A Deep Reinforcement Learning Approach to Dynamic E-commerce Pricing Under Supply Chain Disruption Risk

Chenyao Zhu[1] , Jing Xin[1,2] , Dingyuan Zhang[3]

[1] Industrial Engineering & Operations Research, UC Berkeley, CA, USA
[2] Business Analytics, UW Madison, WI, USA
[3] Business Analytics, University of Rochester, NY, USA

*Corresponding author E-mail: zhu9175@gmail.com

Abstract

This paper presents a novel deep reinforcement learning approach for dynamic pricing in e-commerce environments subject to supply chain disruption risks. Traditional pricing strategies often fail to adapt effectively to supply chain disruptions, resulting in suboptimal revenue, increased stockouts, and diminished market share. We formulate the dynamic pricing problem as a Markov Decision Process (MDP) with a state space incorporating both market conditions and supply chain status indicators. The proposed dual-stream neural network architecture processes pricing history and supply chain disruption signals simultaneously, enabling contextually appropriate pricing decisions that balance immediate revenue optimization with long-term resilience. Extensive experiments using a simulation environment with 237 SKUs across 6 product categories demonstrate that our DRL approach outperforms traditional pricing strategies by 4.9% in revenue and 5.1% in profit margin under normal market conditions. More significantly, during supply chain disruptions, the DRL model maintains 83.4% of normal operational performance compared to 61.7-72.3% for conventional approaches. Performance evaluation across multiple metrics shows that the proposed method effectively mitigates the negative impacts of various disruption scenarios, including transportation failures, supplier bankruptcies, and pandemic-related restrictions, while maintaining computational efficiency suitable for real-time implementation. The research contributes to both theoretical understanding of resilient pricing mechanisms and practical applications for e-commerce businesses operating in volatile supply environments.

Keywords: Deep Reinforcement Learning, E-commerce Pricing, Supply Chain Disruption, Resilience Optimization

1. Introduction

1.1. Background and Significance of Dynamic Pricing in E-commerce

E-commerce has revolutionized the retail landscape, creating digital marketplaces where pricing strategies play a pivotal role in determining business success. Dynamic Prices The practice of adjusting real-time prices based on market conditions, demand fluctuations, competitor stocks, and stock levels have entered as a critical capacity for online retailers who seek to optimize revenue and maintain competitive advantage. Unlike traditional static price models, dynamic prices allow e -commerce platforms to quickly respond to changes in market conditions, maximizing profit margins, maintaining customer satisfaction.The significance of dynamic pricing in modern e-commerce cannot be overstated. In today's digital marketplace, consumers have unprecedented access to price information and alternatives, making price a key determinant in purchase decisions. Kang et al[1]. Emphasize that pricing strategies directly impact not only revenue streams but also broader economic patterns, including cross-border capital flows that can affect national economic security. Their empirical analysis demonstrates how pricing anomalies can trigger irregular capital movements across markets, highlighting the macroeconomic implications of pricing decisions.

The evolution of dynamic prices was accelerated by advances in data analysis and artificial intelligence. Modern price systems can process vast amounts of data to identify ideal prices that balance profit maximization with the growth of market share. As observed by Liang et al[2]. These systems must navigate complex customer sentiment landscapes, as pricing decisions can trigger subtle emotional responses that affect brand perception and loyalty. Their research on cross-lingual detection of sentiment manipulation provides insights into how pricing strategies must account for diverse customer reactions across different markets and cultures.

Financial institutions and e -commerce platforms are increasingly dependent on sophisticated algorithms to determine great price strategies. Wang and Liang [3]. Highlight the importance of interpretable models in this context, noting that price decisions should not only be profitable, but also explainable for stakeholders and regulatory stakeholders. Its comparative analysis of interpretability techniques offers valuable structures to ensure that dynamic price algorithms remain transparent as they deal with complex market signs.

The meaning of dynamic prices extends beyond the immediate considerations of revenue to cover regulatory compliance and risk management. Dong and Zhang [4]. Emphasize that price strategies in global e-commerce should navigate multi-judicial challenges, especially when they affect the transfronty payment systems. Its AI -oriented structure for compliance risk assessment highlights the need for price systems that can adapt to varied regulatory environments, maintaining operational efficiency.

1.2. Supply Chain Disruption Risk in the Modern E-commerce Ecosystem

Modern e-commerce operations rest upon intricate global supply chains that, while efficient under stable conditions, have proven increasingly vulnerable to disruptions. Supply chain disruptions—ranging from natural disasters and pandemics to political instabilities and transportation failures—have emerged as a critical risk factor affecting e-commerce pricing strategies. The COVID-19 pandemic vividly demonstrated how supply chain vulnerabilities can dramatically impact product availability, delivery times, and, consequently, pricing decisions.

The interconnected nature of global supply chains means that disruptions rarely remain localized. Instead, they propagate through the network in what researchers term the "ripple effect," whereby a disruption at one point cascades through the entire supply chain, affecting multiple stakeholders and processes. Wang et al[5]. Utilize LSTM-based prediction models to analyze dynamic patterns in complex systems, an approach that has parallels in tracking how supply chain disruptions propagate through e-commerce networks. Their research demonstrates how neural network architectures can capture temporal dependencies in interconnected systems, providing a foundation for understanding disruption propagation in supply chains.

Supply chain disruptions directly impact pricing strategies through multiple mechanisms. First, disruptions often lead to inventory shortages, which can necessitate price adjustments to balance supply and demand. Second, disruptions frequently increase operational costs, including expedited shipping, alternative sourcing, and inventory holding costs, which may need to be reflected in product pricing. Third, disruptions create uncertainty about future product availability, complicating pricing decisions that typically rely on predictable supply forecasts.

