Supply Chain and Operations Decision Making

www.scodm.reapress.com

Sup. Chain Oper. Decis. Mak. Vol. 2, No. 3 (2025) 156–172.

Paper Type: Original Article

Intelligent Inventory Management in Behpakhsh Company's Supply Chain Using InvAgent: A Large Language Model Based on a Multi-Agent System

Mohammad Bagherian[1,*], Behnam Bagherian[2], Nasim Nahavandi[3]

[1] Student in Industrial Engineering, Tarbiat Modares University, Tehran, Iran; m.bagherian00@gmail.com.

[2] Student in Rail Transportation Engineering, Tehran University of Science and Technology, Iran; Behnam_bagherian@rail.iust.ac.ir

[3] Department of Industrial Engineering, Faculty of Industrial Engineering, Tarbiat Modares University, Tehran, Iran; Nasim_nahavandi@yahoo.com.

Citation:

Received: 14 April 2025 *Revised: 15 June 2025* *Accepted: 26 September 2025*	Bagherian, M., Bagherian, B., & Nahavandi, N. (2025). Intelligent inventory management in Behpakhsh Company's supply chain using Invagent: a large language model based on a multi-agent system. *Supply chain and operations decision making*, *2*(3), 156-172.

Abstract

Behpakhsh Company, as one of the largest distribution and logistics companies in the country, plays a significant role in Supply Chain Management (SCM). SCM involves coordinating and integrating material, information, and financial flows across units to ensure the efficient and effective procurement and distribution of goods. In today's Volatile, Uncertain, Complex, and Ambiguous (VUCA) environment, effective inventory management is essential for the operational success of distribution companies. This paper examines the innovative approach adopted by Behpakhsh Company to leverage InvAgent technology, an artificial intelligence–based language model that uses zero-shot learning to enhance inventory management and reduce costs. By analyzing data and making intelligent decisions under changing conditions, InvAgent improves transparency and adaptability across Behpakhsh's supply chain. The implementation of this model has not only increased efficiency and productivity in Behpakhsh's distribution operations but also helped mitigate the risks of inventory shortages and excessive stockpiling. Ultimately, this study demonstrates that Behpakhsh, through advanced technologies and Large Language Models (LLMs), has achieved improved supply chain performance and enhanced customer satisfaction.

Keywords: Behpakhsh supply chain management, Inventory management, Large language model, Zero-shot learning, InvAgent.

1 | Introduction

Supply Chain Management (SCM) involves the coordination and management of the flows of goods, information, and finances across interconnected and diverse entities, from suppliers to consumers, to deliver products efficiently and effectively. Inventory management, as a vital component of SCM, focuses on monitoring and controlling ordering, storage, and utilization of components and finished products. In today's

Corresponding Author: m.bagherian00@gmail.com

https://doi.org/10.48313/scodm.v2i3.41

Volatile, Uncertain, Complex, and Ambiguous (VUCA) world, effective inventory management is essential for aligning supply and demand, minimizing costs, and enhancing supply chain resilience. It enables firms to remain confident in their ability to adapt to disruptions (Quan et al. [1]), optimize resources (Abaku et al. [2]), and sustain uninterrupted operations (Yasmin [3]) in a highly interconnected, dynamic market environment. Previous studies on inventory management have explored various applications of heuristic approaches, such as the Beer Distribution Game [4–6]. In addition, numerous applications of reinforcement learning–based models have been investigated, including decentralized inventory management (Mousa et al. [7]) and adaptive supply chain synchronization (Kegenbekov and Jackson [8]). However, these approaches often require complex model design and extensive training resources and generally lack explainability. Despite these advances, the application of Large Language Models (LLMs) to address multi-agent Inventory Management Problems (IMP) in the supply chain of Behpakhsh Company has not yet been comprehensively investigated. In this study, InvAgent is introduced as an innovative LLM-based zero-shot multi-agent inventory management system. By leveraging the advanced capabilities of LLMs, this system enhances the operational resilience of Behpakhsh's supply chain and strengthens collaboration among its network components.

The InvAgent approach utilizes the reasoning and decision-making capabilities of LLMs to achieve better coordination, optimize inventory management processes, and enhance transparency and adaptability within Behpakhsh's supply chain. The proposed framework of this paper, illustrated in *Fig. 1*, comprehensively explains the operational mechanism of InvAgent and its impact on improving the efficiency of Behpakhsh's supply chain network.

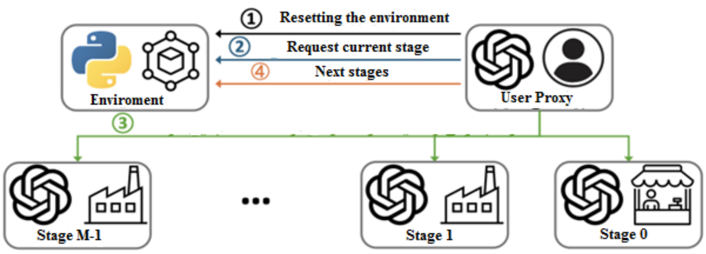

Fig. 1. Illustrates the InvAgent framework.

Fig. 1 illustrates the InvAgent framework, a Zero-Shot LLM-based multi-agent inventory management system. First, the user proxy resets the environment at the beginning of the first episode. In the second step, the user proxy requests the current episode state for each stage from the environment. Then, the user proxy provides the current state to each stage and requests an action from it. Finally, all agents take actions simultaneously and transition to the next state.

Our contributions in this paper, with a focus on Behpakhsh Company, are summarized as follows:

Utilizing large language models for multi-agent inventory management

In this study, the Zero-Shot learning capabilities of LLMs are employed to manage the multi-agent inventory management systems of Behpakhsh Company. This capability enables adaptive and informed decision-making without the need for prior training or explicitly provided examples, thereby enhancing system flexibility.

