected in the data layer and transformed by the AI intelligence layer into predictive distributions of demand, lead time, and disruption risk. These distributions are then passed to the stochastic model predictive control layer, where feasible supply chain decisions are generated under service-level, capacity, risk, budget, and carbon-emission constraints. The selected actions are implemented in the physical supply chain, and realized outcomes such as cost, fill rate, backlog, emissions, and disruption performance are fed back into the digital twin. Safe reinforcement learning updates policy parameters only after feasibility and safety checks, enabling iterative improvement without violating managerial constraints.

6.1. Rolling-Horizon Stochastic Model Predictive Control

At each period, the optimizer solves a finite-horizon stochastic model predictive control problem:

$$\min_{u_{t:t+H}} \mathbb{E}\left[\sum_{\tau=t}^{t+H} C\left(x_\tau, u_\tau, \omega_\tau\right) + V_f\left(x_{t+H+1}\right) \mid x_t, b_t\right]$$

Subject to the state equations and constraints. Only the first action u_t^* is implemented:

$$u_t^{exec} = u_t^*$$

After analyzing demand, deliveries, and any disruptions that have occurred, the state and belief distributions are systematically updated. Consequently, the problem is resolved once again. This receding-horizon mechanism is particularly appealing for supply chains because it effectively combines forward-looking optimization with continual correction and adjustment to real-time conditions. By doing so, it enhances the responsiveness and resilience of the overall supply chain management process.

6.2. Safe Policy Improvement

A reinforcement-learning component is used only for policy improvement in a digital twin. In this study, the reinforcement-learning component is implemented as Constrained Proximal Policy Optimization (C-PPO). The C-PPO agent is trained only in the digital twin and does not directly replace the stochastic MPC controller. Instead, it updates a small set of policy parameters, including the safety-stock multiplier, disruption-scenario weight, and terminal-value coefficient. The RL state consists of inventory, pipeline orders, backlog, capacity, supplier reliability, carbon exposure, realized service level, cost, and emissions. The RL action is a bounded adjustment of the policy-parameter vector. After each candidate update, the stochastic MPC problem is solved again, and the updated policy is accepted only when service, capacity, risk, and carbon-emission constraints remain feasible. It tunes safety-stock multipliers, scenario weights and terminal-value approximations. The reward function is:

$$r_t = -C_t - \lambda_s v_t^s - \lambda_e v_t^e - \lambda_r v_t^r$$

Where the three slack variables represent service shortfall, carbon excess and risk excess, respectively. A candidate policy is accepted only if it passes a safety-improvement test:

$$J_{new} \leq J_{base} - \varepsilon, \quad V_{new}^{viol} \leq V_{base}^{viol}$$

Where J is expected cost and V^{viol} measures service, emissions and operational constraint violations. This design prevents the learning module from replacing governance constraints with unconstrained trial-and-error behavior.

Table 1. Key parameter settings of the C-PPO-Based safe policy improvement module.

Parameter	Setting
RL algorithm	Constrained Proximal Policy Optimization (C-PPO)
Training environment	Offline digital twin
Policy network	Two hidden layers, 64 neurons per layer
Value network	Two hidden layers, 64 neurons per layer
Activation function	ReLU
Learning rate	3×10^{-4}
Discount factor	0.99
GAE parameter	0.95
PPO clipping ratio	0.20
Batch size	256
Training epochs per update	10
Maximum training episodes	1,000
Safety-stock multiplier range	[0.8, 1.5]
Disruption-scenario weight range	[0.5, 2.0]
Terminal-value coefficient range	[0.1, 1.0]
Safety acceptance rule	Reject updates violating service, capacity, risk, or carbon constraints

6.3. Algorithmic Procedure

Table 2. AI-DOM algorithmic procedure.

Step	Module	Input data	Output and data interaction
1	Operational state observation	Inventory, backlog, pipeline orders, capacity, supplier status, transportation status, carbon exposure.	Converts raw records into state x_t and information set I_t.
2	Distributional belief update	Historical demand, lead-time records, disruption indicators, external signals.	Generates predictive distributions and belief state b_t.
3	Scenario construction	Demand, lead-time, and disruption distributions.	Produces rolling-horizon scenarios or uncertainty sets Ω.
4	Stochastic MPC optimization	x_t, b_t, Ω_t, cost and constraint parameters.	Solves the constrained optimization problem and outputs first action u_t.
5	Action execution	Order, production, allocation, backup supply, expediting decision.	Implements the first-period decision in the supply chain system.
6	Outcome observation	Realized demand, arrivals, backlog, service level, emissions, cost.	Feeds operational outcomes back to state updating and digital twin.
7	Safe RL improvement	Digital-twin simulation results and performance diagnostics.	Updates policy parameters only if safety and feasibility tests are satisfied.
8	Managerial reporting	Binding constraints, cost components, risk indicators.	Provides explainable decision rationale for managers.

