IJSOM International Journal of Supply and Operations Management

Reinforcement Learning for Blockchain-Enabled Supply Chain Network Design

Sara Hasheminezhad [1], **Mahsa Moayed** [2], **Ardavan Babaei** [3,4], **Erfan Babaee Tirkolaee** [5,6] ✉

1. Faculty of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran. E-mail: shashemi289@gmail.com
2. Faculty of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran. E-mail: moayedmahsa@gmail.com
3. Faculty of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran. E-mail: ardavan.babaei@kntu.ac.ir
4. Department of Industrial Engineering, Istinye University, Istanbul, Turkey.
5. Corresponding author, Department of Industrial Engineering, Istinye University, Istanbul, Turkey. E-mail: erfan.babaee@istinye.edu.tr
6. Department of Mechanics and Mathematics, Western Caspian University, Baku, Azerbaijan.

Article Info

Article type:
Research Article

Article history:
Received February 12, 2026
Received in revised form May 5, 2026
Accepted May 29, 2026
Available online May 29, 2026

***Keywords*:**
blockchain technology
supply chain management
reinforcement learning
Q-learning
optimization

ABSTRACT

Objective: The objective of this study is to examine the optimal design of blockchain-enabled supply chain networks using reinforcement learning (RL). The goal is to develop an integrated network framework considering simultaneously supply chain connectivity, blockchain node activation and cross-layer integration decisions, to minimize the total network cost under operational constraints.

Methods: The problem is formulated as a combinatorial optimization problem and modeled as a two-layer network of supply chain entities and blockchain nodes. A Q-learning framework is used to explore feasible network configurations in the presence of stochastic costs. The model includes various cost factors, connectivity requirements, constraints for blockchain activation, and penalty mechanisms to ensure feasible solutions.

Results: The results of experiments show that the proposed RL-based method is able to find out the cost-effective network designs in accordance with network connectivity constraints. The convergence behavior is shown to be stable throughout different runs, and the method converges within an earlier stage of training. The sensitivity analysis indicates that the increment of minimum number of active blockchain nodes will increase the cost of the designed network, highlighting the trade-off between blockchain deployment requirements and economic efficiency. The resulting network structure turns out to be feasible, connected, and interpretable.

Conclusion: Based on the findings, the research proves the feasibility of applying reinforcement learning to blockchain-enabled supply chain network design problems. The proposed framework provides an approach that can be utilized evaluating blockchain integration strategies and balancing implementation costs with operational requirements. The research further creates a base for further research involving larger-scale networks, dynamic environments, and multi-objective optimization settings.

Cite this article: Hasheminezhad, S., Moayed, M., Babaei, A., & Tirkolaee, E. (2026). Reinforcement learning for blockchain-enabled supply chain network design. *International Journal of Supply and Operations Management*, *XX*(X), pages. https://doi.org/10.22034/ijsom.2026.111012.3553

Publisher: Kharazmi University
DOI: https://doi.org/10.22034/ijsom.2026.111012.3553

1. Introduction

Statistical reports confirm that world expenditure on blockchain technology totaled $6.6 billion in the year 2021. Market projections indicate the figure will hit over $10 billion in the year 2024, before increasing to a historic high of $19 billion. The technological sector has impressive potential for growth, with a compound annual growth rate (CAGR) of over 69%, as the market capitalization of blockchain technology is poised to hit $25 billion by the year 2025 (Bhutani et al., 2019). In North American regions, blockchain use in agricultural and food supply chains signifies expanding reach, yielding over 31% market share in 2023. The driving force is high-profile incidents surrounding food safety crises and consumers' demands for supply chain transparency. North American markets are subject to stringent regulatory regimes, such as FDA compliance rules, organic labeling requirements, and documentation of sustainability, which in total drive blockchain adoption. The UN Food and Agriculture Organization (FAO) documents that supply chains account for 40% of food wastage in North American communities (Wadhwani et al., 2024).

Logistics blockchain market capitalization stood at USD 19.5 billion in 2023, with growth trajectories expected to reach a CAGR of over 45% from 2024 to 2032. The combination of artificial intelligence with blockchain infrastructure is accelerating the implementation of predictive analytics within logistics processes. Organizations implement AI algorithms to process blockchain datasets to facilitate demand forecasting, inventory coordination, and supply chain disruption prediction (Gujar et al., 2024). Supply chain tracking blockchain solutions amassed $2.1 billion in market size during 2023, with growth opportunities forecasting a CAGR of over 31.9% over the 2024-2032 forecast period. Cloud-based market areas are forecast to expand to $19.2 billion by 2032 (Wadhwani et al., 2024). The 2021 Blockchain Survey revealed that 76% of key executives of high-revenue organizations worldwide attested to the high significance of blockchain to their respective markets in subsequent years. Further, 81% of the world's top 100 market-capitalization publicly listed companies have implemented blockchain technologies, as a strong influence (Deloitte, 2021).

Supply chain constitutes end-to-end processes involving the production and distribution of products or services (Abualigah et al., 2023). Economic considerations were the sole concern for Supply Chain Management (SCM) in the last several decades. Yet, resource scarcity, supplier unreliability, environmental degradation, and social issues have brought sustainability objectives to the center stage (Dobrovnik et al., 2018). Closed-loop supply chain (CLSC) frameworks involve product recovery as well as remanufacturing processes within supply chain operations. These systems aim to minimize waste generation and ensure sustainability by facilitating the reuse of products along with minimizing the use of new raw materials (Goli, 2023). CLSC structures integrate conventional logistics with reverse logistics practices (Zarreh et al., 2024; Bhattacharya et al., 2024). Reverse logistics entails product flow from customers to producers via distribution channels (Rajabi-Kafshgar et al., 2023). Current SCM challenges demand more performance measures, particularly highlighting traceability and transparency initiatives (Marr, 2018).

CLSC models and optimization problem configurations have garnered immense interest when solving remanufacturing system issues optimally. Forward and reverse supply flows complement each other in closed-loop supply chain networks to enhance economic, environmental, and social performance measures. Return processes are managed by manufacturers on their own in CLSC structures. Product remanufacturing, including restoration processes involving the transformation of raw materials to functional condition, is a primary sustainable practice (Asghari et al., 2022). Throughout the process of sustainable development, logistics has been regarded as an important component in supply chain networks. CLSC systems have increasingly dominated, guiding different industries to sustainability goals. Such networks facilitate both forward and reverse movements to meet market demands while preparing returned goods for secondary use (Salehi-Amiri et al., 2021).

