LTIMindtree

WHITEPAPER

Next-Gen Supply Chain Decision-Making through Multi-Agentic RAG

Real-time autonomous supply chains powered by intelligent agents and unified data

Authors

Sujatha Babu | Vijay Kumar

LTIMindtree

Table of contents

Executive summary

There is a growing need for faster, smarter decision-making in supply chains, a shift from static, manual processes to agile, real-time, autonomous processes. Traditional tools and strategies no longer deliver the needed agility and optimization. The advent of Multi-Agentic Retrieval Augmented Generation will enable faster data-driven decisions, reduced latency, enhanced forecast accuracy, operational efficiency, and improved resilience. This will allow planners and supply chain teams to focus on higher-value strategic interventions. The business impact includes cost savings, stronger customer satisfaction, customer retention, successful strategic and new market penetrations through better market responsiveness, improved compliance, and a stronger competitive market edge.

In the supply chain domain, data remains fragmented across legacy systems, ERP platforms, emails, call transcripts, and external unstructured sources such as social media, investor reports, and analyst surveys. This fragmentation is further compounded by tool sprawl from mergers and acquisitions (M&A), fragmented processes, the multiplier impact of latency, and reactive decision-making, thereby hindering agility and decision quality. For example, following a series of acquisitions, a global manufacturing conglomerate found its demand forecast hampered by fragmented data across 12 ERP and supply chain systems – resulting in rising inventory buffers, excess working capital, and service level failures in key markets.

As generative artificial intelligence (GenAI) continues to evolve, a quieter and powerful innovation has emerged in enterprise decision-making: Multi-Agentic Retrieval-Augmented Generation. This framework fuses the real-time context-aware capabilities of Retrieval-Augmented Generation (RAG), drawing from both structured and unstructured data, with autonomous learning agents (Agentic AI systems) that monitor, decide, and iteratively learn. When these specialized agents collaborate to execute complex processes, they create a Multi-Agentic RAG architecture – an intelligent, scalable, and near real-time decision engine for complex enterprise processes.

Multi-Agentic RAG addresses these challenges with its powerful framework by enabling agents to detect market shifts, optimize inventory, and support real-time planning and replanning decisions. It integrates inputs from diverse data sources into a unified intelligent system. For instance, a leading consumer goods company could enhance demand

sensing by ingesting millions of social conversations to track emerging health and lifestyle trends. This real-time consumer insight enables aligning production with emerging trends, accelerating time-to-market for new offerings, and improving market responsiveness.

This paper explores the Multi-Agentic RAG strategic value in supply chains, focusing on planning, thereby offering a scalable blueprint for broader adoption.

Introduction

Supply Chain Management (SCM) covers every step in converting raw materials into final products, covering integrated planning, sourcing, production, network optimization, warehousing, logistics, global trade compliance, distribution, returns, after-sales, and supplier management. Despite its strategic importance, many organizations still operate with multiple siloed tools, from basic spreadsheets to complex planning systems, often inherited through mergers and acquisitions (M&A), alongside unstructured external data such as market trends and business intelligence.

From inconsistent processes and a lack of operational standardization to multiple data sources and gaps in regulatory compliance, our clients often report the same challenge: they lack a unified way to make agile, timely, data-driven decisions.

The advent of Multi-Agentic RAG promises a fundamental shift from static reactive models to agile, real-time, data-driven operations. This shift reduces latency, boosts resilience, and unlocks significant business value through improved forecast accuracy, proactive disruption management, standardized decisions, enhanced shipment tracking, and stronger regulatory compliance.

Artificial Intelligence (AI) has rapidly evolved, with successive developments introducing novel frameworks, innovative training methodologies, and applications. Among these, Agentic RAG, which combines RAG's contextual data integration with autonomous decision-making, stands out. Unlike traditional AI, Agentic AI consists of intelligent agents trained for a specific task and autonomously scan, analyse, decide, act, and continuously learn. When multiple such agents collaborate, they form a scalable Multi-Agentic RAG system.

As global supply chains grow more complex and demand real-time decision-making, Multi-Agentic RAG's ability to integrate structured and unstructured siloed data with domain and industry expertise offers a new approach to operational agility. It empowers supply chain professionals to shift focus from routine tasks to strategic initiatives. While the promise is substantial, successful Multi-Agentic RAG implementation requires addressing challenges such as careful task selection for the agents, agent training, agent collaboration protocols, data security, risk mitigation, integration within the existing technical landscape, and ultimately, the adoption within the organization.