Model explainability and transparency

The proposed model offers a high level of explainability, further enhanced by Chain-of-Thought (CoT) reasoning. This feature facilitates a clearer understanding of the model's behavior and increases trust in its outcomes, making it a more reliable system compared to traditional heuristic and reinforcement learning approaches.

Dynamic adaptation to variable demand

Our model dynamically adapts to different demand scenarios, thereby minimizing costs and preventing inventory shortages. Through extensive evaluation across diverse scenarios, this capability demonstrates improved efficiency in Behpakhsh Company's SCM and optimizes inventory management processes.

These contributions enable Behpakhsh Company to enhance its supply chain productivity by leveraging advanced technologies and to deliver more reliable and transparent operational performance.

2 | Methodology

This section describes the methodological framework for designing and implementing the InvAgent model. This methodology involves defining a multi-layer, multi-period production–inventory system to model and simulate the supply chain processes of the Behpakhsh Company. Subsequently, the proposed InvAgent model is introduced as an LLM-based multi-agent inventory management system, which is designed to optimize supply chain operations

2.1 | Problem Definition

A multi-layer, multi-period production–inventory system is designed to model and simulate a typical multi-stage supply chain for the production and distribution of non-perishable goods. As illustrated in *Fig. 2*, each stage of the supply chain consists of an inventory storage area (for storing essential materials required for production) and a production facility (for converting raw materials into products).

System characteristics:

I. The output materials of stage i are used as the input materials for stage i–1.

II. The stages are indexed in descending order, from stage M–1 (the upstream beginning of the supply chain) to stage 0, which represents the retail level.

III. Production at each stage is constrained by both the production capacity and the available inventory at that stage.

IV. The flow of raw materials starts from upstream stages and eventually reaches the retail stage to satisfy customer demand.

Time periods (T): the supply chain is divided into discrete time periods t. Each simulation starts at t = 0 with initial conditions, and during each time period T, the following events occur:

Shipment review: each stage receives its inbound shipments that arrive after completing the corresponding transportation lead times.

Order and demand review: each stage places replenishment orders to its corresponding suppliers. Orders are placed based on available production capacity and suppliers' inventory levels. Customer demand at the retail stage is fulfilled based on available inventory.

Order and demand fulfillment: each stage delivers products up to its maximum feasible capacity to satisfy incoming demands or replenishment orders. Orders and demands that cannot be fulfilled within the current time period are prioritized in the subsequent period.

Profit calculation: each stage calculates the costs and profits associated with the following activities: product sales, raw material ordering, penalties for delayed fulfillment, and Holding costs for excess inventory.

This multi-layer, multi-period production–inventory system provides the foundational structure for analyzing and simulating the supply chain processes of the Behpakhsh Company. It accurately models transportation, production, and inventory management processes. The proposed InvAgent leverages this structure to enhance supply chain performance and reduce operational costs.

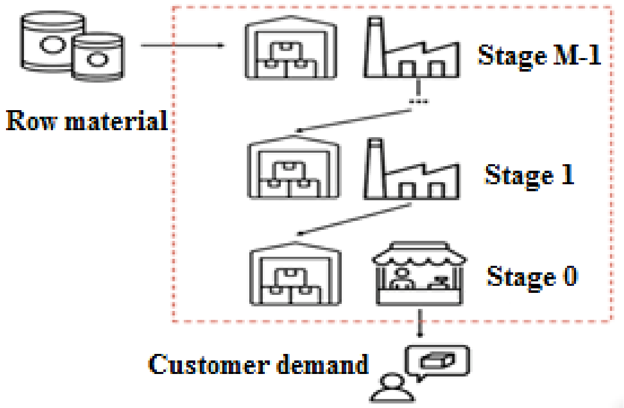

Fig. 2. Flowchart of inventory management of a multi-tier supply chain.

In *Fig. 2*, Raw materials flow through each stage, which includes inventory storage and production facilities. The upstream plant at stage i supplies intermediate products to stage i-1 below, where they are stored as inventory. Stage 0 (retail) supplies finished products to meet customer demand. Considering the items mentioned in *Table 1*, the overall inventory management problem (IMP), which is taken from Hubbs et al. [9], can be described using the following equations:

$$I_{m,t} = I_{m,t-1} + R_{m,t-L_m} - S_{m,t}, \qquad \text{for all } m \in M, \tag{1}$$

$$R_{m,t} = \min\left(B_{m+1,t-1} + O_{m,t,C_m+1}, I_{m+t,t-1} + R_{m+1,t-L_{m+1}}\right), \qquad \text{for all } m = 0, \dots, M-2, \tag{2a}$$

$$R_{M-1,t} = O_{M-1,t}, \tag{2b}$$

$$S_{m,t} = R_{m-1,t}, \qquad \text{for all } m = 1, \dots, M-1, \tag{3a}$$

$$S_{0,t} = \min\left(B_{0,t-1} + D_{t,C_0}, I_{0,t-1} + R_{0,t-L_0}\right), \tag{3b}$$

$$B_{m,t} = B_{M,t-1} + O_{m-1,t} - S_{m,t}, \qquad \text{for all } m = 1, \dots, M-1, \tag{4a}$$

$$B_{0,t} = B_{0,t-1} + D_t - S_{0,t}, \tag{4b}$$

$$P_{m,t} = p_m S_{m,t} - r_m R_{m,t} - k_m B_{m,t} - h_m I_{m,t}, \qquad \text{for all } m \in M. \tag{5}$$

