Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 10, pp 233-245, October, 2026

Reinforcement Learning Optimization of Inventory Financing Decisions Under Demand Uncertainty, Supplier Disruptions, and Variable Borrowing Costs

Ayodeji Ikudabo

Amazon, SPS Finance, Seattle, WA

DOI : https://doi.org/10.5281/zenodo.23229945

ABSTRACT

Inventory financing is increasingly important to supply-chain resilience as firms balance product availability, working-capital constraints, procurement risks, and changing financial conditions. Conventional inventory models often rely on predetermined demand distributions, stable supplier performance, and relatively fixed financing assumptions, limiting their effectiveness in volatile operating environments. This study develops a reinforcement learning framework for optimizing integrated inventory and financing decisions under demand uncertainty, supplier disruptions, and variable borrowing costs. The framework models inventory management as a sequential decision problem in which an intelligent agent dynamically determines replenishment quantities and financing requirements using observed inventory positions, demand patterns, supplier reliability, cash availability, and prevailing borrowing rates. Through repeated interaction with a simulated supply-chain environment, the agent learns policies that minimize cumulative purchasing, holding, shortage, disruption, and financing costs while maintaining service performance and liquidity. The proposed approach is evaluated against conventional reorder-point and optimization-based policies across alternative demand, disruption, and interest-rate scenarios. Performance is assessed using total operating cost, stockout frequency, inventory turnover, financing utilization, service level, and profitability. The study demonstrates how adaptive reinforcement learning can strengthen inventory-financing coordination and support resilient, cost-efficient working-capital decisions under multidimensional uncertainty.

Keywords: Reinforcement Learning; Inventory Financing; Demand Uncertainty; Supplier Disruptions; Borrowing Costs; Supply Chain Optimization

1. INTRODUCTION

1.1 Inventory Financing and the Short-Term Liquidity Challenge

Inventory investment sits at the intersection of operational continuity and short-term financial management because firms must secure sufficient stock while preserving cash for other obligations. For manufacturing, wholesale, retail, and other inventory-intensive firms, replenishment decisions therefore cannot be separated from working-capital financing choices [1]. Inventory protects against demand fluctuations and supply delays, yet larger stocks create storage, obsolescence, insurance, and opportunity costs while extending the cash conversion cycle [2]. When internal liquidity is inadequate, purchases may require bank credit, revolving facilities, or supplier financing, directly linking replenishment volumes to financing costs [3]. The resulting trade-off is significant: insufficient inventory increases stockouts, production interruptions, lost sales, and deterioration in customer service, whereas excessive inventory immobilizes scarce cash and may increase short-term leverage [4]. Effective inventory financing consequently requires coordinated decisions that balance product availability, carrying costs, liquidity preservation, and external capital costs across changing operating and financial conditions over successive time.

1.2 Compounding Effects of Demand, Supply, and Financing Uncertainty

Coordination becomes more difficult when operational and financial uncertainties occur simultaneously. Customer demand is stochastic and can shift unexpectedly, making required replenishment quantities difficult to determine accurately [5]. Meanwhile, supplier disruptions, delays or extended lead times can prevent inventory from arriving when required, increasing the value of precautionary stock [6]. Financing conditions add complexity as changing borrowing rates alter the marginal cost of carrying inventory funded through debt [7]. These uncertainties are interdependent. A demand surge during disruption can accelerate inventory depletion, while higher borrowing rates can make emergency replenishment financially expensive [3]. Optimizing demand response, supplier risk, or financing independently may therefore produce locally efficient but firm-level suboptimal decisions. A coordinated framework must recognize their joint effects on inventory availability, cash balances, borrowing requirements, service performance, and

profitability [8].

1.3 Limitations of Conventional Inventory-Financing Models

Traditional inventory models provide useful decision rules but face limitations in changing environments. Economic order quantity models rely on stable cost and demand assumptions, while reorder-point and safety-stock policies depend on prespecified demand and lead-time distributions [1]. Stochastic programming accommodates uncertainty, but its effectiveness depends on scenario specification, probability estimates, and computational tractability as decision dimensions expand [6]. Conventional optimization can require reformulation when conditions change [4]. Such approaches typically prescribe decisions from parameters rather than learning from realized consequences. This limits their capacity to adapt replenishment and borrowing jointly as demand, supplier reliability, liquidity, and financing costs change [7].

1.4 Research Objective, Questions, and Contributions

This study develops a reinforcement learning framework that coordinates inventory replenishment and financing decisions. An RL agent can observe states, act, receive rewards, and improve its policy through interaction [5]. The study addresses three questions. **RQ1** asks whether RL can reduce inventory-financing costs relative to policies. **RQ2** examines how the optimal policy changes under demand volatility, supplier disruptions, and variable borrowing rates. **RQ3** evaluates whether RL can strengthen liquidity resilience without sacrificing customer service levels [8]. Methodologically, the study integrates inventory, supplier, cash, and borrowing states within one architecture rather than treating them independently [2]. Managerially, it provides a mechanism for coordinating procurement and working-capital decisions. Section 2 consequently examines how inventory, liquidity, supply risk, and financing costs interact.

2. LITERATURE REVIEW AND CONCEPTUAL FOUNDATION

2.1 Inventory Optimization Under Demand Uncertainty

Inventory optimization under uncertain demand requires firms to balance product availability against the financial consequences of carrying excess stock. Stochastic inventory models recognize that realized demand can deviate substantially from expectations, making static replenishment decisions inefficient [6]. Safety stock consequently provides protection against demand variability, although its appropriate level depends on demand dispersion, replenishment lead time, targeted service levels, and shortage penalties [9]. Dynamic replenishment extends this logic by adjusting orders as inventory positions and demand information evolve [7]. Higher inventory can reduce stockouts, production interruptions, and lost sales but simultaneously increases storage, obsolescence, insurance, and opportunity costs [11]. Conversely, lean inventory releases working capital while increasing shortage exposure. The fundamental inventory transition governing these competing effects is expressed as:

$$I_{t+1} = I_t + Q_t - D_t$$

where I_t denotes available inventory, Q_t represents replenishment received, and D_t is realized demand.

