© May 2026 | IJIRT | Volume 12 Issue 12 | ISSN: 2349-6002

Conversational Analytics for Supply Chain Operations using Gen AI and Agents

Siddhartha Shanmugasundaram[1], Dr. D Deepika[2]
[1]*Department of Computational Intelligence School of Computing, College of Engineering and Technology SRM Institute of Science and Technology Kattankulathur, Chennai, India*
[2]*Department of Networking and Communications School of Computing, College of Engineering and Technology SRM Institute of Science and Technology Kattankulathur, Chennai, India*

***Abstract*—A supply chain ecosystem is a vast chain of interconnected links with so many moving parts and data changing all the time that getting timely answers is difficult. Pre-built aggregated views can help address this challenge to a fair degree, but it cannot generate insights on the fly based on a particular situation or simulation. This paper explores a conversational analytics framework that involves GenAI and agentic capabilities allowing Supply chain analysts to query the data through natural and plain language. The response is a structured one that explains the context of the data and provides additional insights. This is accomplished through a combination of agents which analyzes the incoming query (or) analytical workload and through delegation to each specific agent, the response is delivered. The following are the five task driven agents that will perform a series of activities. An intent agent that understands what the user is trying to ask. This will then invoke the context agent which will map the natural language to business specific knowledge. This is then fed to a query agent which generates the technical query to generate the response. The response is then validated through a rule based LLM that acts as a diagnostic agent and an insight agent generates further insights and follow-up questions to the previous chain of thought. This framework integrates live feeds, data from ERP system and external risk related information into Snowflake. Evaluation of this approach showed that there was approximately 60% reduction in lead time, 15-20% improvement in demand forecasting, savings of around 10% and task completion and follow up rose to almost 85%. A conversational framework versus traditional BI dashboards has transformed how analysts operate with data and have resulted in improvement in classification accuracy to 92% and response time to 45 seconds as opposed to 78% & 3 minutes respectively.**

***Index Terms*—Generative AI; Agentic Workflows; Conversational Analytics; Supply Chain Management; Large Language Models; Multi-Agent Systems; Retrieval-Augmented Generation; Snowflake**

I. INTRODUCTION

Supply chain management is a vast ecosystem that involves coordination between multiple business functions like Procurement, manufacturing, logistics and inventory management. Each of these functions generates its own structured and unstructured data. A supply chain analyst may have a question that could involve connecting multiple data sources and generating a response on the fly, but this access is not available to him [1] and hence the bottleneck causes a few hours or even days to respond to a situation or problem.

Conventional BI tools depend on analysts to intermediate, interpret and respond to situations and by the time the data arrives, critical time is lost that does not optimize operations. The response is late to the actual query or prompt [2] from the analyst.

GenAI and LLMs have completely changed this landscape. Systems can interpret natural language, apply reasoning through business context and generate relevant content. Furthermore, they can trigger actions or workflows where AI agents automatically trigger actions based on dynamic changing trends. Rather than having to go through complex dashboards that require a lot of training and navigation [3], conversational GenAI based analytics collaborate autonomously to complete tasks which do not require human intervention [4].

This paper discusses a conversational analytics framework built on the following principles and addresses three structural gaps:

(i) Accessibility gap—most BI tools demand SQL skills creating dependency with Engineering teams
(ii) Integration gap—siloed data that lacks the overall context
(iii) Responsiveness gap—static dashboards do not dynamically generate insights based on latest data

This paper has the following sections. Section II surveys related work. Section III details the system architecture and programmatic agent interactions. Section IV describes implementation, dataset, and evaluation methodology. Section V presents results with statistical analysis. Section VI concludes with limitations and future directions.