In *Eq (1)*, the current inventory at stage m at the end of the current time period t is equal to the ending inventory of the previous period, plus the fulfilled order from period $t - Lm$, minus the sales during the current period. In *Eq (2.a)*, the fulfilled orders at stage m placed during time period t are determined by the minimum of the following three components: 1) the previous backlog at the downstream stage plus the newly requested orders, 2) the production capacity of the upstream stage, and 3) the total available inventory at the upstream stage $m + 1$ at the beginning of time period *t*, which includes the remaining inventory from the previous period and newly arrived orders after accounting for the transportation lead time. The final fulfilled order is the minimum of these three values, ensuring the order does not exceed any of them. It ensures the supply chain operates within inventory and capacity limits, preventing overcommitment and inventory shortages. *Eq (2.b)* indicates that requested orders at the highest stage are always fulfilled, since an unlimited supply of raw materials is assumed at this stage. As shown in *Eq (3.a)*, sales are always equal to the fulfilled orders, except at stage 0 (the retailer). In *Eq. (3.b)*, sales at stage 0 (the retailer) during time period t are determined by the minimum of the following three components: 1) the previous backlog at stage 0 plus the current customer demand, 2) the production capacity at stage 0, and 3) the total available inventory at stage 0 at the beginning of time period t, which consists of the remaining inventory from the previous period and newly fulfilled orders after accounting for transportation lead time. It ensures that retail sales do not exceed total demand, production capacity, or available inventory. In *Eq. (4.a)*, the backlog at stage m during time period t, for all stages except the retailer, is calculated as the sum of the previous backlog at that stage and the orders requested from the downstream stage, minus the sales at stage m. In *Eq. (4.b)*, the backlog at stage 0 (the retailer) during time period t is calculated similarly to *Eq. (4.a)*; however, the requested order is replaced by customer demand, since the retailer interacts directly with customers. Finally, in *Eq. (5)*, the profit at each stage m during time period t is calculated as sales revenue minus procurement costs, unmet demand penalty costs, and inventory holding costs.

Table 1. Parameter symbols and definitions.

Symbol	Definitation
m	Stages $m \in M = \{0,1,2,\dots,M-1\}$
t	$t \in T = \{0,1,2,\dots,T\}$
$I_{m,t}$	Inventory at the end of time period t
$\hat{I}_{m,t}$	Intended inventory at the end of time period t
$O_{m,t}$	Requested orders placed in the time period t
$R_{m,t}$	Orders filled in the time period t
D_t	Customer demand during the time period t
$S_{m,t}$	Sales during the time period t
$B_{m,t}$	Backlog at the end of time period t
L_m	Transition time between stages m+1 and m
L_{max}	Maximum transition time in the system
$P_{m,t}$	Profit at stage m in time period t
c_m	Production capacity at stage m
p_m	Sales price
r_m	Order cost (procurement)
k_m	Penalty for unfilled orders
h_m	Inventory holding cost

2.2 | InvAgent

In this project, we introduce InvAgent, a LLM–based multi-agent inventory management system designed for supply chain optimization. InvAgent consists of several key agents, including a user proxy and one agent for each stage of the supply chain. The user proxy acts as an intermediary between the environment and all supply chain agents, facilitating communication and data exchange management.

The methodological framework of InvAgent is illustrated in *Fig. 1* and proceeds through the following steps:

I. The user proxy resets the environment at the beginning of the first episode.

II. The user proxy requests the current episode state for each stage from the environment.

III. The user proxy provides the state of each stage to the corresponding agent and requests an action from it.

IV. The user proxy sends the agents' actions to the environment and receives the next state and the reward for that step.

V. The user proxy determines whether the simulation has terminated; if not, the simulation returns to Step 2.

At the beginning of the simulation, we generate system messages for the agents (as shown in *Fig. 3*) that provide essential information, such as definitions, roles, and objectives, within the supply chain. The agent state Sm,tS_{m,t}Sm,t and action am,ta_{m,t}am,t are defined as follows:

$$s_{m,t} = \left[c_m, p_m, r_m, k_m, h_m, L_m, I_{m,t-1}, B_{m,t-1}, B_{m+1,t-1}, S_{m,t-L_{max}}, \dots, S_{m,t-1}, 0, \dots, 0, R_{m,t-L_m}, \dots, R_{m,t-1}\right],$$

and

$a_{m,t} = 0_{m,t}$,

The state comprises the current stage, inventory level, backlog, upstream backlog, recent sales, and received deliveries, with zero remaining layers. As illustrated in *Fig. 4*, the prompt is designed to collect state information and action requests from each agent to ensure effective decision-making and transparent communication across the supply chain. It includes contextual information such as the current time step, the stage, and the stage index, which indicate the model's position within the supply chain. The state description (*Fig. 5*) provides a comprehensive snapshot of inventory levels, backlogs, past sales, and incoming deliveries, thereby enabling informed decision-making. Demand details (*Fig. 6*) and downstream orders (*Fig. 7*) help link supply decisions to immediate needs, allowing upstream suppliers to respond rapidly to downstream orders or demand signals. The strategy description (*Fig. 8*) highlights guidelines such as accounting for lead times and avoiding over-ordering to maintain inventory balance. By requiring justification before action selection, the prompt promotes transparency and interpretability in the decision-making process.

System Message:

Retailer: You play a vital role as Stage 1 (Retailer) in a four-stage supply chain. Your objective is to minimize the total cost by effectively managing inventory and orders.

Wholesaler: You play a vital role as Stage 2 (Wholesaler) in a four-stage supply chain. Your objective is to minimize the total cost by effectively managing inventory and orders.

Fig. 3. System messages that provide critical information, such as definitions, roles, and objectives in the supply chain.

Prompt:

You are currently at period {time period}, and you are at stage {stage} out of {number of stages} in the supply chain. Given the current situation:

{State description}

{Demand description} {Downstream order description}

What is your action (order quantity) for this period?

{Strategy description}

Please first explain your reasoning in 1–2 sentences, and then state your action as a non-negative number enclosed in brackets (e.g., **[0]**)

Fig. 4. Prompt for inventory management simulation in LLMs (see Figs. 5–8).