2.2 Supplier Disruptions and Inventory Resilience

Demand uncertainty becomes more consequential when supplier performance is unreliable. Disruptions arising from production failures, transportation interruptions, capacity shortages, quality problems, or external shocks can alter both replenishment quantities and expected lead times [10]. Even relatively infrequent disruptions may materially change optimal inventory policies because delayed deliveries increase stockout exposure during recovery periods [12]. Firms therefore employ resilience strategies including safety inventory, supplier diversification, alternative sourcing, flexible contracts, and redundant capacity [8]. Such protection nevertheless imposes financial consequences. Larger inventory buffers immobilize working capital, while emergency procurement may require higher purchasing expenditure or additional short-term financing [13]. Supplier disruption risk therefore increases the economic value of precautionary inventory while simultaneously creating a liquidity trade-off between operational resilience and the cost of financing additional stock [6].

2.3 Inventory Financing and Variable Borrowing Costs

Inventory decisions are inherently financial because procurement generally requires cash before corresponding sales revenues are collected. Working-capital management therefore connects inventory with cash holdings, accounts receivable, accounts payable, and short-term financing requirements [13]. Longer inventory holding periods can extend the cash conversion cycle and intensify liquidity pressure, particularly for firms with constrained internal funds [9]. Firms may consequently use bank loans, revolving credit facilities, or trade credit to finance replenishment [14]. However, borrowing introduces interest expenses and financing-capacity constraints. Rising interest rates increase the carrying cost of debt-financed inventory and may alter economically desirable order quantities [7]. Conversely, lower financing costs may justify precautionary stock when expected shortage or disruption losses are substantial. Period-specific financing cost is represented as:

$$FC_t = B_t r_t$$

where B_t represents borrowed capital and r_t denotes the applicable period-specific borrowing rate.

2.4 Reinforcement Learning for Sequential Financial and Operational Decisions

Inventory financing constitutes a sequential decision problem because current replenishment and borrowing choices alter subsequent inventory, liquidity, debt, and service positions [15]. Reinforcement learning formalizes this process through a Markov Decision Process comprising states, actions, transition dynamics, rewards, and a decision policy [10]. The state may combine inventory availability, demand conditions, supplier reliability, cash holdings, outstanding borrowing, and financing rates, while actions determine replenishment and financing quantities [12]. Following each action, realized operational and financial outcomes generate a reward that informs subsequent policy improvement. Through repeated interaction, the agent learns to maximize cumulative rather than single-period economic performance [15]. This distinguishes reinforcement learning from supervised forecasting: forecasting estimates future demand, disruptions, or borrowing conditions, whereas RL learns which inventory-financing action should be selected under observed conditions [8]. Consequently, the approach internalizes intertemporal trade-offs between inventory availability, liquidity preservation, financing expense, and service performance.

2.5 Research Gap and Conceptual Integration

Existing scholarship provides substantial foundations for stochastic inventory control, supply-chain resilience, working-capital management, and reinforcement learning, yet these domains are frequently examined independently [11]. Limited integration exists between demand uncertainty, supplier disruption, replenishment decisions, liquidity constraints, and time-varying borrowing costs within a unified adaptive framework [14]. The proposed RL architecture addresses this fragmentation by linking operational states directly with financing consequences and learning coordinated actions dynamically [15]. It therefore bridges operations management and corporate finance while providing an adaptive mechanism for inventory-financing decisions under multidimensional uncertainty.

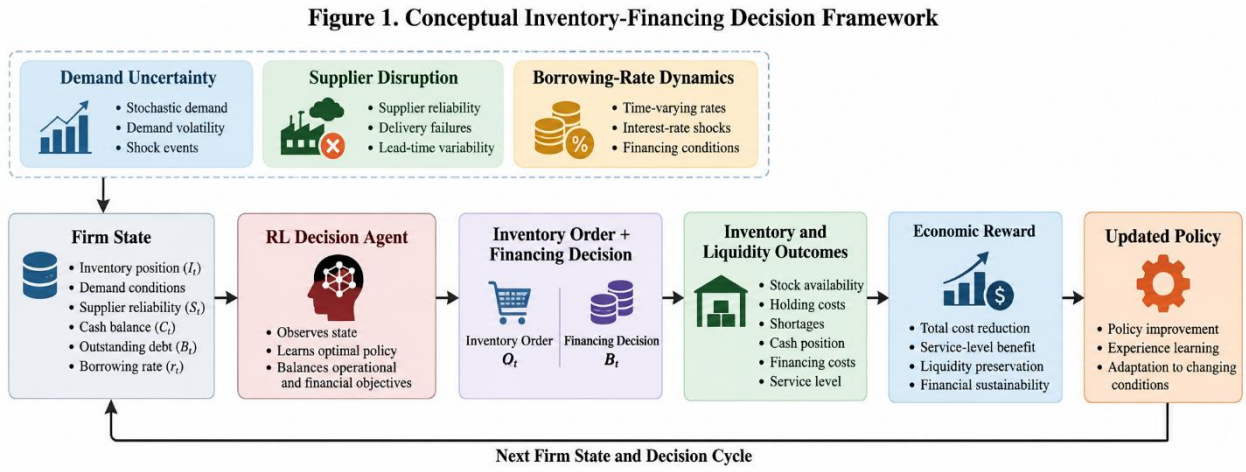


Figure 1. Conceptual Inventory-Financing Decision Framework

Having established how demand uncertainty, supplier disruptions, inventory requirements, and financing conditions interact, Section 3 translates these theoretical relationships into a measurable decision environment by specifying the data architecture, uncertainty processes, integrated inventory-cash dynamics, and experimental dataset required to train and evaluate the reinforcement learning framework.

3. DATA, DECISION ENVIRONMENT, AND PROBLEM FORMULATION

3.1 Data Architecture and Variable Construction

3.1.1 Operational and Financial Inputs

The analytical environment is structured at the firm-period and SKU-period levels, allowing operational inventory movements to be connected with contemporaneous financing conditions. Each observation contains realized demand, opening inventory, replenishment quantities, purchase prices, supplier lead times, and disruption indicators [12]. Financial inputs include cash balances, accounts payable, outstanding short-term debt, available credit, and borrowing rates. Cost variables capture inventory holding, procurement, shortage, disruption, and financing expenses [16]. Service-level measures identify the proportion of demand satisfied without delay. Combining these variables provides the RL agent with a multidimensional representation of operational exposure, liquidity availability, supplier conditions, and financing constraints when decisions are made [14].

3.1.2 Data Preprocessing and Standardization

Before model training, missing observations are treated using variable-appropriate imputation, while extreme values are winsorized or assessed against economically plausible boundaries [17]. Continuous variables are normalized to prevent scale differences from disproportionately influencing learning. Operational and financial observations are temporally aligned so each state reflects information available within the same decision period [13]. Lagged demand, inventory, disruption, and borrowing-rate variables capture temporal dependence. To prevent information leakage, future demand, supplier outcomes, interest rates, and realized costs remain unavailable to the agent at decision time t [18].