II. RELATED WORK

A. Generative AI in Supply Chain Management

Conventional analytics systems no longer offer the same operational efficiency in Supply chain management as GenAI. Wipro [3] has shown that GenAI-based demand planning analytics are better at understanding and responding to external signals and risk factors compared to traditional time forecasting models such as ARIMA and SARIMA. Furthermore, the ability of GenAI to explain forecasts in simple terms and in the current context improves end-user trust and adoption, particularly planners who must use the systems for daily operations. This has also been observed by Carbonneau et al. [21] particularly calling out barriers in adopting conventional models.

Lead-time improvement of up to 60% has been reported from GenAI based automated generation of declarations pertaining to Logistics. McKinsey has reported this, and it also aligns with the existing literature on logistics automation. A review of around 47 papers identifies issues a gap in evaluation frameworks for Supply chain systems that implement Conversational Analytics which is the main essence of this paper.

There are risks to over-reliance of black box models because of explainability issues. Hence this is a good case for hybrid architecture where ARIMA (or) SARIMA is the baseline and LLM based solutions contribute on top of it. This is called out by Carbonneau et al. [21] and Syntetos et al. [23].

B. Role of Agent-based Workflows for Enterprise Analytics

The architectural case for multi-agent specialization over single-model deployment is well-established in the agent literature. Park et al. [24] demonstrate through controlled experiments that specialized agent ensembles outperform single LLMs on complex multi-step reasoning tasks by margins of 12–18%, attributing the gain to reduced context-window saturation and cleaner separation of reasoning concerns. This finding directly motivates the five-agent decomposition in our framework.

Chen et al. [10] characterize effective human-AI analytical workflows in three stages: contextualization, goal formulation, and prompt articulation. Their empirical study of 84 knowledge workers found that systems supporting all three stages reduced task completion time by 34% relative to single-prompt interfaces. Agile Lab's retriever-router-analyzer pattern [11] provides a production-proven implementation of this staged approach for enterprise data environments, with documented deployments in financial services and manufacturing analytics.

Reinforcement learning has also played a significant role in the evolution of systems. Agents learn through prior failure in systems like Reflexion (Shinn et al. [25]). When a failure occurs, the root cause is stored in memory which is recalled thereby reducing repeat failures by approximately 23% mainly through feedback loops.

C. Supply Chain Functions & LLMs

Activities like Demand forecasting benefit significantly from running on LLM. This has resulted in MAPE improvement by 12% over conventional methods. Companies like Procter & Gamble too have implemented LLM demand sensing [12] that works well for different industry segments and product mix.

Advanced warning systems and anomaly detection have led to reduction in cost disruption by 10-12% for Supplier Risk management function (Ivanov and Dolgui [26]).

Additionally, context agent's knowledge graph backed architecture has outperformed conventional RAG in the case of multi-hop reasoning queries. (Jiang et al. [27])

III. SYSTEM ARCHITECTURE

The architecture has a multi-tier agentic framework enabled through Snowflake Cortex. The user interface is an iFrame embedded in Tableau. Alternatively, a Streamlit application service is also enabled in Snowflake to act as an Intelligent data assistant. Snowflake understands the intent and context through its agents and then builds the query based on the user input. The picture below shows the architecture and flow that happens inside Snowflake.