Demand Description:

- **Constant Demand:** The expected demand at the retail stage (Stage 1) is a fixed demand of 4 units for all 12 periods.
- **Variable Demand:** The expected demand at the retail stage (Stage 1) follows a discrete uniform distribution U {0, 4} for all 12 periods.
- **Higher Demand:** The expected demand at the retail stage (Stage 1) follows a discrete uniform distribution U {0, 8} for all 12 periods.
- **Seasonal Demand:** The expected demand at the retail stage (Stage 1) follows a discrete uniform distribution U {0, 4} for the first 4 periods and a discrete uniform distribution U {5, 8} for the remaining 8 periods.
- **Normal Demand:** The expected demand at the retail stage (Stage 1) follows a truncated normal distribution N (4, 22) at 0 for all 12 periods.

Fig. 5. Upon receiving deliveries, we select the next Lm interval.

Status Description:

- Lead time: **{Lead time} cycles**
- Inventory level: **{Inventory} units**
- Current backlog (you owe the downstream stage): **{Backlog} units**
- Upstream backlog (the upstream stage owes you): **{Upstream backlog} units**
- Previous sales (in recent cycles, from oldest to newest): **{Sales}**
- Incoming deliveries (in this cycle and upcoming cycle(s), from nearest to farthest): **{Deliveries}**

Fig. 6. Demand description for the different demand scenarios included in the order.

Bottom Order Description:

Your bottom order from stage {stage-1} for this round is {[stage-1] actions}.

Fig. 7. Downward ordering from the previous to the current stage in a cycle that can transfer downstream information faster.

Strategy Description:

The golden rule of the game: Open orders should always be equal to "Expected down Orders + Backlogs". If open orders are larger than this, inventory will increase (as soon as open orders arrive). If open orders are smaller than this, the backlog will not decrease and may even increase. Please plan in advance when and where to transfer your order. Remember that your upstream stage has its own transfer time, so don't wait until the inventory runs out. Also, avoid ordering multiple units at once. Try to spread your order over multiple rounds to avoid the bullwhip effect. Anticipate future order changes and adjust your orders accordingly to maintain a stable inventory level.

Fig. 8. Strategy description introducing the golden rule and decision-making guidance for the LLM.

Design Features of the Prompt are as follows

Step 1. Zero-shot learning

The prompt we designed operates via zero-shot learning, providing the LLM with all examples. It means the model must generate responses solely based on its pre-existing knowledge and the information provided in the prompt.

Step 2. Demand description

Unlike reinforcement learning, which involves a training process to improve understanding of the environment and demand, since we do not have prior training, providing a detailed and clear demand description is crucial to ensure accurate understanding and effective responses.

Step 3. Downstream orders

The prompt considers downstream orders, where information can be delivered quickly and shared efficiently across different stages.

Step 4. Human-designed strategy

The inherent LLM strategy is generally sufficient for simple scenarios, such as fixed demands. However, for more complex scenarios, like seasonal demand, it is assumed that human-designed strategies can better assist LLMs in decision-making.

Step 5. Chain-of-Thought (CoT)

The CoT approach can improve the explainability of results. By guiding the LLM through a structured reasoning process, CoT helps the model better understand the scenario and enhance its reasoning, ultimately leading to more reliable, accurate outcomes.

3|Experiments

In this section, we evaluate the performance of InvAgent, our proposed LLM based on a multi-agent inventory management system, by describing the experimental scenarios, baseline models, and experimental settings. Then, we present results demonstrating the adaptability and efficiency of InvAgent, summarized with ablation studies to assess the impact of various prompt components in dynamic SCM.

To evaluate InvAgent's performance in a multi-layer supply chain, a set of experimental scenarios was designed. Each scenario provides specific conditions to examine the robustness and adaptability of the proposed model. The goal of these scenarios is to simulate various situations that a supply chain might encounter in real-world conditions.

3.1|Experimental Scenario Features

Dynamic demand scenarios: situations where customer demand changes unexpectedly. Assessing how the model adapts to complex patterns and extreme demand fluctuations.

Resource constraint scenarios: limitations in production capacity, storage space, and transportation. Analyzing model performance under conditions where resources are insufficient to meet demand.

Multi-period scenarios: include delays in transferring materials between different supply chain stages and the evaluation of the impact of scheduling on inventory and production.

Seasonal demand scenarios: demand fluctuations based on seasons, such as increased demand during peak sales periods. Assessing the model's ability to predict and manage inventory for specific periods.

3.2 | Parameter Configuration

Parameters used in these scenarios are summarized in *Table 2* and include:

I. Demand levels: minimum, average, and maximum demand per period.

II. Production capacity: possible production at each stage of the supply chain.

III. Lead times: delays in transferring materials between supply chain stages.

IV. Associated costs: including production, inventory holding, and penalty costs.

Table 2. Parameter arrangement for different supply chain scenarios.

Parameter	Constant	Variable	Larger	Seasonal	Normal
Number of stages	4	4	4	4	4
Number of periods	12	12	12	12	12
Initial inventory	[12,12,12,12]	[12,12,12,12]	[12,12,12,12]	[12,12,12,12]	[12,14,16,18]
Lead times	[2,2,2,2]	[2,2,2,2]	[2,2,2,2]	[2,2,2,2]	[1,2,3,4]
Demand	4	U(0,4)	U(0,8)	C(4,8)	N(4,22)
Production capacities	[20,20,20,20]	[20,20,20,20]	[20,20,20,20]	[20,20,20,20]	[20,22,24,26]
Selling prices	[0,0,0,0]	[0,0,0,0]	[5,5,5,5]	[5,5,5,5]	[9,8,7,6]
Ordering costs	[0,0,0,0]	[0,0,0,0]	[5,5,5,5]	[5,5,5,5]	[5,6,7,8]
Backorder costs	[1,1,1,1]	[1,1,1,1]	[1,1,1,1]	[1,1,1,1]	[1,1,1,1]
Holding costs	[1,1,1,1]	[1,1,1,1]	[1,1,1,1]	[1,1,1,1]	[1,1,1,1]