3.2 Modelling Demand Uncertainty

Demand uncertainty is represented through multiple environments designed to test policy performance beyond average operating conditions. Period demand consists of an expected component and a stochastic disturbance [15]:

$$D_t = \mu_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{D}(0, \sigma_t^2)$$

where μ_t denotes expected demand and σ_t^2 controls variability. Four regimes are considered: stable demand with limited dispersion, moderately volatile demand, highly volatile demand, and shock demand characterized by sudden departures from historical patterns [19]. This scenario structure exposes the agent to progressively greater forecasting and replenishment difficulty. Higher variability accelerates unexpected inventory depletion and increases the probability of shortages when realized demand exceeds available stock [12]. Conversely, overestimating uncertain demand creates excess inventory and holding costs. Demand uncertainty therefore affects not only service performance but also procurement expenditure, cash requirements, financing utilization, and the amount of precautionary inventory economically justified [16].

3.3 Modelling Supplier Disruptions

Supplier uncertainty is incorporated through a stochastic availability indicator that determines whether planned replenishment can be fulfilled within period t [14]:

$$S_t = \begin{cases} 1, & \text{supplier operational} \\ 0, & \text{supplier disrupted} \end{cases}$$

Supplier status is generated according to specified disruption probabilities and recovery dynamics. Disruption scenarios vary in frequency and duration and incorporate complete delivery failures, partial order fulfilment, delayed deliveries, and lead-time expansion [18]. Consequently, the quantity ordered by the agent does not necessarily equal inventory received during that period. Effective replenishment depends on supplier availability, fulfilment proportion, and outstanding orders carried from earlier periods. This distinction prevents the optimization model from implicitly assuming perfect supply reliability [13]. Increasing disruption intensity raises stockout exposure and strengthens incentives to build precautionary inventories, but additional buffers consume working capital. The model therefore captures the financial cost of resilience alongside its operational benefits [17].

3.4 Variable Borrowing-Cost Environment

Borrowing costs are modelled dynamically because firms rarely face invariant financing conditions across planning horizons. The interest-rate process follows a mean-reverting specification:

$$r_{t+1} = \bar{r} + \rho(r_t - \bar{r}) + \eta_t$$

where $\bar{r}$ is the long-run borrowing rate, ρ captures rate persistence, and η_t represents financing shocks [19]. This specification generates periods of relatively stable credit conditions alongside persistent increases or decreases in borrowing costs. The effective financing burden additionally depends on credit utilization because larger inventory purchases can increase outstanding debt when operating cash is insufficient [15]. Higher rates therefore increase the opportunity cost of debt-financed replenishment and reduce available liquidity through interest payments [16]. By observing both borrowing rates and existing credit exposure, the agent can evaluate whether additional inventory protection economically justifies its incremental financing requirement.

3.5 Integrated Inventory-Cash Dynamics

The operational and financing components are integrated through a cash-transition mechanism that translates inventory decisions into liquidity consequences. Period-end cash is represented as:

$$C_{t+1} = C_t + R_t - P_t - H_t - SC_t - r_t B_t + \Delta B_t$$

where C_t denotes opening cash, R_t cash receipts, P_t procurement payments, H_t inventory holding costs, SC_t shortage or disruption-related costs, $r_t B_t$ interest expense, and ΔB_t net additional borrowing [13]. Replenishment therefore affects liquidity through procurement payments and subsequent inventory carrying costs, while shortages influence both service outcomes and associated penalties [18]. Borrowing can relieve immediate liquidity

pressure but creates future financing costs and potentially reduces unused credit capacity. This structure prevents the agent from minimizing inventory costs without considering solvency implications [14]. Similarly, excessive cash preservation cannot be rewarded when underinvestment in inventory produces costly shortages. Inventory availability and liquidity consequently evolve as mutually dependent components of the same sequential decision process [17].

3.6 Experimental Dataset Partitioning

The generated or historical trajectories are partitioned chronologically into training, validation, and testing samples using a 70%/15%/15% allocation [16]. The earliest observations form the training environment in which the agent learns state-action relationships through repeated interactions. The subsequent validation period supports hyperparameter selection, convergence assessment, and model choice without contaminating final evaluation [19]. The remaining 15% constitutes an unseen test environment used exclusively to assess out-of-sample policy performance. Chronological partitioning is preferred to random sampling because inventory, borrowing rates, supplier conditions, and demand exhibit temporal dependence [12]. No future test observations are available during policy training, thereby reducing look-ahead bias and providing a credible assessment of generalization [18].

Table 1. State Variables, Actions, Parameters, and Measurement Definitions

Category	Variable	Symbol	Measurement/Definition	Decision Role
Inventory	Inventory position	I_t	Units available at decision time	State
Demand	Realized/observed demand	D_t	Units requested during period	State/environment
Supply	Supplier status	S_t	Operational or disrupted	State
Supply	Supplier lead time	L_t	Periods until replenishment arrival	State
Finance	Cash balance	C_t	Available monetary liquidity	State
Finance	Outstanding debt	B_t	Short-term borrowing balance	State
Finance	Borrowing rate	r_t	Period-specific financing rate	State
Cost	Holding cost	H_t	Cost of carrying inventory	Parameter/outcome
Cost	Shortage cost	SC_t	Cost associated with unmet demand	Parameter/outcome
Decision	Order quantity	Q_t	Inventory units ordered	Action
Decision	Net borrowing	ΔB_t	Increase or reduction in financing	Action
Performance	Service level	SL_t	Proportion of demand fulfilled	Outcome

4. REINFORCEMENT LEARNING OPTIMIZATION FRAMEWORK

4.1 Markov Decision Process Formulation

The integrated inventory-financing problem is formulated as a Markov Decision Process (MDP), enabling sequential decisions to account for both immediate costs and future economic consequences [18]. The decision environment is represented as:

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$$

where $\mathcal{S}$ denotes the state space containing observable operational and financial conditions, while $\mathcal{A}$ represents feasible replenishment and financing actions [20]. The transition function $P(s_{t+1}|s_t, a_t)$ describes how current states and actions determine subsequent inventory, liquidity, and debt positions under stochastic demand and supplier outcomes. The reward function Rquantifies the economic consequences generated by each decision [22]. Finally, the discount factor $\gamma \in [0,1]$ determines the relative importance assigned to future rewards. The agent therefore seeks a policy $\pi(a|s)$ that

maximizes expected cumulative discounted returns across the planning horizon [19].