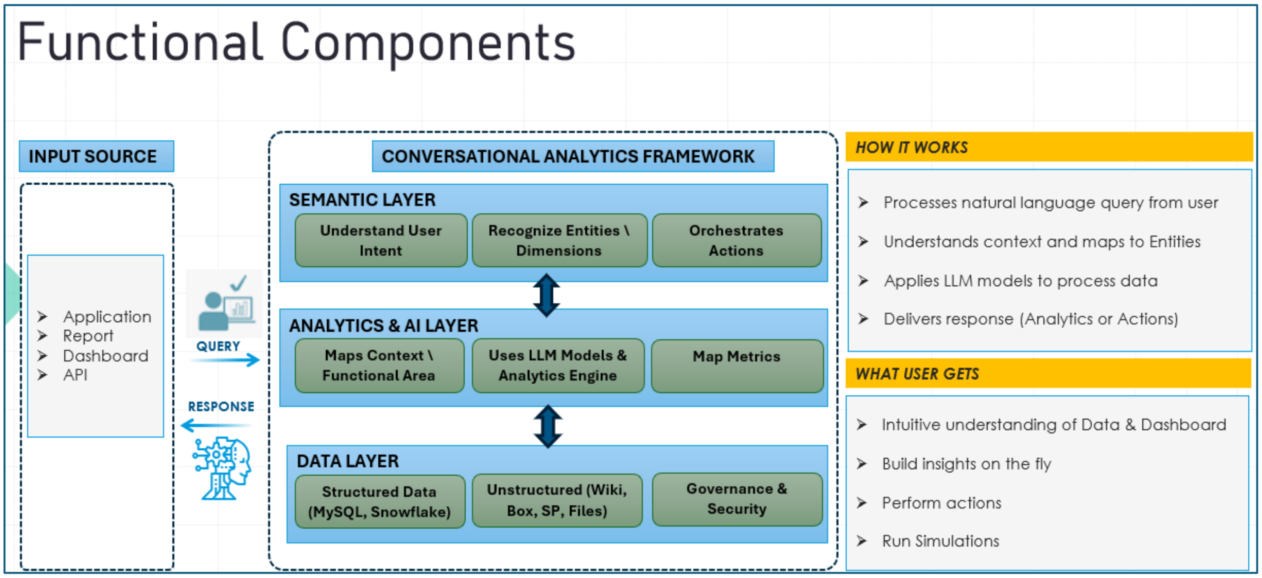

Fig. 1. Functional Components & Architecture

Figure 1. represents the user interaction flow and the functional architectural components of the proposed Conversational Analytics system. Below are details of each component.

Intent Agent. The Intent Agent in Snowflake runs a LoRA-fine-tuned instance of Claude-sonnet that's trained on a labeled corpus of 50,000 supply chain queries across five intent classes: demand forecasting, inventory optimization, supplier assessment, logistics planning, and risk analysis. Based on the confidence level, the agent redirects to the next step in the process.

Context Agent. This is one of the most critical agents that decides the correct response. Context Agent does retrieval of semantics for the domain so that the plain language query can be converted to Business meaning. The context is then serialized as structured JSON payload and passed to the next agent, which is the Query agent.

Query Agent. This agent is responsible for building the query based on the input and context. It uses a structured template and a mix of one-shot and few shot prompting to refine the query output. An example of a structured template is as follows:

SYSTEM: You are a supply chain analyst responsible for Inventory optimization. Generate only safe executable code. Please check for hallucination and avoid if the right context is not available. SCHEMA: <>, CONTEXT: <>, EXAMPLES: few-shot

Diagnostic Agent. Once the query is executed, validation happens in two stages: A rule-based check to match meta-data and ddl definitions and the completeness of the results along with the numerical validation. The second check validates the response with the intent. If all validations pass it goes to the next stage else an escalation is triggered back to the user on what was not validated.

Insight Agent. The Insight Agent formats the final response by factoring RAG, retrieved context and analytical outputs. A temperature setting of 0.3 is applied so that hallucination is reduced.

Data Integration Layer

The following data categories connect to Snowflake.

1) Internal systems like ERP that houses transactional data.
2) Data streams capture IoT data like RFID.

3) External risk factors like Weather, Geopolitical data
4) Supplier performance scorecard and key supplier metrics.

These data feeds are refreshed into Snowflake periodically (every 15 minutes or so) through APIs or data pooling mechanisms.

Snowflake has a semantic layer and schema which has around eight hundred business definitions, around 1200 columns to description mappings and relationship mappings. These help in resolving the incoming plain language query to technically correct structured queries.

D. Analyzer Pattern: Retriever & Router Pattern

Following are the execution steps involved in the process.

Step 1: The intent agent assembles an information package by classifying the query and data is retrieved from pre-materialized views. This is called Retrieval and contextualization.

Step 2: The query agent executes the structured code and validates the completeness of the data. The next step is Routing and Execution.

