Beyond Mere Automation: A Techno-functional Framework for Reimagining Gen-AI in Supply Chain Operations

Sreyoshi Bhaduri
drsre@amazon.com
Amazon
New York, NY, USA

Pavan Nithin Mullapudi
pavmul@amazon.com
Amazon
Seattle, WA, USA

Shannon Dietrich
shadietr@amazon.com
Amazon
Washington DC, USA

Scott DeWaters
sdwaters@amazon.com
Amazon
Nashville, TN, USA

Raj Ratan
ratanraj@amazon.com
Amazon
Seattle, WA, USA

Brajesh Kashyap
bkashya@amazon.com
Amazon
Seattle, WA, USA

Rajanikanth Mandava
rajanikm@amazon.com
Amazon
Seattle, WA, USA

Lu Guo
luguo@amazon.com
Amazon
Seattle, WA, USA

Hungjen Wang
hungjen@amazon.com
Amazon
New York, NY, USA

Vykunth Ashok
vykunth@amazon.com
Amazon
San Francisco, CA, USA

Abhilasha Katariya
abhkata@amazon.com
Amazon
Seattle, WA, USA

Rohit Malshe
malshe@amazon.com
Amazon
Seattle, WA, USA

Ankush Pole
ankupole@amazon.com
Amazon
Seattle, WA, USA

Arkajit Rakshit
rakshit@amazon.com
Amazon
Seattle, WA, USA

Abstract

As Generative AI (Gen-AI) continues to evolve rapidly, its potential to transform supply chain operations remains largely unexplored. Narrowing in on retail supply chain, this paper presents a taxonomy diagram that categorizes trends in Gen-AI adoption across various functions thereby mapping current Gen-AI capabilities and identifying immediate opportunities and potential challenges. We identify several key patterns in Gen-AI integration, including the automation of routine cognitive tasks, and enhancement of human decision-making capabilities. We posit that while Gen-AI shows immense promise in improving supply chain efficiency and resilience, successful implementation requires careful consideration of existing workflows, user capabilities, and organizational readiness. Finally, we present a cohesive vision for scaling Gen-AI in Supply Chain operations. Ultimately, this position paper provides insights for both practitioners looking to implement Gen-AI solutions and researchers exploring the future of AI in and for supply chain management.

Keywords

Generative AI, Supply Chain, Retail

ACM Reference Format:
Sreyoshi Bhaduri, Pavan Nithin Mullapudi, Shannon Dietrich, Scott DeWaters, Raj Ratan, Brajesh Kashyap, Rajanikanth Mandava, Lu Guo, Hungjen Wang, Vykunth Ashok, Abhilasha Katariya, Rohit Malshe, Ankush Pole, and Arkajit Rakshit. 2025. Beyond Mere Automation: A Techno-functional Framework for Reimagining Gen-AI in Supply Chain Operations. In *Proceedings of the 1st Workshop on "AI for Supply Chain: Today and Future" @ 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/XXXXXX.XXXXXX

1 Introduction

An age-old adage warns that if the only tool one has is a hammer, every problem begins to look like a nail. This wisdom feels especially relevant today in retail supply chain management, where the rapid rise of Generative AI (Gen-AI), particularly Large Language Models (LLMs), has sparked both excitement and caution. On one hand, some organizations are eager to wield Gen-AI as a universal solution risking misapplication by force-fitting into complex logistics workflows. On the other, others remain hesitant, wary of the hype or overwhelmed by the pace of change, potentially

KDD '25, August 3, 2025, Toronto, ON, Canada.

ACM ISBN 979-8-4007-1454-2/25/08…$15.00
https://doi.org/10.1145/XXXXXX.XXXXXX

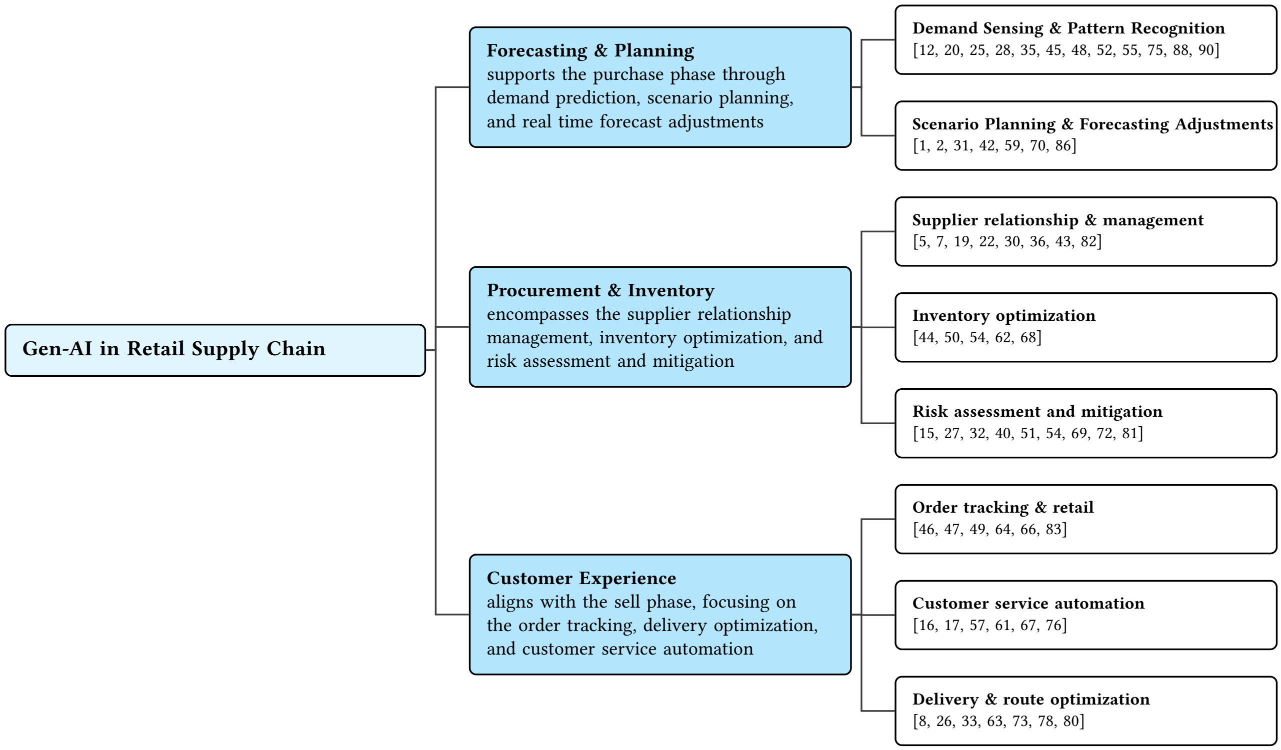

Figure 1: Taxonomy of Gen-AI applications across the retail supply chain, synthesized from a review of 60 empirical papers

missing out on early strategic advantages. The real challenge lies not in a lack of technology or ambition, but in bridging the gap between emerging AI capabilities and the nuanced, domain-specific needs of supply chain operations. What's needed is a structured, techno-functional perspective: one that clarifies *where* Gen-AI can meaningfully contribute across the retail supply chain lifecycle, and *how* organizations can move beyond reactive experimentation toward deliberate, value-driven transformation.