4.2 State-Space Representation

At each decision period, the agent observes a state vector integrating operational, supply, and financing information:

$$s_t = [I_t, D_t, L_t, S_t, C_t, B_t, r_t, \ldots]$$

Here, I_t represents inventory position, D_t observed demand information, L_t supplier lead time, S_t supplier status, C_t available cash, B_t outstanding borrowing, and r_t the prevailing borrowing rate [21]. Additional features can include outstanding orders, credit availability, demand lags, and service performance. Combining these variables is important because replenishment decisions affect financing requirements while liquidity constraints simultaneously restrict feasible inventory investment [18]. The integrated state therefore allows the agent to coordinate operational resilience with financial sustainability rather than optimizing either dimension independently.

4.3 Joint Inventory-Financing Action Space

Given state s_t, the agent simultaneously determines inventory replenishment and financing adjustment through the joint action:

$$a_t = (Q_t, \Delta B_t)$$

where Q_t represents the quantity ordered and ΔB_t denotes additional borrowing or debt repayment [22]. Joint optimization enables the agent to recognize that purchasing inventory may require financing, while financing decisions depend on anticipated inventory needs and available liquidity. For discrete action spaces, order and borrowing quantities can be divided into predefined decision levels, allowing algorithms such as Deep Q-Networks to evaluate alternatives [20]. Continuous action spaces permit more granular quantities and are better suited to actor-critic methods such as Proximal Policy Optimization or Soft Actor-Critic [23].

4.4 Reward Function and Economic Objective

The reward function converts operational and financial outcomes into a common economic objective, allowing the agent to evaluate the consequences of each joint decision [19]. Period reward is specified as:

$$R_t = -(PC_t + HC_t + SC_t + FC_t + DC_t) + \lambda SL_t$$

where PC_t denotes procurement cost, HC_t inventory holding cost, SC_t stockout or shortage cost, FC_t financing cost, DC_t disruption-related cost, and SL_t the service-level benefit [24]. The negative cost component encourages efficient inventory and financing utilization, while the positive service component prevents cost minimization from becoming the sole objective. The weighting parameter λ determines the economic importance attached to customer-service performance relative to direct costs [21]. Without this component, an agent could reduce procurement, holding, and borrowing expenses simply by maintaining inadequate inventory. Incorporating shortage penalties and service rewards instead requires the policy to balance liquidity preservation against product availability, thereby aligning learned behavior with economically sustainable inventory management [18].

4.5 RL Algorithm Selection and Training

Algorithm selection depends principally on whether replenishment and financing actions are represented discretely or continuously. A Deep Q-Network is suitable when Q_t and ΔB_t are restricted to predefined levels because the network approximates state-action values for feasible decisions [20]. For continuous quantities, Proximal Policy Optimization or Soft Actor-Critic can learn flexible policies without enumerating every possible action [23]. Training proceeds through episodes representing sequential planning horizons. During each transition, the agent observes s_t, selects a_t, experiences demand and supplier outcomes, receives R_t, and moves to s_{t+1} [22]. Exploration allows alternative decisions to be tested rather than prematurely exploiting an immature policy. Depending on the algorithm, collected transitions support replay-based value updates or policy-gradient optimization. Repeated episodes progressively adjust network parameters until rewards, costs, service performance, and inventory-financing decisions stabilize across independent training runs [24].

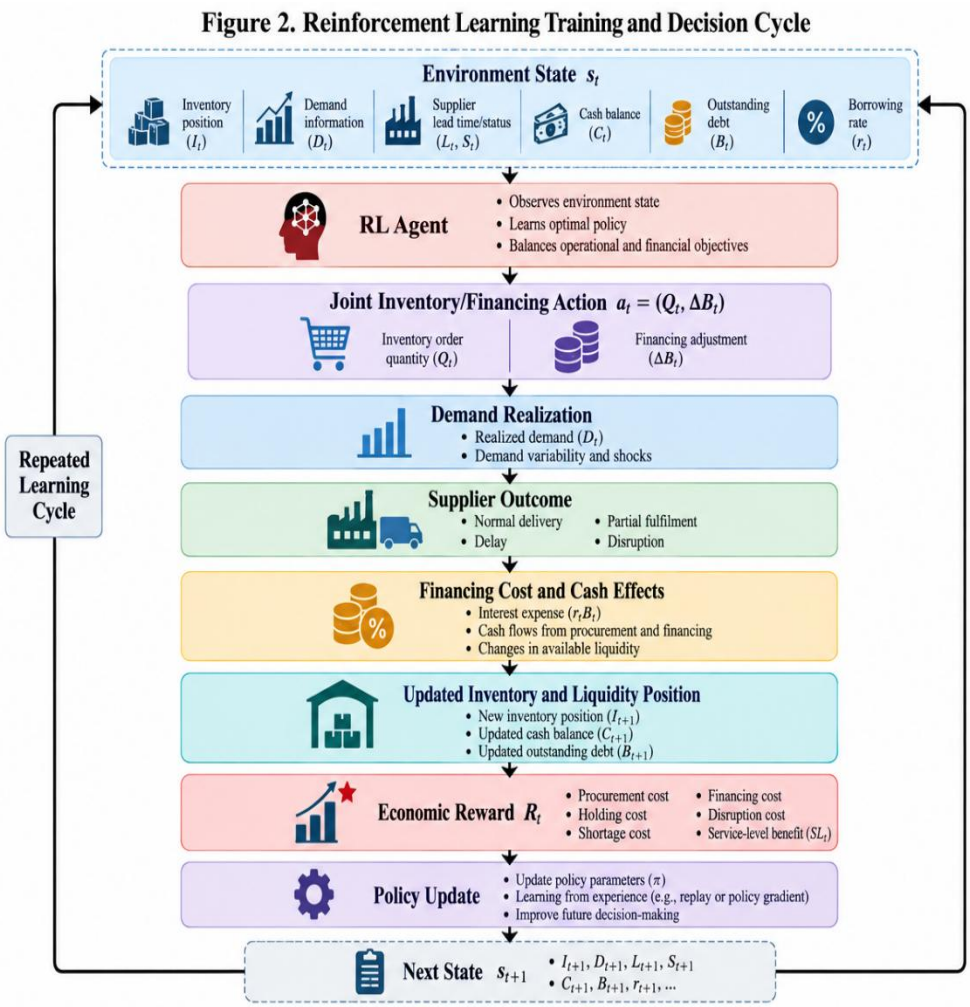


Figure 2. Reinforcement Learning Training and Decision Cycle

4.6 Hyperparameter Tuning and Model Selection

Model performance depends on hyperparameters controlling both learning stability and exploration. Candidate configurations therefore vary the learning rate, discount factor, batch size, neural-network depth and width, exploration schedule, entropy coefficient, and training horizon [21]. For actor-critic models, policy-update frequency and clipping or temperature parameters are additionally evaluated [23]. Each configuration is trained across repeated random seeds to reduce sensitivity to initialization. Model selection relies exclusively on validation-period reward, cost, service level, and stability measures [19]. Test observations remain isolated until final evaluation, preventing performance-driven model selection.