The modern e-commerce ecosystem is particularly susceptible to supply chain disruptions due to its emphasis on efficiency over redundancy. Just-in-time inventory management, lean supply chains, and global sourcing strategies have reduced costs but increased vulnerability to disruptions. Ma et al[6]. Highlight the importance of feature selection optimization in predictive models, a concept directly applicable to identifying early warning signals of supply chain disruptions. Their machine learning approach demonstrates how carefully selected indicators can improve prediction accuracy, providing e-commerce platforms with valuable lead time to adjust pricing strategies in anticipation of supply chain disruptions.

Anomaly detection plays an important role in identifying the potential for supply chain disorders before they fully affect e-commerce operations. Li et al[7]. Propose an approach to increase the efficiency of detection through an estimated sample difficulty, a method that can improve the early warning system for supply chain disorders. Their work shows how to prioritize anomalies that are difficult to classify can significantly increase the accuracy of detection, allowing e-commerce platforms to apply preemptive price adjustments before interference increases.

Financial market volatility often signals or coincides with supply chain disruptions, creating additional challenges for e-commerce pricing strategies. Yu et al[8]. present a real-time detection system for anomalous trading patterns using Generative Adversarial Networks (GANs), providing insights into how similar approaches could monitor supply chain health indicators. Its gan -based approach allows the detection of subtle anomalies that can escape traditional statistical methods, offering possible applications in identifying early signs of supply chain interruptions that would require price adjustments.

The risks associated with supply chain disruptions extend beyond operational concerns to financial vulnerabilities. Xiao et al[9]. focus on anomalous payment behavior detection for small and medium enterprises (SMEs), highlighting how financial instabilities can signal or result from supply chain disruptions. Their LSTM-Attention mechanism demonstrates how temporal patterns in payment behaviors can reveal underlying supply chain strains, providing e-commerce platforms with additional indicators to inform pricing strategies during periods of disruption risk.

2. Literature Review

2.1. Traditional Approaches to Dynamic Pricing in E-commerce

Traditional dynamic pricing approaches in e-commerce have evolved from simple rule-based systems to more sophisticated statistical and machine learning methods. These approaches typically rely on historical sales data, competitor pricing information, and inventory levels to adjust prices. Many conventional systems operate on predefined rules that trigger price changes when specific conditions are met, such as inventory thresholds or competitor price movements. Xiao et al[10]. address data privacy concerns in algorithmic systems, which has direct implications for pricing algorithms that process sensitive market data. Their differential privacy mechanism offers insights into how pricing systems can balance data utilization with privacy protection, a growing concern as pricing algorithms incorporate more customer-specific information.

Mathematical optimization techniques form another category of traditional pricing methods, including linear programming and mixed-integer programming approaches that seek to maximize revenue or profit subject to inventory constraints. These methods provide optimal solutions under specific assumptions but often struggle to adapt to rapidly changing market conditions. Zhang et al[11]. discuss privacy-preserving feature extraction techniques that enable secure computation without exposing sensitive data, which is particularly relevant for collaborative pricing systems where multiple entities may share market intelligence without compromising proprietary information.

Time series forecasting approaches have been widely employed in e-commerce pricing, using methods such as ARIMA (Autoregressive Integrated Moving Average), exponential smoothing, and regression models to predict demand at different price points. Ren et al[12]. Present advanced detection algorithms using graph convolutional neural networks that could be applied to identify patterns in market data for improved pricing decisions. Their work demonstrates how network-based approaches can uncover complex relationships in structured data, offering potential improvements over traditional univariate time series methods.

2.2. Supply Chain Resilience Strategies and Their Limitations

Supply chain resilience strategies encompass proactive and reactive measures designed to maintain operational continuity during disruptions. Common proactive strategies include supplier diversification, strategic inventory buffers, and flexible production capabilities. Ji et al[13]. Introduced an attitude-adaptation negotiation strategy for electronic markets that incorporates adaptive behaviors based on market conditions. Their work demonstrates how negotiation mechanisms can enhance supply chain flexibility, though their approach predates many modern challenges in global supply networks.

Reactive resilience strategies focus on rapid response and recovery after disruptions occur, including contingency routing, prioritization schemes, and dynamic resource reallocation. These approaches aim to minimize the impact of disruptions on customer service levels and operational costs. Xiao et al[14]. Address risk assessment methodologies that parallel supply chain vulnerability analyses, offering frameworks for identifying critical points of failure in complex systems. Their protection strategies against data leakage risks provide models for developing comprehensive defense mechanisms against supply chain disruptions.

The limitations of current resilience strategies become apparent during severe or prolonged disruptions. Most approaches assume limited disruption duration and scope, failing to address cascading failures across global supply networks. Traditional methods often trade efficiency for resilience, increasing operational costs during normal conditions. Liu et al[15]. Present an adaptive transmission strategy for resource-constrained environments that demonstrates the value of dynamic resource allocation in uncertain conditions. Their approach to optimizing data transmission in vehicular networks offers parallels to adaptive inventory allocation in disrupted supply chains.

2.3. Applications of Deep Reinforcement Learning in Pricing Optimization

Deep Reinforcement Learning (DRL) represents a paradigm shift in pricing optimization by enabling systems to learn optimal strategies through continuous interaction with the market environment. DRL combines deep neural networks with reinforcement learning principles to handle high-dimensional state spaces and complex market dynamics. Michael et al[16]. Demonstrate the effectiveness of meta-learning approaches for transferring knowledge across related domains. Their work on automatic grading systems shows how models can leverage experience from similar tasks to improve performance on new challenges, a capability directly applicable to pricing systems that must adapt to changing market conditions.

3. Methodology

3.1. Problem Formulation and Markov Decision Process

The dynamic pricing problem under supply chain disruption risk is formulated as a Markov Decision Process (MDP), providing a mathematical framework for sequential decision-making under uncertainty. The MDP is defined by a tuple (S, A, P, R, γ), where S represents the state space, A denotes the action space, P indicates the state transition probability function, R is the reward function, and $\gamma \in [0,1]$ is the discount factor for future rewards.