Blockchain technology has gained immense attention as the underlying technology for digital currencies such as Bitcoin (Min, 2019). Blockchain technology implementation in SCM is an emerging research area, and very few studies have concentrated primarily on developed economies such as the United States, the United Kingdom, and

Germany (Queiroz et al., 2020; Khokhar et al., 2024). Blockchain technology is a new, decentralized, distributed ledger-based infrastructure that offers security, integrity, and accessibility of data and transactions (Dutta et al., 2020). This technology keeps all the supply chain data in open and immutable structures, free from tampering or interference (Longo et al., 2019). A number of business applications have integrated blockchain technology, including SCM, healthcare settings, and peer-to-peer transactions, to surmount security and privacy limitations of centralized environments while establishing participant trust by removing third parties. In terms of accessing the network, blockchain operates on two significant kinds: public and private systems. Public blockchains do not require permission to be on the network, while private blockchains typically have access controls. Private blockchain models operate based on trust foundations, which enable businesses to develop decentralized applications with real-time finality, improved performance, and improved privacy protection (Fan et al., 2022).

Private blockchains are more efficient, reliable, and secure compared to public blockchains. Ethereum private blockchain deployments are used most often by the majority of enterprises for a range of uses, such as vehicle data certification, financial transactions, and SCM. Proof of Authority (PoA) consensus algorithm rather than Proof of Work (PoW) is usually utilized in private networks due to advantages like participant identity check, deterministic consensus, and operational efficiency (Samuel et al., 2021). Blockchain technology offers high value creation potential across many fields of SCM, finance, logistics, risk management, security, and information management. Blockchain technology also provides more transparency, traceability, efficiency, and information security in SCM processes. Moreover, new-generation technologies such as smart contracts and Internet of Things (IoT) can be integrated with blockchain technology in supply chains to develop additional value (Moosavi et al., 2021).

By eliminating middlemen, like traditional banking institutions, blockchain technology minimizes risks such as cyber-attacks, opacity, privacy violations, and economic or political instability. The technology further increases security controls, reduces transaction costs, enhances supply chain visibility, and creates improved communication among business partners (Min, 2019). Developing blockchain technology-based SCM systems requires more than selecting the most optimal blockchain technology for companies; it requires consideration of characteristics such as data dependability, dual storage architecture deployment, and appropriate product tracking device selection (Azzi et al., 2019). Research indicates decentralization, security, and immutability are of greatest priority blockchain technology features for supply chain applications, while consensus mechanisms rank at the lowest priority. Metrics of supply chain performance guided by blockchain technology features; i.e., transparency, traceability, reliability, asset management effectiveness, and accountability, are ranked higher as more crucial than cost and responsiveness drivers (Çıkmak et al., 2024).

The novelty of this study lies in the integrated formulation of a two-layer supply chain–blockchain network design problem. The proposed model jointly captures supply chain configuration, blockchain node activation, and connectivity decisions within a unified optimization framework. Thus, the main contribution of this work is the problem formulation and integrated modeling structure, while Q-learning is employed as a standard solution strategy to address the resulting combinatorial search problem.

2. Research Background

Pournader et al. (2019) conducted a systematic review with the aim of investigating the current application of blockchain technology in supply chain, logistics, and transport management. Through a co-citation analysis, they identified four prevailing clusters, which they labeled as 'the 4Ts': Technology, Trust, Trade, and Traceability/Transparency. Kamilaris et al. (2019) presented the disruptive possibility of blockchain technology in food and agricultural supply chains, highlighting ongoing projects and initiatives, and addressing general implications, challenges, and outlook. The authors outlined the main benefits, including enhanced traceability at every supply chain phase, better food safety and integrity, good support to small farmers, efficient minimization of waste, greater environmental sensitivity, and more effective supply chain supervision and management. However, they also recorded

a myriad of hindrances and obstacles to further wider utilization, such as technical issues (e.g., the oracle problem, scalability, privacy issues), lack of accessibility and awareness, governance problems, absence of definite regulatory frameworks, and an enormous digital divide between developed and developing countries.

Shoaib et al. (2023) aimed to identify and rank challenges to blockchain adoption in SCM and develop a taxonomic model of the challenges. They employed the fuzzy best-worst method (F-BWM), chosen for its capability to remove data uncertainty and provide more stable results with fewer pairwise comparisons compared to other multi-criteria decision-making (MCDM) methods. The study identified 20 challenges and found that lack of storage capacity/scalability (CH8) and lack of data privacy (CH16) were the most considerable barriers. Dehshiri and Amiri (2024) examined the evaluation of solutions for the implementation of blockchain in sustainable supply chains, particularly in the supply chain of agricultural products. They presented a new hybrid decision method by integrating the Step-wise Weight Assessment Ratio Analysis (SWARA) and Combined Compromise Solution (CoCoSo) methods with the help of Z-numbers to consider ambiguity and reliability in expert judgments. The results of the study highlighted the importance of developing blockchain infrastructure throughout all levels of the supply chain, industrial collaboration, and favorable legal environments, and enhancing collaboration with technical specialists for effective and sustainable adoption of blockchain. This established structure provides accurate, reliable, and flexible results to assess complex technological solutions.

Babaei et al. (2023a) developed a new bi-objective optimization model to design blockchain-based supply chain networks, the first one to include blockchain-sourced transparency in three-level supply chain network planning. The aim was to minimize total costs and enhance transparency. To address the stochasticity of the model and its bi-objectivity, CCP and FGP were used by them, respectively. The authors also designed a superior Branch and Efficiency (B&E) algorithm to contrast solutions based on transparency, cost, and service. Babaei et al. (2024c) researched blockchain's innovative impact on SCM within the oil and gas industry, specifically by developing successful blockchain adoption strategies under risk and uncertain environments. They introduced a data-driven decision-making system, using Data Envelopment Analysis (DEA) models in evaluating various blockchain strategies such as single-use, localization, substitution, and transformation. When applied to the Norwegian oil and gas industry, they concluded that the single-use strategy is the least risky and most cost-effective of those modeled. Babaei et al. (2025a) introduced a two-stage, multi-objective production and distribution planning problem in a two-stage supply chain. The novel model addresses the problem of incomplete information sharing between manufacturers (followers) and distributors (leaders) through the application of Data Envelopment Analysis (DEA). The model took into account significant issues of contemporary times, like traffic congestion, transparency through blockchain technology, uncertain conditions of demand, and post-production operations such as rework and reverse logistics. The research translated this complex bi-level problem into a single-level model via Karush-Kuhn-Tucker (KKT) conditions and solved it via fuzzy goal programming, ultimately evaluating the solutions for both optimality and efficiency.

Babaei et al. (2024b) introduce a tri-objective optimization model for the design of a supply chain-blockchain (SC-BT) network by taking into account transparency, emission, and costs. The model involves a two-step authentication process and takes into account uncertainty in blockchain participants. A Branch and Efficiency (B&E) algorithm was developed to obtain cost-effective, environmentally friendly solutions. In the case study of a three-echelon supply chain, the algorithm yielded a cost reduction of 16% and an emissions decrease of 13%, proving its effectiveness in green blockchain implementation. Babaei et al. (2025c) investigated the effectiveness of blockchain technology implementation within Renewable Energy Supply Chains (RESCs) while focusing on determining and analyzing resulting challenges in ideal and non-ideal situations. They proposed a multi-phase Data Envelopment Analysis (DEA)-based optimization method to identify these challenges in a structured way, working through efficiency measurement, deviation analysis, and super-efficiency ranking. The proposed model was identified as being highly robust and possessing superior discriminatory power compared to existing practices, demonstrating its flexibility and ability to account for uncertainties by acknowledging the direct calculation of optimal efficiency and deviation values.