4.7 Convergence and Policy Stability

Convergence is assessed using moving averages of episodic reward, total inventory-financing cost, service level, and action variability [24]. A policy is considered stable when these measures cease exhibiting systematic improvement across successive training episodes. Repeated runs with alternative random seeds further establish whether learned performance reflects persistent policy behavior rather than favorable initialization [22].

5. EXPERIMENTAL DESIGN, BENCHMARKING, AND EVALUATION

5.1 Benchmark Decision Policies

The reinforcement learning framework is evaluated against three benchmark policies representing progressively sophisticated inventory-control approaches [23]. The first is a traditional reorder-point and safety-stock policy in which replenishment occurs when inventory reaches a predetermined threshold, with buffer stock protecting against expected demand and lead-time uncertainty [26]. The second is a deterministic cost-minimization policy that selects replenishment quantities by minimizing expected procurement, holding, shortage, and financing costs using point estimates of future operating conditions [24]. The third is a stochastic inventory optimization policy that incorporates probability distributions for demand and supply uncertainty but applies predetermined financing rules rather than learning financing actions dynamically [28]. All four policies operate under identical

initial inventory, demand realizations, supplier conditions, cost parameters, credit limits, and borrowing-rate trajectories [25]. This common experimental environment ensures that differences in observed outcomes reflect decision-policy performance rather than unequal exposure to favorable or adverse conditions.

5.2 Experimental Scenarios

Policy resilience is examined using a factorial stress-testing design that systematically varies demand uncertainty, supplier disruption, and borrowing costs [27]. Demand conditions are classified as low or high volatility according to the dispersion of realized demand around expected requirements. Supplier conditions combine low and high disruption probabilities, with disruption events affecting fulfilment quantities and replenishment lead times [23]. Financing conditions similarly distinguish low-cost from high-cost borrowing environments using alternative interest-rate levels and shock processes. Combining these dimensions produces eight principal scenarios, allowing both individual and interaction effects to be evaluated [29]. An additional severe joint-stress scenario simultaneously imposes high demand volatility, elevated supplier disruption, and increased borrowing costs. This scenario is particularly important because operational shocks can increase precautionary inventory requirements precisely when financing that additional inventory becomes more expensive and liquidity becomes more constrained [25].

5.3 Performance Metrics

Performance is evaluated using complementary operational and financial indicators rather than relying solely on accumulated RL reward [24]. Total cost over horizon \(T\) is defined as:

$$TC = \sum_{t=1}^{T} (PC_t + HC_t + SC_t + FC_t + DC_t)$$

capturing procurement, holding, shortage, financing, and disruption-related expenditures. Customer-service performance is measured as:

$$SL = 1 - \frac{\sum_t U_t}{\sum_t D_t}$$

where U_t represents unmet demand and D_t represents total realized demand [26]. Inventory turnover measures how efficiently stock supports sales or consumption, while average inventory indicates the capital committed to buffer and cycle stock. Stockout frequency captures operational failures across periods [30]. Financial performance is further evaluated through average borrowing utilization, cumulative interest expense, and the frequency of liquidity shortfalls in which available cash becomes insufficient to meet required obligations [27]. Cumulative reward summarizes the RL policy's intertemporal objective. Considering these measures jointly prevents apparently low-cost policies from being classified as superior when savings arise from excessive shortages, inadequate service, or unsustainable liquidity positions [28].

5.4 Mean Deviation and Performance Stability

Performance stability is assessed across repeated simulation runs to determine whether observed policy advantages persist under alternative stochastic realizations [29]. For outcome Y_i and corresponding benchmark or reference value $\hat{Y}_i$, mean absolute deviation is calculated as:

$$MD = \frac{1}{N} \sum_{i=1}^{N} \left| Y_i - \hat{Y}_i \right|$$

When policies are compared directly, $\hat{Y}_i$ represents the corresponding outcome generated by the designated reference policy under the identical simulation trajectory [25]. Lower deviation indicates more stable performance relative to the comparison criterion. Standard deviations and confidence intervals are additionally reported for total cost, service level, borrowing utilization, liquidity shortfalls, and cumulative reward, distinguishing systematic performance improvements from outcomes attributable to stochastic variation [23].

5.5 Statistical Comparison and Robustness Protocol

Statistical evaluation uses repeated random seeds while preserving identical stochastic realizations across competing policies within each run [30]. This paired design controls for scenario-specific variation and permits direct comparison of cost, service, inventory, and financing outcomes. Mean policy differences are accompanied by confidence intervals and appropriate paired significance tests, with non-parametric alternatives applied where distributional assumptions are not satisfied [27]. Robustness is further assessed using out-of-distribution environments that differ from conditions encountered during RL training. These tests modify demand distributions, supplier-disruption probabilities and durations, borrowing-rate persistence, financing shocks, and selected cost parameters [24]. Sensitivity to credit limits and shortage penalties is also examined. A policy is considered robust when its relative performance persists across alternative specifications rather than depending on a narrow combination of training assumptions or favorable stochastic realizations [28].