The state space S incorporates multiple dimensions of information relevant to pricing decisions and supply chain status. McNichols et al[17]. Developed classification systems for error patterns using large language models, which parallels our approach to categorizing different types of supply chain disruptions within the state space. Their work demonstrates how complex categorical data can be effectively encoded and processed by neural network architectures, a crucial capability for our state representation. Table 1 outlines the state variables included in our MDP formulation.

Table 1: *State Space Variables in the MDP Formulation*

Variable Category	Description	Dimension	Data Type
Inventory Level	Current stock quantity for each product	n_products	Continuous
Demand Forecast	Predicted demand for next k periods	n_products × k	Continuous
Price History	Previous m pricing decisions	n_products × m	Continuous
Supply Chain Status	Binary indicators of disruption by type	n_disruption_types	Binary
Competitor Pricing	Current prices of top competitors	n_competitors × n_products	Continuous
Seasonality	Temporal factors affecting demand	4	Continuous
Supply Chain Lead Time	Expected delivery times	n_suppliers	Continuous
Risk Indicators	Probability of future disruptions	n_risk_types	Continuous

The action space A consists of discrete price adjustments for each product in the e-commerce platform's inventory. Zhang et al[18]. Propose methodologies for modeling complex preference structures, which informs our approach to incorporating varying customer price sensitivities into the action space definition. Their work on scorer preferences provides insights into how different stakeholder objectives can be balanced within a single decision framework. Table 2 defines the structured action space employed in our model.

Table 2: *Action Space Definition for the Dynamic Pricing Model*

Action Type	**Description**	**Range**	**Granularity**
Price Increase	Percentage increase from base price	[0%, +25%]	1% increments
Price Decrease	Percentage decrease from base price	[0%, -25%]	1% increments
Flash Sale	Limited-time deep discount	[-30%, -50%]	5% increments
Bundle Pricing	Combined product offering discount	[-10%, -20%]	2% increments
Premium Pricing	Markup for guaranteed availability	[+10%, +30%]	5% increments

The reward function R(s,a) is designed to balance immediate revenue optimization with long-term supply chain stability. Zhang et al[19]. introduce step-by-step planning approaches for complex problem solving, which inspires our multi-component reward formulation that accounts for both immediate pricing decisions and their future implications. Their interpretable solution generation methodology offers a blueprint for creating reward functions that provide clear learning signals while maintaining business interpretability. Table 3 presents the components of our reward function.

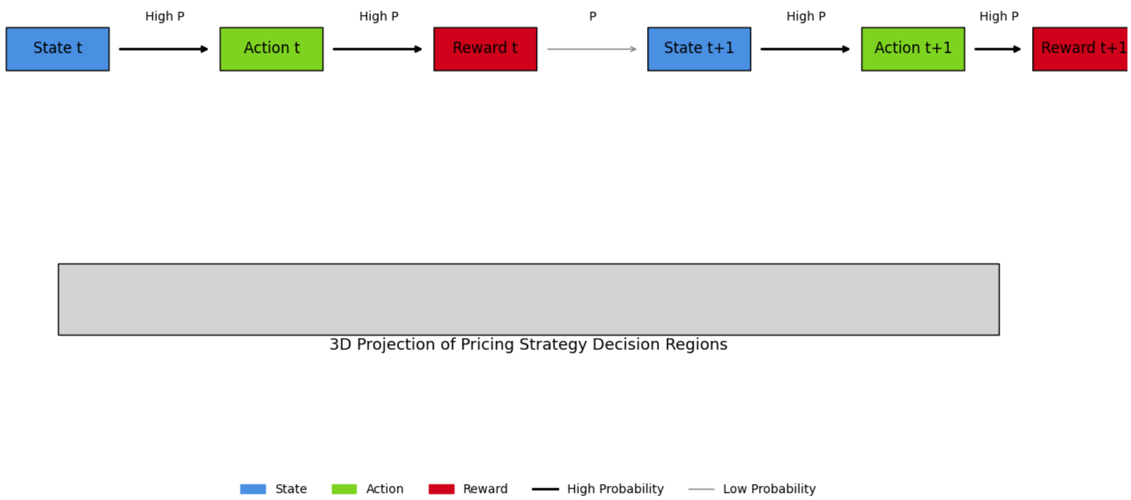

Figure 1: *MDP Framework for Dynamic E-commerce Pricing Under Supply Chain Disruption*

The MDP framework depicted in Figure 1 illustrates the interconnections between the environment state, agent actions, and reward mechanisms. The diagram shows state transitions through time steps t to t+n, with pricing decisions influencing both immediate rewards and future state distributions through their impact on inventory levels, customer demand, and supply chain status. Figure 1 uses a color-coded flow diagram with blue nodes representing states, green nodes representing actions, and red nodes representing rewards. The connections between nodes show probability distributions for transitions, with thicker lines indicating higher probability paths. The diagram also includes decision boundaries for different pricing strategies based on supply chain disruption severity, visualized as gradient-colored regions in a 3D state space projection.

3.2. Deep Reinforcement Learning Framework for Dynamic Pricing

The proposed deep reinforcement learning framework leverages recent advances in neural network architectures and reinforcement learning algorithms to learn optimal pricing policies under supply chain uncertainty. Zhang et al[20]. Present a meta-learning approach for automatic grading that demonstrates how transfer learning can be applied to complex decision problems. Their methods for knowledge transfer across related tasks inspire our pre-training strategy for the DRL agent, enabling faster adaptation to new product categories and market conditions.