Step 3: The last step in when the Insights agent structures the output to a recommendation and validates the coherence of the output (response vs. intent). This is called Analysis and Insight generation. This step also produces the next set of related questions pertaining to the previous context (or) chain of thought.

IV. METHODOLOGY

E. Experimental Setup and Dataset

The evaluation used a dataset from a multinational manufacturing company that is established across various regions of the world. The dataset covers supply chain operations and master data associated with locations, product categories, and suppliers. Characteristics of key datasets are mentioned in the tables below. Table 1 represents the Dimension tables and Table 2 represents the fact tables.

TABLE 1: DIMENSION TABLES OF DATA MODEL

Table	Rows	Purpose
DIM_DATE	1,197	Calendar & fiscal date hierarchy (year, quarter, month, week, fiscal y
DIM_PRODUCT	50	Product master — name, category/sub-category, HS code, unit cos
DIM_LOCATION	20	Warehouses & facilities — name, type, country, region, state, lat/lon
DIM_SUPPLIER	30	Supplier master — name, country, region, risk tier, reliability score

TABLE 2: FACT TABLES OF DATA MODEL

Table	Rows	Purpose
FACT_DEMAND	50,000	Demand actuals vs. forecast by product, location, channel,
FACT_INVENTORY	40,000	Daily inventory snapshots by product, location, supplier, BU
FACT_DISRUPTION_RISK	5,000	Disruption events — natural disasters, severity, geo-impact,
FACT_TRADE_EXPOSURE	20,000	Import/export flows — trade value, HS codes, concentratio

The Data model is represented below in Fig 2. Brief description of the same is as follows.

1) Data Layer: Star schema with 4 Fact & 4 Dimension tables covering the key functions of Supply Chain namely Demand, Inventory, Risk & Trade Exposure.
2) Semantic Layer: The semantic view defines 15 metrics & 30 dimensions, 8 relationships. This enables natural language querying without users having to know the structure of the schema.
3) Application Layer will be either a Tableau Dashboard (or) a Streamlit Chat application.

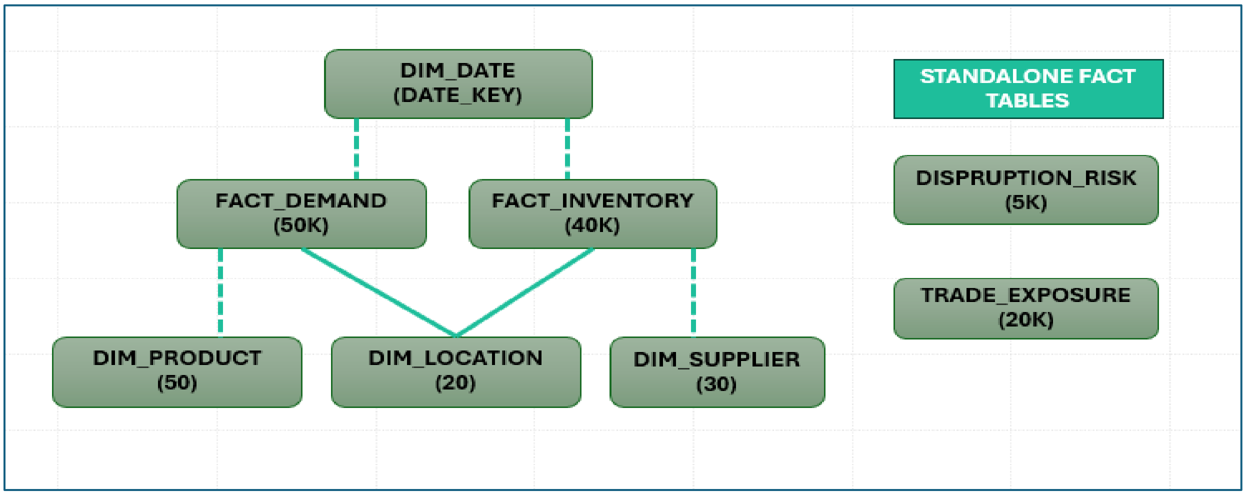

Fig. 2. Data Model

F. Baseline Models

The following baseline systems were evaluated under identical experimental conditions:

Conventional BI (Tableau): Pre-built tableau dashboards connected to specific views in Snowflake. Typically, users interact through dashboard controls, filters and pages. There is no interactivity with the user.