This position paper responds to this challenge of a missing techno-functional paradigm to Gen-AI implementation. While traditional supply chain automation has targeted routine or manual tasks, Gen-AI offers a new paradigm, providing cognitive augmentation for planners and operators, and customer-facing agents across first, middle, and last miles. From forecasting demand and planning procurement, optimizing warehouse workflows, to automating last-mile customer communication, Gen-AI is not one hammer, but a collection of precision instruments waiting to be intentionally and intelligently matched with the right problems. We present a taxonomy and commentary grounded in a review of current empirical work on Gen-AI in supply chain. We describe how Gen-AI can augment the roles of typical supply chain personas such as planners, operators, and analysts across this spectrum. Next, we offer practical insights for integration by leveraging Gen-AI to structure unstructured data, enhance decision quality, and support resilient, adaptive operations at an unprecedented speed and scale.

2 An Age of Generative AI

Since the advent of ChatGPT in late 2022, Generative Artificial Intelligence (Gen-AI), particularly Large Language Models (LLMs), has rapidly transformed multiple industries by demonstrating unprecedented capabilities in understanding, generating, and synthesizing human-like language [9, 56]. This technological breakthrough represents the culmination of several decades of progress in natural language processing, with recent advances in computational resources, dataset availability, and algorithmic innovation converging to enable systems of remarkable linguistic competence [14, 89]. The release of these powerful language models to the public has accelerated their adoption and sparked widespread interest in their potential to augment human capabilities across numerous domains [11].

LLMs, such as GPT-3, GPT-4, and their derivatives, including AI agents employ advanced neural network architectures, notably transformers, enabling them to model extensive contextual relationships within textual data and deliver highly coherent and relevant outputs [10, 77]. These models leverage self-attention mechanisms that allow them to capture long-range dependencies in text, while their massive scale—often comprising hundreds of billions of parameters—provides them with a form of emergent intelligence that surpasses previous approaches [38, 84]. The training process for these models typically involves two phases: pre-training on vast corpora of internet text to develop general language understanding, followed by fine-tuning with human feedback to align outputs with human preferences and safety constraints and in sophisticated methods - use of reinforcement learning to speed up the process [18, 58].

This transformative technology has sparked significant academic and industry interest due to its diverse applications across sectors such as supply chain optimization, customer engagement automation, content creation, and data analytics [23, 89]. For example, educational applications have emerged for personalized tutoring, curriculum development, and research assistance [39, 41]. Even the creative industries have been revolutionized through LLM-powered tools for writing, design ideation, and multimedia content generation [24, 87].

As organizations integrate LLM-driven solutions, the landscape of operational efficiency, strategic decision-making, and innovation will evolve rapidly, necessitating ongoing research and rigorous evaluation to address challenges related to accuracy, ethical considerations, and sustainable deployment [4, 85]. The phenomenon of "hallucination" where models confidently generate plausible but factually incorrect information remains a significant obstacle to reliable deployment in high-stakes domains [37]. Privacy concerns persist regarding the potential for models to memorize and reproduce sensitive training data [13, 53]. Additionally, the environmental impact of training and deploying these increasingly large models raises questions about sustainability [60, 65], while issues of bias and fairness demand careful consideration as these systems become more deeply embedded in societal infrastructure [4, 6]. Despite these obstacles, the rapid pace of innovation in LLM research suggests that we are merely glimpsing the transformative potential of these technologies in reshaping humans interact with information systems. [9, 71].

To navigate the complex and rapidly evolving landscape of AI adoption in supply chain management, it is helpful to think in terms of a layered maturity model of AI capabilities. These range from foundational automation where technology executes repetitive, rules-based tasks to more advanced stages such as predictive modeling, explainability, simulation, and real-time decision support. As supply chain organizations increasingly explore Generative AI, particularly Large Language Models, they often oscillate between two extremes: either overextending the technology's capabilities by using it as a one-size-fits-all hammer, or hesitating to apply it at all due to concerns around trust, complexity, or integration. The challenge lies in recognizing that Gen-AI is not a monolithic solution but a spectrum of tools that can be selectively aligned with specific supply chain functions.

3 Current State of Gen-AI in Supply Chain

The integration of Generative Artificial Intelligence (Gen-AI) into supply chain management represents an emerging frontier characterized by both significant promise and notable research disparities. Similar to adoption patterns observed across other domains, the current body of knowledge in supply chain applications of Gen-AI predominantly consists of narrowly focused empirical investigations alongside a larger number of conceptual position papers that envision and encourage more comprehensive implementations. This uneven bifurcation in literature, between limited experimental studies clustered around select use cases and broader theoretical frameworks, signals the nascent stage of this technological application to the domain.

For this paper, we initially identified relevant literature by conducting systematic searches using academic databases and Google Scholar for terms such as "Gen-AI AND supply chain AND retail" and "LLM AND supply chain AND retail" restricting results to English-language publications with emphasis on empirical studies that elaborated on specific implementations and practical challenges. A critical methodological distinction was the deliberate focus on Generative AI technologies specifically, rather than broader artificial intelligence applications documented extensively in prior literature reviews and position papers [21, 34, 74] in the supply chain domain. Building upon this foundation, the team then employed snowball referencing methodology, examining citations within initial papers to discover additional relevant research, thereby expanding the corpus through iterative citation network exploration. Currently, this work-in-progress review incorporates nearly 50 technical papers spanning the functions across the retail supply chain domain and Gen-AI applications. Similar to other researchers describing supply chain [3, 29, 79] we then organize the current state of research across the three primary domains of retail supply chain: Forecasting & Planning (purchase phase through demand prediction, scenario planning, and real time forecast adjustments), Procurement & Inventory (supplier relationship management, inventory optimization, and risk assessment and mitigation), and Customer Experience (sell phase, including order tracking, delivery optimization, and customer service automation).

3.1 Forecasting & Planning

Notably, several studies focus on demand sensing and pattern recognition, leveraging Gen-AI to improve the granularity and adaptability of forecasts. For example, researchers [75, 88] explored transformer-based architectures for time-series demand forecasting, often comparing their performance to traditional statistical methods. In [59] authors take on a hybrid analytics model to analyze retail order data and better forecast next-time order delivery success with enhanced explain-ability through LLMs. The key here is leveraging LLMs on top of traditional approaches to ensure informed decision making and expedited action. In [1, 2, 48], authors leverage different LLM powered frameworks that input queries in plain text and output insights about the underlying optimization outcomes. The papers highlight how LLMs can simulate demand shocks or test operational strategies without the need for coded simulation environments. These systems allow planners to explore possible future scenarios through natural language interfaces. Finally, real-time forecast adjustment is emerging as a distinct use case, with authors [31] showing how zero-shot forecasting using LLMs can act as a fallback or sanity check for existing models. This real-time capability is particularly valuable for adapting to unforeseen demand changes like weather events, especially in high-velocity retail environments.

