

Review

Not peer-reviewed version

---

# Recent Applications of Artificial Intelligence in Logistics: An Evidence-Oriented Operational Review of Studies Published Since 2024

---

[Byunghyun Ban](#) \*

Posted Date: 9 May 2026

doi: 10.20944/preprints202605.0444.v1

Keywords: artificial intelligence; logistics; supply chain management; vehicle routing; warehouse logistics; port operations; evidence profile; field validation; operational decision support



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Recent Applications of Artificial Intelligence in Logistics: An Evidence-Oriented Operational Review of Studies Published Since 2024

Byunghyun Ban <sup>1,2</sup>

<sup>1</sup> Department of Music, Gyungkook National University; bhban@nanalab.kr

<sup>2</sup> NaNa Lab

## Abstract

Artificial intelligence (AI) research in logistics has rapidly shifted since 2024 toward operationally specific systems for forecasting, routing, warehouse optimization, supply-chain visibility, port scheduling, and smart-port maintenance. This paper presents an evidence-oriented operational review of studies published since 2024 on AI applications in logistics. Instead of classifying the literature by model type alone, the review organizes studies by logistics decision function and evaluates the evidence profile of each application: whether real operational data were used, whether the study relied on simulation or benchmark instances, whether the data scale was reported, and whether field validation was conducted. The analysis shows that recent AI logistics studies increasingly address concrete operational tasks, including demand forecasting, late-delivery prediction, route deviation prediction, dynamic vehicle routing, warehouse order picking, robotic fulfillment scheduling, vessel arrival prediction, berth allocation, quay crane scheduling, container dwell-time prediction, and predictive maintenance at smart ports. However, the evidence base remains uneven. Several studies use real operational or survey data, while many warehouse and routing studies rely on simulation, generated instances, or benchmark settings. Field validation remains rare. The paper argues that the next stage of logistics AI research should move beyond model accuracy and report operational evidence: data provenance, data scale, logistics KPIs, field validation, integration requirements, and human oversight.

**Keywords:** artificial intelligence; logistics; supply chain management; vehicle routing; warehouse logistics; port operations; evidence profile; field validation; operational decision support

## 1. Introduction

Logistics is a natural domain for artificial intelligence because it consists of repeated operational decisions under constraints. Firms must forecast demand, position inventory, assign vehicles, plan routes, schedule warehouse tasks, coordinate robot fleets, predict vessel arrival times, allocate berths, manage cranes, and respond to disruptions. These tasks are data-intensive, time-sensitive, and costly when handled inefficiently.

Since 2024, AI logistics research has become increasingly operational. Recent studies no longer discuss AI only as a broad digital transformation trend. They propose concrete systems for demand forecasting, inventory and risk prediction, graph-based supply-chain visibility, last-mile delivery optimization, route deviation prediction, warehouse order picking, robotic fulfillment systems, vessel arrival prediction, berth allocation, quay crane scheduling, container dwell-time prediction, and port predictive maintenance.

However, not all evidence is equal. A study that uses real terminal data and reports operational effects is different from a study that evaluates an algorithm on generated instances. A survey-based study of AI adoption is different from a simulation study of vehicle routing. A model that improves RMSE is not necessarily validated in field operation. Therefore, recent AI logistics research should

not be evaluated only by algorithm type or reported accuracy. It should also be evaluated by the evidence profile behind each claim.

This paper reviews studies published since 2024 and asks three questions:

1. Which logistics decision functions are being addressed by recent AI studies?
2. What kind of evidence supports each application: real data, simulation, benchmark instances, survey data, panel data, or operational deployment?
3. How often do recent studies report data scale, logistics KPIs, and field validation?

The paper's contribution is not to claim that AI logistics is mature or immature as a whole. Instead, it provides a structured way to read the recent literature: what task was addressed, what data were used, how large the data were, and whether the system was validated beyond simulation.

## 2. Method and Scope

### 2.1. Review Type

This paper is an evidence-oriented operational review of studies published since 2024. It is not a bibliometric study and does not claim exhaustive coverage of all AI logistics publications. The purpose is to synthesize recent application-oriented work according to operational decision function and evidence profile.