Conversational BI (Tableau with user interface): An analytical ecosystem that has a user interface that is connected to LLM for dynamic data insights and generation.

G. Evaluation Metrics and Statistical Validation

The following dimensions were assessed for performance each with defined metrics and statistical tests:

Accuracy: Intent classification & query generation correctness

Efficiency: Query response time was captured for simple and multi-hop questions.

User experience: End-user experience was captured by gathering feedback through demos.

Business impact: KPI improvements were tracked to measure business impact.

H. Prompt Engineering Design

Prompt templates were followed to enable Role based context and constraints. It was specifically instructed not to hallucinate if information is not found.

I. Phased Deployment

Sprint 1 (Weeks 1–4): Data and Schema preparation, defining baseline.

Phase 2 (Weeks 5–8): Tableau Dashboard development, building semantic views in Snowflake and unit testing

Phase 3 (Weeks 9–12): Integration and rollout: Streamlit deployment, Tableau – LLM integration and benchmarking.

V. RESULTS AND DISCUSSION

J. Qualitative Performance

Table 3 represents the various use-cases when it comes to Analytics scenarios and the quick comparison of the implementation time of Conventional (Tableau) vs. Conversational (Streamlit Chat). The key winner here is the ability to do Dynamic data insights, What-if simulations and root cause analysis where Conventional Analytics are found wanting. Follow-up questions based on Insights is another positive evolution of Conversational Analytics.

TABLE 3: ANALYTICS SCENARIOS & COMPARISON TO BUILD

Capability	Tableau Das hboard	Streamlit Chat
Pre-built KPIs	Yes	Yes
Ad-hoc cross-table joins	No (uses flat tables)	Yes (queries r aw facts)
Natural language questions	No	Yes
What-if scenario analys is	No	Yes
Root cause reaso ning	No	Yes
Dynamic follow-up questions	No	Yes

Table 4 below gives a quick glimpse of time taken to build in Conventional vs. Conversational Analytics. Through the proper definition and refinement of Semantic views, the turn-around time to complete various query types is much faster using GenAI based frameworks. Task completion based on insights from LLM showed significant improvement.

TABLE 4: DATA INSIGHTS BUILD TIME

Query Type	B1: Tableau (Conventional)	B2: Conversational
Simple data lookup	2–5 min	45 sec
Complex multi-source analysis	2–4 hours	3 min
What-if scenario modeling	1–2 days	6 min
Root-cause analysis	3–5 days	9 min

Fig 3. shows a clear comparison where the agentic system reduces response time by more than 90% than Traditional BI systems. The comparison is shown for the most probable Query types in the Analytics space.