3.2 Procurement & Inventory

Gen-AI's integration into procurement workflows remains exploratory but promising. In the area of supplier relationship management

LLMs have helped in qualifying vendors, flagging compliance issues, or drafting procurement communications [7, 19, 22, 30]. There is also experimentation with using GenAI to respond to and resolve invoice inquires and discrepancies. These systems reduce manual overhead while supporting more dynamic, data-rich supplier interactions. Inventory optimization is another focal area. Empirical bodies of work in this area demonstrate how generative models support decisions such as reorder timing, quantity planning, and safety stock calibration. Multiple studies explore the use of ChatGPT-like assistants in retail inventory settings [50, 62, 68]. When it comes to risk assessment and mitigation, Gen-AI is used to synthesize multi-source risk data, including geopolitical and environmental disruptions, however, researchers are cautious [32, 81] and discuss broader implications.

3.3 Customer Experience

Gen-AI also plays a growing role in enhancing customer-facing supply chain functions. In [66] for instance, authors build a Retail-GPT that analyzes data from e-commerce transactions, customer interactions, and sales records to predict consumer behavior while other research [49] proposes using Gen-AI for a digital twin system towards a low-carbon integrated freight transportation system. For customer service automation, [17, 57, 67] provide empirical evidence that GPT-based chatbots are innovative tools that can outperform traditional scripted bots in increasing sales and handling complex order queries or troubleshooting, with increased satisfaction and resolution speed. However, [16] caution how ChatGPT can be abused to produce practical and realistic communications used for phishing attacks on customers. Finally, delivery optimization, especially in last mile, is another domain where Gen-AI tools are being embedded. Researchers [8, 26, 63] are actively evaluating transformer-enhanced route planning and real-time dispatch systems that adapt to urban congestion, delivery windows, and dynamic constraints.

3.4 Potential and Missed Opportunities

Despite encouraging use cases, significant gaps remain in the integration of Gen-AI across the retail supply chain, particularly when examined through the lens of first, middle, and last mile operations. These known limitations hinder senior leadership from committing to significant investments while creating uncertainty about which specific skill sets and technologies are essential for implementation. In first mile activities, such as forecasting and strategic planning, Gen-AI is predominantly used for demand pattern recognition and scenario generation. However, few studies explore how these models can be embedded into continuous planning cycles or interact with human decision-makers in live operational contexts. Integration remains largely siloed and simulation-based. Within the middle mile, which includes procurement, inventory, and distribution center operations, adoption is even more limited. While generative models show potential for inventory policy optimization and supplier communications, there is little empirical evidence on their real-time deployment within ERP, warehouse management or financial systems. Chatbots have predominantly been employed to enable inventory management. The gap here lies in customizable execution fidelity and system interoperability. The last mile, often closest to the customer, shows greater Gen-AI experimentation particularly in order tracking, service support, delivery routing, and order-specific notes for delivery drivers. Yet, these deployments borrowing heavily from other contexts tend to remain front-end focused, with limited insight into how they influence back-end supply chain specific logistics or fulfillment strategies.

Collectively, these gaps suggest that while Gen-AI holds transformative promise, current implementations are preliminary, uneven, and sometimes fragmented, requiring deeper system integration and better alignment with operational workflows. Many of the current applications are still disconnected from broader processes and people in the supply chain ecosystem. Initiatives are often driven by AI teams emulating chatbot architectures or recommendation systems from other industries, underscoring a reactive rather than proactive adoption pattern. This reveals a broader issue: the lack of supply chain domain specific techno-functional solutions that align Gen-AI capabilities with the operational realities and performance metrics of supply chain management. For instance Gen-AI systems are frequently evaluated based on generic metrics (e.g., coherence, latency, token usage) instead of supply chain-relevant outcomes such as order fill rate, cost per package, or customer satisfaction scores. Without this alignment, Gen-AI risks being yet another proverbial and expensive hammer applied indiscriminately across different nails in the supply chain space.

4 Vision for Scaling Generative AI in Supply Chain Operations

Our vision for scaling Generative AI across supply chain introduces a Gen-AI-forward architecture allowing techno-functional development (see Figure2). Our framework recommends beginning by ingesting diverse raw data, both structured operational logs and unstructured domain documents, into a sophisticated Knowledge Synthesizer. This component leverages Large Language Models for initial knowledge extraction, which can then be critically refined through Business Intelligence Curation by functional domain experts. The resulting validated knowledge is encoded into a comprehensive, dynamic Knowledge Graph. This graph will serve as a central *brain*, fueling diverse applications like empowering business intelligence teams to build specialized analytical agents, enabling data science to develop advanced attribution models, and generating deep AI insights. All outputs are harmonized by an Orchestrator Agent and surfaced through an intuitive user interface, creating a continuously learning ecosystem that redefines operational intelligence in complex supply chains.

4.1 Foundation: The Four Pillars

Central to our recommended Gen-AI architecture is a robust framework anchored by four interconnected pillars, essential for sustainable AI scaling: Knowledge Framework and Data Organization, Data Architecture and Infrastructure, Domain-specific Agents, and Response Quality and Model Validation.

4.1.1 Knowledge Framework and Data Organization The foundational element of our vision is a sophisticated knowledge framework designed to represent the complex interplay of supply chain people,