In the first scenario, a 4-stage supply chain with a constant demand of 4 units per period is tested over more than 12 time periods, starting with an inventory of 12 units at each stage and a 2-period lead time. This scenario aims to examine the model's basic performance under stable conditions. The second scenario introduces variable demand, uniformly distributed between 0 and 4 units per period, to assess the system's ability to manage fluctuating demand while maintaining efficient inventory levels. The third scenario further increases demand variability, with a uniform distribution between 0 and 8 units per period, and sets the sales and ordering costs at five units per period. This scenario tests the model's capacity to handle high variability and financial impacts. The fourth scenario simulates seasonal demand with a step pattern ranging from 4 to 8 units per period, while maintaining the same financial parameters as in the third scenario, to evaluate system performance under variable but predictable demand patterns. Finally, the fifth scenario involves normally distributed demand with a mean of 4 units and a standard deviation of 2 units per period, along with different lead times, initial inventories, sales prices, and ordering costs at all stages. This scenario examines system performance under more realistic demand fluctuations and varying operational constraints. Collectively, these scenarios provide a testbed for evaluating the efficiency and adaptability of our multi-agent system in dynamic inventory management within a multi-layered supply chain.

3.3 | Experimental Settings

The performance of our model, InvAgent, is evaluated by the cumulative reward across all stages and time periods within a single simulation (episode). The reported numbers for each experiment are averaged over five episodes to reduce uncertainty. For this evaluation, we utilize Python packages such as AutoGen [10], Gymnasium [11], and RLlib [12]. We also leverage LLMs, including GPT-4, GPT-4O, and GPT-4-Turbo [13].

In the constant demand scenario, to improve InvAgent's performance, the final part of the prompt in *Fig. 4* is modified to "([0], [4], or just [8] D)".

The performance of the baseline models is evaluated using the episode reward, averaged over 100 episodes. In the reinforcement learning (RL) section, we explore different hyperparameter settings, including:

I. Number of hidden units: [128,128] and [256,256].

II. Activation function: ReLU.

III. Learning rate: 1e-4, 5e-4, 1e-3.

IV. Training batch size: 500, 1000, 2000.

V. Minibatch size in Stochastic Gradient Descent (SGD): 32, 64, 128.

VI. Number of SGD iterations: 5, 10, 20.

VII. Number of training iterations: 500, 800, 1000, 1500.

VIII. Discount factor: 1.0.

In each experiment, 20 random hyperparameter combinations are selected, and the best configuration is retained. The final hyperparameters used for all scenarios are provided in Appendix C. Reinforcement learning experiments were conducted on an NVIDIA A10 GPU.

3.4 | Experimental Results

The results of the experiments, presented in *Table 3*, highlight the performance of various inventory management models under different demand scenarios. The InvAgent model, in particular, demonstrated remarkable and competitive performance in situations where demand is variable and unpredictable.

In the variable demand scenario, InvAgent (even without using manually designed strategies) achieved the highest average episode reward. Although reinforcement learning models such as MAPPO performed better in some scenarios, the adaptability and zero-shot capability of InvAgent provided unique advantages. This feature enables InvAgent to make reasonable decisions and understand complex concepts without the need for training examples or prior data, functioning much like human intuition.

3.4.1 | Superiority of invagent compared to heuristic models

Compared to heuristic models such as the Base-Stock Policy and Tracking-Demand Policy, InvAgent shows significant Superiority. These traditional models rely on historical or static data and often perform inefficiently when demand patterns exhibit severe, unpredictable changes. In contrast, InvAgent dynamically adapts to real-time conditions and:

I. Minimizes inventory holding costs.

II. Prevents stockouts and supply delays.

The performance of InvAgent under variable-demand conditions was particularly notable, clearly demonstrating its ability to handle complex, unpredictable scenarios. These results indicate that InvAgent not only reduces costs compared to heuristic models but also provides better adaptability.

3.4.2 | Summary of experimental results

InvAgent demonstrates strong capabilities in reducing costs, managing variable demand scenarios, and preventing stock shortages. Compared to heuristic models, it offers higher flexibility and better performance under complex conditions. Compared to reinforcement learning models, it is simpler, more stable, and less costly to implement, though it may achieve lower precision in some cases. These characteristics make InvAgent a suitable and efficient solution for inventory management in Behpakhsh's supply chain, highlighting its ability to optimize supply chain processes and adapt to variable conditions.

Table 3. Avg. episode rewards (±SD) for inventory models across demand scenarios.

Model	Constant	Variable	Larger	Seasonal	Normal
Base-stock	-296.00	-523.69	-392.21	-274.29	-322.44
Demand tracking	-360.00	-412.41	-265.07	-421.90	-232.20
IPPO	-132.17	-389.55	-202.39	-129.73	-102.90
MAPPO	-129.81	-391.53	-106.79	-99.39	-41.98
InvAgent (w/o strategy)	-156.00	-336.60	-350.20	-488.00	-172.60
InvAgent (w/ strategy)	-200.00	-377.60	-357.60	-420.60	-192.40

The comparison of InvAgent's performance in inventory management of BehPakhsh's supply chain indicates that the intrinsic strategy of the Large Language Model (LLM) is generally sufficient for many simpler scenarios, such as constant and variable demand. This capability stems from the LLM's inherent features, which enable dynamic, adaptive decision-making without the need for manually designed strategies.

However, in more complex scenarios, such as seasonal or patterned demands, incorporating human-designed strategies into the model significantly improves decision-making quality and overall performance. It highlights the importance of combining the LLM's intrinsic capabilities with specific strategies to address complex conditions in BehPakhsh's supply chain.

3.5 | Attrition Studies in Behpakhsh Company: Examining the Impact of Prompt Components in the InvAgent Model

To evaluate the impact of different prompt components on the InvAgent model's performance in managing Behpakhsh Company's supply chain under a variable-demand scenario, ablation studies were conducted. Details of this study are presented in *Table 4*.