Table 2. Benchmark Models, Experimental Scenarios, and Evaluation Criteria

Evaluation Component	Specification	Core Assumption/Configuration	Primary Evaluation Criteria	Expected Decision Characteristic
RL policy	Joint adaptive inventory-financing optimization	Learns from evolving operational and financial states	Total cost, service level, liquidity shortfalls, cumulative reward	Adaptive replenishment and financing
Reorder-point policy	Fixed reorder point with safety stock	Predetermined inventory thresholds	Stockouts, average inventory, total cost	Rule-based replenishment
Deterministic optimization	Cost-minimizing inventory policy	Future conditions represented by point estimates	Total and financing costs	Cost-efficient under predictable conditions
Stochastic optimization	Probability-based inventory control	Uncertainty modelled explicitly; financing rule predetermined	Expected cost, service level, inventory utilization	Risk-adjusted replenishment
Demand scenarios	Low/high uncertainty	Alternative demand variances	Service level, stockouts, inventory turnover	Tests demand responsiveness
Supplier scenarios	Low/high disruption	Alternative disruption probabilities/durations	Shortage and disruption costs	Tests supply resilience
Financing scenarios	Low/high borrowing cost	Alternative rate levels and shocks	Interest expense, borrowing utilization	Tests financing responsiveness
Joint-stress scenario	High uncertainty across all dimensions	Demand, supply, and financing deteriorate simultaneously	Total cost, liquidity, service level	Tests integrated resilience
Robustness evaluation	Out-of-distribution conditions	Parameters differ from training environment	Confidence intervals, mean deviation, policy ranking	Tests generalizability

6. RESULTS, ROBUSTNESS, AND DISCUSSION

6.1 Overall Policy Performance

Across the experimental environments, the reinforcement learning policy demonstrates stronger joint operational-financial performance than the benchmark decision rules. The principal improvement is reflected in lower total inventory-financing cost, arising from more selective replenishment and borrowing decisions rather than uniform inventory reductions [31]. Relative to reorder-point and deterministic policies, the RL agent maintains fewer stockout periods while sustaining higher service levels, particularly when demand and supply conditions vary over time [34]. Inventory turnover also improves because replenishment responds dynamically to inventory positions instead of relying exclusively on predetermined thresholds [36]. Financing expenditure declines as borrowing is concentrated in periods when the expected operational benefit of additional inventory exceeds its incremental funding cost. Consequently, average borrowing utilization and liquidity-shortfall frequency remain comparatively controlled [32]. Repeated-run comparisons indicate that these differences are statistically persistent and economically meaningful rather than products of isolated favorable simulation trajectories [38].

6.2 Performance Under Individual and Joint Shocks

Policy differences become more pronounced under adverse conditions. During demand shocks, the RL agent increases replenishment selectively when expected shortage costs exceed additional holding and financing costs [33]. Under supplier disruption, it builds precautionary inventory before persistent shortages emerge and moderates orders when replenishment becomes temporarily ineffective [35]. Interest-rate shocks produce a different response: expensive debt encourages delayed nonessential purchases, reduced borrowing intensity, and greater preservation of available cash [31]. Under simultaneous demand, supply, and financing shocks, the agent does not maximize inventory protection unconditionally. Instead, it balances service continuity against liquidity exposure, accepting additional stock only when expected shortage and disruption losses justify financing costs [37]. This adaptive behavior contrasts with benchmarks whose predetermined rules respond less effectively to interacting shocks.