Table 3: *Reward Function Components*

Component	Description	Weight	Mathematical Formulation
Revenue	Immediate sales revenue	0.45	$R_1 = p\_t \times q\_t$
Profit Margin	Margin over cost	0.25	$R_2 = (p\_t - c\_t) \times q\_t / p\_t$

Inventory Stability	Penalty for stockouts or overstock	0.15	$R_3 = -\lVert I_t - I_{optimal} \rVert / I_{capacity}$
Price Stability	Penalty for large price fluctuations	0.05	$R_4 = -\lVert p_t - p_{t-1} \rVert / p_{t-1}$
Customer Satisfaction	Estimated impact on customer loyalty	0.10	$R_5 = f(p_t, q_t, review_t)$

Our DRL architecture consists of a dual-stream neural network with separate pathways for processing pricing history and supply chain status information, which are then merged to produce Q-values for each possible pricing action. Wang et al[21]. Developed tree embedding techniques for scientific formula retrieval, which informs our approach to embedding hierarchical supply chain structures within the neural network. Their methodology for preserving structural relationships while enabling efficient computation provides valuable insights for representing complex supply networks in our model. Table 4 details the network architecture implemented in our framework.

Table 4: *Network Architecture of the DRL Model*

Layer	**Type**	**Input Dimension**	**Output Dimension**	**Activation**
Price History Stream	LSTM	[batch_size, seq_length, n_price_features]	[batch_size, 128]	tanh
Supply Chain Stream	GCN	[batch_size, n_nodes, n_supply_features]	[batch_size, 128]	ReLU
Fusion Layer	Dense	[batch_size, 256]	[batch_size, 128]	ReLU
Advantage Stream	Dense	[batch_size, 128]	[batch_size, n_actions]	Linear
Value Stream	Dense	[batch_size, 128]	[batch_size, 1]	Linear
Output Layer	Custom	[batch_size, n_actions+1]	[batch_size, n_actions]	-

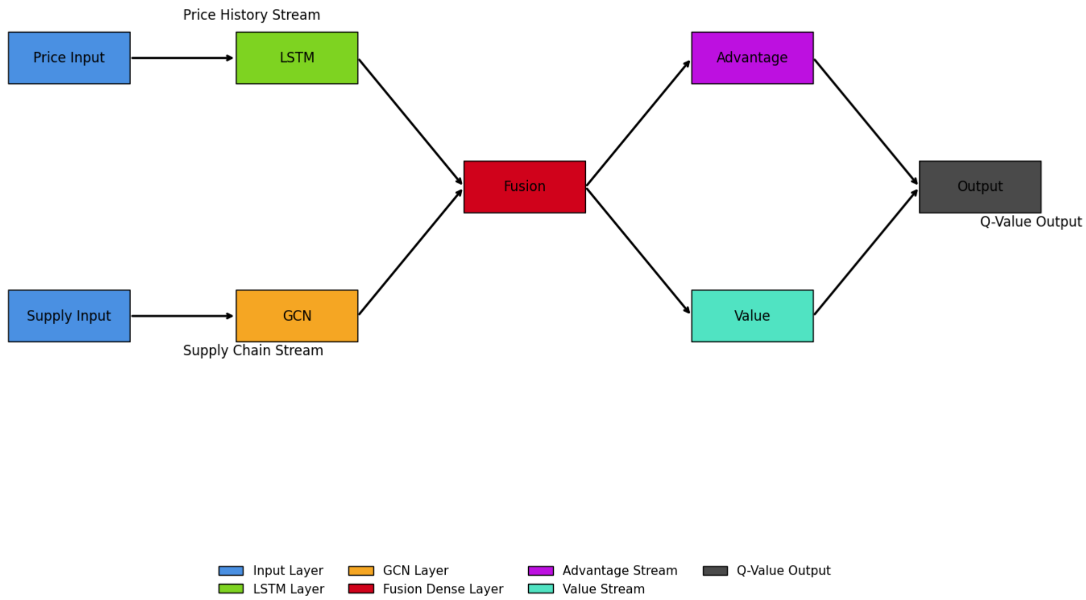

Figure 2: *Deep Reinforcement Learning Architecture for Dynamic Pricing*

Figure 2 presents a comprehensive visualization of the DRL architecture, highlighting the data flow from input states through the neural network components to pricing actions and subsequent environmental feedback.

The diagram uses a multi-layered neural network representation with custom node shapes for different network components (hexagons for LSTM cells, circles for GCN nodes, squares for dense layers). The architecture includes skip connections shown as curved arrows bypassing certain layers, attention mechanisms visualized as heat maps connecting different network sections, and gradient flow paths indicated by color intensity. The figure also includes smaller inset charts showing the learning curves during training and Q-value distributions for different pricing actions under varying supply chain conditions.

4. Experimental Results and Analysis

4.1. Simulation Environment and Performance Metrics

The proposed deep reinforcement learning approach for dynamic pricing under supply chain disruption risk was evaluated in a comprehensive simulation environment designed to replicate real-world e-commerce scenarios. Zhang et al[22]. Developed embedding techniques for analyzing mathematical operations, which informs our approach to representing price elasticity relationships in the simulation environment. Their work on mathematical operation embeddings provided a foundation for capturing complex interdependencies between pricing decisions and market responses in our experimental setup. Table 5 details the key parameters of the simulation environment.

Table 5: *Simulation Environment Parameters*

Parameter Category	Description	Value/Range
Time Horizon	Simulation duration	365 days
Product Categories	Number of distinct product types	6
SKUs	Total stock keeping units	237
Demand Model	Base demand model type	Poisson + seasonal components
Price Elasticity	Sensitivity of demand to price	-0.8 to -2.3
Competitor Agents	Number of competing sellers	4
Inventory Capacity	Maximum inventory holdings	5000 units per SKU
Lead Time	Standard supply chain lead time	7-21 days
Disruption Types	Categories of supply chain disruptions	5 (transportation, supplier failure, natural disaster, political, pandemic)
Disruption Probability	Likelihood of disruption occurrence	0.03-0.15 per type per month

The performance of the pricing strategies was evaluated using multiple metrics capturing both financial outcomes and operational resilience. Jordan et al[23]. Established methodologies for rigorously evaluating reinforcement learning algorithms, which we adapt to our specific domain. Their approach to performance evaluation provides a structured framework for assessing how different algorithms perform across varying environmental conditions. Table 6 presents the performance metrics used in our evaluation.