Choi (2023) analyzed supply chain finance problems in the setting of fashionable product supply chains, comparing blockchain-based versus traditional bank-based systems. Based on analytical models using the fundamental newsvendor problem setting with one manufacturer and retailer who encounter a revenue sharing agreement and Nash bargaining, the study applied mean-risk analysis to measure performance. The research demonstrated analytically that the blockchain-based supply chain has a lower level of operational risk compared to its traditional counterpart. The study also noted that blockchain leads to an optimum product quantity and inventory service level that is lower, which may lead to a higher likelihood of stockout for the consumer, but confirmed the robustness of its findings based on other common supply chain contracts and various measures of risk. Babaei et al. (2025b) proposed four blockchain adoption models for green supply chains. They classified the approaches based on tracing and authority type. The results showed that link-based/component tracing is cost-effective, which supports informed decision-making for blockchain integration. Rahmanzadeh et al. (2019) explored the alignment of new product development and supply chain strategic planning by using a blockchain-based platform to address intellectual property (IP) protection problems in open innovation (OI). They formulated a fuzzy mathematical model to optimize strategic decisions with environmental epistemic uncertainty and created a blockchain-based idea registration mechanism, safeguarding the innovators' rights. A home appliance case study revealed that companies were able to achieve good designs through an investment of approximately 1% of the total supply chain cost, and the registration mechanism significantly reduced the cost of using non-original designs by more than 41%. Recently, Santoso et al. (2025) investigated how private blockchain technology strengthens trust among supply chain partners by improving transparency, data security, and record immutability. Using qualitative case studies of major firms (e.g., Amazon, Maersk, Microsoft, Walmart, Alibaba), the research showed the blockchain enhances collaboration, operational efficiency, and competitive advantage.

3. Materials and Methods

The integration of blockchain technology into supply chain networks is an optimization problem that has several conflicting objectives and operational constraints that need to be met. The preliminary decision for organizations is how to choose which supply chain entities are to be connected with blockchain nodes, how blockchain nodes are to be connected, and which blockchain nodes should be enabled to achieve cost-effective network configurations. The problem is complex due to several reasons: interdependence of operations of the supply chain and blockchain infrastructure, stochasticity of the cost of connection and operations, necessity of network connectivity and operability, and necessity of sufficient activation of blockchain nodes to make the network robust. The multi-dimensional aspect of this problem prevents traditional optimization techniques due to exponential growth in the number of feasible configurations and dynamic supply chain environments.

This study formulates the optimal design of a blockchain-enabled supply chain network as a combinatorial optimization problem, with the objective of determining the optimal configuration to maximize the overall cost of the network in relation to the operational constraints. An agent-based type of reinforcement learning (RL), specifically Q-learning, is employed to address this complex optimization problem, with flexibility in handling random cost variations as well as systematic search of the solution space. To formally describe the network design problem addressed in this study, the mathematical formulation of the proposed model is presented below.

3.1 Problem Formulation

In this study, the system is modeled as a two-layer network composed of supply chain (SC) nodes and blockchain (BC) nodes. The main objective is to determine an efficient connectivity structure among these nodes while minimizing the overall cost of establishing and operating the network.

In order to facilitate the understanding of the mathematical model, Table 1 summarizes the notations, i.e., sets, parameters, and decision variables, used to represent the structure of the network as well as cost components in both layers.

Table 1. Notations and descriptions.

Sets	
i, k	Index of supply chain entities (suppliers, manufacturers, distributors, retailers); $i \neq k$
j, l	Index of blockchain participants (validators, miners, stakeholders); $j \neq l$
Parameters	
C_{ik}^{SS}	Costs for establishing/maintaining connections between nodes *i* and *k* (transportation, communication, coordination)
C^{A}	Costs for onboarding blockchain node *j* (hardware, software, security, training)
C_{jl}^{BB}	Costs for connections between nodes *j* and l (network communication, consensus, computational resources)
C_{ij}^{SB}	Costs for connections between supply chain node *i* and blockchain member *j* (integration, data transmission)
Decision variables	
a_j	1, if node *j* is included in the blockchain network; 0 otherwise
Y_{ik}	1, if a connection is established between supply chain nodes *i* and *k*; 0 otherwise
Z_{jl}	1, if a connection is established between blockchain nodes *j* and l; 0 otherwise
X_{ij}	1, if a connection is established between supply chain member *i* and blockchain node *j*; 0 otherwise

The objective is to minimize the total network cost, including connection costs and blockchain activation costs, as displayed by Eq. (1):

$$\text{minimize } Z = \sum_{i,j} C_{ij}^{SB} X_{ij} + \frac{1}{2}\sum_{i,k} C_{ik}^{SS} Y_{ik} + \frac{1}{2}\sum_{j,l} C_{jl}^{BB} Z_{jl} + \sum_{j \in B} C^{A} a_j \quad (1)$$

The first term represents the connection cost between supply chain nodes and blockchain nodes. The second and third terms represent internal connectivity costs within each layer. The final term represents the cost of activating blockchain nodes.

3.1.1 Computational Complexity

The proposed network design problem is a binary combinatorial optimization model in which all decision variables; i.e., x_{ij}, y_{ik}, z_{jl}, and a_j, are binary. Let $|S|$ and $|B|$ denote the number of supply chain nodes and blockchain nodes, respectively. The total number of binary variables is therefore computed by Eq. (2):

$$|S| \times |B| + \frac{|S|(|S|-1)}{2} + \frac{|B|(|B|-1)}{2} + |B|. \quad (2)$$

Since each variable can take two possible values, the size of the feasible state space grows exponentially as $O(2^n)$, where n is the total number of binary decisions. Such exponential growth makes traditional exact optimization methods computationally intractable even for moderately sized networks. Moreover, the feasibility of a candidate solution depends on several interacting connectivity constraints, which intensify the combinatorial explosion of the decision space. As a result, classical optimization techniques cannot efficiently explore all feasible configurations. In contrast, RL can navigate this exponentially large space through iterative exploration and exploitation, without the need to enumerate all possible solutions.

RL is particularly suitable for the present problem since the decision process is inherently sequential and dynamic, and decisions are made iteratively based on the current system state and feedback from the environment. Unlike

metaheuristic approaches such as genetic algorithms (GAs) or particle swarm optimization (PSO), which repeatedly search over a population of candidate solutions, RL learns an adaptive decision policy through continuous interaction with the problem environment. Furthermore, although mixed-integer linear programming (MILP) can provide exact optimization under explicit mathematical formulations, it often becomes computationally prohibitive for large-scale and highly dynamic instances. Therefore, the present study focuses on establishing an RL-based baseline to demonstrate feasibility, while an extensive comparative analysis against GA, PSO, MILP, and other optimization paradigms is reserved for future research.