### 2.2. Time Boundary

This review covers studies published from January 2024 onward. The phrase "since 2024" includes studies published in 2024, 2025, and early 2026. The boundary was chosen because recent AI logistics literature has increasingly focused on generative AI, graph-based supply-chain visibility, reinforcement-learning-based warehouse and routing methods, and AI-enabled smart-port operations.

### 2.3. Inclusion Criteria

Studies were included if they satisfied the following criteria:

1. publication year 2024 or later;
2. direct relevance to logistics, supply-chain operations, transportation, warehousing, or port operations;
3. use of AI, machine learning, deep learning, reinforcement learning, generative AI, graph learning, or intelligent decision-support methods;
4. focus on a concrete operational decision task;
5. sufficient methodological information to classify data type and validation approach.

### 2.4. Exclusion Criteria

The review excluded:

1. corporate press releases;
2. consulting reports;
3. business news;
4. generic AI papers without logistics application;
5. logistics papers without identifiable AI or machine-learning components.

Review articles were not used as the main evidence base. They may inform background awareness, but the synthesis below focuses on application-oriented studies that report a specific AI method, operational task, and evaluation setting.

### 2.5. Evidence Profile Framework

Each study was evaluated using five evidence-profile questions.

**Table 1. Evidence-profile questions used in this review.**

Evidence dimension	Question
Data type	Did the study use real operational data, survey data, panel data, benchmark data, simulated data, or generated instances?
Data scale	Did the study report the number of orders, samples, records, vessels, routes, containers, respondents, or instances?
Simulation dependence	Was the main evaluation simulation-based or benchmark-based?
Field validation	Was the method tested in a real operational environment?
Logistics KPI linkage	Did the study report logistics-relevant outcomes such as delivery time, fill rate, route cost, order throughput, vessel waiting time, relocations, downtime, or service level?

This framework replaces vague claims about maturity. The key question is not whether an application sounds advanced, but whether the evidence is operationally strong.

### 3. Supply Chain Planning and Visibility

#### 3.1. Demand Forecasting and Inventory Support

Demand forecasting remains a central AI task in logistics. Na proposed an LSTM-based demand prediction model for supply chains and evaluated it through simulation experiments [1]. The paper constructs objective functions and constraints for demand forecasting, builds an LSTM model, and compares predicted and actual values. Its evidence profile is therefore simulation-oriented: the study addresses a concrete logistics function, but field deployment and real operational validation are not clearly established.

Sattar et al. provide a stronger data-scale example. Their 2025 study uses the DataCo Smart Supply Chain dataset, drawing from ERP and logistics databases covering 180,519 orders over a full calendar year [2]. The study compares ARIMA, XGBoost, Random Forest, RNN, and other models for demand forecasting, inventory optimization, late delivery prediction, fraud detection, and risk mitigation. It also reports logistics-relevant outcomes such as fill rate, stockouts, and holding/ordering costs.

This contrast is useful. Both studies address supply-chain forecasting, but their evidence profiles differ. One is mainly simulation-based; the other uses a large open operational dataset and connects model outputs to downstream inventory and risk decisions.

#### 3.2. Supply-Chain Visibility and Relationship Inference

Zheng and Brintrup proposed a graph-based machine learning approach for supply-chain visibility using federated learning and graph convolutional networks [3]. The study aims to predict hidden relationships in supply-chain knowledge graphs while preserving privacy. It reports validation with real-world supply-chain data, although the supporting data are not publicly available because of commercial sensitivity.

A related work by Zheng and Brintrup uses generative AI-enhanced embeddings and machine-learning models for relationship prediction in supply-chain knowledge graphs [4]. The paper addresses the problem that link prediction alone may identify possible relationships but fail to capture contextual information such as products or supply locations. In this case, generative AI is used not as a free-form decision-maker, but as a semantic representation layer for supply-chain visibility.

These studies suggest an important recent direction: AI in supply chains is moving from pure forecasting toward relationship inference and visibility enhancement. The evidence is promising because real or case-based data are used, but the operational validation remains limited. The studies improve analytical visibility, but do not yet show broad field deployment across production logistics systems.