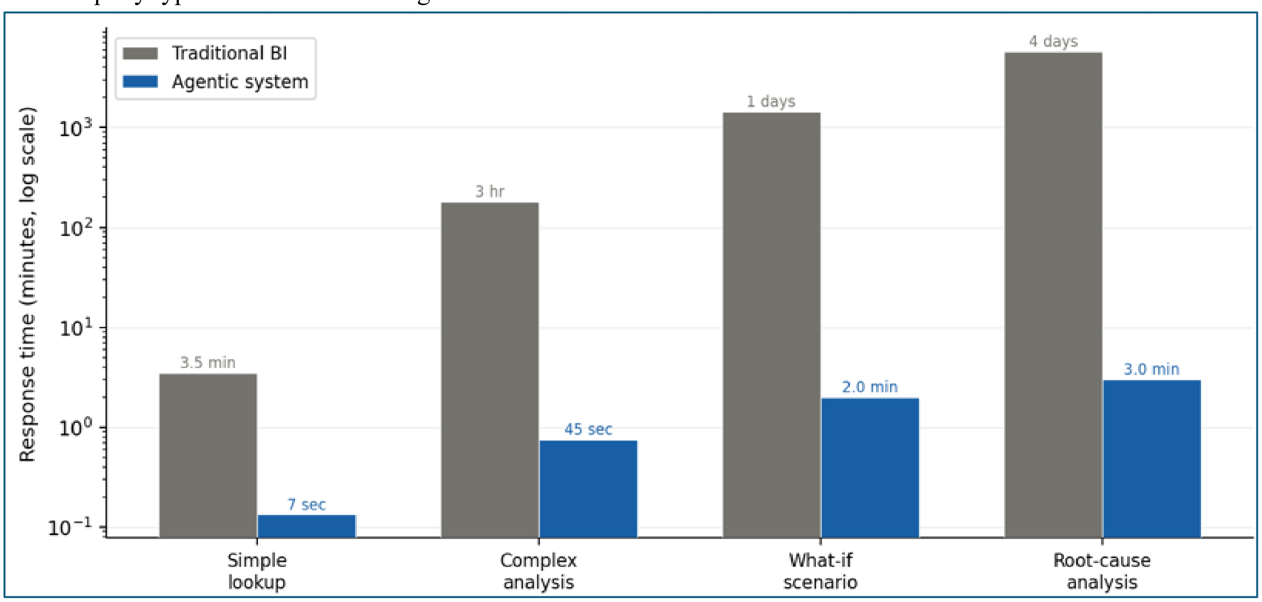

Fig. 3. Query Response Time Comparison

K. User Experience Findings

Fig 4 represents the most common roles in Supply chain domain. The System usability scale (SUS) is a representation of how easy and convenient it is to derive meaning out of the information presented and how simplicity in navigation of the information. Conversational Analytics clearly showed a significant jump in SUS benchmark. Supply chain planners recorded the highest ratings from a user-experience standpoint proving that natural language querying significantly improved end-user experience.