3.2 Network Architecture

It is composed of two six-node layers that refer to different functional dimensions of the integrated system. The Operational layer is made up of operational nodes like manufacturers, suppliers, warehouses, retailers, and logistics. The nodes in this layer undertake physical goods transfer, information sharing, and operational decision-making processes inherent in supply chain operations. The Blockchain layer manages data openness, transaction permission, unalterable record keeping, and smart contract regulation, and provides the technological foundation necessary for safe and open operations.

There are three forms of connections in such architecture, all of which serve various purposes in the overall network. SC-SC connections represent node-to-node connections in the supply chain, through which operational connections, information sharing, and physical products flow between various entities within the supply chain. They enable coordination and cooperation among suppliers, producers, and distributors. BC-BC connections are those connections between blockchain nodes through which synchronization of information, consensus actions, and distributed ledger management are carried out within the blockchain network. They impose identical information on all the blockchain nodes and enable them to participate in the consensus process. SC-BC relationships are blockchain node-supply chain node relationships providing data inputs of data, transactional recording, and verification processes linking the physical and virtual spaces of the coupled system.

Every type of connection has incurred costs that are a function of distance, data volume, transaction amount, and security requirements. They are modeled as random variables for the sake of capturing real-world randomness and variability of operational conditions. In addition, activation of blockchain nodes has incurred fixed costs of deploying infrastructure, maintenance, and operational requirements, which are the investment in deploying and maintaining blockchain infrastructure within the network.

3.3 Operational Constraints

Several operational constraints are incorporated to ensure the feasibility and logical consistency of the proposed network design. First, the model requires that a minimum number of blockchain nodes be activated in the network in order to guarantee sufficient blockchain infrastructure. This requirement is expressed by ensuring that the total number of activated blockchain nodes is at least equal to a predefined threshold L, as denoted by Constraint (3):

$$\sum_{j \in B} a_j \geq L. \tag{3}$$

Furthermore, if a blockchain node is activated, it must be connected to at least one supply chain node. This condition guarantees that every active blockchain node actually participates in the network and supports at least one entity within the supply chain layer. Without such a constraint, a blockchain node could be activated without providing any functional connectivity, as shown by Constraint (4):

$$\sum_{i \in S} X_{ij} \geq a_j \qquad \forall j \geq B. \tag{4}$$

In addition, the model ensures that there is at least one connection between the supply chain layer and the blockchain layer. This requirement prevents the formation of isolated layers and guarantees that the blockchain infrastructure is effectively integrated with the supply chain network (Constraint (5)).

$$\sum_{i \in S} \sum_{j \in B} X_{ij} \geq 1. \tag{5}$$

To maintain logical consistency in the network structure, self-connections within each layer are not permitted. In other words, a supply chain node cannot be connected to itself, and similarly, a blockchain node cannot establish a link with itself (Constraint (6)).

$$Y_{ii} = 0 \quad \forall i \in S; Z_{jj} = 0 \quad \forall j \in B. \tag{6}$$

Finally, all decision variables in the model are defined as binary variables. This means that each connection or activation decision can take only two possible values: 1 if the connection or activation exists, and 0 otherwise. This binary structure reflects the discrete nature of network design decisions in the proposed blockchain-enabled supply chain system (Constraint (7)).

$$X_{ij}, Y_{ik}, Z_{jl}, a_j \in \{0,1\} \quad \forall i, j, k, l. \tag{7}$$

3.4 Cost Model

The total network cost is comprised of four components with clear weighting to reflect network characteristics and operating realities. SC-BC connectivity cost is given full cost weighting to reflect the critical importance of supply chain-blockchain integration toward achieving the overall objectives of transparency, traceability, and operational efficiency. SC-SC connection costs are 0.5-weighted due to their symmetry in operating relationships, where any connection between two supply chain partners is a two-way street and provides mutual benefit and collaborative operating value. BC-BC connection costs are also 0.5-weighted due to symmetry in blockchain network connections, where each connection provides bidirectional data sharing and consensus participation. BC node activation charges represent sunk cost infrastructure installation and upkeep charges, reflecting the intensive capital required to install blockchain functionality within the network.

To ensure network feasibility and impose operational constraints, penalty mechanisms are employed as a penalty for infeasible configurations. A $1,000,000 penalty cost is incurred if fewer than, say, three blockchain nodes are up, and this ensures that the network is still sufficiently redundant and fault-tolerant. Similarly, a penalty of $1,000,000 is incurred for not associating active blockchain nodes with any supply chain node to prevent isolated blockchain infrastructure with no operational value and no active SC-BC connections, ensuring that the integration between the supply chain layer and blockchain layer is maintained. Such heavy penalties ensure that the optimization process maintains viable configurations as an utmost priority and explores cost-saving alternatives within realistic network constraints.

3.5 Reinforcement Learning Framework

RL is employed in this study as a decision-making framework to address the combinatorial nature of the proposed blockchain enabled supply chain network design problem. As discussed in the computational complexity analysis, the number of feasible configurations grows exponentially with the number of nodes and potential connections, which makes traditional exact optimization approaches computationally expensive as the network size increases. Instead of attempting to enumerate all possible configurations, RL enables an agent to explore the solution space and learn cost effective network structures through interaction with the environment.

Within this framework, the network design problem is formulated as a sequential decision-making process in which the agent incrementally constructs the network by making decisions regarding node activation and connectivity. At

each step, the agent observes the current configuration of the network, including the status of activated blockchain nodes and the connections established between the supply chain and blockchain layers. Based on this information, the agent selects an action that modifies the configuration, such as activating a blockchain node or establishing a connection between nodes. The environment then updates the network structure and returns feedback in the form of a reward signal.

The reward mechanism reflects the objective of the network design model. The agent is encouraged to minimize total operational and connection costs while maintaining a feasible and functional network structure. Configurations that satisfy connectivity requirements and ensure sufficient blockchain node activation receive higher rewards, whereas infeasible or inefficient configurations are penalized. Through repeated interactions across multiple learning episodes, the agent gradually learns which decisions lead to lower overall costs and feasible network configurations.

By modeling network design as a sequential learning problem, the RL agent can effectively navigate the exponentially large configuration space and identify cost efficient network structures in blockchain enabled supply chain systems. This approach aligns well with the problem structure because network construction occurs sequentially and the quality of each decision can be evaluated through environmental feedback, making RL a natural baseline for capturing adaptive decision-making behavior in dynamic settings.