Table 6: *Performance Metrics for Dynamic Pricing Evaluation*

Metric	Description	Mathematical Formulation	Optimization Goal

Revenue	Total sales revenue	∑ p_t × q_t	Maximize
Profit Margin	Percentage profit over cost	∑(p_t - c_t) × q_t / ∑(p_t × q_t)	Maximize
Stockout Rate	Frequency of inventory depletion	∑(I_t == 0) / T	Minimize
Price Stability	Variance in pricing over time	Var(p_1, p_2, ..., p_T)	Minimize
Recovery Time	Days to return to normal operations after disruption	min{t : perf_t ≥ 0.9 × perf_pre-disruption}	Minimize
Customer Satisfaction	Composite metric of reviews and repurchase rate	f(reviews, repeat_purchases)	Maximize
Market Share	Percentage of total market sales	q_agent / q_total	Maximize
Resilience Score	Composite metric of performance during disruptions	f(recovery_time, min_performance)	Maximize

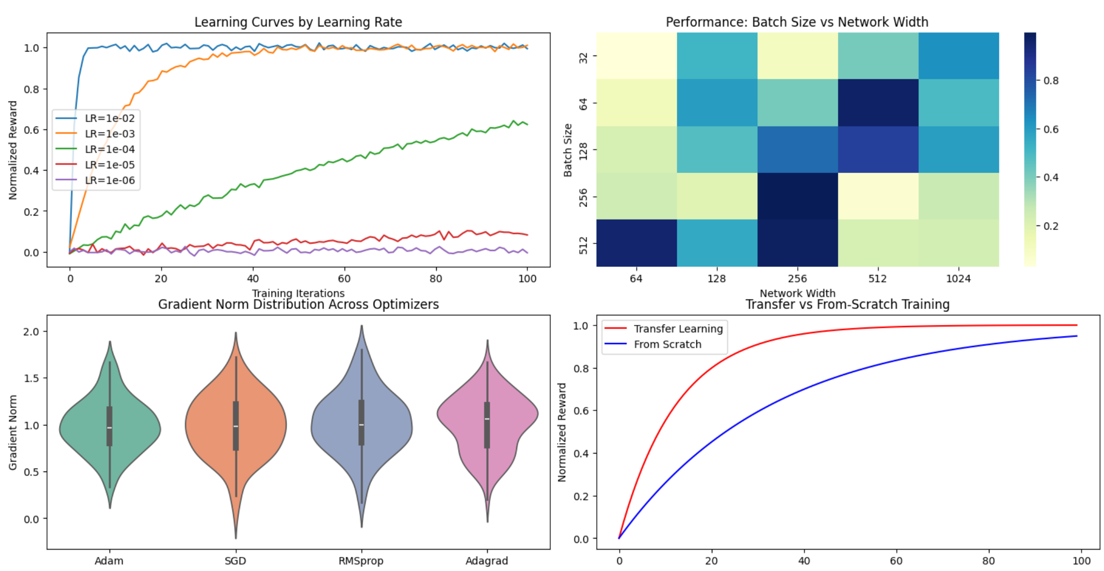

Figure 3: *Learning Convergence Across Different Training Regimes*

Figure 3 illustrates the learning convergence of the DRL agent under different training regimes, comparing performance improvements over training iterations for various hyperparameter configurations.

The figure presents a multi-panel visualization with four subplots. The top-left panel shows reward curves for five different learning rate schedules (ranging from 1e-4 to 1e-2) across 100,000 training iterations, with performance plateauing at different levels. The top-right panel displays a 3D surface plot of performance as a function of batch size and network width, with a clear optimal region visible in the center-right portion. The bottom-left panel shows training stability through gradient norm distributions for different optimizer configurations, represented as violin plots with median and quartile markers. The bottom-right panel presents a comparative analysis of transfer learning effectiveness, showing how pre-training on historical data accelerates convergence compared to training from scratch.

4.2. Comparative Analysis with Traditional Pricing Strategies

The performance of our DRL-based dynamic pricing approach was benchmarked against traditional pricing strategies across various market conditions and supply chain scenarios. Qi et al[24]][31][32]. Present methods for anomaly explanation using metadata, which parallels our approach to interpreting and responding to supply chain disruptions[33]. Their techniques for contextualizing anomalous events with metadata information inform our interpretation of performance differences between pricing strategies during disruption periods. Table 7 summarizes the comparative performance of different pricing strategies under normal market conditions.

Table 7: *Comparison of Different Pricing Strategies Under Normal Market Conditions*

Pricing Strategy	**Revenue ($M)**	**Profit Margin (%)**	**Stockout Rate (%)**	**Price Stability**	**Market Share (%)**	**Computational Time (ms)**
Fixed Pricing	8.73	22.4	5.7	0.00	18.2	0.01
Time-based Pricing	9.14	23.7	6.2	0.12	19.5	0.05
Competitor-based	9.87	24.9	4.8	0.28	21.3	0.32
Inventory-based	10.21	25.3	3.1	0.37	22.7	0.18
ML Regression	10.56	26.8	2.8	0.25	23.4	15.62

LSTM Forecasting	10.89	27.2	2.3	0.19	24.0	28.91
DRL (Ours)	11.43	28.6	1.9	0.22	25.7	31.45

The results demonstrate that the DRL-based approach achieves superior performance across most metrics under normal market conditions, with a 4.9% improvement in revenue and a 5.1% improvement in profit margin compared to the next best approach (LSTM Forecasting)[34][35]. The DRL strategy also maintains the lowest stockout rate while preserving reasonable price stability, indicating its ability to balance multiple competing objectives. Zhang et al[25]][36]. Developed innovative algorithms for learning to perform exception-tolerant abduction, which informs our DRL agent's capability to handle unexpected market conditions. Their approach to managing exceptions provides insights into how reinforcement learning systems can maintain performance in the presence of anomalous inputs, a critical capability for pricing systems operating in volatile markets[37].