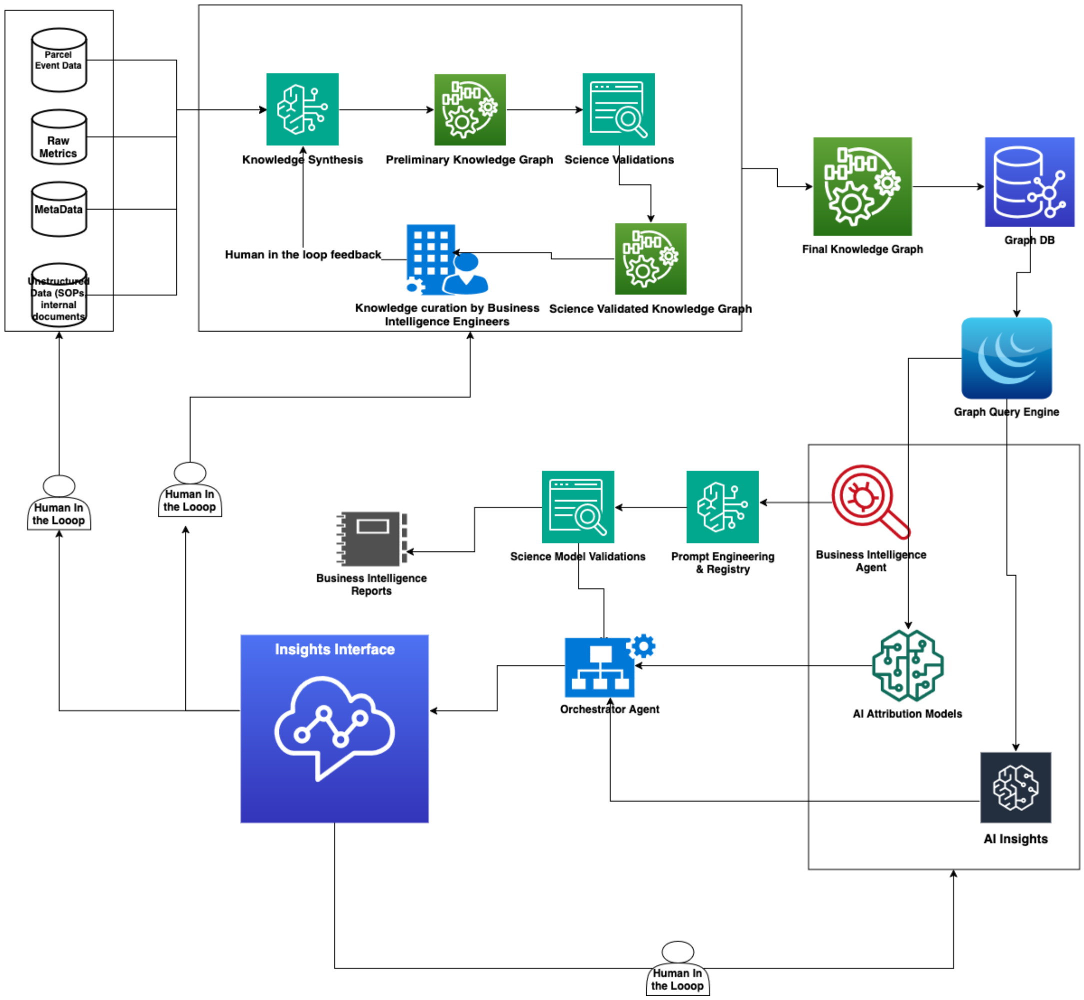

Figure 2: This Architecture diagram presents our original Gen-AI architecture for supply chain operations, designed to move beyond siloed deployments and enable intentional, end-to-end integration. At its core, the architecture addresses the current fragmentation in Gen-AI adoption where isolated agents or models are deployed without systemic coordination or alignment to business goals. Our framework is structured around interconnected layers: (1) a Knowledge Synthesizer that ingests structured and unstructured data using LLMs; (2) a Business Intelligence Curation layer that encodes expert reviewed outputs into a dynamic, graph-based knowledge system; and (3) an Orchestration Engine that powers domain-specific agents for Speed, Cost, Quality related KPIs responsible for optimizing discrete trade-offs. These agents surface insights through a unified interface, with built-in human-in-the-loop feedback mechanisms to ensure transparency and alignment with operational KPIs. What differentiates this architecture is its focus on techno-functional alignment: it is not just technically feasible, but operationally grounded, bridging AI capabilities with the rhythms and realities of supply chain work.

processes, and metrics. We employ a property graph-based knowledge representation system that surpasses traditional relational data structures by modeling intricate connections and dependencies inherent in supply chain operations. At its core, this system is structured around a carefully constructed ontology, delineating crucial classes such as Stages, representing various lifecycle steps of a parcel; Events, capturing status changes and critical transitions; Metrics, encompassing both predictive and lagging performance indicators; and Entities, identifying different actors and objects within the supply chain ecosystem.To ensure data integrity and consistency, we implement a robust Master Data Management (MDM) framework that establishes golden records for key entities and maintains standardized taxonomies across systems. The framework incorporates a sophisticated versioning system that tracks changes to both the knowledge structure and its contents over time. This includes version control for ontology updates, maintaining audit trails of relationship changes, and ensuring backward compatibility as the knowledge base evolves. Relationships among these classes are meticulously defined and visualized as graph edges, enhancing the AI systems' ability to accurately interpret context and causality. This structured representation effectively mitigates risks of hallucinations and significantly improves response accuracy from Gen-AI systems.

4.1.2 Data Architecture and Infrastructure Complementing the knowledge framework is a scalable and performant data architecture tailored for extensive operational demands, with stringent latency requirements (sub-second for critical operations, <10 seconds for analytical queries). The infrastructure encompasses near real-time data pipelines leveraging AWS's distributed infrastructure, enabling rapid event processing across multiple availability zones. It features scalable storage solutions (utilizing services like S3, RDS, and DynamoDB) optimized to handle the diverse needs of transactional and analytical workloads simultaneously. Advanced feature engineering pipelines systematically transform raw data into formats optimized for AI applications, ensuring readiness for sophisticated modeling and predictive analysis. Data tables at various aggregations are stored with appropriate encryption and access controls, following AWS Well-Architected Framework principles. Additionally, comprehensive data governance frameworks are integral to our approach, ensuring robust security protocols and stringent compliance standards (including GDPR and CCPA) across all data operations

4.1.3 Domain-specific Agents Recognizing that a generalized approach can dilute effectiveness, our framework emphasizes the development of specialized AI agents that address targeted supply chain functions. Traditionally, supply chain optimization has been framed around the triad of speed, cost, and quality. However, these priorities are often inversely competing, since improving one can come at the expense of the others. For example, improving delivery speed may increase operational costs or degrade packaging quality. We propose a Gen-AI system where each priority is represented by a domain-specific agent, e.g., a Speed Agent, Cost Agent, and Quality Agent, each equipped with the ability to analyze context-specific data and recommend optimal strategies within its area of focus. Crucially, these agents are coordinated by a Supervisor Agent or directed by a human user, who defines the business objective for a given period (e.g., maximizing speed during strategic weeks or reducing cost during off-peak operations). This supervisor role is not just about setting parameters, but also about receiving transparent insights when trade-offs are made. Rather than hidden optimizations, the system provides explainable outputs: why a delivery delay was accepted, why costs were higher for a particular route, or how quality metrics were preserved under constrained timelines. This kind of balance with goal-aligned orchestration, currently difficult for supply chain leaders to achieve manually, can become tractable and actionable through AI-driven collaboration between specialized agents and human oversight.

4.1.4 Response Quality and Trust Calibration Ensuring consistent reliability and accuracy is paramount, hence our model validation pillar integrates extensive quality control mechanisms. Systematic prompt engineering and robust management processes guarantee consistency and precision in AI responses across diverse supply chain scenarios. Detailed response logging and comprehensive analytics further enhance transparency and facilitate continual improvement. Automated benchmarking against established performance metrics ensures objective measurement of accuracy and efficiency, complemented by regular human-in-the-loop review processes to maintain quality and instill trust in AI outputs.

4.2 Recommended Operational Framework

The successful deployment of our Gen-AI vision necessitates clearly defined roles and responsibilities across organizational teams. The engineering team maintains and manages the core infrastructure and data pipelines, ensuring seamless operational integrity. The business intelligence team develops and sustains the domain-specific agents, embedding functional expertise directly into the AI tools. Science teams are tasked with model development and rigorous validation, continuously enhancing model performance. Lastly, the product team oversees the human-in-the-loop review processes and drives comprehensive adoption of the system across organizational units. In all these processes feedback and feed-forward mechanisms are used in all possible setups. This framework provides a concrete and more traditional roadmap for leadership to consider and plan for investment.