Scenario study: demand randomly varies between 0 and 4 units per period. These variations are repeated over 12 periods to simulate demand dynamics and uncertainty.

3.5.1 | Results of ablation studies

Prompt without strategy section: This performed best in this scenario. Other models were compared using this prompt as the reference.

Impact of different prompt components: this component of the prompt played a key role in improving model performance by providing precise information about demand fluctuations. This component optimized the model's performance in coordinating with downstream levels of the supply chain, thereby reducing delays. By utilizing agents' historical activity, the model was able to make more informed decisions. This feature allowed the model to leverage all previous messages throughout the episode. Reasoning through the CoT made the model's decision-making more structured and transparent. This process played an important role in inventory management, helping prevent stockouts and overstocking.

3.5.2 | Prompt examples and model responses

An example of a prompt and the InvAgent model's response for a constant demand scenario using GPT-4 is shown in *Fig. 9*. The prompt included a detailed description of the current state, a demand description specifying expected retail demand, and a description of a strategy aligning open orders with expected downstream orders while considering lead times and the bullwhip effect. In response, the retail agent, given that the current inventory is sufficient to meet maximum demand for up to 3 periods with a 2-period lead time, decides not to place an order in this period to prevent excess inventory accumulation, as explained in the agent's reasoning.

3.5.3 | Analysis of the invagent application in behpakhsh

This example demonstrates how InvAgent can make intelligent decisions across various supply chain scenarios for Behpakhsh using precise prompts and logical analysis. In this example:

I. Prevention of inventory accumulation: the model's decision helps reduce holding costs and prevents resource wastage.

II. Consideration of lead times and bullwhip effect: By accounting for lead times, InvAgent aligns new orders with current conditions.

III. Logical and transparent response: transparency in decision-making logic increases trust in the model and enables review and evaluation of responses.

This approach not only demonstrates InvAgent's high adaptability but also highlights its role in improving inventory management and reducing costs in Behpakhsh's supply chain.

Evaluation of base exploratory variables results

The evaluation results of base stock policies and demand-tracking policies are presented in *Table 5*. These evaluations show how different inventory policies perform under various demand conditions. In the first part of the table, the performance of the base-stock policy is shown for different desired inventory levels, based on production capacity. The results indicate that maintaining a lower inventory level generally performs better under varying demand conditions. In the second part, five types of demand-tracking policies are presented, defined using various formulas including sales, lead time, and backlog. While none of these policies consistently outperform the others, averaging sales is generally effective in managing variable demand in most scenarios.

3.6 | Case Studies at Behpakhsh: Evaluating InvAgent Performance under Different Demand Scenarios

This section presents two case studies of InvAgent's performance in managing inventory in Behpakhsh's supply chain. These studies examine variable demand scenarios without a strategy and seasonal demand with a strategy. The supply chain structure in both scenarios includes four stages: retailer, wholesaler, distributor, and manufacturer, flowing from downstream to upstream.

Variable demand scenario: innovative performance analysis in Behpakhsh's supply chain

Under variable demand, InvAgent's performance in managing demand changes, inventory, accumulation, and profit in Behpakhsh's supply chain was examined. As shown in *Fig. 10*, the model responded to demand dynamics through real-time understanding and proper timing.

Supply chain response to demand changes

I. At the start of the simulation, demand suddenly increases, and the retailer, as the first point of contact with customers, faces inventory depletion.

II. Retailer action: to meet customer demand, the retailer places new orders with the wholesaler.

III. Inventory stabilization: due to supply chain lead times (from manufacturer to retailer), the retailer's inventory is not immediately stabilized.

IV. Wholesaler and distributor actions: as intermediary links, they fulfill orders for upstream customers and replenish their inventory after receiving orders from retailers.

V. Manufacturer: at the top of the supply chain, manufacturers process incoming orders and convert raw materials into products.

Inventory accumulation at the distributor

Midway through the simulation, inventory accumulation at the distributor reached its peak due to:

I. Previous inventory depletion: In period 6, the distributor's inventory is exhausted because new orders were either not placed on time or delayed by transfer lead times.

II. High accumulation in period 7: Newly arrived inventory in period 7 causes accumulation at the distributor.

Lessons learned and recommendations

I. Lead time management: to prevent accumulation or shortages, the distributor should place orders considering supply chain lead times, stabilizing inventory flow.

II. Demand forecasting: accurate forecasts from InvAgent can optimize inventory provisioning across supply chain levels.

III. Supply chain coordination: all levels, from manufacturer to retailer, must coordinate orders to ensure timely inventory transfer.

Seasonal demand scenario: Inventive performance analysis at Behpakhsh

Under seasonal demand, InvAgent reacts to changing demand patterns and their impact on inventory, accumulation, and profit. This scenario is designed based on different demand patterns across time periods. *Fig. 11* shows how the model places orders and manages inventory.

Demand pattern in the scenario

I. Periods 1–4: demand follows a uniform distribution, ranging from 0 to 4 units per period.

II. Periods 5–12: demand shifts to a different uniform distribution, indicating a peak period.

Model performance across periods 1–4: increased downstream orders

I. Retailers and wholesalers anticipate increased future demand and place larger orders with the distributor.

II. Distributor and manufacturer: attempt to procure large amounts from upstream suppliers, quickly depleting manufacturer inventory and increasing new orders from the distributor.

Period 4: accumulation and reduced manufacturer profit

I. Inventory accumulation: increased order volume in period 4 significantly reduces manufacturer inventory and increases supply chain accumulation.

II. Profit reduction: due to inventory management costs and poor coordination, the manufacturer's profit reaches a minimum.

Periods 5–12: restoring balance and flexibility

Order management: the manufacturer starts ordering raw materials and gradually restores inventory and orders to balance.