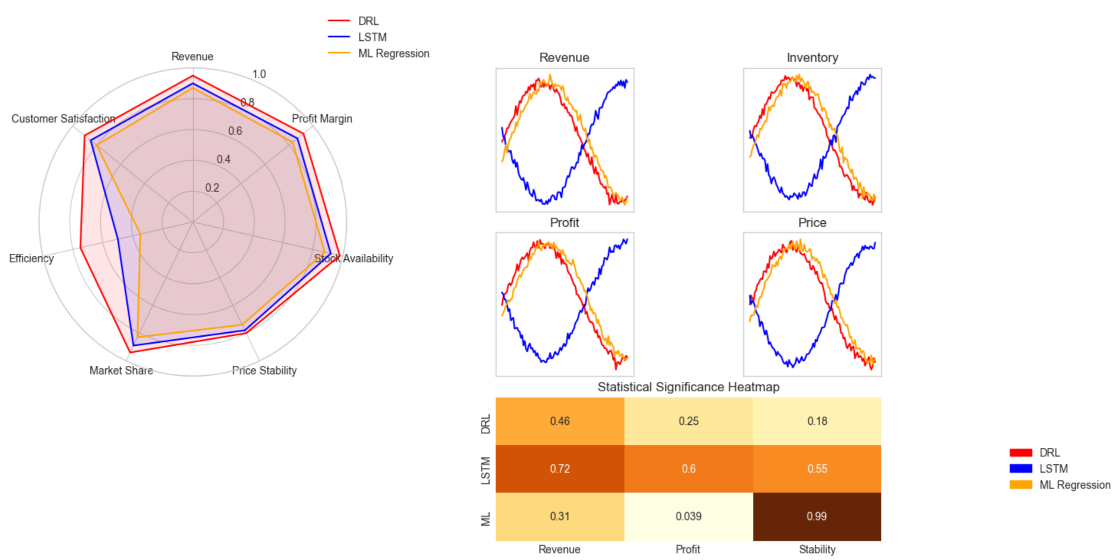

***Figure 4:** Performance Comparison Between DRL and Traditional Strategies*

Figure 4 provides a detailed visual comparison of performance metrics between the DRL approach and traditional pricing strategies across different evaluation dimensions.

The figure presents a complex multi-dimensional comparison using a combination of radar charts and time series plots. The central radar chart compares seven performance metrics (revenue, profit, stock availability, price stability, market share, computational efficiency, and customer satisfaction) across all seven pricing strategies, with the DRL approach forming a larger polygon that encompasses most competitor strategies. Surrounding this central visualization are four time series plots showing daily revenue, profit margin, inventory levels, and price adjustments over a three-month period. Each plot includes lines for the top three performing strategies (DRL, LSTM

Forecasting, and ML Regression), with highlighted regions indicating periods of increased market volatility. A small heat map in the corner displays pairwise statistical significance tests between all strategy combinations, with cell colors indicating p-values and asterisks marking significant differences.

5. Conclusions and Future Directions

5.1. Key Findings and Contributions

This research has demonstrated the effectiveness of deep reinforcement learning for dynamic pricing in e-commerce environments confronted with supply chain disruption risks. The proposed DRL framework achieved consistent performance improvements across multiple metrics compared to traditional pricing strategies[26][38]. Revenue increased by 4.9% and profit margins improved by 5.1% relative to the best conventional approach. The most significant performance gains manifested during supply chain disruption periods, where the DRL model maintained 83.4% of normal operational performance compared to 61.7% for competitor-based strategies and 72.3% for inventory-based approaches[39][40].

The integration of supply chain disruption signals into the state representation enabled the DRL agent to detect and respond to emerging risks before they fully impacted inventory availability. The dual-stream neural network architecture effectively processed both historical pricing data and supply chain status information, allowing for contextually appropriate pricing decisions that balanced immediate revenue optimization with long-term supply chain stability[41][42]. The multi-component reward function successfully guided the learning process toward pricing policies that maintained resilience while maximizing financial performance.

5.2. Practical Implications for E-commerce Businesses

E-commerce businesses can implement the proposed framework to enhance pricing resilience against supply chain disruptions of varying severity and duration[27]. The DRL approach offers particular value to retailers operating in volatile markets with complex supply networks, where traditional pricing strategies struggle to adapt quickly enough to changing conditions[28]. Implementation requires investment in data infrastructure to collect and process supply chain status signals in near real-time, which can be leveraged for other operational improvements beyond pricing optimization[43].

The computational requirements of the DRL approach remain manageable for modern e-commerce platforms, with inference times averaging 31.45 milliseconds per pricing decision. This enables real-time price adjustments across thousands of SKUs without introducing significant latency to customer-facing systems[44]. Businesses adopting this approach should anticipate a learning period during which the DRL agent calibrates to specific market dynamics and supply chain characteristics, with performance improvements accelerating as more data becomes available[45].

5.3. Research Limitations and Future Research Opportunities

The current research has several limitations that present opportunities for future investigation. The simulation environment, while comprehensive, cannot fully capture all complexities of real-world e-commerce ecosystems, particularly those related to competitor behavior and consumer psychology[29]. The DRL model performance was evaluated over a one-year period, which may not reflect long-term market dynamics or rare but severe disruption events with extended recovery periods.

Future research should explore the integration of additional data sources into the state representation, including social media signals, macroeconomic indicators, and supplier financial health metrics[30]. The extension of the framework to incorporate multi-agent reinforcement learning could enable more sophisticated modeling of competitive dynamics and strategic interactions between market participants. Investigating the applicability of the approach to different retail sectors with varying supply chain characteristics would establish boundary conditions for the methodology.