3.6 Q-Learning Framework

The Q-learning algorithm is run with selected hyperparameters optimized to balance the effectiveness of learning with the stability of convergence. A 0.1 learning rate facilitates stepwise learning without deteriorating the stability of the algorithm while incrementally enhancing decision-making without drastic oscillations. A discount factor of 0.9 places importance on the consideration of reward in the future, allowing the algorithm to make decisions maximizing long-term network performance at the cost of short-term maximization. An exploration rate of 0.1 offers a good balance between exploitation behavior and exploration behavior, and the algorithm will keep exploiting the known good configurations and exploring new solutions. The ε-greedy policy is a method in which, with probability (1-ε), the agent chooses the action with the highest Q-value and with probability ε, a random action to guarantee exploitation of the best-known actions and exploration of superior potential actions in balance. Training is done for 10,000 episodes with at most 50 steps per episode for convergence without providing high computational overhead. State space is created in a complete way to cover all of the interest network setup data. Binary matrices allow for the depiction of SC-BC connections as to which supply chains and blockchain nodes are connected. Symmetric matrices depict SC-SC and BC-BC connections, showing two-way node relations within a layer. A binary vector represents the activation states of BC nodes, indicating whether each node is active and ready for network operation. Node or connection activation status switching is performed as operations with random selection are utilized during exploration steps to discover new network configurations.

Q-value update utilizes the Bellman equation, as denoted by Eq. (8):

$$v(s) = \max_{a}\big(R(s,a) + \gamma v(\acute{s})\big), \tag{8}$$

where $v(s)$ is the estimated total reward from the current state, γ stands for a discount between 0 and 1 that determines the relative value of delayed reward. A less discount factor assigns greater weight to future reward and more to immediate reward. Furthermore, a denotes the action in the current state, and s displays the current state or position in the decision-making process. Moreover, $\acute{s}$ is the successor state reached by taking action a in state s. Finally, $R(s,a)$ is the reward or short-term payoff of performing action a in state s.

3.7 Implementation

The implementation is done in Python 3.9.6 using NumPy for numerical operations and NetworkX for graph representation and visualization. The implementation includes cost functions to compute costs, validation functions to

validate constraints, and visualization tools to visualize networks. The configuration with the lowest cost is selected as the solution upon training and visualized as a network graph with active nodes and connections (Figure 1).

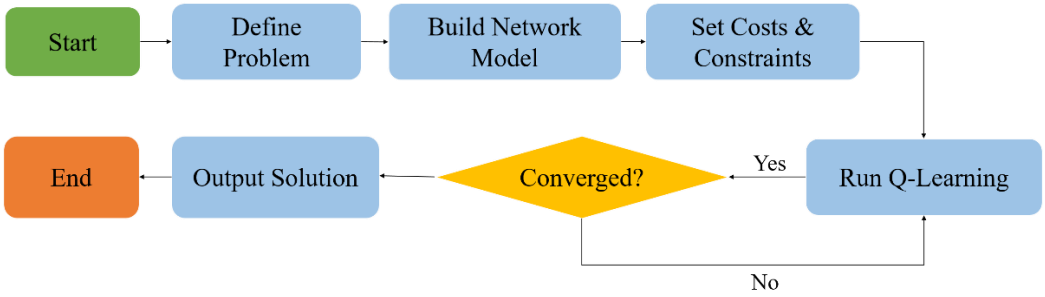

Figure 1. Algorithm execution process.

Validation experiments include constraint satisfaction verification, cost function correctness checking, and convergence testing to ensure that the algorithm is converging to stable solutions. The approach enables systematic exploration of optimal configurations under realistic constraints and can be extended to larger networks and dynamic environments.

Convergence is the point at which the Q-learning algorithm's performance stabilizes and does not improve significantly anymore. The algorithm checks if the total network cost has been steady in consecutive episodes (typically 100-500 episodes) or if the maximum training episodes have been attained. If Converged (Yes), then training stops, and the best network configuration is printed out as the optimal solution. If Not Converged (No), the algorithm continues training, looking for new configurations and updating Q-values to find better solutions. This ensures the final blockchain-supply chain network design is a cost-effective, reliable solution rather than an interim result.

3.8 Methodological Justification

This methodology combines combinatorial optimization and RL in solving the complex problem of blockchain-based supply chain network design. The Q-learning framework provides flexibility toward random cost variation and offers convergence properties, while the combined constraint system ensures real-world practicality. The penalty mechanisms further guide the optimization toward feasible and implementable configurations. Through this formulation, the approach enables rigorous examination of optimal network structures under practical constraints and remains applicable to larger networks and dynamic environments where traditional optimization methods face computational infeasibilities.

Within this methodological setting, the present study is intended to establish and validate an RL-based baseline for the considered problem. The work is framed as a proof-of-concept investigation designed to demonstrate that the proposed Q-learning framework can effectively operate within the blockchain-enabled supply chain network design setting. Accordingly, the main objective is to provide a numerical illustration of the learning capability, feasibility, and convergence behavior of the approach, rather than to claim that it consistently yields globally optimal solutions or outperforms all alternative optimization methods. In this context, the random-configuration baseline serves as a neutral reference point, allowing the analysis to highlight the ability of the learning agent to improve network design decisions over uninformed configurations while confirming the practical functionality of the proposed framework.

The experimental design is intentionally compact to support controlled validation of the proposed framework. The two-layer architecture corresponds directly to the two essential components of the problem, the supply chain layer and the blockchain layer, so increasing the number of layers would alter the problem scope and shift the analysis away from the intended blockchain-enabled supply chain setting. Within this controlled structure, the selected network size is sufficient for evaluating the feasibility, convergence characteristics, and structural behavior of the RL-based method.

Finally, it should be clarified that the aim of this study is not to develop a new RL algorithm. The methodological contribution lies in the integrated formulation of the two-layer blockchain-enabled supply chain network design problem, including its decision structure, cost representation, and constraint modeling. Within this formulation, RL functions as the solution strategy for navigating the resulting combinatorial decision space. Given the discrete variables and the controlled scale of the experimental setting, a standard tabular Q-learning implementation is sufficient for demonstrating the feasibility and convergence of the proposed formulation.

4. Results and Discussion

This section presents and explains the outcomes of the Q-learning-based optimization model for the formulation of a blockchain-based supply chain network (see Figure 2). Multiple runs were tested to analyze the convergence behavior of the model, sensitivity to the design constraints, and solution stability across repeated runs.