This paper explores the strategic value of Multi-Agentic RAG in supply chains, focusing on demand planning as a key use case. The intent is to highlight the strategic role of Multi-Agent RAG in shaping autonomous supply chain decisions, not the mechanics behind them.

Examples of current adoption of agentic RAG in planning tools

As enterprises seek to modernize planning capabilities, several leading solution providers have begun integrating Agentic RAG into their platforms, reflecting a clear shift towards intelligent, autonomous, and interoperable planning systems. As highlighted by ISG's 2025 market report, there has been a rise in Agentic AI use cases across various industries[i], underscoring the growing relevance of these capabilities in supply chain optimization.

One of the key highlights of SAP's integrated AI capabilities is enabling a new interoperability protocol for Agent2Agent (A2A)[ii] through its partnership with Google Cloud. This provides the foundation for secure collaboration between AI agents across different platforms, thus authenticating a proper multi-agent system. SAP is also advancing multimodal RAG, especially for video-based content. Joule[iii], grounded by SAP Knowledge Graph and SAP Business Data Cloud, can activate AI agents to execute complex workflows.

o9 Solutions' digital brain platform already uses AI and machine learning (ML). o9 is integrating RAG technology into its digital brain platform to help businesses with their supply chain decision-making. This could include intelligent automation through translating natural language queries into o9's proprietary language and determining the response from one or more of the knowledge bases, through data retrieval, by accessing domain knowledge, or by hyper-automating the workflows[iv]. In a recent development, o9 showcased its "Decision Replay System"[v], akin to post-game analysis in sports. The self-learning model can use multi-layer causal analysis to understand the delta between planning and outcomes and feed this intelligence back into the planning loop.

As per a press release at the beginning of the year, Oracle Fusion Cloud SCM[vi] now has embedded more than 20 new role-based AI agents to help organizations achieve new levels of productivity and growth across the supply chain, spanning procurement, planning, product life cycle management, etc. The agents include a procurement policy advisor, a sustainability policy guide, etc.

Kinaxis[vii], through its Maestro, has introduced AI agents that users can interact with to monitor, predict, and act in real-time, thereby automating key tasks like inventory management and disruption mitigation. Its Agentic AI framework will also enable companies to create AI agents on Maestro easily. This allows for AI-driven supply chain orchestration that is accessible to businesses at any stage of AI maturity.

These initiatives signal growing confidence in the potential of Agentic RAG to transform traditional planning models. For instance, businesses can deploy agents to continuously monitor commodity markets, detect pricing trends, and forecast supply shifts. This real-time sensing enables proactive adjusting of procurement strategies – thereby increasing supply chain reliability, reducing cost volatility, and avoiding stockouts. As capabilities mature, the next wave of adoption will likely focus on scalable, cross-functional orchestration – paving the way for more dynamic and resilient supply chains.

Designing the Multi-Agentic RAG framework for planning

How does this translate into real-world implementation for our clients? Consider key level 1 supply chain processes such as demand and supply planning, inventory management, production planning, warehousing, logistics, shipping, and compliance. Figure 1 (High-level overview of agents in Supply Chains) illustrates the primary agent(s) in each of these processes. Solid and dashed arrows indicate the communication between agents within and between processes.

The data can be structured (from ERPs, POS, etc.) or unstructured (from emails, news feeds, social media, weather channels, maritime information, investor reports, etc.), and from internal and external sources. While RAG refers to training data as internal data and all other data as external data, internal and external data here refer to proprietary enterprise data and publicly or commercially available data, respectively. RAG provides relevancy and accuracy by connecting to multiple data sets. For example, it can combine historic data from ERP systems and real-time sentiment from social media feeds for conversations around current trends. This and the agent's embedded domain expertise help the agent sense demand spikes and arrive at a more accurate forecast.