4.3 Scaling Strategy

Our proposed scaling strategy leverages advanced cloud infrastructure capabilities to ensure seamless growth and sustained high performance. The strategy includes automated scaling of computational resources, accommodating increased data volumes through distributed processing frameworks. Load balancing mechanisms effectively distribute computational demands across multiple specialized AI agents, maintaining optimal system responsiveness. Furthermore, geographic distribution of infrastructure ensures low latency and high availability, supporting global operational requirements. In some cases the scaling strategies also involve simpler models distilled from more complex models and routing mechanisms to decide which agents are called at any given point. Additionally, intelligent caching strategies and real-time performance monitoring enable proactive scaling decisions, ensuring consistent performance as the system grows from small-scale deployments to enterprise-wide implementations.

4.4 Success Metrics

Evaluating the effectiveness of our Gen-AI scaling framework involves a suite of quantifiable success metrics. Improvements must be systematically tracked across operational efficiency metrics, number of decisions made, decision-making response times, quality of decisions, prediction accuracy, recommendation relevance, adoption rates of AI-powered tools, and tangible financial outcomes such as cost reductions and service quality enhancements. This structured, comprehensive approach ensures continuous assessment and improvement, promoting long-term operational excellence and sustainable scaling of Gen-AI capabilities within complex supply chain networks.

5 Recommendations for the Field

To drive innovation and progress in this field, we recommend organizations collectively invest in a new class of co-designed Gen-AI systems (LLM models, and AI agents) that embed both technical and functional expertise. These systems are as fluent in inventory logic and logistics scheduling as in natural language or transformer based technologies. This requires deep collaboration between AI researchers and supply chain practitioners, alongside clear frameworks that translate supply chain KPIs into Gen-AI design and evaluation. The following recommendations provide the techno-functional roadmap for strategic Gen-AI integration across supply chain.

5.1 Align Technical Design & Functional Realities

Many current Gen-AI applications are built in technical silos, designed to push the boundary of model capability rather than solve field-based supply chain problems. This mismatch limits business leadership enthusiasm and operational adoption. Gen-AI solutions must instead be designed to mirror the natural decision cadence of supply chain professionals— like daily replenishment tasks, weekly forecast adjustments, or quarterly scenario planning exercises. Tools should seamlessly plug into existing systems, embedding AI-generated recommendations into familiar planning or fulfillment workflows. Importantly, interpret-ability and accuracy must be prioritized over technical novelty. A Gen-AI tool should not only generate recommendations but explain them in ways that align with operational language and mental models as accurately as possible. For instance, an LLM-based replenishment assistant should justify reorder suggestions based on anticipated lead times, current inventory turnover, and service-level targets—contextualizing the decision in real business logic.

5.2 Democratize Role-Centric Empowerment

Widespread adoption of Gen-AI requires that the technology be accessible to every actor in the supply chain ecosystem—not just data scientists. Democratization of AI must empower planners, buyers, warehouse supervisors, and customer agents to envision themselves as users and co-designers of intelligent tools. Interfaces must be intuitive, prompt formats adapted to real-world language, and outputs tailored to diverse roles and levels of decision authority. This is not just a matter of accessibility; it's about changing how people view their responsibilities. A planner should see see Gen-AI not as an unfamiliar science initiative, but as their day to day job assistant, which they are part owner in developing and enhancing. Their assistant can help sanity-check assumptions, simulate alternative demand conditions, or draft communications to suppliers. The planner must begin to provide valuable guidance to improve the outputs of the AI models/agents. This shift requires cultural transformation, embedding AI into the everyday identity of supply chain professionals.

5.3 Build Just-In-Time AI Education

Given the real-time and high-pressure nature of supply chain work, structured AI training sessions are often impractical. What's needed instead is a shift toward just-in-time, embedded education. AI learning should happen in the flow of work, through micro-lessons, contextual tips, and task-aware prompts. For instance, an inventory analyst using a Gen-AI dashboard might receive interactive suggestions on how to refine input parameters or interpret results, while a warehouse manager using an AI-Agent-based assistant might be guided to phrase operational queries more effectively. These interventions must be designed not as one-off trainings, but as learning scaffolds that build intuition incrementally. Over time, this model cultivates a more confident and AI-literate workforce without disrupting operations.

5.4 Embed Technical Teams Within Functional Domains

Achieving meaningful techno-functional alignment requires that AI developers work side-by-side with supply chain professionals, and supply chain professionals embrace AI in all possible aspects. Rather than building tools in isolation, technical experts must be embedded within operations teams to observe, co-design, and iterate on solutions. This model turns AI engineers into empathetic problem solvers, who understand functional pain points and regulatory constraints, and who can tailor model behavior accordingly. An engineer embedded within a logistics team, for instance, is far more likely to design a routing assistant that accounts for vehicle capacity, cut-off times, or carrier preferences. Such embedded collaborations also foster mutual trust essential for enabling iterative model refinement and sustainable adoption.

5.5 Translate KPIs into Gen-AI Objectives

To ensure that Gen-AI tools deliver meaningful outcomes, they must be evaluated based on supply chain-specific performance metrics. Current practice often centers on AI-native benchmarks like BLEU scores or model perplexity, which are poorly aligned with business impact. Instead, organizations must define clear mappings between Gen-AI output and operational KPIs such as forecast accuracy, order fill rate, inventory turns, on-time delivery, and customer satisfaction. These KPIs should inform model fine-tuning, reward mechanisms in reinforcement learning loops, and even the framing of Gen-AI prompts. By doing so, AI systems will be optimized not just for language fluency, but for supply chain efficacy.

5.6 Launch Cross-Functional Pilots at All Scales

Effective Gen-AI integration does not require massive infrastructure overhauls. Organizations can start small, with focused cross-functional pilots that pair a single AI engineer with a domain expert such as a buyer, dispatcher, or inventory planner. These pilots should focus on addressing specific, high-impact pain points, with clear success metrics and rapid iteration. As trust grows and outcomes are demonstrated, solutions can scale to broader teams. This approach encourages grassroots innovation, accelerates feedback cycles, and lowers the barrier to experimentation. Moreover, it provides the organization with reusable blueprints—models, prompts, and interface components that can be adapted for other functions.

5.7 Foster a Culture of AI Curiosity and Agency

Ultimately, the shift to Gen-AI is not just technological but cultural. Organizations must foster curiosity, experimentation, and a sense of shared ownership over the evolution of AI tools. Supply chain professionals should be encouraged to challenge AI outputs, propose new use cases, and customize interfaces to better suit their workflows. Leaders should recognize and reward not just model accuracy, but experimentation at the frontlines whether it's a planner discovering a more effective prompt for demand simulation or a warehouse lead using LLM models or AI agentic workflows to onboard a new employee. When AI tools are shaped by users, and when users are empowered to shape them, adoption becomes organic and innovation becomes continuous.

6 Closing Thoughts

To move forward, what is needed is a new class of co-designed Gen-AI systems that embed both technical and functional expertise systems that are as fluent in inventory logic and logistics scheduling as they are in natural language. This calls for deeper collaboration between AI researchers and supply chain experts, alongside clearer frameworks to translate supply chain KPIs into Gen-AI training and evaluation protocols. Only then can Gen-AI deliver the systemic impact it promises across the full retail supply chain lifecycle. To realize the full potential of Generative AI in supply chain operations, organizations must thus move beyond reactive experimentation and toward intentional, techno-functional integration. This means aligning technological capabilities with the functional realities, constraints, and rhythms of supply chain work—from forecasting and procurement to fulfillment and customer service. Achieving this vision requires a collaborative approach that unites the perspectives and skills of engineering, data science, and supply chain management.

