Vol. 8, No. 2, April 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Large Language Models for Intelligent Decision Support in Inventory and Supply Chain Operations: A Systematic Literature Review

Yusuf Durachman[1], Eva Khudzaeva[2], Naura Aulia[3]

[1,2,3]Department of Information Systems, Syarif Hidayatullah State Islamic University Jakarta, Jakarta, Indonesia

Received:
October 26, 2025
Revised:
March 10, 2026
Accepted:
April 1, 2026
Published:
April 12, 2026

Corresponding Author:
Author Name*:
Yusuf Durachman
Email*:
yusuf_durachman@uinjkt.ac.id.

DOI:
10.63158/journalisi.v8i2.1540

Abstract. Generative Artificial Intelligence (GenAI), particularly large language models (LLMs), is increasingly explored to strengthen decision support in supply chain and inventory management by improving interpretability and access to analytics. However, prior work is scattered across optimization, simulation, logistics, and governance discussions, limiting clear system design guidance. This study conducts a Systematic Literature Review (SLR) following PRISMA 2020 across IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Google Scholar, yielding 200 records, of which 34 studies were included in qualitative synthesis. Results show that LLMs are predominantly positioned as orchestration and explanatory layers operating alongside structured components such as optimization solvers, simulation engines, and digital twins, rather than as autonomous decision-makers. Governance, organizational readiness, and trust emerge as central considerations for operational deployment. This review provides an evidence map linking LLM roles and integration architectures across supply chain and inventory contexts. While LLMs offer strong augmentation capabilities, direct empirical validation for specific contexts such as web-based inventory systems remains limited; design implications for such systems are derived from the broader corpus, underscoring the need for standardized evaluation benchmarks and targeted empirical studies.

Keywords: Systematic Literature Review, Inventory Analytics, Supply Chain Decision Support, LLM Integration Architecture, PRISMA.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Inventory is an operational asset that absorbs working capital and directly affects customer service levels, stockout risk, and storage costs. In many organizations, especially micro, small, and medium enterprises (MSMEs), inventory management still faces persistent challenges such as inconsistent transaction logging, low stock visibility, and reporting delays that lead to inaccurate ordering decisions [1]. Web-based inventory systems are a widely adopted response to these challenges, offering improved transaction traceability, data standardization, and cross-device access in real-time. Empirical evidence from Indonesian MSMEs confirms that web-based digitalization supports more structured recording and reporting for business operations [2]

However, the availability of a web-based inventory system alone is insufficient to address the challenges of demand dynamics and supply chain uncertainty. Beyond transaction recording, organizations increasingly require intelligent analytical capabilities — such as demand forecasting, stockout risk prediction, safety stock calculation, and reorder policy recommendations to support adaptive decision-making. The literature documents widespread adoption of machine learning and deep learning models for these inventory management problems, with diverse evaluation metrics reflecting the importance of a strong analytical foundation in modern inventory systems. [3], [4]. A recent systematic review confirms that machine learning approaches are increasingly applied across inventory control settings, spanning reinforcement learning, optimization, and hybrid methods, though standardized evaluation protocols remain limited [5]

While traditional machine learning models have significantly advanced analytical forecasting, the emergence of Large Language Models (LLMs) introduces a broader shift toward natural language-driven decision support across supply chain and inventory ecosystems. LLMs can function as orchestration layers and natural-language interfaces interpreting operational conditions, generating insight summaries, and enriching complex decision-making workflows [6]. However, deploying LLMs in operational contexts introduces critical reliability challenges, particularly the risk of hallucination: generating convincing but factually incorrect outputs that pose severe risks if used directly for procurement or reordering decisions without robust verification mechanisms [7].

Despite rapid technological advancement, existing literature reviews have predominantly addressed either the general adoption of AI in supply chains or broad governance issues, leaving a gap in the systematic mapping of how LLMs are specifically architected, integrated, and evaluated for inventory-related decision support across the broader supply chain and operations management literature [8]. This Systematic Literature Review (SLR) addresses that gap by synthesizing evidence on LLM-enabled decision support in inventory management and adjacent supply chain operations during the 2020–2026 period.

The corpus of 34 included studies spans supply chain operations, manufacturing, logistics, and warehouse management broadly. Implications for web-based inventory system design are therefore presented throughout this paper as derived design directions informed by the broader evidence base, rather than as findings directly validated in web-based inventory deployments. This scoping is intentional and reflects the current state of the literature, where inventory-specific LLM research remains nascent.

The novelty of this study lies in its synthesis of scattered evidence across optimization, simulation, digital twins, and logistics operations into a cohesive integration framework with explicit inventory-system design implications. Specifically, this review contributes by: (1) identifying operational and system characteristics relevant to LLM-based analytics integration as reported across the corpus; (2) analyzing dominant stock analysis methods and their evaluation metrics; (3) mapping patterns of LLM/GenAI deployment as decision support and the main challenges and mitigation strategies; and (4) assessing integration architectures and the overall empirical maturity of the field. Ultimately, this study provides evidence-grounded design directions for developing reliable, human-in-the-loop inventory analytics systems, with explicit acknowledgment of where these directions require further empirical validation.

2. METHODS

2.1. Review Design and Reporting

This study employed a Systematic Literature Review (SLR) to identify, evaluate, and synthesize research on the use of Large Language Models (LLMs) and Generative AI (GenAI) for decision support in inventory management and related supply chain operations. The review protocol was designed to ensure rigor, transparency, and reproducibility, drawing on established SLR guidelines in software engineering and information systems research [9], [10]. To strengthen reporting quality, the review process also followed the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework, particularly for the identification, screening, eligibility, and inclusion stages [11].

Although the review was initially motivated by challenges associated with web-based inventory systems, the final body of literature extended beyond that narrow focus and included studies in logistics, warehousing, manufacturing, and broader supply chain operations. Accordingly, the implications discussed for web-based inventory system design are presented as derived design directions based on cross-study synthesis, rather than as direct empirical findings uniformly reported across the reviewed studies.

2.2. Research Questions

This review aims to systematically map the literature on LLM-enabled decision support in inventory management and adjacent supply chain contexts published between 2020 and 2026. The study addresses the following research questions:

1) RQ1: What operational and system characteristics are commonly reported in inventory and supply chain contexts relevant to LLM-based analytics integration?
2) RQ2: What methods are used for stock analysis, demand fulfillment, and optimization, and what evaluation indicators or metrics are most frequently reported?
3) RQ3: What patterns of LLM/GenAI deployment are reported for decision support in inventory and supply chain settings, and what key challenges and mitigation strategies are identified?
4) RQ4: What integration and architectural patterns are used to operationalize LLM/GenAI for inventory-related analytics?
5) RQ5: What is the overall maturity of evidence across the literature, and what research gaps and future directions emerge for inventory decision support?

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.3. Scope Definition Using PICOC

To establish a consistent and reproducible review scope, the PICOC framework (Population, Intervention, Comparison, Outcome, Context) was applied as follows:

1) Population (P): Organizations and industries implementing inventory management, warehouse management, or supply chain operations, including web-based inventory information systems.
2) Intervention (I): The application of LLMs and/or GenAI, including use cases such as natural language querying, recommendation generation, retrieval-augmented generation, and agent-based architectures.
3) Comparison (C): Traditional approaches or conventional machine learning methods compared with LLM/GenAI-enhanced approaches, where such comparisons were explicitly reported.
4) Outcome (O): Reported evidence of performance, including predictive accuracy, reduction of stockouts or overstock, service-level improvement, decision quality, optimization outcomes, and organizational adoption or readiness.
5) Context (C): Inventory management and related supply chain activities, including replenishment, procurement, warehousing, reporting, and risk management, with particular attention to implications for web-based inventory systems.