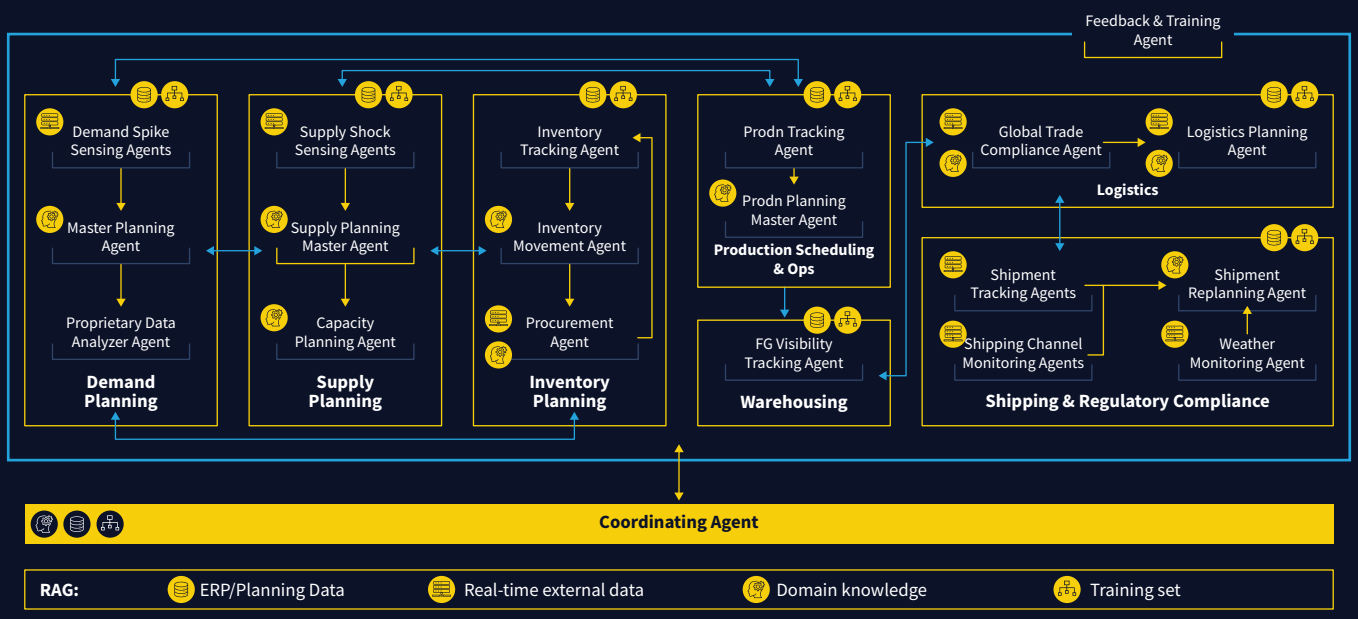

Figure 1: High-level overview of agents in Supply Chains

Process	Current state	Future state			
	Role of a demand planner	Role of agent(s)	Description	Demand planner's role	% Automation expected
Data collation for demand forecasting	• Collect input data • Adjust inputs	Demand Spike **Sensing Agent**, ***Collating Agent*** *Collate data that impacts the demand forecast of a product/product category.*	• Agents automatically poll real-time external and internal data sources impacting demand. • Collating Agent collates and reconciles responses from sensing agents and filters outliers.	-	100%
Forecast generation and consensus demand finalization	• Generate system forecast • Analyse, adjust and finalize	**Forecasting Agent** *Generate a forecast based on collated external and internal data using statistical and AI/ML algorithms.*	• Agent uses an AI/ML algorithm to arrive at a consensus demand forecast from collated data. This includes segmentation, basic time series-based statistical forecast, driver-based forecast using AI/ML, NPI forecast using like item, cannibalization impact, sell-out forecast, and sell-out to sell-in conversion.	-	100%

			• Agent can set up segmentation parameters to group products into different classes based on sales volume or variability, apply different forecasting models, and use segments as filters for exception-based planning.		
Plan publishing	Publish and archive the plan	**Forecasting Agent** *Automatically publish the plan along with the basis for the plan if it does not deviate beyond a pre-agreed norm. Trigger alerts to demand planner in case of deviations.*	• If the consensus demand plan deviation is within a pre-defined threshold, the Agent will automatically publish the plan along with the basis for the plan. This may include disaggregation of forecasts, Publish-to-Supply Plan, Publish-to-Integrated Business Plan, and snapshots. • If the deviation is beyond the threshold, automatically alert the Planner along with the plan details, underlying data used to arrive at the plan, the significant driver(s) that are leading to the deviation, and assumptions, if any. The Demand Planner can then manually intervene to adjust and publish the plan.	Adjust plan, if needed	75%
Post Game Analysis	Conduct post-plan review	**Monitoring Agent** *Retrain based on deviations between actuals, published plan, and plan adjustments made by the demand planner.*	Agent compares the published plan to actuals as well as the manual adjustments by the demand planner, analyses deviations, and re-trains.	Analyse and retrain	75%

Table 1: Current and future state in demand planning

The agent personas that emerge are detailed in Figure 2. (Supply chain agent personas). This proposed framework will be demonstrated for demand planning, though the approach applies equally to other supply chain modules.