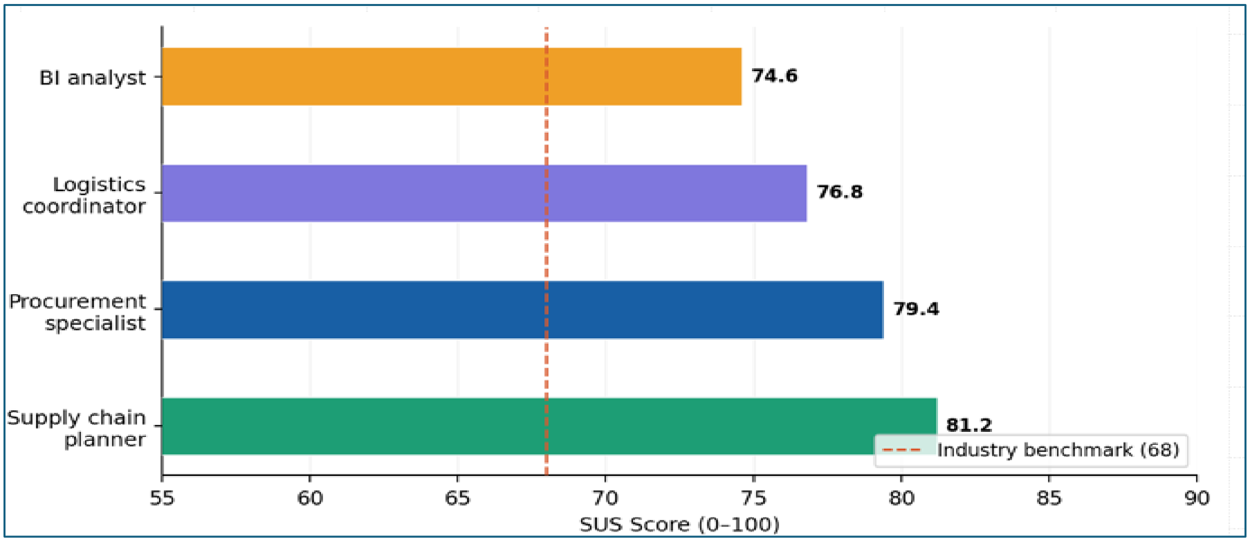

Fig. 4. System Usability Scale (SUS) scores by User Role

L. Benchmarking and Comparative Analysis

Following parameters showed significant improvement from end-user experience as well as quality of data output. Furthermore, the dynamic data insights part significantly improved self-service analytics, reducing dependency on Engineering teams to keep modifying the dashboards. Table 5 below is the summary of the performance & accuracy indicators, and it is evident that the proposed system outperforms the conventional analytics approach.

TABLE 5: PERFORMANCE & ACCURACY INDICATORS

Dimension	Conventional	Conversational
Intent accuracy	0.41	0.78
Data quality	3.1	3.8
Dynamic Data insights	None	Partial
Multi-source synthesis	Manual	Limited
Hallucination rate	N/A	8.4%
Task completion rate	65%	74%

M. Limitations

Context building takes time. Also, LLMs can be expensive if the user payloads are more. There is significant effort involved in Context setting, prompt building and validation of the output with respect to conventional systems. Hallucination is the biggest risk and to limit the same, it is critical to keep the context relevant.

VI. CONCLUSION

This paper presented a comparison of Conventional as opposed to Conversational Analytics framework which was built by integrating BI-layer to LLM based systems that can also perform Agentic actions. By defining the appropriate context, metadata definition, the analytical capabilities can be significantly increased. Following are the key takeaways.

- Effectiveness of GenAI depends largely on Semantic & Business definition.
- Implementation on Dashboards can be spun off in parallel.
- Linking the semantic definition & underlying data can enable Cross functional GenAI & Data Insights
- Plug into Application (that generates data for Dashboards)
- Knowledge engine (Docs & Processes) can be incorporated.
- Integrated environment to monitor KOIs & Trigger actions.

Future Analytics space will undergo a major revamp where the user will not have to go through dashboards but a summary of KPIs would be presented intuitively and the user can ask more questions to get a deeper understanding of the data and its context. Unlike conventional dashboards where the interaction is read-only, agentic actions can be enabled where users can initiate transactions through Analytical systems.

ACKNOWLEDGEMENT

I would like to thank my guide Dr. D Deepika, Assistant Professor, for her timely and valuable guidance in helping shape this project and paper. Her critical insights have gone a long way in establishing this framework.

REFERENCES

[1] E. Hofmann and E. Rüsch, “Industry 4.0 and the current status as well as future prospects on logistics,” Computers in Industry, vol. 89, pp. 23–34, 2017.

[2] Lumi AI, “Agentic workflows in analytics: The next era of enterprise analytics,” Nov. 2025. [Online]. Available: https://www.lumi-ai.com/ai-glossary/agentic-workflows-in-analytics

[3] Wipro, “How GenAI improves supply chain management,” Dec. 2023. [Online]. Available: https://www.wipro.com/retail/articles/how-genai-improves-supply-chain-management/

[4] IBM, “What are agentic workflows?” Mar. 2025. [Online]. Available: https://www.ibm.com/think/topics/agentic-workflows

[5] McKinsey & Company, “Beyond automation: How gen AI is reshaping supply chains,” Apr. 2025. [Online]. Available: https://www.mckinsey.com/capabilities/operations/our-insights/beyond-automation-how-gen-ai-is-reshaping-supply-chains

[6] EY, “How supply chains benefit from using generative AI,” Jul. 2024. [Online]. Available: https://www.ey.com/en_gl/insights/supply-chain/how-generative-ai-in-supply-chain-can-drive-value

[7] L. Zhang et al., “The potential of large language models in supply chain management,” arXiv preprint arXiv:2501.15411, Jan. 2025.