2.4. Search Strategy

2.4.1. Databases and Time Window

The literature search was conducted up to February 2026 to capture recent developments in this rapidly evolving area. To maximize coverage, the search included both structured academic databases and broad academic search platforms. Structured databases comprised IEEE Xplore and the ACM Digital Library, both of which support advanced Boolean syntax and field-level filtering. Additional sources included ScienceDirect, SpringerLink, and Google Scholar, which differ in indexing practices, retrieval constraints, and search syntax capabilities. The review focused on publications from 2020 to 2026, reflecting the period in which GenAI- and LLM-related research began to expand substantially in enterprise and operational domains. Early 2026 publications were included to capture newly available and early-access work. A summary of the data sources, query adaptations, and retrieval counts is provided in Table 3.

2.4.2. Search Keywords and Query String

The search strategy was organized around four core concept groups:

1) inventory and warehouse systems,
2) stock analysis and decision support,
3) LLM/GenAI technologies, and
4) analytics, optimization, and decision-making.

To improve retrieval sensitivity, the search incorporated synonyms and related terms commonly used in the literature. These included terms such as *inventory system*, *inventory management*, *warehouse management*, *supply chain*, *logistics*, *decision support*, *analytics*, *optimization*, *stock analysis*, *large language model*, *LLM*, *generative AI*, *GenAI*, and *foundation model*. The core Boolean query applied to structured databases was as follows:

("inventory system" OR "inventory management" OR "warehouse management" OR "supply chain" OR "logistics") AND ("web-based" OR "information system" OR "digital twin") AND ("large language model" OR "LLM" OR "generative AI" OR "GenAI" OR "foundation model") AND ("decision support" OR "analytics" OR "optimization" OR "stock analysis")

This search string was adapted to the syntax and field restrictions of each platform, including title, abstract, and keyword fields where supported. For Google Scholar, where lengthy Boolean expressions are less stable and ranking is relevance-based, the search was simplified and limited to the first 95 most relevant records within the 2020–2026 publication window.

2.5. Eligibility Criteria and Study Selection

2.5.1. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria used in this review are summarized in Table 1 and Table 2. Studies were included if they met all of the following conditions: (1) they addressed inventory, warehouse, or stock management, including the implementation of inventory information systems and web-based systems; (2) they discussed stock analysis for decision support, such as stock availability, stock sufficiency, adequacy evaluation, allocation or fulfillment, optimization, or other stock-related decision rules within inventory or supply chain operations; (3) they examined the use of GenAI/LLMs for analytical or decision-support purposes in inventory, warehouse, or supply chain contexts; and (4) they were published as peer-reviewed journal articles, academic books, or book chapters, written in English or Indonesian, with accessible full text, and published between 2020 and 2026. Studies were excluded if they were not relevant to inventory, warehouse, or stock management; were not peer-reviewed journal articles, academic books, or book chapters; were duplicate or incomplete records; lacked accessible full text; or were published outside the specified time range.

Table 1. Inclusion criteria

	Inclusion Criteria
1)	The study discusses inventory/warehouse/stock management and/or the implementation of inventory information systems (including web-based ones).
2)	Studies discuss stock analysis for decision support (e.g. stock availability/stock sufficiency, stock adequacy evaluation, allocation/fulfillment of needs, optimization/stock-related decision rules) in the context of inventory or supply chain operations.
3)	The study discusses the use of GenAI/LLM (e.g., large language model, generative AI, foundation model) on tasks of analysis or decision-making related to inventory, supply chain, or warehouse management.
4)	Peer-reviewed journals or academic books/book chapters; in English/Indonesian; Full text accessible.
5)	Published in 2020–2026

Table 2. Exclusion Criteria

	Exclusion Criteria
1)	Studies are irrelevant to inventory/warehouse/stock management
2)	Publications are not peer-reviewed journals or academic books/book chapters
3)	Duplicates, incomplete notes, or full-text are inaccessible.
4)	Published outside the range of 2020–2026.

2.5.2. Screening Procedure

All retrieved records were exported to Mendeley and Microsoft Excel for reference management and screening. Duplicate entries were first identified using Mendeley's

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

duplicate detection feature and then manually verified in Excel to ensure accuracy. Study selection was conducted in two stages. The first stage consisted of title and abstract screening to remove records that were clearly irrelevant to the review scope. The second stage involved full-text screening based on the predefined eligibility criteria. Records whose full texts could not be retrieved, including inaccessible articles behind paywalls and abstract-only entries, were documented and excluded prior to the final eligibility assessment. The overall study selection process is illustrated in the PRISMA flow diagram in Figure 1.

2.6. Quality Appraisal

To contextualize the strength and trustworthiness of the included evidence, each selected study was assessed using a concise quality appraisal checklist. The appraisal considered whether the study:

1) clearly described its purpose and operational context;
2) employed appropriate and transparent methods;
3) presented explicit supporting evidence, such as evaluation results, empirical findings, or case-based observations; and
4) reported limitations or threats to validity.

The quality appraisal informed interpretation of the findings rather than functioning as a strict exclusion mechanism. Accordingly, studies were not excluded solely because of lower quality unless their methodological reporting was critically insufficient to address any of the review questions. Where lower-quality studies are cited, their evidential limitations are taken into account in the synthesis.

2.7. Data Extraction

A standardized data extraction form was used to ensure consistent capture of relevant information across all included studies. The extracted information comprised:

1) bibliographic details, including author, year, and publication venue;
2) study context, including industry or application domain and whether a web-based inventory setting was involved;
3) operational focus, such as stock availability, stock sufficiency, allocation, optimization, or decision rules;

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

4) the role of LLM/GenAI, including decision support, narrative analysis, recommendation generation, or system integration;
5) evaluation details, including datasets, experimental settings, and reported performance metrics where available; and
6) stated limitations, challenges, and research gaps.

The use of a structured extraction template supported comparability across studies and enabled both descriptive and thematic synthesis.

2.8. Data Synthesis and Coding Procedures

2.8.1. Thematic Synthesis

The qualitative synthesis employed an inductive thematic analysis approach. Extracted data were reviewed iteratively to identify recurring themes related to system context, stock analysis functions, LLM roles, evaluation practices, and implementation patterns. This process generated:

1) a taxonomy of operational characteristics relevant to inventory analytics;
2) a mapping of relationships among themes, such as the relationship between LLM roles and integration architectures; and
3) a set of recurring methodological patterns, limitations, and research gaps.

Where implications for web-based inventory system design are discussed, these are explicitly framed as synthesis-derived design directions rather than as direct findings consistently evaluated across the corpus.

2.8.2. Coding and Consistency Checks

To address RQ4, a structured coding process was used to classify how LLMs were operationalized in relation to inventory and supply chain systems. Coding was conducted iteratively, and ambiguous or overlapping cases were resolved by revisiting the full text and reassessing the study against the PICOC framework. This process was intended to improve internal consistency and reduce category overlap. The coding resulted in four architectural categories:

1) embedded operational integration, where LLMs are directly incorporated into an operational system or application;
2) tool-augmented configurations, where LLMs are combined with optimization solvers, simulators, or digital twins;

3) standalone analytical approaches, where LLM-based methods are presented without a complete operational system context; and
4) conceptual frameworks, where no implemented operational system is reported. These categories were derived inductively from the reviewed studies rather than imposed a priori.

2.9. Study Selection Outcome

The search process identified 200 records across all sources. After deduplication, 198 unique records remained for title and abstract screening. At this stage, 75 records were excluded because they were clearly irrelevant to the review scope. The remaining 123 records were considered for full-text retrieval; however, 65 reports could not be obtained in full text, primarily because access was restricted or only abstract-level information was available. As a result, 58 full-text studies were assessed for eligibility.

During the eligibility assessment, 24 studies were excluded because they did not satisfy the inclusion criteria, for example because they focused on financial stock trading rather than physical inventory management or did not explicitly involve LLM/GenAI integration. Ultimately, 34 studies met all eligibility requirements and were included in the final qualitative synthesis. The complete selection process is presented in Figure 1, and the search-source summary is provided in Table 3.