### 3.3. AI Adoption and Supply-Chain Resilience

Some recent studies evaluate AI adoption at the organizational level rather than the algorithm level. Pan et al. surveyed 231 Chinese manufacturing senior executives and supply-chain managers and used partial least squares structural equation modeling to examine the effect of AI usage on supply-chain resilience [5]. The study reports that AI usage has a positive effect on supply-chain resilience, with efficiency and collaboration acting as mediating factors.

Guo et al. examine generative AI adoption and supply-chain resilience using panel data from Chinese A-share listed firms from 2017 to 2022 [6]. This is not a logistics algorithm paper, but it contributes evidence about organizational-level AI adoption and resilience.

The evidence profiles of these studies differ from forecasting or routing papers. They do not validate a logistics algorithm in a warehouse or transport network. Instead, they provide survey or panel-data evidence on AI adoption and supply-chain outcomes. These studies are useful for understanding organizational impact, but they should not be interpreted as field validation of a specific AI logistics system.

## 4. Transportation and Last-Mile Delivery

### 4.1. Vehicle Routing Approximation

Vehicle routing remains a core AI logistics problem. Ammon et al. proposed a supervised machine-learning approach for the capacitated vehicle routing problem using graph neural networks trained on solved instances generated by OR-Tools [7]. The study uses generated problem instances rather than real delivery data. Its contribution is algorithmic: learning to approximate routing solutions for a classical combinatorial optimization problem.

This kind of study is useful for methodological development, but its evidence profile should be read carefully. It demonstrates computational feasibility, not field validation. Data scale should be described in terms of generated instances and problem sizes rather than real orders or deliveries.

### 4.2. Dynamic Last-Mile Delivery

Jiang and Chang proposed ML-CALMO, a machine-learning-enhanced cloud-assisted last-mile optimization framework that integrates LSTM demand forecasting, CNN traffic prediction, and DQN-based routing with queueing theory [8]. The authors provide public code and data for reproducibility, including Docker containers, random seeds, and hyperparameters. Importantly, they also acknowledge the limitation of simulation-based evaluation and the need for field validation.

This makes ML-CALMO a useful example of transparent but still simulation-dependent research. The paper is valuable because it integrates multiple AI modules into a logistics decision pipeline, but the evidence remains short of real-world deployment.

### 4.3. Route Deviation Prediction

Konovalenko et al. study last-mile route deviation prediction using machine learning [9]. The goal is to predict whether drivers will deviate from optimized planned routes and to quantify the degree of deviation. This is an operationally meaningful problem because real drivers often deviate from planned routes based on local knowledge, constraints, or situational judgment.

Unlike pure route optimization, route deviation prediction supports dispatch monitoring and risk anticipation. Its value lies in identifying where planned routes are likely to fail in practice.

However, evidence strength depends on the quality and scale of historical planned-versus-actual route data, which should be explicitly reported in future work.

#### 4.4. Interpretation

Transportation studies published since 2024 show two evidence patterns. Some works use generated or simulated instances to test routing algorithms. Others use historical route or traffic data to predict deviations or operational outcomes. For logistics adoption, the second type is generally closer to field relevance, but both types require additional validation before being treated as deployment-ready.

## 5. Warehouse Logistics and Robotics

### 5.1. Dynamic Order Picking

Warehouse order picking is one of the most important intralogistics tasks. Mahmoudinazlou et al. proposed a deep reinforcement learning framework for dynamic order picking in a single-block warehouse with an autonomous picking device [10]. The study formulates the problem as a Markov decision process and evaluates performance against benchmark algorithms. The authors report that under high order arrival rates, the DRL approach can achieve approximately 98% order fulfillment compared with 82% for benchmark algorithms.

The evidence profile is computational. The study uses generated or benchmark-style warehouse instances rather than field deployment in a live warehouse. Its operational relevance is high because order picking is a real warehouse bottleneck, but the system is not field-validated.

### 5.2. Real-Time Task Planning

Zhang et al. examine real-time task planning for order picking in intelligent logistics warehousing [11]. The paper jointly considers order assignment and path planning, which are often separated in warehouse research. This is operationally important because assignment and path planning interact in real warehouses.