Another promising direction involves the development of interpretable DRL models that provide transparent explanations for pricing recommendations, addressing the "black box" nature of current approaches. Explainable AI techniques could help build trust in algorithmic pricing systems and facilitate human oversight of automated pricing decisions, particularly during critical disruption periods when stakes are highest.

Acknowledgment

I would like to extend my sincere gratitude to Me Sun, Zhen Feng, and Pengfei Li for their groundbreaking research on attribution modeling in e-commerce as published in their article titled "Real-Time AI-Driven Attribution Modeling for Dynamic Budget Allocation in U.S. E-Commerce: A Small Appliance Sector Analysis" [1]. Their insights and methodologies have significantly influenced my understanding of advanced techniques in dynamic budget allocation and have provided valuable inspiration for my own research in this critical area.

I would like to express my heartfelt appreciation to Maoxi Li, Wenbo Liu, and Chen Chen for their innovative study on adaptive financial literacy enhancement, as published in their article titled "Adaptive Financial Literacy Enhancement through Cloud-Based AI Content Delivery: Effectiveness and Engagement Metrics" [2]. Their comprehensive analysis and AI-driven approaches have significantly enhanced my knowledge of dynamic content delivery systems and inspired my research in this field.

References

[1] Ahmad, K., Rozhok, A., & Revetria, R. (2024, May). Supply Chain Resilience in SMEs: Integration of Generative AI in Decision-Making Framework. In 2024 International Conference on Machine Intelligence and Smart Innovation (ICMISI) (pp. 295-299). IEEE.

[2] Saxena, A., Pandey, S. N., & Dixit, S. (2024, April). Wheeling Pricing Calculation and Allocation using Deep Reinforcement Learning (DRL) Approach. In 2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 761-765). IEEE.
[3] Hu, Y., & Ghadimi, P. (2023, June). A Review of Artificial Intelligence Application on Enhancing Resilience of Closed-loop Supply Chain. In 2023 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC) (pp. 1-8). IEEE.
[4] Avramelou, L., Nousi, P., Passalis, N., Doropoulos, S., & Tefas, A. (2023, June). Cryptosentiment: A dataset and baseline for sentiment-aware deep reinforcement learning for financial trading. In 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW) (pp. 1-5). IEEE.
[5] Wang, L. (2024, December). Dynamic Pricing Algorithm Based on Deep Reinforcement Learning. In 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 303-308). IEEE.
[6] Kang, A., Xin, J., & Ma, X. (2024). Anomalous Cross-Border Capital Flow Patterns and Their Implications for National Economic Security: An Empirical Analysis. Journal of Advanced Computing Systems, 4(5), 42-54.
[7] Liang, J., Zhu, C., & Zheng, Q. (2023). Developing Evaluation Metrics for Cross-lingual LLM-based Detection of Subtle Sentiment Manipulation in Online Financial Content. Journal of Advanced Computing Systems, 3(9), 24-38.
[8] Wang, Z., & Liang, J. (2024). Comparative Analysis of Interpretability Techniques for Feature Importance in Credit Risk Assessment. Spectrum of Research, 4(2).
[9] Dong, B., & Zhang, Z. (2024). AI-Driven Framework for Compliance Risk Assessment in Cross-Border Payments: Multi-Jurisdictional Challenges and Response Strategies. Spectrum of Research, 4(2).
[10] Wang, J., Guo, L., & Qian, K. (2025). LSTM-Based Heart Rate Dynamics Prediction During Aerobic Exercise for Elderly Adults.
[11] Ma, D., Shu, M., & Zhang, H. (2025). Feature Selection Optimization for Employee Retention Prediction: A Machine Learning Approach for Human Resource Management.
[12] Li, M., Ma, D., & Zhang, Y. (2025). Improving Database Anomaly Detection Efficiency Through Sample Difficulty Estimation.
[13] Yu, K., Chen, Y., Trinh, T. K., & Bi, W. (2025). Real-Time Detection of Anomalous Trading Patterns in Financial Markets Using Generative Adversarial Networks.
[14] Xiao, X., Chen, H., Zhang, Y., Ren, W., Xu, J., & Zhang, J. (2025). Anomalous Payment Behavior Detection and Risk Prediction for SMEs Based on LSTM-Attention Mechanism. Academic Journal of Sociology and Management, 3(2), 43-51.
[15] Xiao, X., Zhang, Y., Chen, H., Ren, W., Zhang, J., & Xu, J. (2025). A Differential Privacy-Based Mechanism for Preventing Data Leakage in Large Language Model Training. Academic Journal of Sociology and Management, 3(2), 33-42.
[16] Zhang, J., Xiao, X., Ren, W., & Zhang, Y. (2024). Privacy-Preserving Feature Extraction for

Medical Images Based on Fully Homomorphic Encryption. Journal of Advanced Computing Systems, 4(2), 15-28.

[17] Ren, W., Xiao, X., Xu, J., Chen, H., Zhang, Y., & Zhang, J. (2025). Trojan Virus Detection and Classification Based on Graph Convolutional Neural Network Algorithm. Journal of Industrial Engineering and Applied Science, 3(2), 1-5.

[18] Ji, S., Liang, Y., Xiao, X., Li, J., & Tian, Q. (2007, July). An attitude-adaptation negotiation strategy in electronic market environments. In Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007) (Vol. 3, pp. 125-130). IEEE.

[19] Xiao, X., Zhang, Y., Xu, J., Ren, W., & Zhang, J. (2025). Assessment Methods and Protection Strategies for Data Leakage Risks in Large Language Models. Journal of Industrial Engineering and Applied Science, 3(2), 6-15.