Table 3. Search Sources, Query Adaptation, and Records Retrieved

Source	Source Type	Query Adaptation	Retrieved
IEEE Xplore	Structured database	Full Boolean string applied to title, abstract, and keyword fields	20
ACM Digital Library	Structured database	Full Boolean string applied to abstract fields (2020–2026)	0
ScienceDirect	Publisher platform	Full Boolean string applied to title, abstract, and keyword fields; filtered to peer-reviewed and review articles (2020–2026)	51

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

SpringerLink	Publisher platform	Full Boolean string applied with minor syntax adjustments based on platform constraints (2020–2026)	34
Google Scholar	General academic search engine	Simplified query using core concept pairs; first 95 most relevant records extracted (2020–2026)	95
		Total	**200**

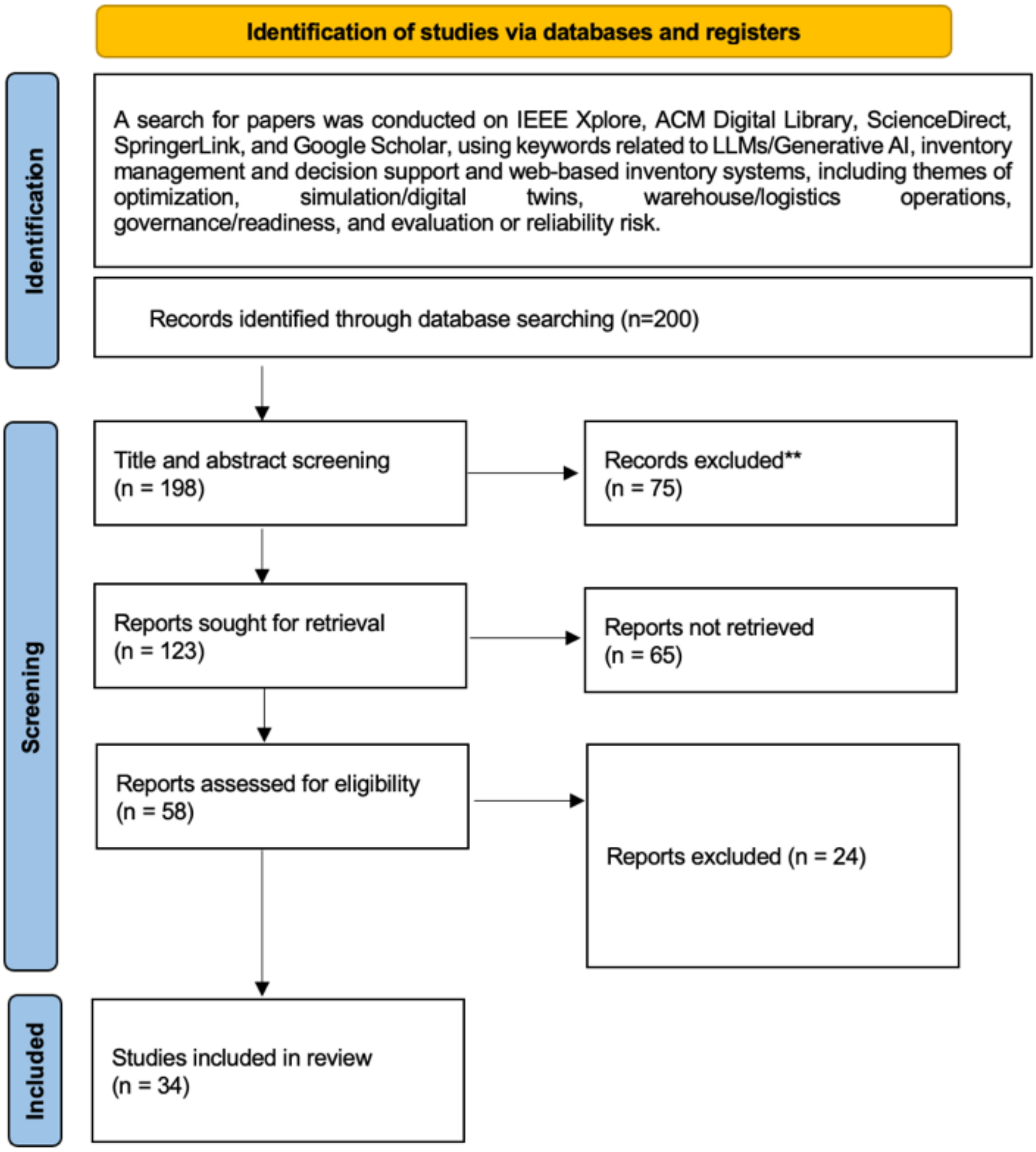

Figure 1. PRISMA Flow Diagram

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3. RESULTS AND DISCUSSION

3.1. Overview of Data

This section presents the synthesis results based on the 34 studies included in the final review corpus [12]–[44]. The findings are organized in two stages. First, the included studies are profiled descriptively in terms of publication year, research method, inventory process area, LLM/GenAI use-case, outcome category, and integration architecture. This descriptive mapping provides an overview of the current state of the literature and establishes the empirical context for the subsequent synthesis. Second, the findings are interpreted in relation to RQ1–RQ5, covering operational and system characteristics relevant to LLM-based analytics integration, stock analysis methods and evaluation indicators, deployment patterns of LLM/GenAI in inventory and supply chain settings, architectural integration approaches, and the overall maturity of evidence in the field.

Overall, the corpus reflects a rapidly expanding but still early-stage research area. The literature is heavily concentrated in recent years, with a sharp increase in publications after 2023, indicating growing interest in applying LLMs and GenAI to operational decision support. At the same time, the evidence base remains uneven: many studies focus on conceptual framing, benchmarking, or architecture proposals, while comparatively fewer report real-world implementations or rigorous quantitative evaluations. Across the corpus, research attention is concentrated on inventory optimization, risk and resilience planning, governance, and simulation-related applications, whereas routine inventory functions such as reporting, monitoring, and procurement remain much less explored. These patterns suggest that current research is moving quickly toward intelligent operational support, but practical deployment in day-to-day inventory systems is still emerging. Where implications for web-based inventory system design are discussed, they are framed as derived design directions from the broader literature rather than as direct empirical findings consistently reported across all included studies.

3.1.1. Overview of Included Studies

A total of 34 studies met the eligibility criteria and were included in the final qualitative synthesis. The full list of selected studies is presented in Table 4 [12]–[44]. Collectively, these studies span a range of contexts related to inventory management, warehousing, logistics, supply chain coordination, simulation environments, and broader operations

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

management. Although the review was motivated by the design of web-based inventory systems, the final corpus extends beyond that specific application area and includes adjacent domains in which LLMs or GenAI are used to support analytical, optimization, explanatory, or decision-making functions.

Table 4. Articles selected for review

Articles
Jackson et al., 2024 [12]
Shamsuddoha et al., 2025 [13]
Teixeira et al., 2025 [8]
Gezdur et al., 2025 [14]
Ieva et al., 2025 [15]
Nicoletti et al., 2025 [16]
Fan et al., 2025 [17]
Du et al., 2025 [18]
Wang et al., 2025 [19]
Elbasheer et al., 2025 [20]
Li et al., 2025 [21]
Huang et al., 2025 [22]
Chen et al., 2025 [23]
Cimino et al., 2025 [24]
Zhao et al., 2025 [25]
Li et al., 2024 [26]
Dubey et al., 2024 [27]
Malhotra et al., 2025 [28]
Singh et al., 2025 [29]
Jaouhari et al., 2025 [30]
Moica et al., 2025 [31]
Ji et al., 2025 [32]
Wu et al., 2025 [33]
Sharma et al., 2025 [34]
Boone et al., 2025 [35]
Handler et al., 2024 [36]
Srivastava et al., 2024 [37]

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Articles
Fosso Wamba et al., 2024 [38]
Li et al., 2024 [39]
Jackson et al., 2024 [40]
Kmiecik, 2025 [41]
Berlec et al., 2025 [42]
Wu, 2026 [43]
Fosso Wamba et al., 2023 [44]

This breadth is important because it shows that the literature on LLM-enabled decision support in inventory-related settings is not confined to a single technical implementation pathway. Rather, it encompasses conceptual frameworks, optimization support tools, digital twin integrations, governance-oriented studies, and embedded system applications. As a result, the corpus offers a broad basis for identifying common patterns, recurring challenges, and emerging design directions, even though not all studies directly evaluate a web-based inventory platform.