The demand planning use case

Table 1 (Current and future state in demand planning) outlines the four key processes in demand planning, the current role of the Demand Planner, the envisaged future state within Multi-Agentic RAG implementation, and the estimated automation level for each, thus providing a clear view of how amenable each process is to Agentic AI implementation. The core agents in demand planning are indicated in Table 2 (Core agents in demand planning).

Signal Intelligence agent persona

Continuously monitors designated channel(s) to detect real-time signals or events. (e.g., tracking inclement weather, maritime disruptions, competitor activity, and critical investor updates)

Adaptive Planning Strategist persona

Harness real-time internal and external data to **design, optimize, and adapt plans**. (e.g., agile response for forecasting demand or emerging disruptions)

Strategic Alignment Trainer persona

Re-train by identifying gaps between planned strategy and actuals, driving **continuous alignment** and smarter future planning.

Orchestration agent persona

Coordinates across all agents for cohesive decision-making and a feedback loop for continuous alignment.

Figure 2: Supply chain agent personas

Agent	Based on	Description
Sensing Agent	Signal Intelligence agent persona	*External sensing agents* gather insights to identify market shifts or competitor activity (social media, investor reports, industry sources such as S&P Global), detect demand spikes by analysing event data (e.g., from event booking sites), and health indicators in life sciences and pharma, etc. *Internal sensing agents* analyse purchase patterns (POS data), current promotions (Sales, Marketing), dynamic pricing strategies (Finance), and order data (Sales, Customer Service).
Collating Agent	Orchestration agent persona	Aggregates and reconciles insights from sensing agents to create a unified view.
Forecasting Agent	Adaptive Planning Strategist persona	Uses collated data to generate forecasts using AI/ML models (e.g., establishing segmentation logic, applying segment-based filters for exception-based planning, etc).
Monitoring Agent	Strategic Alignment Trainer persona	Incorporates continuous learning by analysing variances between actual outcome, agent's published plan, and the Demand Planner's adjusted plan.

Table 2: Core agents in demand planning

Let us focus on the data collation and forecast generation processes listed in Table 1, walk through the agent(s) and RAG considerations, and list the value realization and guardrails.

Consider a major promotion planned around an upcoming outdoor event. A demand planner would have generated a forecast in advance based on the expected lift from the promotion. However, the estimates must be adjusted promptly if severe weather is later predicted. Without timely insight into the changing conditions, the planner relies on outdated assumptions, leading to suboptimal planning decisions. Now imagine a system where one agent continuously monitors weather conditions and another generates forecasts. When detecting an adverse weather pattern, the sensing agent alerts the forecasting agent, who autonomously adjusts based on severity and probability.

More broadly, demand planners collate input from internal and external channels via pull or push mechanisms. Forecasts are generated by coalescing and reconciling data and manually adjusting the plan iteratively due to the stochastic nature of data—often without real-time visibility into emerging disruptions.

A Multi-Agentic RAG framework illustrated in Figure 3 (Multi-Agentic RAG framework) addresses this complexity through intelligent agents.

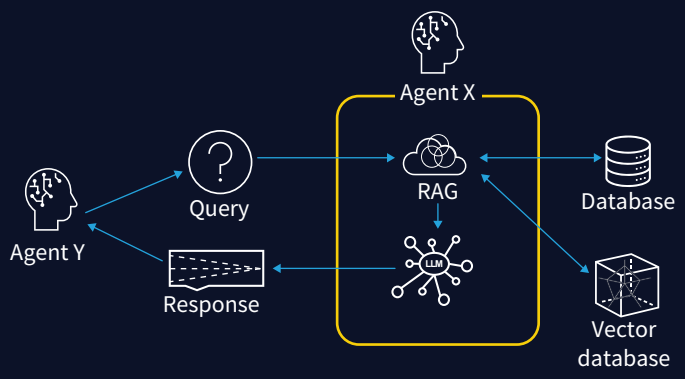

Figure 3: Multi-Agentic RAG framework

Internal sensors
Example: channel inventory, production uptime

External sensors
Example: competitor social media, analyst survey, investor reports

Structured first-party data
Example: historical sales data, current order data

Unstructured first-party data
Example: opportunities in emails and call transcripts, contract penalties

Figure 4: RAG components

The RAG components can be managed by one or more RAG agents based on the nature of data retrieval complexity that the agent can handle, as indicated in Figure 4 (RAG components).