The evidence profile appears algorithmic and experimental rather than field deployment-based. Therefore, the study is best interpreted as an intelligent planning contribution that needs additional validation in real warehouse systems.

### 5.3. Robotic Mobile Fulfillment Systems

Zhu et al. study robotic mobile fulfillment systems and focus on order allocation, shelf selection, and coordinated robot scheduling [12]. Their work introduces deep reinforcement learning and optimization techniques into robotic mobile fulfillment operations. This line of research is important because warehouses increasingly rely on robot fleets, and robot scheduling affects throughput, congestion, and order completion.

The evidence profile is model- and experiment-based. It addresses realistic warehouse decision tasks but does not, based on the accessible source, establish field validation in an operating warehouse.

### 5.4. AGV Path Control

Recent AGV studies also show the increasing use of AI in warehouse and manufacturing logistics. Umakarthy et al. propose an intelligent path-control approach for autonomous AGVs in flexible manufacturing systems using dynamic programming and reinforcement-learning-related decision structures [13]. The study focuses on path selection and obstacle-aware navigation.

This type of work is relevant to material handling logistics. However, the evidence profile again depends on whether the study uses simulated environments, laboratory settings, or field trials. For

adoption, future studies should report operating environment, number of AGVs, obstacle scenarios, path length, collision rate, completion time, and field-test conditions.

### 5.5. Interpretation

Warehouse AI research since 2024 is rich in algorithmic innovation but still often simulation-heavy. Recent studies address real operational problems--order picking, task planning, robot scheduling, and AGV navigation--but field validation remains limited. Stronger evidence would require deployment in actual warehouses, integration with warehouse management systems, and reporting of logistics KPIs such as order throughput, picking time, robot idle time, congestion, labor productivity, and error rate.

## 6. Port Logistics and Smart Port Operations

### 6.1. Vessel Arrival Prediction and Berth Allocation

Zhang et al. propose dynamic berth allocation based on vessel arrival time prediction [14]. The study uses real-time AIS data to predict vessel estimated time of arrival and incorporates those predictions into a berth allocation optimization model. The paper reports AIS data collected for container vessel arrivals at a specific port from May 27 to June 27, 2024. This is a relatively strong operational data profile because it uses real AIS data and links prediction to downstream berth planning.

The study is important because it demonstrates a useful pattern in port AI: prediction alone is not the final goal. ETA prediction becomes operationally valuable when integrated into berth allocation decisions that reduce waiting time and improve berth utilization.

### 6.2. Quay Crane Scheduling

You et al. propose a deep reinforcement learning approach for quay crane scheduling using proximal policy optimization and a hybrid metaheuristic algorithm [15]. Quay crane scheduling is a central port operation because crane allocation affects vessel turnaround time and terminal productivity.

The evidence profile is computational and optimization-focused. The study contributes to adaptive scheduling, but further evidence is needed on real terminal integration, data scale, crane fleet size, vessel schedules, and operational constraints.

### 6.3. Low-Carbon Berth and Crane Scheduling

Jo et al. propose hierarchical reinforcement learning for real-time berth allocation, quay crane assignment, and scheduling under uncertainty and carbon-emission considerations [16]. This study is notable because it connects AI scheduling with environmental objectives. It addresses not only operational efficiency but also low-carbon planning.

The evidence profile should be evaluated according to whether real port data are used and whether carbon-emission outcomes are connected to operational decision variables. Even when field validation is absent, the study is valuable because it shows that recent port AI research increasingly includes sustainability constraints.

### 6.4. Generative AI for Container Dwell-Time Prediction

Kim et al. propose a framework combining generative AI and machine learning for import container dwell-time prediction in port logistics [17]. The framework uses generative AI to standardize unstructured owner and cargo information into standard codes, then uses the standardized data in machine-learning models. The authors report experiments on real container terminal data, a 13.88% improvement in mean absolute error compared with models without

standardized information, and up to 14.68% reduction in container relocations when improved predictions are applied to stacking strategies.

This is one of the strongest examples in the reviewed set because it connects AI to an operational port KPI: relocation reduction. It also shows a practically important role for generative AI. The generative model is not used to make unconstrained decisions; it is used as a data-standardization layer that improves downstream prediction.