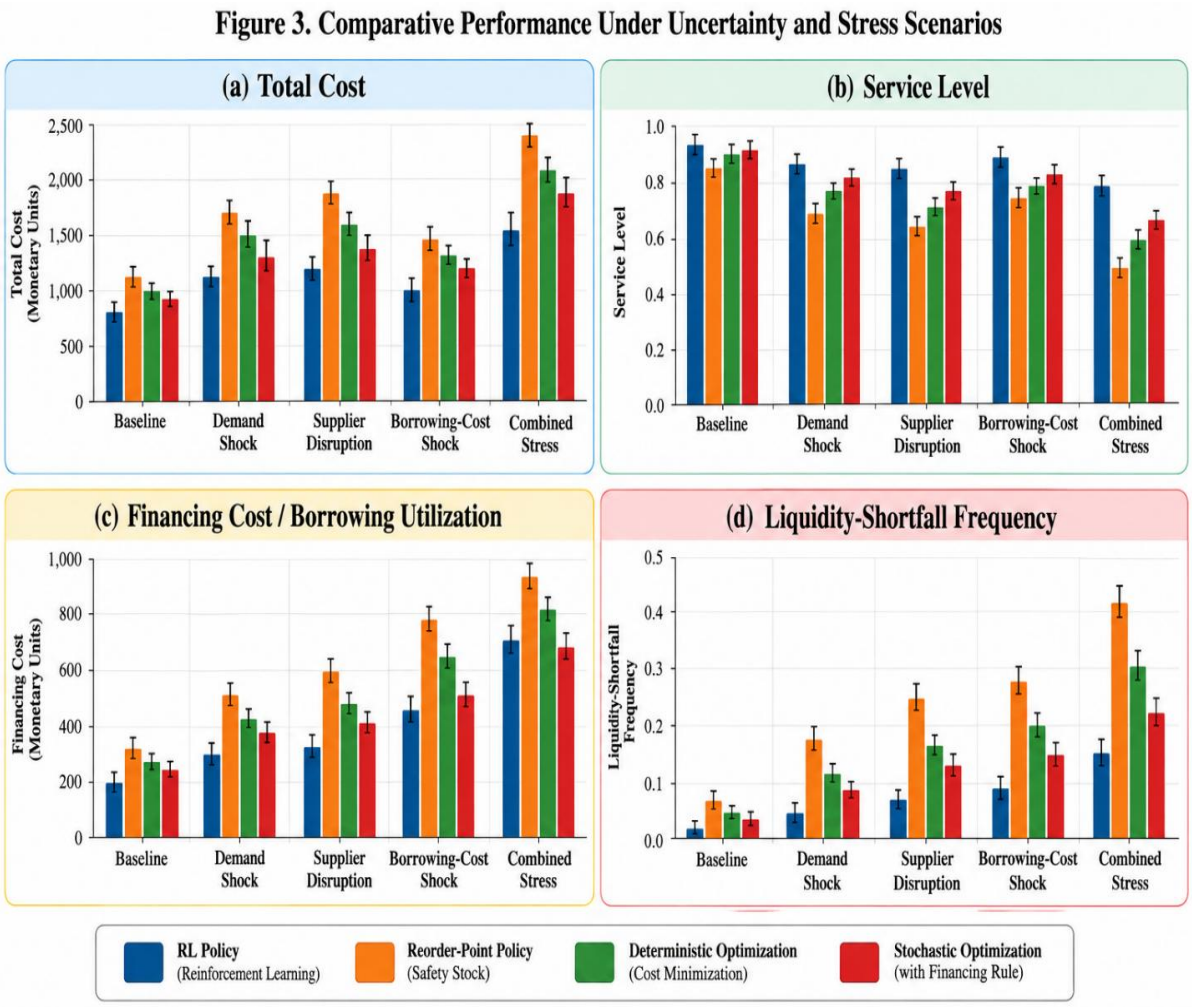


Figure 3. Comparative Performance Under Uncertainty and Stress Scenarios

6.3 Sensitivity and Robustness Analysis

Sensitivity analysis indicates whether the learned policy remains effective when environmental parameters depart from training conditions. Alternative tests vary supplier-disruption probabilities, demand-volatility levels, interest-rate trajectories, shortage penalties, holding costs, and available credit limits [36]. Increasing shortage penalties encourages greater precautionary inventory, whereas higher holding costs discourage persistent excess stock [32]. Tighter credit constraints strengthen cash conservation and reduce the feasibility of debt-funded replenishment. Similarly, greater disruption probabilities increase the value of inventory buffers, particularly when borrowing remains affordable [38]. Across out-of-distribution conditions, policy rankings and confidence intervals provide evidence of stability. Performance deterioration under extreme parameter shifts is interpreted as a boundary of policy generalizability rather than ignored [34].

6.4 Interpretation of the Learned Inventory-Financing Strategy

The learned strategy is economically important because it treats inventory as both an operational buffer and a use of financial capital [35]. When supplier risk rises, the agent may rationally accept higher holding costs because additional stock reduces expected disruption and shortage losses. However, the same inventory accumulation becomes less attractive when borrowing rates rise sharply [37]. The policy therefore changes financing intensity alongside replenishment rather than applying a fixed funding rule. Cash is preserved when marginal financing costs exceed expected inventory benefits, while borrowing expands when additional liquidity prevents economically larger stockout losses [33]. This conditional behavior explains the policy's advantage.

6.5 Managerial and Theoretical Implications

The findings imply that procurement, treasury, and supply-chain planning should not operate as isolated decision functions [31]. Procurement managers require visibility into financing conditions, while treasury teams benefit from understanding how liquidity decisions influence inventory resilience and customer service [36]. For CFOs, the framework provides a mechanism for evaluating working capital as a dynamic operational-financial system. Theoretically, it extends inventory optimization by embedding financing constraints and adaptive learning directly within sequential replenishment

decisions under uncertainty [38].

7. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

7.1 Principal Findings and Contributions

This study develops an integrated reinforcement learning framework in which inventory replenishment and financing are optimized as interconnected sequential decisions. By incorporating demand uncertainty, supplier disruptions, liquidity positions, and variable borrowing costs within a common decision environment, the framework extends inventory optimization beyond conventional operational boundaries. The findings demonstrate that adaptive policies can reduce aggregate inventory-financing costs, moderate unnecessary borrowing, preserve liquidity, and maintain customer service under multidimensional uncertainty. Importantly, these improvements arise from dynamically balancing inventory protection against financing exposure rather than pursuing minimum inventory or minimum borrowing as independent objectives across periods.

7.2 Limitations

Several limitations qualify these findings. Simulated environments may not fully reproduce real-world behavioural and institutional complexities, while transaction costs and supplier interdependence remain simplified. Model performance also depends on data quality, parameter assumptions, and environmental stability. Furthermore, limited reinforcement-learning interpretability may constrain managerial acceptance and implementation within highly regulated financial settings.

7.3 Future Research and Closing Implication

Future research should extend the framework to multi-supplier networks, multiple SKUs, dynamic credit limits, inflationary environments, and trade-credit arrangements. Explainable reinforcement learning could improve decision transparency, while ESG-related supplier risks may broaden disruption modelling. Validation using firm-level operational and financial data would further establish practical applicability. Ultimately, inventory management should be understood as a dynamic operational-financial decision system, rather than solely an operational optimization problem.

REFERENCE

1. Feng L. Joint optimization algorithm for vehicle scheduling and supply chain inventory management based on multi-agent deep reinforcement learning. Neural Computing and Applications. 2025 Dec;37(34):28643-69.

2. Oluwatobi Alebiosu. Quantifying regulatory delay risk transmission across utility-scale renewable energy projects in hydrocarbon-dominant economies. Int J Hydropower Civ Eng 2023;4(1):35-47. DOI: 10.22271/27078302.2023.v4.i1a.89

3. Nozari H, Nassar S, Szmelter-Jarosz A. Fuzzy multi-objective optimization model for resilient supply chain financing based on blockchain and IoT. Digital. 2025 Jul 31;5(3):32.

4. Belhadi A, Kamble SS, Mani V, Benkhati I, Touriki FE. An ensemble machine learning approach for forecasting credit risk of agricultural SMEs' investments in agriculture 4.0 through supply chain finance. Annals of Operations Research. 2025 Feb;345(2):779-807.

5. Kalaichelvan K, Ramalingam S, Dhandapani PB, Leiva V, Castro C. Optimizing the economic order quantity using fuzzy theory and machine learning applied to a pharmaceutical framework. Mathematics. 2024 Mar 11;12(6):819.

6. Filani OM, Nwokocha GC, Alao OB. Predictive vendor risk scoring model using machine learning to ensure supply chain continuity and operational resilience. management. 2021 Jul;8:9.

7. Chukwunweike J. Coordinating PLC-based load shedding and variable-speed drives to prevent transformer overloading during industrial production peaks. Int J Electr Data Commun. 2025;6(2 Pt B):118-135. doi:10.22271/27083969.2025.v6.i2b.109.

8. Guo H, Song L, Yang F, Guo X. Supply chain finance: A strategic enabler for purchasing and supply chain management. IEEE Transactions on Engineering Management. 2025 Oct 7.

9. Yilmaz Eroglu D. Tariff-sensitive global supply chains: Semi-Markov decision approach with reinforcement learning. Systems. 2025 Aug 1;13(8):645.

10. Bhetwal S, Bastakoti R, Acharya N, Gupta GK, Bhandari AB. Benchmarking Local LLMs for Natural-Language-to-SQL Querying in Biopharmaceutical Manufacturing: An Empirical Benchmark on Consumer-Grade Hardware. arXiv preprint arXiv:2606.01338. 2026 May 31.

11. Bussieweke F, Mula J, Campuzano-Bolarin F. Optimisation of recovery policies in the era of supply chain disruptions: a system dynamics and reinforcement learning approach. International Journal of Production Research. 2025 Mar 4;63(5):1649-73.