At the core of this transformation lies a pressing need to democratize AI. Today, most Gen-AI tools are confined to technical teams or innovation pilots, far removed from the daily workflows of planners, buyers, and warehouse supervisors. For Gen-AI to deliver systemic value, it must be usable not just by AI scientists, but by everyone including non-technical supply chain professionals. This democratization entails more than simply making tools accessible. It requires a cultural shift where supply chain professionals see themselves not merely as executors of plans, but as collaborators who enhance their workflows through AI systems. For example, a demand planner should begin to see a large language model not as a black-box forecaster built by an AI consultant, but as a human-in-the-loop reasoning augmenter that can help simulate across scenarios, cross-check logic or human error, or interrogate assumptions.

However, we must acknowledge a key barrier in this vision: time and operational bandwidth. Supply chain is a high-pressure environment where decisions must be made quickly, often with incomplete information. It is not always feasible for frontline professionals to pause and enroll in structured AI training programs. As such, AI education must be just-in-time, embedded, and contextual. Organizations should prioritize development of modular micro-learning resources, interactive prompts, and AI co-pilot systems that teach users while they work. Instead of static tutorials or centralized training, learning should be integrated into the task flow. For example, helping a planner learn how to refine a Gen-AI-generated scenario, or showing a warehouse supervisor how to frame a question to a GPT-based inventory assistant. These moments of learning should be designed not as isolated classroom exercises, but as productivity-enhancing on-the-job tools that can help build intuition over time.

Simultaneously, technical teams must meet operations functional teams where they are. This means hiring embedded science and AI teams that work shoulder-to-shoulder with functional operators, not as external consultants, but as integrated problem-solvers. These embedded roles act as bridges, translating functional pain points into AI design opportunities and iterating solutions in real-world environments. Rather than building generic Gen-AI tools in isolation, AI developer teams must collaborate directly with demand planners, logistics coordinators, or customer service leads to re-imagine how AI can serve their specific decisions and constraints. This integrated model allows AI solutions to evolve in tandem with business logic, regulatory shifts, and operational nuance. Importantly, this techno-functional approach should not be reserved for large enterprise teams. Even mid-size and regional supply chains can adopt this model incrementally starting with small, cross-functional pilots that pair AI engineers with inventory planners or last-mile teams. Over time, these collaborations help seed a new kind of organizational literacy: one where supply chain operators are not only AI users, but active participants in shaping what Gen-AI can do for their function.

In this next phase of Gen-AI integration, success will hinge less on model sophistication and more on thoughtful orchestration—of people, processes, AI-Agents, and perspectives. Engineering, science, and operations must work in concert, not in silos, to build Gen-AI systems that are technically sound, functionally relevant, and human-centric in design. Only then can supply chains move from fragmented experiments to strategic transformation.

References

[1] Mahsa Abbaszadeh Nakhost, Alexander Bloemer, and Stefan Minner. 2025. Full automation or just a copilot? A comparison study of Large Language Models on spare parts demand forecasting. *A comparison study of Large Language Models on spare parts demand forecasting (April 29, 2025)* (2025).

[2] Mahdi Abolghasemi, Odkhishig Ganbold, and Kristian Rotaru. 2025. Humans vs. large language models: Judgmental forecasting in an era of advanced AI. *International Journal of Forecasting* 41, 2 (2025), 631–648.

[3] James B Ayers and Mary Ann Odegaard. 2017. *Retail supply chain management.* CRC Press.

[4] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021), 610–623.

[5] Tomaž Berlec, Marko Corn, Sergej Varljen, and Primož Podržaj. 2025. Exploring Decentralized Warehouse Management Using Large Language Models: A Proof of Concept. *Applied Sciences* 15, 10 (2025), 5734.

[6] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. *arXiv preprint arXiv:2005.14050* (2020).

[7] Bloomberg. 2023. Walmart Is Using AI to Negotiate the Best Price With Some Vendors. https://www.bloomberg.com/news/articles/2023-04-26/walmart-uses-pactum-ai-tools-to-handle-vendor-negotiations

[8] Aigerim Bogyrbayeva, Meraryslan Meraliyev, Taukekhan Mustakhov, and Bissenbay Dauletbayev. 2022. Learning to solve vehicle routing problems: A survey. *arXiv preprint arXiv:2205.02453* (2022).

[9] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258* (2021).

[10] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems* 33 (2020), 1877–1901.

[11] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv preprint arXiv:2303.12712* (2023).

[12] Ricardo Caetano, José Manuel Oliveira, and Patrícia Ramos. 2025. Transformer-Based Models for Probabilistic Time Series Forecasting with Explanatory Variables. (2025).

[13] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. 2021. Extracting Training Data from Large Language Models. In *30th*

USENIX Security Symposium (USENIX Security 21). 2633–2650.

[14] Qingqing Chang, Jinze Li, Damai Xu, Guangzhi Wang, Yunfan Feng, Lingpeng Wang, and Yong Chen. 2023. A Survey on In-context Learning. *arXiv preprint arXiv:2301.00234* (2023).

[15] Wenwen Chen, Yangchongyi Men, Noelia Fuster, Celia Osorio, and Angel A. Juan. 2024. Artificial Intelligence in Logistics Optimization with Sustainable Criteria: A Review. *Sustainability* 16, 21 (2024), 1–22.

[16] Minhaz Chowdhury, Nafiz Rifat, Shadman Latif, Mostofa Ahsan, Md Saifur Rahman, and Rahul Gomes. 2023. Chatgpt: The curious case of attack vectors' supply chain management improvement. In *2023 IEEE International Conference on Electro Information Technology (eIT)*. IEEE, 499–504.

[17] N Chowdhury, O Abdoulsalam Awais, and S Aktar. 2023. Improving customer care with ChatGPT: A case study. (2023).

[18] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems*. 4299–4307.

[19] CNBC. 2023. A.I. Could 'Remove All Human Touchpoints' in Supply Chains. Here's What that Means. https://www.cnbc.com/2023/06/19/supply-chains-how-ai-could-remove-all-human-touchpoints.html

[20] Rares Cristian, Pavithra Harsha, Clemente Ocejo, Georgia Perakis, Brian Quanz, Ioannis Spantidakis, and Hamza Zerhouni. 2024. Inter-series transformer: Attending to products in time series forecasting. *arXiv preprint arXiv:2408.03872* (2024).

[21] Giovanna Culot, Matteo Podrecca, and Guido Nassimbeni. 2024. Artificial intelligence in supply chain management: A systematic literature review of empirical studies and research directions. *Computers in Industry* 162 (2024), 104132.

[22] DHL. 2023. ChatGPT and the Like: Artificial Intelligence in Logistics. https://dhl-freightconnections.com/en/trends/chatgpt-and-the-like-artificial-intelligence-in-logistics/

[23] Yogesh K Dwivedi, Nir Kshetri, Laurie Hughes, Emma L Slade, Anand Jeyaraj, Arpan Kumar Kar, Abdullah M Baabdullah, Alex Koohang, Vishnupriya Raghavan, Mona Ahuja, et al. 2023. So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management* 71 (2023), 102642.