#### 6.5. Predictive Maintenance in Smart Ports

Aslam et al. examine machine-learning-based predictive maintenance at smart ports using IoT sensor data [18]. The study applies models such as ANN, decision tree, random forest, XGBoost, and Gaussian Naive Bayes to port machinery maintenance. The work is relevant because port equipment failure can delay terminal operations.

The evidence profile is stronger than pure simulation if the IoT sensor data are drawn from actual port equipment, but field deployment still needs to be distinguished from offline model evaluation. Future work should report equipment type, number of sensors, time span, number of failure events, model latency, edge-device constraints, and maintenance decision outcomes.

#### 6.6. Interpretation

Port AI studies show a strong trend toward linking prediction to operational decisions. ETA prediction feeds berth allocation. Dwell-time prediction feeds stacking decisions. Predictive maintenance supports equipment planning. Reinforcement learning supports berth and crane scheduling. The strongest studies are those that report real data and connect model outputs to port KPIs such as waiting time, crane relocation, equipment downtime, berth utilization, or turnaround time.

## 7. Evidence Profile of Recent AI Logistics Studies

Table 2. Evidence profile of selected AI logistics studies published since 2024.

Study	Domain	AI function	Data type	Data scale reported	Field validation?	Logistics KPI linkage
Na [1]	Supply-chain demand	Prediction	Simulation experiment	Not clearly reported in accessible source	No	Forecast error
Sattar et al. [2]	SCM forecasting/risk	Prediction / decision support	DataCo ERP and logistics dataset	180,519 orders over one calendar year	No direct field deployment	Fill rate, stockouts, holding/ordering cost, late delivery
Zheng & Brintrup [3]	Supply-chain visibility	Relationship prediction	Real-world supply-chain data	Not publicly available due to commercial sensitivity	No field deployment reported	Visibility, risk management support
Zheng & Brintrup [4]	GenAI visibility	Relationship prediction	Real case-study data	Not clearly reported in accessible source	No field deployment reported	Supply-chain relationship inference
Pan et al. [5]	Supply-chain resilience	Organizational decision support	Survey data	231 executives/managers	No	Resilience, efficiency, collaboration

Study	Domain	AI function	Data type	Data scale reported	Field validation?	Logistics KPI linkage
Guo et al. [6]	GenAI adoption	Organizational decision support	Panel data	Chinese A-share firms, 2017-2022	No	Supply-chain resilience
Ammon et al. [7]	Vehicle routing	Optimization	Solver-generated instances	Problem-instance scale to be reported	No	Route cost / tour quality
Jiang & Chang [8]	Last-mile delivery	Prediction + routing	Simulation / generated benchmark with public code	Repository available; exact scale to be reported	No; field validation identified as needed	Delivery time, service efficiency
Konovalenko et al. [9]	Last-mile delivery	Route deviation prediction	Historical planned/actual route data	To be verified	No	Route deployment risk reported
Mahmoudina zlou et al. [10]	Warehouse order picking	DRL routing	Generated/benchmark warehouse instances	Arrival-rate scenarios reported	No	Fulfillment rate, throughput time, unfulfilled orders
Zhang et al. [11]	Warehouse task planning	Assignment/path planning	Algorithmic experiment	To be verified	No	Picking deployment / planning performance reported
Zhu et al. [12]	Robotic fulfillment	DRL + scheduling	Model experiment	To be verified	No	Order deployment / shelf selection, robot scheduling reported
Umakarthy et al. [13]	AGV path control	Path planning	Experimental or simulated flexible environment	To be verified	Field validation unclear	Path cost, obstacle response
Zhang et al. [14]	Port berth allocation	ETA prediction + optimization	AIS data from a specific port	May 27-June 27, 2024	No	Waiting time, berth deployment / utilization reported
You et al. [15]	Quay crane scheduling	DRL + metaheuristic	Computational experiment	To be verified	No	Scheduling deployment / efficiency reported
Jo et al. [16]	Berth/crane scheduling	Hierarchical RL	Optimization experiment	To be verified	No	Berth/crane deployment / efficiency, emissions reported
Kim et al. [17]	Port dwell-time prediction	GenAI + ML	Real container terminal data	To be verified from full text	No	MAE deployment improvement

Study	Domain	AI function	Data type	Data scale reported	Field validation?	Logistics KPI linkage
Aslam et al. [18]	Smart-port maintenance	ML classification	IoT sensor data	To be verified from full text	ent, relocation reported	Predictive maintenance / failure detection unclear

This table avoids calling applications mature or immature. Instead, it shows the evidence behind each study. The pattern is clear: real data are increasingly used, but field validation remains rare. Data scale is often underreported or difficult to extract from abstracts. Many studies report model metrics, but fewer report direct logistics KPIs.