3.1.2. Publication Year Trend

Figure 2 shows the distribution of included studies by publication year. The results reveal a clear upward trend in research activity, with 1 study in 2023, 8 studies in 2024, 24 studies in 2025, and 1 study in 2026. This distribution indicates that the field has expanded rapidly within a short period, particularly during 2024–2025, when LLM and GenAI technologies began to gain stronger traction in enterprise, industrial, and supply chain applications.

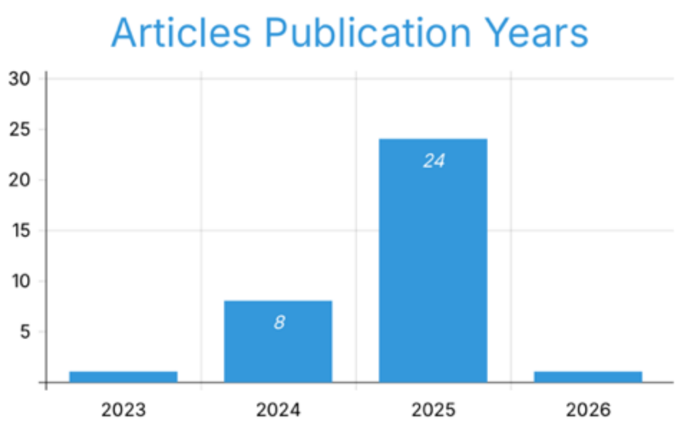

Figure 2. Distribution of included studies by publication

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The concentration of publications in 2025 suggests that the field is currently in an accelerated growth phase rather than a mature consolidation phase. This pattern is consistent with broader developments in GenAI research, where early conceptual interest has quickly evolved into domain-specific experimentation and architecture development. In the context of inventory and supply chain management, this growth signals increasing recognition of the potential value of LLMs for decision support, simulation, interpretability, and knowledge-intensive operational tasks. At the same time, the recentness of the literature also implies that many studies remain exploratory and that the long-term evidence base is still developing.

3.1.3. Research Method Distribution

As summarized in Table 5, the included studies display a methodologically diverse profile. The largest category consists of benchmark or evaluation studies (n = 9), followed by conceptual or framework papers (n = 8). Other methodological categories include case studies (n = 5), survey or empirical quantitative studies (n = 4), simulation or experimental studies (n = 3), systematic or bibliometric reviews (n = 3), and system implementation or prototype studies (n = 2).

Table 5. Distribution of Method

Methods Category	n
Benchmark / evaluation study	9
Conceptual / framework	8
Case study (industry)	5
Survey / empirical quantitative	4
Simulation / experimental	3
Systematic review / bibliometric review	3
System implementation / prototype	2

This distribution indicates that the literature remains at a relatively early and exploratory stage. A substantial proportion of studies are devoted to conceptualization, benchmarking, or model discussion, while only a small number report implemented systems or deployed prototypes. In other words, the field is still building foundational understanding of how LLMs and GenAI might support inventory-related decision-making, rather than providing extensive evidence from mature real-world deployments. The presence of benchmark and conceptual studies as the dominant categories also suggests that researchers are still actively defining feasible use-cases, evaluating performance boundaries, and exploring suitable integration strategies before large-scale operational adoption becomes commonplace.

3.1.4. Inventory Process Areas Distribution

Table 6 presents the distribution of studies by inventory process area. Most studies focus on inventory optimization and policy design (n = 15), followed by risk and resilience inventory strategy (n = 9). Smaller numbers of studies examine delivery scheduling or transportation interfaces (n = 3) and warehousing operations, such as putaway, picking, storage, and allocation (n = 3). More limited attention is given to training or knowledge management for the supply chain workforce (n = 2), while inventory reporting and monitoring (n = 1) and purchasing or procurement (n = 1) are only minimally represented.

Table 6. Distribution Inventory Process Areas

Research Method Category	n
Inventory optimization / policy design (OR/heuristics)	15
Risk & resilience inventory strategy	9
Delivery scheduling / transportation interface	3
Warehousing operations (putaway/picking/storage/allocation)	3
Training / knowledge management (SCM workforce)	2
Inventory reporting & monitoring (dashboards/alerts/summarization)	1
Purchasing / procurement	1

This distribution highlights a strong concentration of current research on high-complexity and decision-intensive functions, particularly those involving optimization, policy modeling, and strategic resilience. These are areas where LLMs are often combined with advanced analytics, simulation, or solver-based environments. By contrast, routine but operationally essential functions, such as reporting, dashboard interpretation, alert summarization, and procurement support, receive comparatively little attention. This imbalance is notable because these lower-complexity but high-frequency activities are central to many web-based inventory systems. The findings therefore point to an important research opportunity: while current work emphasizes advanced optimization

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

and strategic reasoning, the practical use of LLMs for everyday inventory monitoring and operational reporting remains underdeveloped.

3.1.5. LLM/GenAI Use-Case Distribution

As shown in Table 7, the most common LLM/GenAI use-case category is governance, risk, and evaluation guidance (n = 13). This is followed by digital twin or simulation integration (n = 8). Other categories include benchmarking LLMs for supply chain management tasks (n = 4), explanation or interpretability support (n = 4), and optimization modeling or solver assistance, such as natural-language-to-operations-research or code generation (n = 3). Only 1 study addresses autonomous agents or multi-agent decision-making, and only 1 study focuses on summarization or reporting generation.

Table 7. Distribution of LLM/GenAI use-cases

LLM/GenAI use-cases Category	n
Governance / risk / evaluation guidance	13
Digital twin / simulation integration	8
Benchmarking LLMs for SCM tasks	4
Explanation / interpretability (narrative explanations)	4
Optimization modeling / solver assistance (NL2OR/code generation)	3
Autonomous agents / multi-agent decision-making	1
Summarization / reporting generation	1

These results suggest that current research is using LLMs less as standalone operational decision engines and more as supportive layers around decision-making, evaluation, explanation, and simulation. The prominence of governance and risk-related studies indicates that issues of trust, oversight, policy, and responsible deployment are central in the current discourse. Likewise, the relatively high number of simulation- and digital-twin-related studies shows that LLMs are often being tested in controlled or hybrid environments rather than in fully autonomous operational roles. The very limited number of studies on autonomous agents and routine reporting applications implies that truly end-to-end operational automation and everyday business assistance remain underexplored. This pattern reinforces the view that the field is still experimenting with bounded and assistive roles for LLMs before wider adoption in routine inventory workflows.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.1.6. Outcome Categories Distribution

Table 8 summarizes the outcome categories reported across the corpus. The largest group of studies falls into the not evaluated (conceptual/review) category, accounting for 19 studies. Among the studies that do report outcomes, the most common focus is on decision quality, solution optimality, or KPI improvement (n = 6), followed by organizational adoption, readiness, or performance outcomes (n = 5). Only 2 studies report accuracy-related metrics, such as MAE, RMSE, or MAPE, and another 2 studies report cost or efficiency outcomes, such as time savings, resource utilization, or distance reduction.