Reduced accumulation: improved order management and better supply chain coordination reduce inventory accumulation and enhance profitability.

Example

Reason: By forecasting future demand for the next two periods (considering lead time) and noting that current inventory is sufficient to meet maximum demand for three periods, I decided not to place any orders this period to avoid excessive inventory buildup.

Action: [0]

Currently, this is the first period, and you are at stage 1 of 4 in the supply chain. Based on your current status:

- Lead time: 2 periods
- Current inventory: 12 units
- Current backlog (your debt to downstream): 0 units
- Upstream backlog (upstream debt to you): 0 units
- Past sales (in recent periods, from oldest to newest): [0, 0]
- Incoming deliveries (in this period and subsequent periods, from near to far): [0, 0]

The forecasted demand at the retailer (stage 1) is a discrete uniform distribution U {0,4} for all 12 periods. What will be your action (order quantity) for this period?

Golden rule of this game: Open orders should always equal "expected downstream orders + backlog." If open orders exceed this amount, inventory will increase (once orders arrive). If open orders are less than this amount, the backlog will not decrease and may even increase. Please consider lead time and pre-place your orders. Note that your upstream stage also has its lead time, so do not wait until your inventory runs out. Also, avoid placing large orders at once. Try to spread your orders over multiple periods to prevent the bullwhip effect. Forecast future demand changes and adjust your orders accordingly to maintain stable inventory levels.

First, state your reasoning in 1–2 sentences, then enter your action as a non-negative number in brackets (e.g., [0])

Retailer Agent (via User Interface):

Reason: By forecasting future demand over the next two periods (considering the lead time) and noting that the current inventory is sufficient to meet maximum demand for the next three periods, I decided not to place any order this period to prevent excessive inventory accumulation.

Fig .9. Example of InvAgent input and response for a constant demand scenario in the company to broadcast using GPT-4.

Table 5. Evaluation results (average episode rewards and standard deviations) for a variety of norm-based models in different demand scenarios for broadcasting.

Inventory	Constant	Variable	Larger	Seasonal	Normal
$0.8c_m$	-208.00	-435.69	-234.28	-207.75	-150.67
$0.9c_m$	-252.00	-479.69	-310.74	-229.08	-226.31
c_m	-296.00	-523.69	-392.21	-274.29	-322.44
$S_{m,t-1}L_m + B_{m,t-1}$	-364.00	-390.17	-393.31	-525.84	-283.39
$S_{m,t-1}(L_m + 1) + B_{m,t-1}$	-120.00	-395.68	-470.55	-524.26	-351.23
$\bar{S}_{m,t-1}L_m + B_{m,t-1}$	-360.00	-412.41	-265.07	-421.90	-232.20
$\bar{S}_{m,t-1}(L_m + 1) + B_{m,t-1}$	-252.00	-382.77	-489.75	-610.03	-177.54
$1.2\bar{S}_{m,t-1}L_m + B_{m,t-1}$	-361.00	-397.22	-325.81	-479.07	-218.98

Table 6. Hyperparameters for independent neighbor policy optimization (IPPO) with parameter sharing as a basis for comparison in broadcast.

Parameter	Constant	Variable	Larger	Seasonal	Normal
Hidden units	[128,128]	[256,256]	[128,128]	[128,128]	[128,128]
Activation function	ReLU	ReLU	ReLU	ReLU	ReLU
Learning rate	0.0001	0.0001	0.001	0.0005	0.0005
Training batch size	1000	1000	2000	2000	1000
SGD mini-batch size	128	128	128	128	128
SGD iterations	5	10	5	5	5
Training episodes	1000	1500	1000	800	500

Table 7. Hyperparameters for the basic multi-agent proximal policy optimization (mappo) model.

Parameter	Constant	Variable	Larger	Seasonal	Normal
Hidden units	[128,128]	[128,128]	[128,128]	[256,256]	[128,128]
Activation function	ReLU	ReLU	ReLU	ReLU	ReLU
Learning rate	0.0001	0.0001	0.001	0.0001	0.0001
Training batch size	500	2000	2000	1000	500
SGD mini-batch size	128	32	32	128	128
SGD iterations	10	5	10	10	10
Training episodes	500	500	800	1500	500

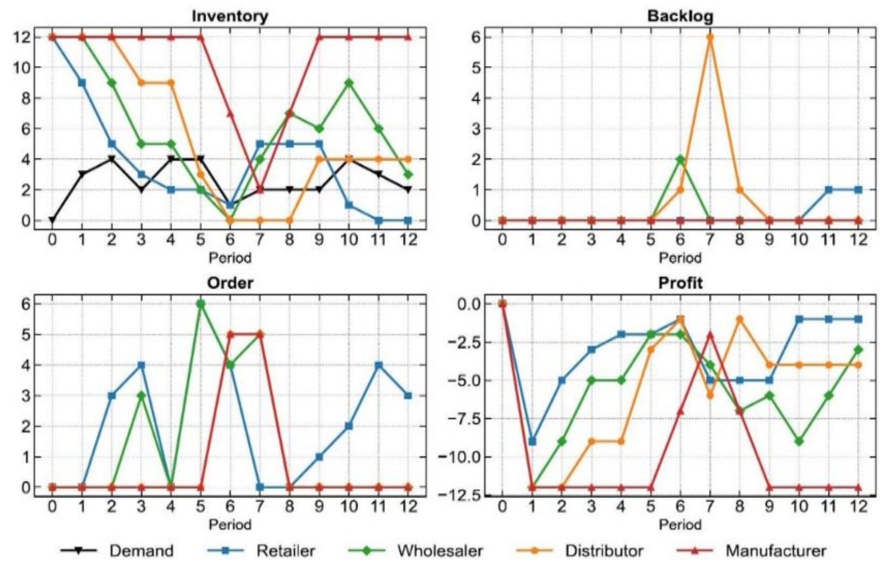

Fig. 10. Inventory, backlog, orders, and profit under variable demand (InvAgent, no strategy).