## 8. Cross-Domain Patterns

### 8.1. Prediction Is Often Better Evidenced than Automation

Demand forecasting, ETA prediction, dwell-time prediction, route deviation prediction, and predictive maintenance often use historical or sensor data. These tasks are relatively easier to evaluate because predictions can be compared against actual outcomes. However, prediction studies still need stronger reporting of data scale, time span, missing data handling, and operational KPI linkage.

### 8.2. Optimization Studies Remain Simulation-Heavy

Vehicle routing, warehouse order picking, robot scheduling, berth allocation, and crane scheduling often rely on generated instances, computational experiments, or simulated environments. This is understandable because controlled evaluation is easier in optimization research. However, simulation-heavy evidence does not prove operational performance in live logistics environments.

### 8.3. Generative AI Appears Most Convincing as a Structured Data Layer

The strongest generative AI use cases in the reviewed set do not use GenAI as an unconstrained decision-maker. Instead, GenAI is used to standardize unstructured information, create embeddings, enrich knowledge graphs, or support relationship prediction. This is operationally safer than free-form automation because the generative component is constrained by downstream predictive or classification tasks.

### 8.4. Field Validation Is the Main Missing Layer

Few studies report actual deployment in a live logistics environment. Even when real data are used, the model is typically evaluated offline. For logistics adoption, the critical next step is field validation: integration with TMS, WMS, ERP, robot-control systems, port community systems, or terminal operating systems.

## 9. Discussion

### 9.1. Why Evidence Profile Matters

AI logistics research can look strong when evaluated only by model performance. A model may improve RMSE, accuracy, or route cost in a benchmark. But logistics managers need different evidence. They need to know whether the data were real, whether the data scale was sufficient, whether the system was tested in operations, and whether the output improved logistics KPIs.

This is why the evidence profile is useful. It separates four levels of evidence:

1. simulation or generated instances,
2. historical operational data,
3. real-world data with logistics KPI linkage,
4. field validation in a live operating environment.

Most recent studies fall into the first three levels. Very few reach the fourth.

### 9.2. Data Scale Should Be Standard Reporting

Recent studies do not consistently report data scale in a way that is easy to compare. Some studies provide clear numbers, such as the DataCo dataset with 180,519 orders or a survey with 231 respondents. Others describe real-world or case-study data but do not make the scale visible in accessible abstracts or summaries.

Future AI logistics papers should report:

- number of orders, routes, containers, vessels, robots, sensors, or records;
- time span of the data;
- geographic or operational scope;
- whether data are public or proprietary;
- missing data rates;
- train/test split or validation method.

Without this information, operational generalizability is difficult to assess.

### 9.3. Logistics KPIs Should Accompany Model Metrics

Model metrics are not enough. Forecasting error, classification accuracy, or route-cost reduction should be connected to logistics KPIs.

- demand forecasting should report stockouts, fill rate, inventory cost, or service level;
- route optimization should report delivery time, vehicle utilization, empty mileage, or driver workload;
- warehouse AI should report throughput, picking time, unfulfilled orders, congestion, or robot idle time;
- port AI should report vessel waiting time, berth utilization, crane relocations, turnaround time, or equipment downtime.

The Kim et al. container dwell-time study is a useful example because it links prediction improvement to relocation reduction [17]. More studies should make this operational bridge explicit.

### 9.4. Human Oversight Remains Necessary

The reviewed studies support the value of AI in logistics, but they do not support full autonomy. Logistics remains physical, multi-stakeholder, and exception-heavy. AI can assist prediction and optimization, but humans must handle exceptions, safety, governance, customer negotiation, regulatory constraints, and trade-offs under uncertainty.