Table 8. Distribution of Outcome Categories

Outcome Categories	n
Not evaluated (conceptual/review)	19
Decision quality / solution optimality / KPI improvement	6
Adoption/readiness/performance (organizational)	5
Accuracy metrics (MAE/RMSE/MAPE/etc.)	2
Cost/efficiency (time, distance, resource utilization)	2

This pattern underscores the limited empirical maturity of the field. Although the literature is growing quickly, many studies remain conceptual, strategic, or framework-oriented and do not yet provide formal evaluation evidence. Even among evaluated studies, hard quantitative metrics are relatively uncommon compared with broader claims about decision quality or organizational readiness. This does not reduce the importance of the conceptual contributions, but it does suggest that the evidence base for operational effectiveness remains fragmented. In practical terms, the literature appears to be ahead in proposing roles for LLMs and GenAI, yet still behind in systematically validating those roles using robust operational metrics, cross-context comparisons, or longitudinal deployment evidence.

3.1.7. Integration and Architecture Distribution

As summarized in Table 9, the most common integration pattern is embedded operational integration, with 16 studies reporting LLM/GenAI capabilities incorporated directly into an operational system or application. This is followed by tool-augmented configurations (n = 11), in which LLMs are combined with optimization solvers, simulators, or digital twins. Smaller portions of the literature present standalone analytical methods without a full operational system context (n = 4) or remain at the level of conceptual frameworks with no reported implementation (n = 3).

Table 9. Distribution of Integration/Architecture

Integration/Architecture Category	n
Embedded in operational system/app (e.g., web inventory system)	16
Tool-augmented (LLM + optimizer/simulator/digital twin)	11
Standalone analytical method (no system)	4
Conceptual framework (no implementation)	3

These findings indicate that LLM adoption in inventory-related settings is increasingly being conceptualized as part of a broader socio-technical and architectural ecosystem, rather than as an isolated model component. The dominance of embedded and tool-augmented configurations suggests that researchers view LLMs as complementary technologies that enhance existing systems, analytical pipelines, or simulation environments. This has important implications for system design, particularly in web-based inventory contexts, where the value of LLMs may lie not only in prediction or generation, but also in their ability to mediate between users, operational data, optimization modules, and reporting interfaces. At the same time, the relatively small number of fully implemented and evaluated operational systems confirms that architecture proposals currently outpace large-scale validated deployment.

3.1.8. Summary of Corpus Characteristics

Taken together, the descriptive results show that the literature on LLM/GenAI for inventory and related supply chain decision support is expanding rapidly, but remains methodologically and empirically immature. The field is dominated by recent publications, exploratory methods, optimization-oriented applications, governance and simulation use-cases, and embedded or hybrid system architectures. In contrast, routine operational functions, quantitative evaluations, and production-level implementations remain limited.

These corpus characteristics provide important context for the subsequent synthesis under RQ1–RQ5. They suggest that the most reliable conclusions from the current literature concern emerging patterns, architectural tendencies, and research priorities, rather than settled evidence about mature best practices. They also indicate that future research should move beyond conceptual enthusiasm toward more rigorous evaluation of real-world deployment, especially in practical inventory settings where routine monitoring, procurement support, explainability, and human-centered system interaction are central.

3.2. Research Question

1) **RQ1: What operational and system characteristics are commonly reported in inventory and supply chain contexts relevant to LLM-based analytics integration?**

Across the included corpus (n = 34), detailed specifications of classic web-based inventory interface features — such as role-based access control, alert configurations, or KPI dashboards — are inconsistently reported and often omitted, particularly in conceptual and framework-oriented studies (n = 8). Instead, the literature characterizes integration readiness primarily through broad architectural orientations: an "analytics-in-the-loop" positioning (16/34) and interoperability with external analytical back-ends (11/34). While specific interface features like role-based access control are often omitted, the field has converged on a hybrid integration logic where the LLM functions as a layer connecting user intent to structured analytical components. Detailed feature-level reporting for web-based inventory interfaces specifically remains a significant gap in the current literature; this gap represents a derived implication for future system-level research rather than a finding directly evidenced by the corpus.

2) **RQ2: What methods are used for stock analysis, demand fulfillment, and optimization, and what evaluation indicators/metrics are most commonly reported?**

The dominant methodological stream in the corpus emphasizes inventory optimization and policy design rooted in operations research and heuristic decision rules (15/34), followed by risk- and resilience-oriented inventory strategies (9/34). Routine operational domains, such as warehousing (3/34), inventory reporting and monitoring (1/34), and procurement (1/34) are rarely the primary focus of included studies. In terms of evaluation, the evidence base is notably weak regarding standardized, inventory-specific metrics. A large portion of studies relies on conceptual discussions without empirical evaluation (19/34). Among empirically evaluated works, outcomes are largely qualitative, focusing on broad decision quality (6/34) or organizational adoption readiness (5/34). Direct quantitative reporting remains severely limited: only 2/34 studies explicitly report accuracy metrics (e.g., MAE/RMSE/MAPE) and 2/34 report cost or efficiency outcomes. This evaluation gap is a derived implication for system designers: standardized, inventory-specific benchmarking protocols need to be developed before LLM-based analytics can be reliably adopted in operational deployments.

3) RQ3: What are the patterns of LLM/GenAI deployment as decision support in inventory and supply chain settings, and what are the main challenges and mitigation strategies?

The most frequent LLM/GenAI use-case in the corpus is governance, risk, and evaluation guidance (13/34), indicating that the literature currently prioritizes responsible deployment and accountability over widespread operational execution. The second most prevalent pattern is digital twin and simulation integration (8/34), where LLMs support decision-making through simulation-backed reasoning rather than unconstrained free-form outputs, a direct architectural response to hallucination risk. Other patterns include LLM benchmarking for SCM tasks (4/34) and interpretability/explainability (4/34), while autonomous agent-based decision-making (1/34) and automated reporting generation (1/34) remain marginal. These patterns suggest that the primary mitigation strategy for LLM unreliability is structural: LLMs are constrained by embedding them within governance frameworks or by coupling their outputs to verifiable deterministic systems. For web-based inventory system designers, this implies that any near-term integration of LLM capabilities should be accompanied by explicit validation layers, human oversight mechanisms, and accountability structures, a design requirement derived from the governance-first orientation consistently evidenced across the corpus.

4) RQ4: What integration and architecture patterns are used to operationalize LLM/GenAI for inventory-related analytics?

The evidence base reveals four recurring integration patterns. Embedded operational integration is the most prevalent approach (16/34), where LLM capabilities are integrated directly into an application layer to support operational interpretation and user interaction. Tool-augmented architectures are also widely adopted (11/34), positioning LLMs as reasoning layers connected to verifiable analytical back-ends such as optimizers,

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

simulation engines, or digital twins. A smaller subset proposes standalone analytical methods without a full operational system context (4/34), and a minority remains at the conceptual framework level with no reported implementation (3/34). The near-absence of standalone LLM deployments and purely conceptual frameworks among the majority of studies reflects a field-wide consensus that LLMs require structural anchoring to be operationally reliable. For web-based inventory systems specifically, the architecture pattern evidence supports a derived design principle: LLM components should be architected as augmentation layers interfacing with, rather than replacing, existing operational logic and data validation pipelines.

5) **RQ5: What is the overall maturity of evidence across the corpus, and what research gaps and future directions emerge for inventory decision support?**

The overall evidence maturity indicates that the field is still transitioning from conceptual framing to operational validation. More than half of the included studies are not empirically evaluated, focusing instead on conceptual discussions or systematic reviews (19/34). Among the evaluated studies, outcomes primarily emphasize general decision quality (6/34) and organizational readiness (5/34). Rigorous quantitative reporting remains extremely limited: only 2/34 studies report accuracy metrics and 2/34 report direct cost or efficiency improvements. These gaps collectively suggest that the field is progressing from exploratory demonstration toward measurement standardization but has not yet achieved methodological consolidation. For the practical design of web-based inventory systems, future research should prioritize measuring actual operational impact (e.g., service-level attainment, stockout reduction) in live environments rather than relying solely on conceptual frameworks — a direction derived from the evidence gaps identified in the corpus rather than from direct corpus findings.