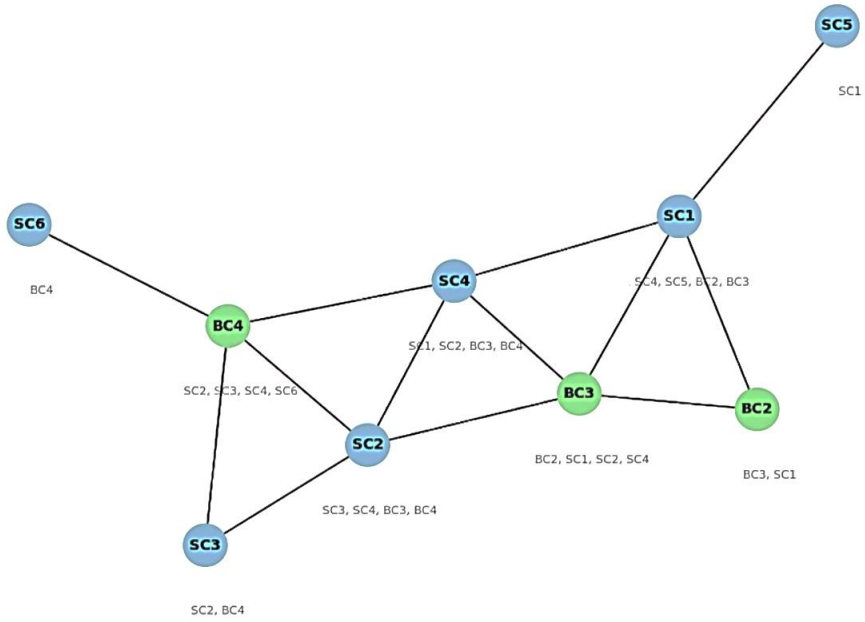

Figure 2. A blockchain-integrated supply chain network.

4.1 Convergence Behavior

In the first experiments, the algorithm was executed for 10,000 episodes over several independent trials with *min_active_BC* = 3, where *min_active_BC* specifies the minimum number of blockchain nodes that must be active in any valid network setting. As indicated by Figure 3, the overall network cost always reduced during the initial training period, specifically in the first 5,000 episodes, before reaching convergence.

This decreasing trend demonstrates the capacity for learning by the model through effective exploration and exploitation.

During runs, best costs increasingly improved over time (e.g., from 528 to 478 for Run 1 and from 552 to 503 for Run 2), indicating incremental discovery of better network structures. The closeness of learning trajectories in different runs is evidence of reproducibility and stability of learning with the same hyperparameters.

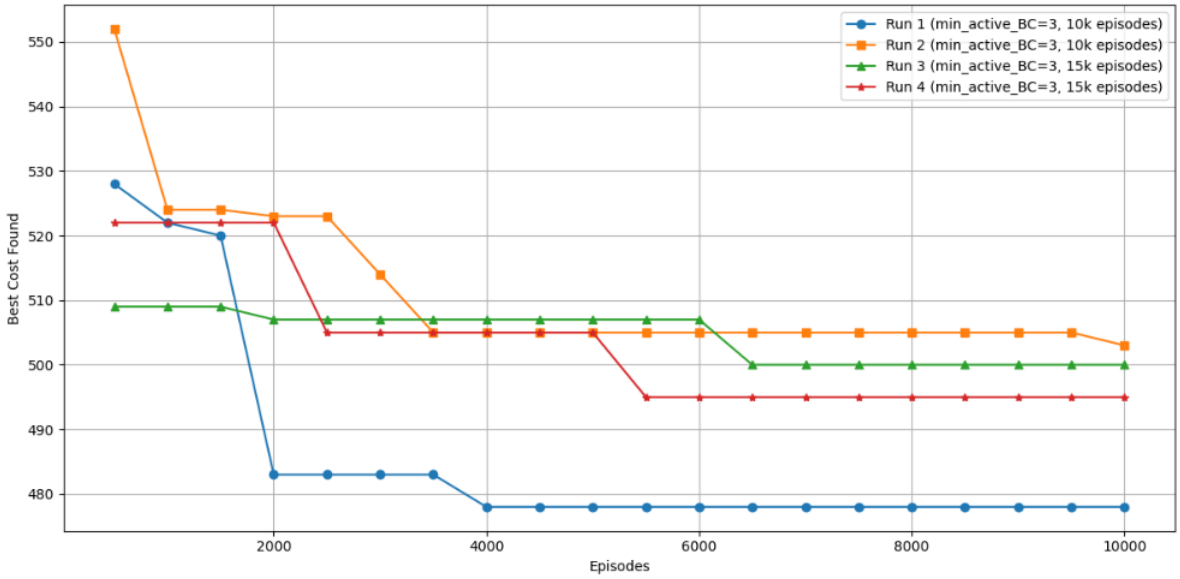

Figure 3. Trend of network cost over Q-learning episodes.

To probe the effect of longer training horizons, the number of episodes was increased to 15,000 in subsequent runs from 10,000. As shown in Figure 3 (Run 3 and Run 4), the best costs obtained were 500 and 495, respectively, similar to shorter runs. This effect suggests that training beyond 10,000 episodes provides marginal improvements only, and convergence occurs at episode 5,000 as well.

4.2 Sensitivity Analysis: Minimum Active Blockchain Nodes

To investigate the cost impact of tighter activation restrictions, the *min_active_BC* parameter was increased stepwise from 3 to 6. The best fit costs achieved under each condition are summarized in Table 1 and graphed in Figure 4.

Table 1. Best cost found under varying values of *min_active_BC*

min_active_BC	Best Cost Found
3	496
4	575
5	683
6	819

As expected, the overall cost increases dramatically as the minimum number of functional blockchain nodes increases. This is not just because of the constant node-activation cost, but also because more SC-BC connections must be established and maintained in order to make this possible. Figure 4 reflects the trend firmly in this sense, indicating the rise in expense over isolated runs based on *min_active_BC*. The performance demonstrates the appropriateness of the model in performing cost-policy sensitivity analysis for strategic decision-making in blockchain infrastructure development.

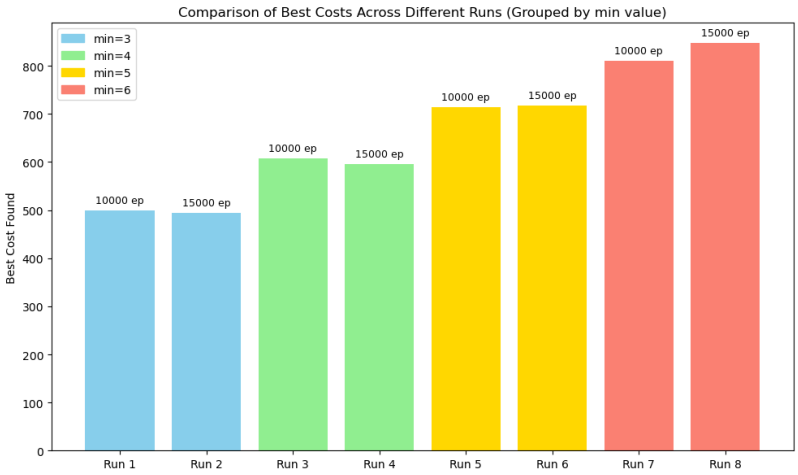

Figure 4. Comparison of Best Costs Across Different Runs (Grouped by min value)

For each *min_active_BC* value, multiple independent training runs were undertaken. In both cases, the model was continually converging to cost values within a particular range, like 712 and 683 for *min_active_BC = 5*, and 819 and 857 for *min_active_BC* = 6. The observations substantiate the consistency of the Q-learning process despite its underlying randomness and confirm its ability to accurately converge to nearly optimal network parameters repeatedly.