Note. Table 2 clarifies not only the algorithmic sequence but also the data input, output variables, and interaction logic among forecasting, optimization, execution, feedback, and safe learning modules.

7. Numerical Experiment

7.1. Experimental Setting

A numerical experiment was designed to evaluate the decision value of the proposed model. The experiment represents a single-product replenishment system with one regular supplier and an emergency expediting option. Demand follows a seasonal stochastic process with temporary shock events. Regular orders have a nominal lead time of two periods, but disruption risk can delay arrivals by one additional period. The cost structure includes ordering cost, expediting cost, holding cost, backlog penalty and carbon cost. The simulation horizon is 36 periods, and the results are averaged over 1,000 independent scenarios generated with a fixed random seed.

The experiment compares four policies:

Table 3. Compared policies.

Policy	Description
Static base-stock	A fixed order-up-to rule using a constant target inventory level.
Deterministic RHO	A rolling-horizon optimizer using mean forecasts and no explicit risk buffer.
Stochastic AI-MPC	A rolling stochastic MPC policy using distributional forecasts and a risk-adjusted buffer.
Proposed AI-DOM	The proposed model with distributional updating, constrained rolling optimization and safe policy tuning.

Note. The benchmark policies are arranged from static non-adaptive control to the proposed closed-loop AI-DOM. This comparison isolates the incremental value of mean forecasting, distributional uncertainty modeling, constrained optimization, and safe policy improvement.

7.2. Results

Table 4. Numerical comparison of policies over 1,000 simulated scenarios.

Policy	Total cost	Fill rate	Stockout-period ratio	Emissions	Max backlog	Cost reduction
Static base-stock	31574.8	0.574	0.743	1406.2	198.1	0.0%
Deterministic RHO	13387.3	0.968	0.095	1346.5	68.7	57.6%
Stochastic AI-MPC	14585.5	0.984	0.040	1363.0	55.5	53.8%
Proposed AI-DOM	14108.0	0.983	0.046	1354.2	56.0	55.3%

Note. Total cost, emissions, and maximum backlog are scenario averages over 1,000 simulations. Fill rate and stockout-period ratio measure service reliability. Cost reduction is calculated relative to the static base-stock benchmark.

The static base-stock policy performs poorly because it cannot adapt to seasonal demand changes and temporary shocks. Deterministic rolling-horizon optimization sharply reduces total cost, but it achieves a lower fill rate and a higher stockout-period ratio because it uses mean forecasts without explicit risk protection. Stochastic AI-MPC improves service performance by incorporating predictive distributions and disruption buffers, but it tends to preserve a relatively conservative inventory position. The proposed AI-DOM policy achieves a cost reduction of 55.3% relative to the static benchmark while maintaining a fill rate of 0.983. Compared with stochastic AI-MPC, it reduces total cost and emissions by tuning risk buffers and expediting decisions through safe policy improvement.

7.3. Interpretation

The results support three observations.

First, the value of AI increases when predictions are expressed as distributions and used in constrained decisions.

Second, the lowest-cost policy is not necessarily the most appropriate policy if it sacrifices service reliability.

Third, safe policy improvement can refine the cost-service-emissions trade-off without permitting uncontrolled learning in the operational system.

8. Management Implications

The model has several implications for supply chain management.

First, firms should evaluate AI forecasts by their operational decision value rather than by forecast error alone. A slightly less accurate but better calibrated distribution may produce superior inventory and expediting decisions.

Second, digital supply-chain initiatives should integrate forecasting, optimization and execution systems rather than building isolated dashboards.

Third, the proposed framework supports explainable decision making because it reports binding constraints and cost components.

Fourth, resilience and sustainability can be incorporated as decision constraints rather than treated as ex-post performance reports.

For implementation, a firm can begin with a restricted pilot, such as one product family, one warehouse network or one supplier category. The required data include demand history, inventory snapshots, open orders, supplier lead times, expediting records, backlog events and carbon-intensity coefficients. The rolling-horizon optimizer can be integrated into existing planning cycles, while the safe learning component can be trained offline in a digital twin before controlled deployment.

9. Conclusion

This paper developed an AI-enabled dynamic optimization model for supply chain management under demand and disruption uncertainty. The study formalized the supply chain as a controlled stochastic system, introduced distributional AI updating, constructed a stochastic model predictive control formulation and added safe reinforcement-learning-based policy improvement. A reproducible numerical experiment showed that the proposed AI-DOM model improves the cost-service-emissions trade-off relative to static base-stock and conventional stochastic AI-MPC policies. The findings suggest that the managerial value of AI lies in closed-loop sequential decision making rather than prediction alone.