3.3. Discussion

The synthesis indicates that LLM/GenAI-enabled decision support in inventory and adjacent operations contexts is predominantly conceptualized as augmentation rather than autonomy. Across studies that connect LLMs to simulation and digital-twin ecosystems, generative models function as natural-language interfaces that translate user intent into structured artefacts or interactions while execution and validity remain anchored in formal engines and operational representations [12], [20], [23], [24], [40]. A similar pattern appears in optimization-oriented contributions, where LLMs reduce modelling barriers through natural-language-to-OR modelling, automated optimization modelling, or multi-agent optimization support, while solution feasibility and correctness continue to rely on solver- or algorithm-based reasoning [19], [21], [22], [43]. This augmentation-first positioning is also consistent with adjacent industrial decision-making research in manufacturing, where LLM-based agents and fine-tuned frameworks are proposed to support structured decision tasks while remaining embedded within engineered systems rather than replacing them entirely [17], [18]. The corpus implies that the near-term contribution of LLMs lies less in replacing analytical decision logic than in improving accessibility, interaction, and interpretability within structured decision environments. For inventory decision support specifically, this suggests that the most feasible near-term role for LLMs is as an interpretive and interaction layer atop existing systems — a derived design direction, not a finding directly validated in web-based inventory contexts by the included studies.

A second key finding is the prominence of governance, readiness, and risk-oriented discourse. Many studies frame GenAI value realization as contingent on organizational capabilities, implementation pathways, and accountability arrangements, rather than on model performance alone [8], [27], [33], [34], [35], [38], [40], [44]. Complementary discussion of AI-driven intelligent automation and real-time information flow further reinforces that operational transformation depends on system integration and process readiness [9]. Empirical contributions also associate technology readiness and GenAI usage intensity with performance outcomes through mechanisms such as coordination and capability development [26], [28], [39]. Taken together, adoption is framed as a socio-technical change process that reshapes decision authority, transparency expectations, and workflow design.

A third observation concerns process coverage. The literature is concentrated around modelling, optimization, and resilience-oriented decision support, while routine operational workflows such as monitoring, reporting, and exception communication receive comparatively less direct attention. Although warehouse and logistics contexts demonstrate feasibility for assistant-based support, algorithm-supported scheduling, and alternative management architectures [15], [31], [41], [42], explicit inventory reporting and summarization appears only in limited instances [32]. This imbalance suggests that smart web-based inventory environments may offer value not only through improved decision policies but also through user-facing interpretive support that strengthens coordination and decision comprehension.

Governance and organizational readiness emerge as central themes, with the literature specifying conditions under which GenAI can be responsibly embedded into operations and supply chain decision environments. Conceptual frameworks emphasize structured implementation pathways, capability development, and managerial alignment to avoid ad hoc deployment and to clarify accountability [27], [33], [35], [38], [40], [44] Bibliometric and systematic reviews reinforce that implementation challenges and governance concerns are recurrent across the intelligent supply chain discourse, indicating that organizational capability building is treated as a prerequisite rather than a peripheral concern [8], [30], [34]. Empirical evidence further indicates that technology readiness and the depth of GenAI use are associated with performance outcomes, often via mediating mechanisms such as coordination and operational capability [26], [28], [39]. At the operational layer, LLM applications in employee training illustrate how readiness also involves workforce enablement and knowledge development, complementing organizational readiness arguments in adoption-oriented research [14]. In parallel, procurement-focused discussion highlights that as GenAI expands into upstream decision processes, boundary-setting and decision-right delineation become increasingly salient aspects of responsible integration [29]. Overall, the corpus positions governance and readiness not as compliance add-ons but as core design dimensions shaping the legitimacy, reliability, and sustainability of GenAI-enabled decision support.

The corpus exhibits a strong optimization-centric orientation, consistent with the operations research foundations of inventory decision-making. Several studies develop natural-language-driven optimization modelling, automated optimization model generation, or multi-agent optimization processes that lower the expertise threshold required to formulate, test, and revise decision policies [19], [21], [22]. Inventory-focused work under disruption similarly evaluates LLM-enabled, collaborative approaches to inventory optimization, reinforcing the role of LLMs as support mechanisms within structured optimization pipelines [43]. In these contributions, LLMs primarily function as modelling and interaction enablers that support problem formulation, exploration, and explanation, while feasibility and correctness are preserved by solvers, algorithmic components, or simulation-based evaluation. However, the modelling-centric emphasis may narrow research attention to model enhancement and solver interaction, leaving operationally consequential dimensions such as monitoring design, exception handling, user cognition, and cross-functional workflow coordination less developed despite their importance in web-based inventory environments.

A prominent integration pathway in the corpus is the coupling of LLMs with simulation and digital-twin infrastructures to anchor generative outputs in operational realism. Studies on natural-language-driven simulation modelling demonstrate how LLMs can translate user descriptions into executable simulation artefacts, reducing modelling effort and adoption barriers while keeping scenario evaluation within formal simulation dynamics [12], [20]. Complementary contributions argue that LLMs can enhance digital-twin ecosystems by improving interaction, orchestration, and multi-level decision support in Industry 5.0 contexts, while also highlighting challenges and opportunities for implementation [23], [24]. Related conceptual discussion on foundation models and digital engineering in logistics and supply chains provides additional support for treating these integrations as part of broader digital engineering architectures rather than isolated AI add-ons [16]. Together, these studies suggest that future inventory and logistics decision-support systems may increasingly rely on hybrid architectures where LLMs mediate human intent and interpretation, whereas operational validity remains governed by simulation and digital-twin representations [40].

Despite the natural fit of LLMs for explanation and summarization, comparatively few contributions in the corpus focus on routine inventory monitoring, reporting, and exception communication workflows. Warehouse and logistics studies illustrate the feasibility of LLM-enabled assistance and new management architectures [15], [41], [42]. and case-based work highlights practical integration with enterprise systems [31]. However, explicit inventory-status summarization is represented only in limited form through lightweight automated generation of headline-like status messages [32]. The fact that only 1/34 studies specifically address inventory reporting and monitoring despite this being one of the most common daily use-cases of web-based inventory systems indicates a structural misalignment between where LLM research attention is currently concentrated and where operational value for inventory system users is most immediately needed.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The evaluation landscape described in the corpus is still consolidating. Decision-support scholarship notes that LLMs introduce new questions for how decision support systems should be designed and assessed [36]. Benchmarking work contributes task-specific evidence by comparing LLM performance for supply-chain decision tasks such as risk identification [25]. Mapping and synthesis papers document rapid growth in GenAI-related publications but heterogeneous evaluation standards and fragmented empirical evidence base [8], [30], [34], [37]. Empirical studies connect GenAI readiness and usage depth to performance outcomes, including sustainability-linked outcomes and performance effects mediated by coordination and capability development [26], [28], [39]. Nevertheless, standardized inventory-specific indicators and longitudinal field evaluations remain limited in the abstract-level evidence, suggesting that the field is progressing from exploratory demonstrations toward measurement standardization but has not yet achieved methodological consolidation.

Reliability and trust considerations recur across the literature, reflecting concerns about controllability, consistency, and decision accountability when LLM outputs inform operational decisions. Adoption-focused studies highlight that while GenAI can support operational gains, deployment risks and governance challenges remain salient in operations and supply chain contexts [38], [44]. Resilience-oriented work similarly recognizes risks and emphasizes the need for responsible integration when GenAI is used in decision processes that shape operational continuity [35]. Benchmarking evidence implies variation in model behaviour and output quality across tasks, reinforcing the need for validation when LLMs are embedded in decision-support pipelines [25]. Consequently, multiple contributions converge on collaborative paradigms in which LLMs provide interpretive or advisory support while decision authority remains with humans or deterministic components. This collaborative orientation is particularly explicit in disruption-focused inventory optimization, where human–AI collaborative approaches are evaluated as a means of balancing flexibility and control [43]. Overall, the corpus positions trust as a design property that must be engineered through governance-aligned safeguards and evaluation.