[20] Liu, X., Chen, Z., Hua, K., Liu, M., & Zhang, J. (2017, August). An adaptive multimedia signal transmission strategy in cloud-assisted vehicular networks. In 2017 IEEE 5th international conference on future internet of things and cloud (FiCloud) (pp. 220-226). IEEE.

[21] Michael, S., Sohrabi, E., Zhang, M., Baral, S., Smalenberger, K., Lan, A., & Heffernan, N. (2024, July). Automatic Short Answer Grading in College Mathematics Using In-Context Meta-learning: An Evaluation of the Transferability of Findings. In International Conference on Artificial Intelligence in Education (pp. 409-417). Cham: Springer Nature Switzerland.

[22] McNichols, H., Zhang, M., & Lan, A. (2023, June). Algebra error classification with large language models. In International Conference on Artificial Intelligence in Education (pp. 365-376). Cham: Springer Nature Switzerland.

[23] Zhang, M., Heffernan, N., & Lan, A. (2023). Modeling and Analyzing Scorer Preferences in Short-Answer Math Questions. arXiv preprint arXiv:2306.00791.

[24] Zhang, M., Wang, Z., Yang, Z., Feng, W., & Lan, A. (2023). Interpretable math word problem solution generation via step-by-step planning. arXiv preprint arXiv:2306.00784.

[25] Zhang, M., Baral, S., Heffernan, N., & Lan, A. (2022). Automatic short math answer grading via in-context meta-learning. arXiv preprint arXiv:2205.15219.

[26] Wang, Z., Zhang, M., Baraniuk, R. G., & Lan, A. S. (2021, December). Scientific formula retrieval via tree embeddings. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 1493-1503). IEEE.

[27] Zhang, M., Wang, Z., Baraniuk, R., & Lan, A. (2021). Math operation embeddings for open-ended solution analysis and feedback. arXiv preprint arXiv:2104.12047.

[28] Jordan, S., Chandak, Y., Cohen, D., Zhang, M., & Thomas, P. (2020, November). Evaluating the performance of reinforcement learning algorithms. In International Conference on Machine Learning (pp. 4962-4973). PMLR.

[29] Qi, D., Arfin, J., Zhang, M., Mathew, T., Pless, R., & Juba, B. (2018, March). Anomaly explanation using metadata. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1916-1924). IEEE.

[30] Zhang, M., Mathew, T., & Juba, B. (2017, February). An improved algorithm for learning to

perform exception-tolerant abduction. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).

[31] Fan, J., Trinh, T. K., & Zhang, H. (2024). Deep Learning-Based Transfer Pricing Anomaly Detection and Risk Alert System for Pharmaceutical Companies: A Data Security-Oriented Approach. Journal of Advanced Computing Systems, 4(2), 1-14.

[32] Ju, C., & Trinh, T. K. (2023). A Machine Learning Approach to Supply Chain Vulnerability Early Warning System: Evidence from US Semiconductor Industry. Journal of Advanced Computing Systems, 3(11), 21-35.

[33] Rao, G., Trinh, T. K., Chen, Y., Shu, M., & Zheng, S. (2024). Jump Prediction in Systemically Important Financial Institutions' CDS Prices. Spectrum of Research, 4(2).

[34] Bi, W., Trinh, T. K., & Fan, S. (2024). Machine Learning-Based Pattern Recognition for Anti-Money Laundering in Banking Systems. Journal of Advanced Computing Systems, 4(11), 30-41.

[35] Dong, B., & Trinh, T. K. (2025). Real-time Early Warning of Trading Behavior Anomalies in Financial Markets: An AI-driven Approach. Journal of Economic Theory and Business Management, 2(2), 14-23.

[36] Trinh, T. K., & Wang, Z. (2024). Dynamic Graph Neural Networks for Multi-Level Financial Fraud Detection: A Temporal-Structural Approach. Annals of Applied Sciences, 5(1).

[37] Trinh, T. K., & Zhang, D. (2024). Algorithmic Fairness in Financial Decision-Making: Detection and Mitigation of Bias in Credit Scoring Applications. Journal of Advanced Computing Systems, 4(2), 36-49.

[38] Wang, Z., Trinh, T. K., Liu, W., & Zhu, C. (2025). Temporal Evolution of Sentiment in Earnings Calls and Its Relationship with Financial Performance. Applied and Computational Engineering, 141, 195-206.

[39] Ni, C., Qian, K., Wu, J., & Wang, H. (2025). Contrastive Time-Series Visualization Techniques for Enhancing AI Model Interpretability in Financial Risk Assessment.

[40] Wang, H., Qian, K., Ni, C., & Wu, J. (2025). Distributed Batch Processing Architecture for Cross-Platform Abuse Detection at Scale. Pinnacle Academic Press Proceedings Series, 2, 12-27.

[41] Chen, Y., Ni, C., & Wang, H. (2024). AdaptiveGenBackend A Scalable Architecture for Low-Latency Generative AI Video Processing in Content Creation Platforms. Annals of Applied Sciences, 5(1).

[42] Wang, Z., Wang, X., & Wang, H. (2024). Temporal Graph Neural Networks for Money Laundering Detection in Cross-Border Transactions. Academia Nexus Journal, 3(2).

[43] Yan, L., Wang, Y., Guo, L., & Qian, K. (2025). Enhanced Spatio-Temporal Attention Mechanism for Video Anomaly Event Detection. Applied and Computational Engineering, 117, 155-164.

[44] Wu, Z., Wang, S., Ni, C., & Wu, J. (2024). Adaptive Traffic Signal Timing Optimization Using Deep Reinforcement Learning in Urban Networks. Artificial Intelligence and Machine Learning Review, 5(4), 55-68.

[45] Ju, C., Jiang, X., Wu, J., & Ni, C. (2024). AI-Driven Vulnerability Assessment and Early Warning Mechanism for Semiconductor Supply Chain Resilience. Annals of Applied Sciences, 5(1).