This is particularly important for port operations, warehouse robotics, and last-mile delivery. A model may recommend a route, berth allocation, or robot schedule, but the operational context may require human judgment.

## 10. Recommendations for Future AI Logistics Studies

Table 3. Recommended reporting checklist for AI logistics studies.

Reporting item	Why it matters
Data source	Distinguishes real operational data from simulation or generated instances.
Data scale	Helps assess generalizability.
Time span	Indicates whether seasonality, disruption, or drift can be evaluated.
Operational context	Clarifies whether the study applies to warehouse, port, last-mile, SCM, or other logistics settings.
Baseline methods	Allows fair comparison against non-AI or conventional methods.
Model metrics	Shows technical performance.
Logistics KPIs	Shows operational value.
Field validation	Indicates readiness for deployment.
Integration requirements	Identifies whether the model can connect to real logistics systems.
Human oversight	Clarifies who is responsible for decisions and exceptions.

## 11. Managerial Implications

For logistics managers, the findings suggest five practical lessons.

First, do not evaluate AI claims only by algorithm name. Ask what data were used, how large the data were, and whether the system was tested in an operational environment.

Second, distinguish simulation success from operational readiness. Simulation studies can be useful, but they do not guarantee performance under real traffic, warehouse congestion, equipment failure, or port disruption.

Third, require logistics KPIs. A model that improves prediction accuracy should also show whether it reduces cost, delay, stockouts, relocations, downtime, or workload.

Fourth, treat generative AI cautiously. The most convincing recent GenAI logistics studies use it for constrained tasks such as standardization, embedding, and relationship inference, not autonomous operational decision-making.

Fifth, plan for human oversight. Logistics AI should support human decision-makers, not remove responsibility for exceptions and high-risk decisions.

## 12. Limitations

This paper has several limitations.

First, it is an evidence-oriented operational review rather than a full systematic review. The goal is classification and synthesis, not exhaustive coverage.

Second, the review focuses on studies published since 2024. This strengthens recentness but excludes foundational older work.

Third, some included studies have limited accessible detail in abstracts or public pages. Where data scale or validation setting could not be verified from accessible information, the table marks this as “to be verified” rather than inferring unsupported details.

Fourth, the review excludes corporate reports and news articles. This improves academic focus but may underrepresent actual industrial deployments.

Fifth, the evidence-profile framework is interpretive. Other reviewers may classify some studies differently depending on access to full text, supplementary materials, or implementation details.

### 13. Conclusions

AI logistics research published since 2024 increasingly addresses concrete operational tasks: demand forecasting, supply-chain visibility, route optimization, route deviation prediction, dynamic order picking, robot scheduling, berth allocation, quay crane scheduling, container dwell-time prediction, and smart-port maintenance.

The most important question is no longer simply whether AI can improve model performance. The more important question is what evidence supports the claimed operational value. Did the study use real operational data? How large were the data? Was the system tested only in simulation? Were logistics KPIs reported? Was there field validation?

The reviewed studies show clear technical progress, but also a major evidence gap. Real data are increasingly used, and some studies link AI performance to logistics KPIs. Yet field validation remains rare, and many studies still rely on simulation, generated instances, or offline evaluation.

The future of logistics AI research should therefore be evidence-oriented. Strong papers should report not only model accuracy, but also data provenance, data scale, operational KPIs, integration requirements, field validation, and human oversight.