Synthesizing the reviewed evidence suggests that smart web-based inventory decision-support systems should be designed around three interdependent principles: analytical grounding, interpretive interaction, and reliability-by-design. Analytical grounding is

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

supported by integration patterns that anchor LLM outputs within optimization, simulation, or digital-twin infrastructures, thereby maintaining decision validity while improving accessibility [12], [21], [22], [23]. Interpretive interaction leverages LLM capabilities to translate complex analytical outputs into understandable narratives and actionable workflows, an area evidenced in operational assistance and inventory-status summarization but still unevenly developed across the corpus [15], [32]. Reliability-by-design requires governance-aligned safeguards, including validation, accountability arrangements, and structured oversight, to mitigate known risks and maintain trust in decision processes [25], [36], [38]. Aligning these principles with organizational capability development and readiness is essential for moving beyond proof-of-concept deployments toward sustainable operational use [28], [39], [40].

Future work should prioritize end-to-end validation of LLM-enabled inventory decision support in real operational environments and develop standardized benchmarking tasks that reflect inventory decision workflows. Existing benchmarking demonstrates feasibility for evaluating LLMs in adjacent decision tasks, but the literature continues to call for clearer agendas, stronger empirical grounding, and more consistent measurement approaches [8], [25], [33], [34]. Inventory-focused research could advance by explicitly measuring operational outcomes such as service-level attainment, stockout reduction, and overstock mitigation, particularly under disruption conditions where collaborative approaches appear promising [35], [43]. In addition, upstream process coverage remains a clear opportunity. Procurement-oriented discussion indicates that GenAI is expected to expand into sourcing and purchasing decisions, suggesting the need to study boundary-setting, accountability, and workflow integration in these contexts [29]. Finally, broader automation and real-time information flow perspectives imply that integration with operational data streams and intelligent automation architectures will remain important for translating analytics into routine execution [13]..

This review is constrained by the heterogeneity of study designs and the predominance of conceptual and framework-oriented contributions relative to longitudinal, inventory-specific field evaluations [8], [34], [36]. Moreover, the rapid evolution of LLM capabilities and integration practices implies that architectural patterns and governance mechanisms may shift, requiring periodic re-evaluation of evidence standards and design principles as the field matures [33], [38].

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

4. CONCLUSION

This review synthesizes evidence from 34 included studies spanning the 2020–2026 period to map the current landscape of LLM-enabled decision support in inventory management, revealing three overarching findings. First, LLMs are predominantly positioned as augmentation layers—functioning as natural-language interfaces and orchestration components—while decision correctness remains anchored in structured analytical tools such as optimization solvers, simulation engines, and digital twins to mitigate hallucination risks. Second, governance and organizational readiness are central design requirements; because erroneous reordering or procurement decisions carry direct financial consequences, human oversight and accountability arrangements are functional prerequisites for safe deployment. Third, the field remains in an early maturity phase, characterized by a scarcity of hard quantitative evidence on inventory-specific outcomes, such as stockout reduction or cost efficiency, and a lack of standardized evaluation benchmarks. It is important to note that current design implications for web-based inventory systems are primarily derived from the broader supply chain corpus rather than directly validated within specific web-based deployments. Therefore, to enable confident integration into routine inventory workflows at an operational scale, future research must move beyond conceptual framing by validating end-to-end LLM integrations in real environments, developing evaluation benchmarks grounded in operational KPIs, establishing transparent reliability controls, and explicitly targeting web-based inventory system design as an empirical research context.

REFERENCES

[1] L. Chen, A. Bhaumik, X. Wang, and J. Wu, "The Effects of Inventory Management on Business Efficiency," *J. Multidiscip. Res.*, p., 2023, doi: 10.36948/ijfmr.2023.v05i04.4877.

[2] K. Angellin, R. Oetama, and M. Amri, "Web-based inventory and sales information system: Indonesian micro small medium enterprise case study," *JOINS (J. Inf. Syst.)*, vol. 8, no. 1, pp. 57–66, 2023, doi: 10.33633/joins.v8i1.7977.

[3] Ö. Albayrak Ünal, B. Erkayman, and B. Usanmaz, "Applications of Artificial Intelligence in Inventory Management: A Systematic Review of the Literature,"

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Arch. Comput. Methods Eng., vol. 30, no. 4, pp. 2605–2625, 2023, doi: 10.1007/s11831-022-09879-5.

[4] K. Douaioui, R. Oucheikh, O. Benmoussa, and C. Mabrouki, "Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review.," *Appl. Syst. Innov. (ASI)*, vol. 7, no. 5, 2024, doi: 10.3390/asi7050093.

[5] R. Bergsma, C. de Ruijt, and S. Bhulai, "A systematic review of machine learning approaches in inventory control optimization," *Oper. Res. Perspect.*, vol. 15, p. 100367, 2025, doi: 10.1016/j.orp.2025.100367.

[6] Z. Bahroun, A. Saihi, R. As'ad, and M. Tanash, "A systematic analysis of generative artificial intelligence for supply chain transformation," *Supply Chain Anal*, vol. 13, p. 100188, 2026, doi: 10.1016/j.sca.2025.100188.

[7] D. Anh-Hoang, V. Tran, and L.-M. Nguyen, "Survey and analysis of hallucinations in large language models: attribution to prompting strategies or model behavior," *Front. Artif. Intell.*, vol. Volume 8-2025, 2025, doi: 10.3389/frai.2025.1622292

[8] A. R. Teixeira, J. V. Ferreira, and A. L. Ramos, "Intelligent Supply Chain Management: A Systematic Literature Review on Artificial Intelligence Contributions," *Information*, vol. 16, no. 5, 2025, doi: 10.3390/info16050399.

[9] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, Aug. 2015, doi: 10.1016/j.infsof.2015.03.007.

[10] B. Kitchenham, P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering-A systematic literature review," *Inf. Softw. Technol.*, vol. 51, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[11] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, 2021, doi: 10.1371/journal.pmed.1003583.

[12] I. Jackson, M. Jesus Saenz, and D. Ivanov, "From natural language to simulations: applying AI to automate simulation modelling of logistics systems," *Int. J. Prod. Res.*, vol. 62, no. 4, pp. 1434–1457, Feb. 2024, doi: 10.1080/00207543.2023.2276811.

[13] M. Shamsuddoha, E. A. Khan, M. M. H. Chowdhury, and T. Nasir, "Revolutionizing Supply Chains: Unleashing the Power of AI-Driven Intelligent Automation and Real-Time Information Flow," *Information*, vol. 16, no. 1, 2025, doi: 10.3390/info16010026.

[14] A. Gezdur and J. Bhattacharjya, "Innovators and transformers: Enhancing supply chain employee training with an innovative application of a large language model," *Int. J. Phys. Distrib. Logist. Manag.*, vol. 55, no. 4, pp. 394–408, Jan. 2025, doi: 10.1108/IJPDLM-12-2023-0492.

[15] S. Ieva *et al.*, "Enhancing Last-Mile Logistics: AI-Driven Fleet Optimization, Mixed Reality, and Large Language Model Assistants for Warehouse Operations," *Sensors*, vol. 25, no. 9, 2025, doi: 10.3390/s25092696.

[16] B. Nicoletti and A. Appolloni, "The Impact of AI Foundation models on the future of digital engineering for logistics and supply Chain," *Digit. Eng.*, vol. 7, p. 100058, 2025, doi: 10.1016/j.dte.2025.100058.

[17] H. Fan, X. Liu, J. Y. H. Fuh, W. F. Lu, and B. Li, "Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics," *J. Intell. Manuf.*, vol. 36, no. 2, pp. 1141–1157, 2025, doi: 10.1007/s10845-023-02294-y.