4 | Conclusion

In this study, the effectiveness of using LLMs as autonomous agents in multi-agent inventory management for optimizing the Behpaksh supply chain was investigated. Our innovative model, InvAgent, was able to make adaptive and informed decisions without the need for prior training by leveraging the zero-shot learning capabilities of LLMs. One of the outstanding features of InvAgent was its integration of structured reasoning via the CoT method, which improved the model's explainability and transparency. This feature made InvAgent a more reliable system than traditional heuristic and reinforcement learning models.

Key results

I. Competitive performance: Invariant demonstrated superior performance and lower costs across various demand scenarios compared to heuristic policies.

II. Adaptability: the model adapted to complex, unpredictable demand changes and minimized shortages.

III. High potential: the use of LLMs for intelligent SCM highlighted reductions in inventory costs and improved decision-making.

Future outlook

To expand the capabilities of InvAgent and improve its performance in the Behpaksh Company's supply chain, the following axes will be considered in future research:

I. Use of reinforcement learning: tuning the model to improve decision-making capabilities through iteration and learning optimal strategies.

II. Evaluation with real-world data: 1) using real supply chain data, including seasonal and variable demand patterns, to practically evaluate the model's performance, and 2) decomposing data into level, trend, and seasonal components to improve forecast accuracy.

III. Combining human-designed strategies with the inherent capabilities of LLMs: improving the management of complex and unpredictable demand patterns by combining these two approaches.

Final conclusion

The InvAgent model demonstrated how new technologies, such as LLMs, can make Behpaksh Company's supply chain more efficient and flexible. This research is an important step towards the digitalization and intelligence of SCM and enables the use of advanced solutions to manage future challenges.

Table 8. Ablation of InvAgent prompts in variable demand; rewards averaged over 5 trials (±SD) with % change from first result.

Model	Reward	Performance change
GPT-4	-336.60	0.00%
GPT-4	-377.60	-12.18%
GPT-4	-349.40	-3.80%
GPT-4	-419.00	-24.48%
GPT-4	-379.40	-12.72%
GPT-4	-339.20	-0.77%
GPT-4	-369.80	-9.86%
GPT-4	-387.40	-15.09%
GPT-4o	-405.00	-20.32%
GPT-4-Turbo	-636.40	-89.07%

Acknowledgments

The authors would like to express their sincere appreciation to all individuals whose valuable comments and support contributed to the successful completion of this study.

Funding

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

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Quan, Y., Pothen, A., & Montreuil, B. (2023). Predictive demand disruption signals for supply chain networks. *IISE annual conference and expo.* IISE. https://www.researchgate.net/

[2] Abaku, E. A., Edunjobi, T. E., & Odimarha, A. C. (2024). Theoretical approaches to AI in supply chain optimization: Pathways to efficiency and resilience. *International journal of science and technology research archive*, *6*(1), 92–107. https://doi.org/10.53771/ijstra.2024.6.1.0033

[3] Yasmin, G. (2024). Supply chain management: ensuring seamless operations. *Journal of management science research review*, *2*(1), 55–66. https://www.jmsrr.com/index.php/Journal/article/view/25

[4] Oroojlooyjadid, A., Nazari, M., Snyder, L. V, & Takáč, M. (2022). A deep q-network for the beer game: Deep reinforcement learning for inventory optimization. *Manufacturing & service operations management*, *24*(1), 285–304. https://doi.org/10.1287/msom.2020.0939

[5] Goodwin, J. S., & Franklin, S. G. (1994). The beer distribution game: using simulation to teach systems thinking. *Journal of management development*, *13*(8), 7–15. https://doi.org/10.1108/02621719410071937

[6] Edali, M., & Yasarcan, H. (2014). A mathematical model of the beer game. *Journal of artificial societies and social simulation*, *17*(4), 2. https://doi.org/10.18564/jasss.2555

[7] Mousa, M., van de Berg, D., Kotecha, N., del Rio Chanona, E. A., & Mowbray, M. (2024). An analysis of multi-agent reinforcement learning for decentralized inventory control systems. *Computers & chemical engineering*, *188*, 108783. https://doi.org/10.1016/j.compchemeng.2024.108783

[8] Kegenbekov, Z., & Jackson, I. (2021). Adaptive supply chain: Demand--supply synchronization using deep reinforcement learning. *Algorithms*, *14*(8), 240. https://doi.org/10.3390/a14080240

[9] Hubbs, C. D., Perez, H. D., Sarwar, O., Sahinidis, N. V, Grossmann, I. E., & Wassick, J. M. (2020). Or-gym: A reinforcement learning library for operations research problems. *ArXiv preprint arxiv:2008.06319*. https://doi.org/10.48550/arXiv.2008.06319

[10] Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., ... & Wang, C. (2024). Autogen: Enabling next-gen LLM applications via multi-agent conversations. *First conference on language modeling*. 1-46. https://openreview.net/pdf?id=BAakY1hNKS

[11] Towers, M., Terry, J. K., Kwiatkowski, A., Balis, J. U., de Cola, G., Deleu, T., Goulão, M., Kallinteris, A., KG, A., Krimmel, M., Perez-Vicente, R., Pierré, A., Schulhoff, S., Tai, J. J., Shen, A. T. J., & Younis, O. G. (2023). *Gymnasium*. Zenodo. https://doi.org/10.5281/zenodo.8127025

[12] Liang, E., Liaw, R., Nishihara, R., Moritz, P., Fox, R., Goldberg, K., ... & Stoica, I. (2018). RLlib: Abstractions for distributed reinforcement learning. In *International conference on machine learning* (pp. 3053-3062). PMLR. https://doi.org/10.48550/arXiv.1712.09381

[13] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*. https://doi.org/10.48550/arXiv.2303.08774