### References

1. N. Na, "Research on Supply Chain Demand Prediction Model Based on LSTM," *Procedia Computer Science*, vol. 243, pp. 313-322, 2024. DOI: 10.1016/j.procs.2024.09.039.
2. M. U. Sattar et al., "Enhancing Supply Chain Management: A Comparative Study of Machine Learning Techniques with Cost-Accuracy and ESG-Based Evaluation for Forecasting and Risk Mitigation," *Sustainability*, vol. 17, no. 13, 5772, 2025. DOI: 10.3390/su17135772.
3. G. Zheng and A. Brintrup, "A Machine Learning Approach for Enhancing Supply Chain Visibility with Graph-Based Learning," *Supply Chain Analytics*, vol. 11, 100135, 2025. DOI: 10.1016/j.sca.2025.100135.
4. G. Zheng and A. Brintrup, "Enhancing Supply Chain Visibility with Generative AI: An Exploratory Case Study on Relationship Prediction in Knowledge Graphs," *International Journal of Production Research*, 2025. DOI: 10.1080/00207543.2025.2543964.
5. H. Pan, N. Zou, R. Wang, J. Ma, and D. Liu, "Artificial Intelligence Usage and Supply Chain Resilience: An Organizational Information Processing Theory Perspective," *Systems*, vol. 13, no. 9, 724, 2025. DOI: 10.3390/systems13090724.
6. J. Guo, F. Jia, and L. Chen, "How Generative AI Adoption Affects Supply Chain Resilience: An Operations and Supply Chain Management Perspective," *Technological Forecasting and Social Change*, vol. 224, 124446, 2026. DOI: 10.1016/j.techfore.2025.124446.
7. S. Ammon, F. Phillipson, and R. J. Almeida, "A Supervised Machine Learning Approach for the Vehicle Routing Problem," 2024. DOI: 10.5220/0012430000003639.
8. T.-H. Jiang and Y.-C. Chang, "Machine Learning-Enhanced Last-Mile Delivery Optimization: Integrating Deep Reinforcement Learning with Queueing Theory for Dynamic Vehicle Routing," *Applied Sciences*, vol. 15, no. 21, 11320, 2025. DOI: 10.3390/app152111320.
9. A. Konovalenko, L. M. Hvattum, and K. A. H. Iversen, "Predicting Last-Mile Delivery Route Deviations Using Machine Learning," *Expert Systems with Applications*, vol. 298, 129921, 2026. DOI: 10.1016/j.eswa.2025.129921.
10. S. Mahmoudiazlou, A. Sobhanan, H. Charkhgard, A. Eshragh, and G. Dunn, "Deep Reinforcement Learning for Dynamic Order Picking in Warehouse Operations," *Computers & Operations Research*, vol. 182, 107112, 2025. DOI: 10.1016/j.cor.2025.107112.
11. S. Zhang et al., "Real Time Task Planning for Order Picking in Intelligent Logistics Warehousing," *Scientific Reports*, vol. 15, 7331, 2025. DOI: 10.1038/s41598-025-88305-9.
12. Z. Zhu, S. Wang, and T. Wang, "Optimizing Robotic Mobile Fulfillment Systems for Order Picking Based on Deep Reinforcement Learning," *Sensors*, vol. 24, no. 14, 4713, 2024. DOI: 10.3390/s24144713.

13. S. Umakarthykeyan and S. Priyadharshini, "Intelligent Path Control of Autonomous AGVs in Flexible Manufacturing Systems: A Reinforcement Learning Approach," *Scientific Reports*, 2025. DOI: 10.1038/s41598-025-33531-4.
14. H. Zhang, J. Guo, and S. Guo, "An Optimization Study on Dynamic Berth Allocation Based on Vessel Arrival Time Prediction," *Ocean Engineering*, vol. 332, 121321, 2025. DOI: 10.1016/j.oceaneng.2025.121321.
15. L. N. B. Long, S.-S. You, T. N. Cuong, and H.-S. Kim, "Optimizing Quay Crane Scheduling Using Deep Reinforcement Learning with Hybrid Metaheuristic Algorithm," *Engineering Applications of Artificial Intelligence*, vol. 143, 110021, 2025. DOI: 10.1016/j.engappai.2025.110021.
16. S. Jo et al., "A Hierarchical Reinforcement Learning Approach for Real-Time Berth Allocation and Quay Crane Assignment and Scheduling," *International Journal of Production Research*, 2025. DOI: 10.1080/00207543.2025.2542518.
17. M. Kim, T. Kim, T. Park, H. Park, and H. Bae, "Generative AI and Machine Learning Collaboration for Container Dwell Time Prediction via Data Standardization," arXiv:2602.20540, 2026.
18. S. Aslam et al., "Machine Learning-Based Predictive Maintenance at Smart Ports Using IoT Sensor Data," *Sensors*, vol. 25, no. 13, 3923, 2025. DOI: 10.3390/s25133923.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.