[18] K. Du *et al.*, "LLM-MANUF: An integrated framework of Fine-Tuning large language models for intelligent Decision-Making in manufacturing," *Adv. Eng. Inform.*, vol. 65, p. 103263, 2025, doi: 10.1016/j.aei.2025.103263.

[19] Y. Wang, J. Wang, and Z. Chu, "Multi-agent large language models as evolutionary optimizers for scheduling optimization," *Comput. Ind. Eng.*, vol. 206, p. 111197, 2025, doi: 10.1016/j.cie.2025.111197.

[20] M. Elbasheer *et al.*, "Natural language-driven production planning: integrating large language models with automatic simulation model generation in manufacturing systems," *J. Intell. Manuf.*, 2025, doi: 10.1007/s10845-025-02732-z.

[21] J. Li, R. Wickman, S. Bhatnagar, R. K. Maity, and A. Mukherjee, "Abstract Operations Research Modeling Using Natural Language Inputs," *Information*, vol. 16, no. 2, 2025, doi: 10.3390/info16020128.

[22] C. Huang *et al.*, "Orlm: A customizable framework in training large models for automated optimization modeling," *Oper. Res.*, 2025, doi: 10.1287/opre.2024.1233.

[23] C. Chen, K. Zhao, J. Leng, C. Liu, J. Fan, and P. Zheng, "Integrating large language model and digital twins in the context of industry 5.0: Framework, challenges and opportunities," *Robot. Comput. Integr. Manuf.*, vol. 94, p. 102982, 2025, doi: 10.1016/j.rcim.2025.102982.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[24] A. Cimino, F. Longo, V. Solina, and P. Veltri, "Integrating large language models with industrial simulation for multi-level decision support: an innovation management perspective in Industry 5.0," *Eur. J. Innov. Manag.*, vol. 29, no. 11, pp. 27–53, Dec. 2025, doi: 10.1108/EJIM-10-2024-1246.

[25] M. Zhao, O. Hussain, Y. Zhang, M. Saberi, and A. Leshob, "Benchmarking large language models for supply chain risk identification: an extended evaluation within the LARD-SC framework," *Serv. Oriented Comput. Appl.*, 2025, doi: 10.1007/s11761-025-00474-7.

[26] L. Li, W. Zhu, L. Chen, and Y. Liu, "Generative AI usage and sustainable supply chain performance: A practice-based view," *Transp. Res. E Logist. Transp. Rev.*, vol. 192, p. 103761, 2024, doi: 10.1016/j.tre.2024.103761.

[27] R. Dubey, A. Gunasekaran, and T. Papadopoulos, "Benchmarking operations and supply chain management practices using Generative AI: Towards a theoretical framework," *Transp. Res. E Logist. Transp. Rev.*, vol. 189, p. 103689, 2024, doi: 10.1016/j.tre.2024.103689.

[28] G. Malhotra and R. Manzoor, "Generative artificial intelligence adoption for achieving supply chain efficiency, circularity and sustainability," *J. Enterp. Inf. Manag.*, pp. 1–23, Sep. 2025, doi: 10.1108/JEIM-02-2025-0072.

[29] V. Singh, L. Hughes, M. A. Albashrawi, I. Jeon, and Y. K. Dwivedi, "Generative artificial intelligence (GenAI) in procurement and supply chain management: applications, opportunities and challenges," *J. Syst. Inf. Technol.*, vol. 28, no. 1, pp. 123–144, Sep. 2025, doi: 10.1108/JSIT-03-2025-0139.

[30] A. El jaouhari, A. Samadhiya, A. Kumar, and S. Luthra, "Integrating generative artificial intelligence into green logistics: A systematic review and policy-oriented research agenda," *J. Clean. Prod.*, vol. 519, p. 146018, 2025, doi: 10.1016/j.jclepro.2025.146018.

[31] S. Moica, T. Lucian, V. Kostopoulos, A. Gligor, and N. A. Mostafa, "GenAI Technology Approach for Sustainable Warehouse Management Operations: A Case Study from the Automative Sector," *Sustainability*, vol. 17, no. 20, 2025, doi: 10.3390/su17209081.

[32] B. Ji, Y. Hwang, D. Kim, J. Park, M. Ryu, and Y. Cho, "A Lightweight AI System to Generate Headline Messages for Inventory Status Summarization," *Systems*, vol. 13, no. 9, 2025, doi: 10.3390/systems13090741.

[33] H. Wu, G. Li, and D. Ivanov, "The transformative power of generative AI for supply chain management: Theoretical framework and agenda," *Front. Eng. Manag.*, vol. 12, no. 2, pp. 425–433, 2025, doi: 10.1007/s42524-025-4240-x.

[34] A. Sharma, P. Kakkar, S. Hooda, and M. S. Kaswan, "How Does Gen- AI Reshape Supply Chain, Logistics, and Manufacturing? A Systematic Bibliometric Review," *Oper. Res. Forum*, vol. 7, no. 1, p. 5, 2025, doi: 10.1007/s43069-025-00582-2.

[35] T. Boone, B. Fahimnia, R. Ganeshan, D. M. Herold, and N. R. Sanders, "Generative AI: Opportunities, challenges, and research directions for supply chain resilience," *Transp. Res. E Logist. Transp. Rev.*, vol. 199, p. 104135, 2025, doi: 10.1016/j.tre.2025.104135.

[36] A. Handler, K. R. Larsen, and R. Hackathorn, "Large language models present new questions for decision support," *Int. J. Inf. Manage.*, vol. 79, p. 102811, 2024, doi: 10.1016/j.ijinfomgt.2024.102811.

[37] S. K. Srivastava, S. Routray, S. Bag, S. Gupta, and J. Z. Zhang, "Exploring the Potential of Large Language Models in Supply Chain Management:," *J. Glob. Inf. Manag.*, vol. 32, no. 1, 2024, doi: 10.4018/JGIM.335125.

[38] S. Fosso Wamba, C. Guthrie, M. M. Queiroz, and S. Minner, "ChatGPT and generative artificial intelligence: an exploratory study of key benefits and challenges in operations and supply chain management," *Int. J. Prod. Res.*, vol. 62, no. 16, pp. 5676–5696, Aug. 2024, doi: 10.1080/00207543.2023.2294116.

[39] L. Li, Y. Liu, Y. Jin, T. C. E. Cheng, and Q. Zhang, "Generative AI-enabled supply chain management: The critical role of coordination and dynamism," *Int. J. Prod. Econ.*, vol. 277, p. 109388, 2024, doi: 10.1016/j.ijpe.2024.109388.

[40] I. Jackson, D. Ivanov, A. Dolgui, and J. Namdar, "Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation," *Int. J. Prod. Res.*, vol. 62, no. 17, pp. 6120–6145, Sep. 2024, doi: 10.1080/00207543.2024.2309309.

[41] M. Kmiecik, "Creating a genetic algorithm for third-party logistics' warehouse delivery scheduling via a large language model," *J. Model. Manag.* vol. 20, no. 4, pp. 1138–1162, Jan. 2025, doi: 10.1108/JM2-06-2024-0192.

[42] T. Berlec, M. Corn, S. Varljen, and P. Podržaj, "Exploring Decentralized Warehouse Management Using Large Language Models: A Proof of Concept," *Appl. Sci.*, vol. 15, no. 10, 2025, doi: 10.3390/app15105734.

[43] R. Wu, "Inventory optimization under supply chain disruptions: Leveraging large language models for human-AI collaborative decision-making," *J. King Saud Univ. Comput. Inf. Sci.*, 2026, doi: 10.1007/s44443-025-00458-9.

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

[44] S. Fosso Wamba, M. M. Queiroz, C. J. Chiappetta Jabbour, and C. (Victor) Shi, "Are both generative AI and ChatGPT game changers for 21st-Century operations and supply chain excellence?," *Int. J. Prod. Econ.*, vol. 265, p. 109015, 2023, doi: 10.1016/j.ijpe.2023.109015.