[24] Kathleen Elkins and Jingwen Chun. 2023. Can large language models transform computational social science? *Patterns* 4, 8 (2023).

[25] Mohd Faizan. 2025. AI AND SUPPLY CHAIN: A COMPREHENSIVE ANALYSIS. *International Journal of Development Research* (2025).

[26] Getu Fellek, Ahmed Farid, Goytom Gebreyesus, Shigeru Fujimura, and Osamu Yoshie. 2023. Graph transformer with reinforcement learning for vehicle routing problem. *IEEJ Transactions on Electrical and Electronic Engineering* 18, 5 (2023), 701–713.

[27] Jiawei Feng, Mengsi Cai, Fangze Dai, Tianci Bu, Xiaoyu Zhang, Huijun Zheng, and Xin Lu. 2024. Modeling Supply Chain Interaction and Disruption: Insights from Real-world Data and Complex Adaptive System. *arXiv preprint arXiv:2405.10818* (2024).

[28] Colin Gavaghan, Alistair Knott, and James Maclaurin. 2021. The Impact of Artificial Intelligence on Jobs and Work in New Zealand. (2021).

[29] Deng Ge, Yi Pan, Zuo-Jun Shen, Di Wu, Rong Yuan, and Chao Zhang. 2019. Retail supply chain management: a review of theories and practices. *Journal of Data, Information and Management* 1, 1 (2019), 45–64.

[30] A Shaji George, AS Hovan George, and T Baskar. 2023. Leveraging AI to Revolutionize Procurement: ChatGPT's Hidden Potential to Transform the Procurement Iceberg. *Partners Universal International Innovation Journal* 1, 6 (2023), 110–131.

[31] Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew G Wilson. 2023. Large language models are zero-shot time series forecasters. *Advances in Neural Information Processing Systems* 36 (2023), 19622–19635.

[32] Christian Hendriksen. 2023. Artificial intelligence for supply chain management: Disruptive innovation or innovative disruption? *Journal of Supply Chain Management* 59, 3 (2023), 65–76.

[33] Business Insider. 2025. *Uber Freight is betting that AI will transform the trucking industry by 2025*. https://www.businessinsider.com/ai-trucking-logistics-uber-freight-tech-optimize-routes-2025-4

[34] Ilya Jackson, Dmitry Ivanov, Alexandre Dolgui, and Jafar Namdar. 2024. Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. *International Journal of Production Research* 62, 17 (2024), 6120–6145.

[35] Md Abrar Jahin, Asef Shahriar, and Md Al Amin. 2024. Mcdfn: Supply chain demand forecasting via an explainable multi-channel data fusion network model integrating cnn, lstm, and gru. *arXiv e-prints* (2024), arXiv–2405.

[36] Valeria Jannelli, Stefan Schoepf, Matthias Bickel, Torbjørn Netland, and Alexandra Brintrup. 2024. Agentic LLMs in the Supply Chain: Towards Autonomous Multi-Agent Consensus-Seeking. *arXiv preprint arXiv:2411.10184* (2024).

[37] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *Comput. Surveys* (2023).

[38] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling Laws for Neural Language Models. *arXiv preprint arXiv:2001.08361* (2020).

[39] Enkelejda Kasneci, Kathrin Sëgitz, Gjergji Kasneci, Katrin Deiss, Michael Färber, Wolfgang Gaissmaier, Carsten Kovacs, Roland Schatz, Frederic Toussaint, Jörg Lischeid, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023), 102274.

[40] Mohsen Khabir, Gholam Reza Emad, Mehrangiz Shahbakhsh, and Maxim A. Dulebenets. 2025. A Strategic Pathway to Green Digital Shipping. *Logistics* (2025).

[41] Rubaina Khan, Sreyoshi Bhaduri, Tammy Mackenzie, Animesh Paul, Sankalp KJ, and Indrani Sen. 2024. Path to personalization: A systematic review of genai in engineering education. In *KDD AI4Edu Workshop*.

[42] Naghmeh Khosrowabadi, Kai Hoberg, and Christina Imdahl. 2022. Evaluating human behaviour in response to AI recommendations for judgemental forecasting. *European Journal of Operational Research* 303, 3 (2022), 1151–1167.

[43] Mariusz Kmiecik. 2023. ChatGPT in third-party logistics–The game-changer or a step into the unknown? *Journal of Open Innovation: Technology, Market, and Complexity* 9, 4 (2023), 100174.

[44] Mariusz Kmiecik. 2025. Creating a genetic algorithm for third-party logistics' warehouse delivery scheduling via a large language model. *Journal of Modelling in Management* (2025).

[45] Carsten Krause. 2024. Case Study: Amazon's AI-Driven Supply Chain: A Blueprint for the Future of Global Logistics. *The CDO TIMES* (August 2024). cdotimes.com

[46] Anuj Kumar, Nimit Gupta, and Gautam Bapat. 2024. Who is making the decisions? How retail managers can use the power of ChatGPT. *Journal of Business Strategy* 45, 3 (2024), 161–169.

[47] Abderrahim Lakehal, Adel Alti, and Boubakeur Annane. 2025. CORES: Context-Aware Emotion-Driven Recommendation System-Based LLM to Improve Virtual Shopping Experiences. *Future Internet* 17, 2 (2025), 94.

[48] Beibin Li, Konstantina Mellou, Bo Zhang, Jeevan Pathuri, and Ishai Menache. 2023. Large language models for supply chain optimization. *arXiv preprint arXiv:2307.03875* (2023).

[49] Xueping Li, Haowen Xu, Jose Tupayachi, Olufemi Omitaomu, and Xudong Wang. 2024. Empowering cognitive digital twins with generative foundation models: Developing a low-carbon integrated freight transportation system. *arXiv preprint arXiv:2410.18089* (2024).

[50] Zhihong Li, Albaraa Ksibi, and Xiaoying Xu. 2024. Optimizing inventory management using a multi-agent LLM system. (2024).

[51] Weiqiang Lin. 2022. Automated infrastructure: COVID-19 and the shifting geographies of supply chain capitalism. *Prog Hum Geogr.* (2022).

[52] SKR Malikireddy. 2023. Enhancing retail supply chain resilience with generative AI. *World Journal of Advanced Engineering Technology and Sciences* 9 (2023), 399–409.

[53] Milad Nasr, Nicholas Carlini, Briland Hitaj, Matthew Jagielski, Florian Tramèr, and Nicolas Papernot. 2023. Scalable Extraction of Training Data from (Production) Language Models. *arXiv preprint arXiv:2311.17035* (2023).

[54] Benjamin Nitsche. 2021. *Artificial Intelligence and Sustainability: Innovations in Business and Managerial Practices*.

[55] José Manuel Oliveira and Patrícia Ramos. 2024. Evaluating the Effectiveness of Time Series Transformers for Demand Forecasting in Retail. *Mathematics* 12, 17 (2024), 2728.