01

Collating Agent:

Periodically or in real-time, this agent consolidates and reconciles insights from other agents and performs outlier correction.

02

Forecasting Agent:

Generate forecast using reconciled insight iteratively—through historic patterns, adjustments for pre-set deviations, or alerting Demand Planner for intervention if deviation exceeds pre-set limits—until plan finalization.

03

Feedback and Training Agent:

Evaluates agent-generated plans and adjustments by Demand Planner, including:

- Recurring outliers
- Frequency of manual intervention
- Magnitude, root causes, and direction of deviations between agent-generated plan and adjusted forecasts (under- or over-prediction)
- Obsolescence of current forecasting drivers and emergence of new drivers
- Impact of the segmentation rules

This closed-loop approach enhances agility, accuracy, and adaptability in demand planning, empowering planners to focus on strategic decision-making instead of reactive corrections.

The 3Es (Enablement, Enhancement, and Effectiveness)

To effectively deploy Multi-Agentic RAG in supply chains, organizations should adopt a structured approach based on the 3Es, as shown in Figure 5 (Structured implementation approach).

Figure 5: Structured implementation approach

The advantage of this framework extends beyond automation. It enables real-time access to diverse data sources, dynamic workflows, adaptive reasoning, and autonomous decision-making. Key ROI drivers include:

Forecast accuracy improvements (by 15-25%) through real-time integration of structured and unstructured data, enabling faster reaction to disruptions (e.g., market shifts, shipment delays, commodity volatility, tariffs, customer feedback, competitor activity, etc).

Reduction in inventory holding costs (by 10-20%) through improved demand visibility and optimized working capital.

On-time in-full (OTIF) improvement of up to 99% through more responsive and accurate planning.

Planner productivity gains (by 25-40%) through minimizing manual activities and accelerating planning cycle times.

Fewer stockouts and missed opportunities through proactive detection of demand shifts and risks

Lean high-performing planning teams enabled by agent-based decision support.

Stronger competitive edge through faster market responsiveness.

Continuous learning embedded in the planning ecosystem.

The result is an agile supply chain with better working capital utilization through optimized inventory, lower safety stock, smarter pricing, better assortments, increased on-time delivery, and positive market engagement. However, adoption hurdles exist and include:

- Complexity of domain-specific workflows.
- Encoding domain expertise.
- ERP's restriction of external agents' direct interaction.
- Effective feedback loops and re-training cycles.
- Legal, compliance, and data governance.
- Scalability due to limited robust plug-and-play agents.
- Organizational change management, shifting from a UI- or application-backend mindset to a process-first agent-narrative approach.

Conclusion

When you have optimized your supply chain to the optimum as per traditional strategies, Agentic RAG takes you even further with exemplary ROI that current systems and platforms cannot. Multi-Agentic RAG achieves this through integrating diverse data, autonomous decision-making, and enabling adaptive learning – freeing planners to focus on strategic oversight.

To scale its adoption, organizations need

- Generic plug-and-play agents
- Customizable industry-specific agents
- Real-time training algorithms
- Clear performance metrics
- Standardized interfaces for real-time data
- Human-in-the-loop processes, guardrails
- Governance protocols and legal clarity around autonomous actions

Success depends on high-quality data, feedback systems, and coordinated organizational change management. With measurable gains in agility, resilience, and decision speed, Multi-Agentic RAG is moving from concept to competitive advantage. The shift towards autonomous supply chain decision-making is underway and accelerating.

Glossary

Term	Description
Adaptive Planning Strategist persona	Harness real-time data from internal and external sources to design, optimize, and adapt plans.
Agent personas	Personas of agents identified for the process under consideration
Agentic RAG	Autonomous agents orchestrate retrieval across diverse information sources in a multi-step workflow, enabling iterative decision making.
Agents	Autonomous AI software entities that perform tasks on behalf of users or systems by perceiving the environment, making autonomous decisions, and taking actions. They can operate independently or in collaboration with other agents.
Consensus demand forecast	A unified forecast derived from multiple sources—system-generated, sales, marketing, finance, and customer inputs. It is finalized through collaborative planning and serves as the baseline for supply and financial planning.
Demand Planning	A strategic process that uses historical data, market intelligence, and statistical models to forecast future customer demand. It helps align inventory levels with anticipated demand to improve service levels and reduce costs.
Forecasting techniques	Time Series-Based Forecasting: Uses historical data observed at regular intervals to predict future demand. Assumes that past patterns (trend, seasonality) will persist. Common methods include moving averages, exponential smoothing, and ARIMA models. Driver-Based Forecasting (Causal Forecasting): Relies on known external or internal variables (e.g., population, promotions) to predict demand. Regression analysis is often used to model the relationship between the demand and the independent drivers.