[8] Y. Chen, M. Wang, and R. Kumar, “Agentic workflows for conversational human-AI interaction,” arXiv preprint arXiv:2501.18002, Jan. 2024.

[9] AgileLab, “Building conversational analytics for enterprises: AI agents that make data speak,” Jun. 2025. [Online]. Available: https://www.agilelab.it/blog/building-conversational-analytics-for-enterprises

[10] Y. Chen, S. Park, and J. Kim, “Design patterns for agentic AI systems in enterprise analytics,” Proc. ACM Conf. Human Factors in Computing Systems, pp. 1–15, 2024.

[11] AgileLab, “Enterprise conversational data access: The retriever-router-analyzer pattern,” Technical Report, Jun. 2025.

[12] J. Raper, “The role of large language models (LLMs) in supply chain management,” LinkedIn Pulse, Sep. 2024. [Online]. Available: https://www.linkedin.com/pulse/role-large-language-models-llms-supply-chain-management-jason-raper

[13] Supply Chain Brain, “How large language models are transforming supply chain management,” Apr. 2025. [Online]. Available: https://www.supplychainbrain.com/blogs/1-think-tank/post/41602

[14] Datup, “Generative AI in the supply chain: How to use in 2026,” Feb. 2026. [Online]. Available: https://datup.ai/en/blog/generative-ai-in-supply-chain

[15] ABBYY, “6 AI trends reshaping supply chains in 2026,” Supply & Demand Chain Executive, Feb. 2026.

[16] IBM, “What is conversational analytics?” Aug. 2023. [Online]. Available: https://www.ibm.com/think/topics/conversational-analytics

[17] Nextiva, “What is conversation analytics? A best-practice guide for 2026,” Dec. 2025. [Online]. Available: https://www.nextiva.com/blog/conversation-analytics.html

[18] Couchbase, “What is conversational analytics? Plus examples and tools,” Sep. 2025. [Online]. Available: https://www.couchbase.com/blog/conversational-analytics/

[19] AIMultiple, “Top 16 supply chain AI use cases with examples,” Jan. 2026. [Online]. Available: https://research.aimultiple.com/supply-chain-ai/

[20] Coursera, “GenAI for supply chain optimization,” Course Materials, Mar. 2025. [Online]. Available:

https://www.coursera.org/learn/genai-for-supply-chain-optimization
[21] R. Carbonneau, K. Laframboise, and R. Vahidov, "Application of machine learning techniques to forecasting in supply chain management: A systematic review," Expert Syst. Appl., vol. 34, no. 6, pp. 3336–3341, 2008.
[22] M. Attaran, "Artificial intelligence and its applications in industry," Advances in Artificial Intelligence, vol. 2023, Art. no. 4504754, 2023.
[23] A. A. Syntetos, M. Z. Babai, J. E. Boylan, S. Kolassa, and K. Nikolopoulos, "Some key directions for research in forecasting," Int. J. Forecasting, vol. 32, no. 1, pp. 3–11, 2016.
[24] J. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in Proc. 36th ACM Symp. User Interface Software and Technology (UIST), 2023, pp. 1–22.
[25] N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan, and S. Yao, "Reflexion: Language agents with verbal reinforcement learning," in Advances in Neural Information Processing Systems (NeurIPS), vol. 36, 2023.
[26] D. Ivanov and A. Dolgui, "A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0," Production Planning & Control, vol. 32, no. 9, pp. 775–788, 2021.
[27] Z. Jiang, F. F. Xu, L. Gao, Z. Sun, Q. Liu, J. Dwivedi-Yu, Y. Yang, J. Callan, and G. Neubig, "Active retrieval augmented generation," in Proc. 2023 Conf. Empirical Methods in Natural Language Processing (EMNLP), 2023, pp. 7969–7992.