4.3 Structural Analysis of Optimal Networks

The optimal topologies attained at every trial were depicted in graph representations. Systematically, the following structural characteristics were noted:

1. BC nodes' partial activation: not all the nodes available were used, even in more restricted instances;
2. SC-BC connections were preponderant in the design, emphasizing the importance of cross-layer integration;
3. The resulting networks were connected, sparse, and acyclic, reflecting cost-restricted but feasible designs.

These topological characteristics make it possible for the model to find efficient and interpretable structures that adhere to both economic and operational standards for decentralized supply chain planning.

5. Limitations and Future Directions

The proposed Q-learning method is demonstrated to work well for small-sized supply chain networks, but exhibits immense scalability problems as the network size increases. The state space expands exponentially with the number of nodes, rendering computation intractable for large systems. In addition, the current implementation faces limitations in its exploration strategy: the application of a basic ε-greedy mechanism with random action selection may result in slow convergence in complex environments. Incorporating domain-specific heuristics or more advanced exploration strategies would therefore significantly improve learning efficiency and solution quality. Furthermore, the static nature of the current model and its single-objective focus on cost minimization may not adequately represent the dynamic and multi-dimensional character of real-world supply chain operations, in which costs, demands, and operational priorities evolve continuously. Building on these observations, the present study still relies on a synthetic experimental testbed rather than empirical supply chain data. Future work should extend the proposed framework to larger-scale network instances and assess its performance using real-world case studies and operational datasets.

Subsequent research may also conduct systematic benchmark comparisons between the proposed RL approach and other optimization techniques, such as GA, PSO, and MILP, to evaluate their relative performance across alternative supply chain configurations. Follow-up investigations should further explore dynamic environments in which the

Q-learning agent learns to adapt to new conditions and address multiple objectives, including environmental sustainability, resilience, social impact, and service quality. Integrating multi-objective RL formulations would yield a more comprehensive framework aligned with modern supply chain priorities and enable decision-makers to balance economic efficiency with broader organizational and societal goals. While a tabular Q-learning implementation remains adequate for the current problem scale and discrete decision structure, future studies could explore more advanced RL architectures, including deep RL, to effectively handle larger network scales and high-dimensional state spaces.

6. Conclusion

This research successfully demonstrated the effectiveness of Q-learning RL for blockchain-based supply chain network optimization under standard cost and connectivity scenarios. The proposed methodology tackled the complex combinatorial optimization problem of meeting blockchain implementation expenditures and operational efficiency, with resultant guaranteed cost-minimizing network configurations under diverse experimental setups. The sensitivity analysis illustrated qualitatively significant trade-offs between network resilience and economic viability, with costs rapidly increasing as progressively demanding minimum blockchain node standards are imposed.

The ability of the framework to treat stochastic cost variability while preserving tractable topologies is particularly valuable to organizations that have to make blockchain adoption decisions subject to changing supply chain environments. By virtue of its contribution of a systematized quantitative approach to network design optimization, this work contributes to the evidence-based application strategies for blockchain integration deployable across larger, real-life supply chain networks, potentially in the future, to more efficient, transparent, and cost-saving decentralized supply networks.

Author Contributions

Sara Hasheminezhad: Writing – Original Draft, Investigation, Project Administration, Methodology, Software, Writing – review & editing. Mahsa Moayed: Formal analysis, Writing - Original Draft, Validation, Methodology, Software, Visualization. Ardavan Babaei: Writing – review & editing, Conceptualization, Data Curation, Validation. Erfan Babaee Tirkolaee: Writing – review & editing, Investigation, Supervision, Validation.

Data Availability Statement

Data will be provided upon a request from the corresponding author.

Acknowledgements

The authors would like to thank anonymous reviewers for their valuable suggestions in manuscript revision.

Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and any form of misconduct.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare no conflict of interest.

Declaration of Generative AI and AI-assisted technologies in the writing process

Not applicable.

References

Abualigah, L., Hanandeh, E. S., Zitar, R. A., Thanh, C. L., Khatir, S., & Gandomi, A. H. (2023). Revolutionizing sustainable supply chain management: A review of metaheuristics. *Engineering Applications of Artificial Intelligence, 126*, 106839. https://doi.org/10.1016/j.engappai.2023.106839

Asghari, M., Afshari, H., Mirzapour Al-E-Hashem, S. M. J., Fathollahi-Fard, A. M., & Dulebenets, M. A. (2022). Pricing and advertising decisions in a direct-sales closed-loop supply chain. *Computers & Industrial Engineering*, *171*, 108439. https://doi.org/10.1016/j.cie.2022.108439

Azzi, R., Chamoun, R. K., & Sokhn, M. (2019). The power of a blockchain-based supply chain. *Computers & Industrial Engineering*, *135*, 582-592. https://doi.org/10.1016/j.cie.2019.06.042

Babaei, A., Khedmati, M., Akbari Jokar, M. R., & Tirkolaee, E. B. (2023). Designing an integrated blockchain-enabled supply chain network under uncertainty. *Scientific Reports*, *13*(1), 3928. https://doi.org/10.1038/s41598-023-30439-9

Babaei, A., Tirkolaee, E. B., & Ali, S. S. (2024b). Innovative supply chain network design with two-step authentication and environmentally-friendly blockchain technology. *Annals of Operations Research*, 1-31. https://doi.org/10.1007/s10479-024-05950-5

Babaei, A., Tirkolaee, E. B., & Amjadian, A. (2024c). Crafting efficient blockchain adoption strategies under risk and uncertain environments. *Alexandria Engineering Journal*, *103*, 137-147. https://doi.org/10.1016/j.aej.2024.05.106

Babaei, A., Khedmati, M., & Jokar, M. R. A. (2025a). A new model for production and distribution planning based on data envelopment analysis with respect to traffic congestion, Blockchain technology and uncertain conditions. *Annals of Operations Research*, *348*(3), 1145-1181. https://doi.org/10.1007/s10479-023-05349-8

Babaei, A., Khedmati, M., Jokar, M. R. A., & Tirkolaee, E. B. (2025b). Product tracing or component tracing? Blockchain adoption in a two-echelon supply chain management. *Computers & Industrial Engineering*, *200*, 110789. https://doi.org/10.1016/j.cie.2024.110789

Babaei, A., Tirkolaee, E. B., Kia, R., & Sezer, N. S. (2025c). Assessing the efficiency of blockchain adoption in renewable energy supply chains in ideal and non-ideal scenarios. *Energy*, *330*, 136782. https://doi.org/10.1016/j.energy.2025.136782