Integrated Business Planning (IBP)	A holistic planning approach that aligns strategic, financial, and operational plans across the enterprise. IBP integrates demand, supply, and financial forecasts to support decision-making and performance tracking
Internal and external data	Internal data is proprietary information generated within an organization. External data is third-party or publicly available information that is often used to enrich context or validate insights.
New Product Introduction (NPI)	The process of launching new products into the market. It is critical for managing ramp-up and lifecycle transitions. In demand planning, this involves forecasting demand for new items using analogues, cannibalization impact, and placeholder items.
Orchestration agent persona	Coordinates across all agents to ensure cohesive decision-making and a feedback loop for continuous alignment.
Post-game analysis	A retrospective review of forecast accuracy and planning effectiveness. It involves comparing actual outcomes to planned forecasts, identifying root causes of deviations, and refining models.
Retrieval-Augmented Generation (RAG)	An AI technique that combines the retrieval of both structured and unstructured data to generate more accurate responses.
Signal Intelligence agent persona	Continuously monitors targeted channel(s) to detect specific signals or events in real-time.
Strategic Alignment Trainer persona	Re-trains by identifying gaps between planned strategy versus actual execution, driving continuous alignment and smarter future planning.
Structured and unstructured data	Structured data is information organized in predefined formats and easily searchable (e.g., tables and databases). Unstructured data is raw or loosely organized content lacking a consistent schema (e.g., emails and images).
Supply Chain Management (SCM)	The coordination of sourcing, production, and distribution activities to deliver products efficiently and cost-effectively. SCM integrates suppliers, manufacturers, warehouses, and retailers to optimize the flow of goods, information, and finances.

References

i *State of the Agentic AI Market Report, ISG, June 2025:*
https://isg-one.com/advisory/ai-advisory/state-of-the-agentic-ai-market-report-2025

ii *How SAP and Google Cloud Are Advancing Enterprise AI Through Open Agent Collaboration, Model Choice, and Multimodal Intelligence, SAP, April 2025:*
https://news.sap.com/2025/04/sap-google-cloud-enterprise-ai-open-agent-collaboration-model-choice-multimodal-intelligence/

iii *Joule Agents, SAP:*
https://www.sap.com/products/artificial-intelligence/ai-agents.html

iv *o9 Expands Its Collaboration With Microsoft to Advance Generative AI Capabilities in the o9 Digital Brain Planning Platform With Microsoft Azure OpenAI Service, o9 Solutions, Apr 2024:*
https://o9solutions.com/news/o9-expands-its-collaboration-with-microsoft-to-advance-generative-ai-capabilities-in-the-o9-digital-brain-planning-platform-with-microsoft-azure-openai-service/

v *o9 CEO Charts the Next Agentic AI Frontier in Enterprise Planning and Execution, o9 Solutions, June 2025:*
https://o9solutions.com/articles/o9-co-founder-and-ceo-charts-the-next-ai-driven-frontier-in-enterprise-planning-and-execution/

vi *Oracle AI Agents Help Transform Supply Chain Workflows, Oracle, January 2025:*
https://www.oracle.com/in/news/announcement/oracle-ai-agents-help-transform-supply-chain-workflows-2025-01-30/

vii *Kinaxis to Unveil the Next Phase of AI Innovation at Kinexions 2025, Kinaxis:*
https://www.kinaxis.com/en/news/press-releases/2025/kinaxis-unveil-next-phase-ai-innovation-kinexions-2025

Authors

Sujatha Babu

Dr. Sujatha Babu is the principal of the supply chain advisory practice. She has over 20 years of experience leading digital transformation initiatives for clients across various sectors, including Fortune 500 companies. She holds a PGDM from IIM Calcutta and a Ph.D. from IIT Madras and has published five peer-reviewed research papers, co-authored a book, and contributed to a book chapter.

Vijay Kumar

Vijay Kumar is a supply chain specialist with 9 years of experience in Supply Chain Management. He has overseen global implementations of o9 planning solutions for FMCG, manufacturing, and retail sectors. Currently, he is focusing on supply chain digital transformation and providing data-driven solutions to enhance accuracy, agility, and operational efficiency.