[56] OpenAI. 2023. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774* (2023).

[57] Michał Orzoł and Katarzyna Szopik-Depczyńska. 2023. ChatGPT as an innovative tool for increasing sales in online stores. *Procedia Computer Science* 225 (2023), 3450–3459.

[58] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training Language Models to Follow Instructions with Human Feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744.

[59] Cagatay Ozdemir, Sezi Cevik Onar, and Ömer Ekmekcioğlu. 2024. Innovative Order Delivery Success Prediction in Online Retail: Integrating ML and LLM to Gain Actionable and Understandable Insights. In *International Conference on Intelligent and Fuzzy Systems*. Springer, 169–176.

[60] David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon Emissions and Large Neural Network Training. *arXiv preprint arXiv:2104.10350* (2021).

[61] Artur M Pereira, J Antao B Moura, Evandro De B Costa, Thales Vieira, Andre RDB Landim, Eirini Bazaki, and Vanissa Wanick. 2022. Customer models for artificial intelligence-based decision support in fashion online retail supply chains. *Decision Support Systems* 158 (2022), 113795.

[62] Yinzhu Quan and Zefang Liu. 2024. Invagent: A large language model based multi-agent system for inventory management in supply chains. *arXiv preprint arXiv:2407.11384* (2024).

[63] Baptiste Rabecq and Rémy Chevrier. 2022. A deep learning Attention model to solve the Vehicle Routing Problem and the Pick-up and Delivery Problem with Time Windows. *arXiv preprint arXiv:2212.10399* (2022).

[64] Mohammad Reshadati. 2024. Creating a Gen-AI based Track and Trace Assistant MVP (SuperTracy) for PostNL. *arXiv preprint arXiv:2409.02711* (2024).

[65] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. 2020. Green AI. *Commun. ACM* 63, 12 (2020), 54–63.

[66] Farooq Shareef, Rishi Ajith, Parth Kaushal, and Karthik Sengupta. 2024. RetailGPT: A Fine-Tuned LLM Architecture for Customer Experience and Sales Optimization. In *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*. 1390–1394. https://doi.org/10.1109/ICSSAS64001.2024.10760685

[67] Keivan Simetgo, Apostolos N Giovanis, and Dimitris Kallivokas. 2025. THE ROLE OF CHATGPT AND ARTIFICIAL INTELLIGENCE IN CUSTOMER MANAGEMENT STRATEGY TRANSFORMATION: A SYSTEMATIC. (2025).

[68] Damian Skórnóg and Mariusz Kmiecik. 2023. SUPPORTING THE INVENTORY MANAGEMENT IN THE MANUFACTURING COMPANY BY CHATGPT. *Logforum* 19, 4 (2023).

[69] Kazi Abdul Wadud Sohag. 2025. Integrating AI and Sustainable Practices in the Supply Chain of the Garment Industry in Bangladesh. (2025).

[70] Matthew J Spaniol and Nicholas J Rowland. 2023. AI-assisted scenario generation for strategic planning. *Futures & Foresight Science* 5, 2 (2023), e148.

[71] Alex Tamkin, Miles Brundage, Jack Clark, and Deep Ganguli. 2021. Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. *arXiv preprint arXiv:2102.02503* (2021).

[72] The ILS Company. 2025. Unlocking the Potential of AI in Logistics: Benefits, Challenges Best Practices. *ILS Company Blog* (2025).

[73] Yonglin Tian, Xuan Li, Hui Zhang, Chen Zhao, Bai Li, Xiao Wang, and Fei-Yue Wang. 2023. VistaGPT: Generative parallel transformers for vehicles with intelligent systems for transport automation. *IEEE Transactions on Intelligent Vehicles* 8, 9 (2023), 4198–4207.

[74] Reza Toorajipour, Vahid Sohrabpour, Ali Nazarpour, Pejvak Oghazi, and Maria Fischl. 2021. Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research* 122 (2021), 502–517.

[75] Iván Vallés-Pérez, Emilio Soria-Olivas, Marcelino Martínez-Sober, Antonio J Serrano-López, Juan Gómez-Sanchís, and Fernando Mateo. 2022. Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications* 201 (2022), 116993.

[76] Varun Vasudevan, Faezeh Akhavizadegan, Abhinav Prakash, Yokila Arora, Jason Cho, Tanya Mendiratta, Sushant Kumar, and Kannan Achan. 2025. LLM-driven Constrained Copy Generation through Iterative Refinement. *arXiv preprint arXiv:2504.10391* (2025).

[77] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Advances in Neural Information Processing Systems*. 5998–6008.

[78] Chandra Sekhar Veluru. 2023. A comprehensive study on optimizing Delivery Routes through Generative AI using Real-Time Traffic and Environmental Data. *Journal of Scientific and Engineering Research* 10, 10 (2023), 168–175.

[79] Chandra Sekhar Veluru. 2024. Investigating the Impact of Artificial Intelligence and Generative AI in E-Commerce and Supply Chain: A Comprehensive. *European Journal of Advances in Engineering and Technology* 11, 4 (2024), 131–143.

[80] Matt Villano. 2025. How Uber Freight is leveraging AI to make truck routes more efficient. *Business Insider* (April 2025).

[81] Samuel Fosso Wamba, Maciel M Queiroz, Charbel Jose Chiappetta Jabbour, and Chunming Victor Shi. 2023. Are both generative AI and ChatGPT game changers for 21st-Century operations and supply chain excellence? *International Journal of Production Economics* 265 (2023), 109015.

[82] Shangcheng Wang. 2024. Forecasting Demand and Controlling Inventory Using Transformer-Based Neural Network Optimized by Whale Optimization Algorithm. In *Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence*. 915–923.

[83] Wang Wang, Ping Zhang, Changxia Sun, and Dengchao Feng. 2024. Smart customer service in unmanned retail store enhanced by large language model. *Scientific Reports* 14, 1 (2024), 19838.

[84] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022. Emergent Abilities of Large Language Models. *arXiv preprint arXiv:2206.07682* (2022).

[85] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, et al. 2022. A Taxonomy of Risks Posed by Language Models. *arXiv preprint arXiv:2112.04359* (2022).

[86] Aleksandr Yankelevich, R.V. Rikard, Travis Kadylak, Michael J. Hall, Elizabeth A. Mack, John P. Verboncoeur, and Shelia R. Cotten. 2018. Preparing the workforce for automated vehicles. *American Center for Mobility* (2018).

[87] Andy Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In *27th International Conference on Intelligent User Interfaces*. 841–852.

[88] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are transformers effective for time series forecasting?. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 11121–11128.

[89] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A Survey of Large Language Models. *arXiv preprint arXiv:2303.18223* (2023).

[90] Beilei Zhu and Fan Li. 2025. Unveiling the Unexpected: Transformer Outperforms Specialized Architectures in Volatile Supply Chain Forecasting. In *Computational Science and Computational Intelligence*, Hamid R. Arabnia, Leonidas Deligiannidis, Farid Ghareh Mohammadi, Soheyla Amirian, and Farzan Shenavarmasouleh (Eds.). Springer Nature Switzerland, Cham, 113–126.