Future research can extend the model to multi-echelon supply chains, strategic supplier behavior, multi-agent coordination, carbon-market mechanisms and empirical validation with firm-level data. Additional work may also examine how managers interpret and accept AI-generated optimization recommendations when service guarantees and sustainability targets are simultaneously binding.

References

[1] H. Min. Artificial intelligence in supply chain management: Theory and applications. *International Journal of Logistics Research and Applications*, vol. 13, no. 1, pp. 13-39, 2010.

[2] G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, vol. 176, pp. 98-110, 2016.

[3] D. Ivanov. Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak. *Transportation Research Part E*, vol. 136, Article 101922, 2020.

[4] D. Ivanov and A. Dolgui. OR-methods for coping with the ripple effect in supply chains during COVID-19 pandemic: Managerial insights and research implications. *International Journal of Production Economics*, vol. 232, Article 107921, 2021.

[5] C. S. Tang. Perspectives in supply chain risk management. *International Journal of Production Economics*, vol. 103, no. 2, pp. 451-488, 2006.

[6] L. V. Snyder and Z.-J. M. Shen, Fundamentals of Supply Chain Theory, 2nd ed. *Hoboken, NJ: Wiley*, 2019.

[7] S. Axsater, Inventory Control, 3rd ed. *Cham: Springer*, 2015.

[8] W. B. Powell. Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions. *Hoboken, NJ: Wiley*, 2022.
[9] R. Bellman. Dynamic Programming. Princeton. *NJ: Princeton University Press*, 1957.
[10] D. P. Bertsekas. Dynamic Programming and Optimal Control, 4th ed. Belmont, *MA: Athena Scientific*, 2017.
[11] C. E. Garcia, D. M. Prett, and M. Morari. Model predictive control: Theory and practice - A survey. *Automatica*, vol. 25, no. 3, pp. 335-348, 1989.
[12] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, vol. 54, no. 1, pp. 150-168, 2006.
[13] W. Klibi, A. Martel, and A. Guitouni. The design of robust value-creating supply chain networks: A critical review. European Journal of Operational Research, vol. 203, no. 2, pp. 283-293, 2010.
[14] J. Gijsbrechts, R. N. Boute, J. A. Van Mieghem, and D. J. Zhang. Can deep reinforcement learning improve inventory management? Performance and implementation of dual sourcing-mode problems. *Manufacturing & Service Operations Management*, vol. 24, no. 3, pp. 1349-1368, 2022.
[15] B. Rolf, I. Jackson, M. Muller, S. Lang, T. Reggelin, and D. Ivanov. A review on reinforcement learning algorithms and applications in supply chain management. *International Journal of Production Research*, vol. 61, no. 20, pp. 7151-7179, 2023.
[16] R. S. Sutton and A. G. Barto. Reinforcement Learning: An Introduction, 2nd ed. *Cambridge, MA: MIT Press*, 2018.
[17] R. Fildes, S. Ma, and S. Kolassa. Retail forecasting: Research and practice. *International Journal of Forecasting*, vol. 38, no. 4, pp. 1283-1318, 2022.
[18] S. Makridakis, E. Spiliotis, and V. Assimakopoulos. M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, vol. 38, no. 4, pp. 1346-1364, 2022.
[19] B. Wilder, B. Dilkina, and M. Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 1658-1665, 2019.
[20] M. Christopher and H. Peck. Building the resilient supply chain. *The International Journal of Logistics Management*, vol. 15, no. 2, pp. 1-14, 2004.
[21] H. L. Lee. The triple-A supply chain. *Harvard Business Review*, vol. 82, no. 10, pp. 102-112, 2004.
[22] M. S. Sodhi and C. S. Tang. Supply chain management for extreme conditions: Research opportunities. *Journal of Supply Chain Management*, vol. 57, no. 1, pp. 7-16, 2021.
[23] R. Dekker, J. Bloemhof, and I. Mallidis. Operations research for green logistics - An overview of aspects, issues, contributions and challenges. *European Journal of Operational Research*, vol. 219, no. 3, pp. 671-679, 2012.
[24] S. Seuring and M. Muller. From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, vol. 16, no. 15, pp. 1699-1710, 2008.
[25] F. Sahin and E. P. Robinson. Flow coordination and information sharing in supply chains: Review, implications, and directions for future research. *Decision Sciences*, vol. 33, no. 4, pp. 505-536, 2002.