12. Asante, A., Daramola, G., Davis, R. W., & Kumar, S. (2025). Environmentally Friendly Chelation for Enhanced Algal Biomass Deashing. *Phycology*, *5*(3), 32. https://doi.org/10.3390/phycology5030032

13. Wu J, Zhang Z, Zhou SX. Credit rating prediction through supply chains: A machine learning approach. Production and Operations Management. 2022 Apr;31(4):1613-29.

14. Zhu Y, Zhou L, Xie C, Wang GJ, Nguyen TV. Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. International journal of production economics. 2019 May 1;211:22-33.

15. Eseoghnene OM. Applying blockchain-based fraud prevention mechanisms for securing digital banking transactions and strengthening customer trust in financial institutions worldwide. International Journal of Research Publication and Reviews. 2025;6(9):351-67.

16. Wei Y. [Retracted] A Machine Learning Algorithm for Supplier Credit Risk Assessment Based on Supply Chain Management. International Transactions on Electrical Energy Systems. 2022;2022(1):4766597.

17. Swarnali SH. Predictive Financial Volatility Analytics Using Machine Learning Pipelines in ERP-Integrated Enterprises. American Journal of Interdisciplinary Studies. 2024 Dec 3;5(04):134-81.

18. Oluwatobi Alebiosu. Comparative analysis of public-private partnership legislation driving electricity infrastructure expansion, renewable energy investment, and economic resilience nationally. Int J Foreign Trade Int Bus 2025;7(2):239-249. DOI: 10.33545/26633140.2025.v7.i2c.262

19. Giannelos S. Reinforcement learning in energy finance: A comprehensive review. Energies. 2025 May 23;18(11):2712.

20. Hongjin S. Analysis of risk factors in financial supply chain based on machine learning and IoT technology. Journal of Intelligent & Fuzzy Systems. 2021 Apr 12;40(4):6421-31.

21. Daramola G. Reducing environmental compliance failures through automated regulatory reporting and real-time risk escalation in industrial operations. Int J Res Publ Rev. 2025;6(3):10328-10340. doi:10.55248/gengpi.06.0326.2410.

22. Zhang W, Yan S, Li J, Peng R, Tian X. Deep reinforcement learning imbalanced credit risk of SMEs in supply chain finance. Annals of Operations Research. 2024 Mar 20:1-31.

23. Cui Y, Yao F. Retracted article: Integrating deep learning and reinforcement learning for enhanced financial risk forecasting in supply chain management. Journal of the Knowledge Economy. 2024 Dec;15(4):20091-110.

24. Kwaku Gyamfi Boamah. AI-Driven Threat Intelligence Adoption in Security Operations Centers: Human Skill Erosion or Analyst Augmentation? A Comparative Case Study of Resource-Constrained SOCs in Emerging Markets. International Journal of Science and Research Archive, 2026, 19(03), 025-033. Article DOI: https://doi.org/10.30574/ijsra.2026.19.3.1226.

25. Taiwo S. Financial risk optimization in consumer goods using Monte Carlo and machine learning simulations. World Journal of Advanced Research and Reviews. 2022 Jan 1.

26. Badakhshan E, Ball P. Applying digital twins for inventory and cash management in supply chains under physical and financial disruptions. International Journal of Production Research. 2023 Aug 3;61(15):5094-116.

27. Oluwatobi Alebiosu. Engineering stakeholder alignment frameworks for resolving land rights, environmental compliance, and project delivery conflicts. Int J Appl Res 2024;10(12):401-412. DOI: 10.22271/allresearch.2024.v10.i12e.13917

28. Hasan MR, Rahman MA, Gomes CA, Nitu FN, Gomes CA, Islam MR, Shawon RE. Building robust AI and machine learning models for supplier risk management: A data-driven strategy for enhancing supply chain resilience in the USA. Advances in Consumer Research. 2025 Jul 1;2(4):1152-71.

29. Lu Q, Zhou W. Cost Optimization Method for Procurement and Inventory of NEV Manufacturers Based on Deep Reinforcement Learning. Electrical & Electronic Engineering Research.;5(1):41.

30. George Daramola. Assessing environmental impacts of rapid data centre expansion on energy consumption, water resources, and emissions. Int. J.

Res. Eng. 2023;5(1):38-47. DOI: 10.33545/26648776.2023.v5.i1a.191

31. Zhang Y, He L, Zheng J. A deep reinforcement learning-based dynamic replenishment approach for multi-echelon inventory considering cost optimization. Electronics. 2024 Dec 27;14(1):66.

32. Yang Y, Wang M, Wang J, Li P, Zhou M. Multi-agent deep reinforcement learning for integrated demand forecasting and inventory optimization in sensor-enabled retail supply chains. Sensors. 2025 Apr 11;25(8):2428.

33. Alebiosu O. AI and cybersecurity governance and legal compliance challenges affecting digital energy infrastructure, smart grids, and critical utility resilience systems. Int J Comput Appl Technol Res. 2026;15(8):1-14. doi:10.7753/IJCATR1508.1001.

34. Lu X, Wang H, Peng Z, Liao C, Liu C. Dynamic optimization of multi-echelon supply chain inventory policies under disruptive scenarios: A deep reinforcement learning approach. Symmetry. 2025 Dec 4;17(12):2078.

35. Kumar L, Khedlekar UK. A hybrid AI and blockchain framework for intelligent supply chain optimization with reinforcement learning-based inventory management, stochastic demand forecasting, and sustainable supplier selection. Soft Computing. 2026 Jan;30(1):419-47.

36. Olatunji AO, Gyamfi LN, Abubakar SA, Mirambo JB, Kim K, Jide-Jegede M, Shehu TS, Giwa AA. A Deep Reinforcement Learning Framework for Dynamic Inventory Control and Cost Optimization Under Demand Uncertainty. IEEE Access. 2026 Apr 13.

37. Bhandari, A. B. (2026). Quality Control Frameworks for Electronic Health Record Data Workflows in Resource-Constrained Settings: A Structured Review with Implications for Global Health Information Management. Current Journal of Applied Science and Technology, 45(8), 1–16. https://doi.org/10.9734/cjast/2026/v45i84728

38. Badakhshan A, Badakhshan E, Saad S, Bahadori R. From Heuristics to Reinforcement Learning: Integrated Operational–Financial Control of Supply Chains Under Demand Disruption. Applied Sciences. 2026 Jun 5;16(11):5712.