Bhutani. A., & Wadhwani. P. (2019, November 13). Global Blockchain Market Size to hit $25 Bn by 2025. Gminsights.https://www.gminsights.com/pressrelease/blockchainmarket?utm_source=PrNewswire.com&utm_medium=referral&utm_campaign=Paid_PrNewswire

Choi, T. M. (2023). Supply chain financing using blockchain: Impacts on supply chains selling fashionable products. *Annals of Operations Research*, *331*(1), 393-415. https://doi.org/10.1007/s10479-020-03615-7

Çıkmak, S., Kantoğlu, B., & Kırbaç, G. (2024). Evaluation of the effects of blockchain technology characteristics on SCOR model supply chain performance measurement attributes using an integrated fuzzy MCDM methodology. *International Journal of Logistics Research and Applications*, *27*(6), 1015-1045. https://doi.org/10.1080/13675567.2023.2193736

Dehshiri, S. J. H., & Amiri, M. (2024). Evaluation of blockchain implementation solutions in the sustainable supply chain: A novel hybrid decision approach based on Z-numbers. *Expert Systems with Applications*, *235*, 121123. https://doi.org/10.1016/j.eswa.2023.121123

Deloitte. (2021). "Deloitte's 2021 Global Blockchain Survey". Deloitte. https://www2.deloitte.com/us/en/insights/topics/understanding-blockchain-potential/global-blockchain-survey.html

Dobrovnik, M., Herold, D. M., Fürst, E., & Kummer, S. (2018). Blockchain for and in Logistics: What to Adopt and Where to Start. *Logistics*, *2*(3), 18. https://doi.org/10.3390/logistics2030018

Dutta, P., Choi, T. M., Somani, S., & Butala, R. (2020). Blockchain technology in supply chain operations: Applications, challenges and research opportunities. *Transportation Research Part E: Logistics and Transportation Review*, *142*, 102067. https://doi.org/10.1016/j.tre.2020.102067

Fan, C., Lin, C., Khazaei, H., & Musilek, P. (2022, August). Performance analysis of hyperledger besu in private blockchain. In 2022 IEEE international conference on decentralized applications and infrastructures (DAPPS) (pp. 64-73). IEEE https://doi.org/10.1109/DAPPS55202.2022.00016

Goli, A. (2023). Integration of blockchain-enabled closed-loop supply chain and robust product portfolio design. *Computers & Industrial Engineering*, *179*, 109211. https://doi.org/10.1016/j.cie.2023.109211

Gujar. S., Kshirsagar. R. (2024, August). Blockchain in Logistics Market Size. Gminsights. https://www.gminsights.com/industry-analysis/blockchain-in-logistics-market

Kamilaris, A., Fonts, A., & Prenafeta-Boldύ, F. X. (2019). The rise of blockchain technology in agriculture and food supply chains. *Trends in Food Science & Technology*, *91*, 640-652. https://doi.org/10.1016/j.tifs.2019.07.034

Khokhar, R. H., Rankothge, W., Rashidi, L., Mohammadian, H., Ghorbani, A., Frei, B., ... & Freitas, I. (2024). A survey on supply chain management: Exploring physical and cyber security challenges, threats, critical applications, and innovative technologies. *International Journal of Supply and Operations Management*, *11*(3), 250-283. https://doi.org/10.22034/IJSOM.2024.110219.2975

Longo, F., Nicoletti, L., & Padovano, A. (2017). Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & industrial engineering*, *113*, 144-159. https://doi.org/10.1016/j.cie.2017.09.016

Marr, B. (2018). How blockchain will transform the supply chain and logistics industry. Retrieved February, 22, 2018.

Min, H. (2019). Blockchain technology for enhancing supply chain resilience. *Business Horizons*, *62*(1), 35-45. https://doi.org/10.1016/j.bushor.2018.08.012

Moosavi, J., Naeni, L.M., Fathollahi-Fard, A.M. *et al.* RETRACTED ARTICLE: Blockchain in supply chain management: a review, bibliometric, and network analysis. *Environ Sci Pollut Res* **33**, 5760 (2026). https://doi.org/10.1007/s11356-021-13094-3

Pournader, M., Shi, Y., Seuring, S., & Koh, S. L. (2020). Blockchain applications in supply chains, transport and logistics: A systematic review of the literature. *International Journal of Production Research*, *58*(7), 2063-2081. https://doi.org/10.1080/00207543.2019.1650976

Queiroz, M. M., Telles, R., & Bonilla, S. H. (2020). Blockchain and supply chain management integration: A systematic review of the literature. *Supply chain management: An International Journal*, *25*(2), 241-254. https://doi.org/10.1108/SCM-03-2018-0143

Rahmanzadeh, S., Pishvaee, M. S., & Rasouli, M. R. (2020). Integrated innovative product design and supply chain tactical planning within a blockchain platform. *International Journal of Production Research, 58*(7), 2242-2262. https://doi.org/10.1080/00207543.2019.1651947

Rajabi-Kafshgar, A., Gholian-Jouybari, F., Seyedi, I., & Hajiaghaei-Keshteli, M. (2023). Utilizing hybrid metaheuristic approach to design an agricultural closed-loop supply chain network. *Expert Systems with Applications, 217*, 119504. https://doi.org/10.1016/j.eswa.2023.119504

Salehi-Amiri, A., Zahedi, A., Akbapour, N., & Hajiaghaei-Keshteli, M. (2021). Designing a sustainable closed-loop supply chain network for walnut industry. *Renewable and Sustainable Energy Reviews, 141*, 110821. https://doi.org/10.1016/j.rser.2021.110821

Samuel, C. N., Glock, S., Verdier, F., & Guitton-Ouhamou, P. (2021, May). Choice of ethereum clients for private blockchain: Assessment from proof of authority perspective. In 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC) (pp. 1-5). IEEE. https://doi.org/10.1109/ICBC51069.2021.9461085

Santoso, J. T., Wibowo, M. C., & Raharjo, B. (2025). Trustworthiness in supply chains: Leveraging private blockchain solutions. *International Journal of Supply and Operations Management*, *12*(2), 123-148. https://doi.org/10.22034/ijsom.2024.110372.3087

Shoaib, M., Zhang, S., Ali, H., Akbar, M. A., Hamza, M., & Rehman, W. U. (2024). Robust framework to prioritize blockchain-based supply chain challenges: The fuzzy best-worst approach for multiple criteria decision-making. *Kybernetes, 53*(10), 3326-3347. https://doi.org/10.1108/K-01-2023-0046

Wadhwani. P., & Ambekar. A. (2024, June). Blockchain in Agriculture and Food Supply Chain Market Size. Gminsights. https://www.gminsights.com/industry-analysis/blockchain-in-agriculture-and-food-supply-chain-market

Zarreh, M., Khandan, M., Goli, A., Aazami, A., & Kummer, S. (2024). Integrating Perishables into Closed-Loop Supply Chains: A Comprehensive Review. *Sustainability, 16*(15), 6705. https://doi.org/10.3390/su16156705