CAST AI: AI-Powered Inventory Forecasting and Strategic Decision Support for SMEs

Prepared by: Abdullah Moghal, Imthiaz Faizal, Zayd Alnachef
Course: INSY-4321-001
Date: April 28, 2025

Problem Context and Purpose

Small and medium-sized enterprises (SMEs) face inventory management as one of their most important operational difficulties. Many SMEs face problems with demand forecasting and inventory management because they do not have the advanced tools that larger corporations possess. The challenges result in negative inventory decisions that cause overstocking, stockouts, and missed sales opportunities.

This project aims to solve the identified challenges through predictive analytics, together with generative AI, to develop a system that predicts inventory demand while explaining its predictions in a clear and actionable way. Through this method, data-driven recommendations become accessible and easily understood by managers, which enables SMEs to make confident decisions.

Problem Identification & Business Impact

Challenges Faced by SMEs:

- Overstocking: Excess inventory requires capital investment that increases storage costs while simultaneously risking product waste.
- Understocking: Results in lost sales, together with customer dissatisfaction.
- Lack of Advanced Forecasting Tools: SMEs usually do not possess advanced data analytical tools and AI-based forecasting systems.
- Difficulty Interpreting Model Outputs: The technical nature of data models prevents managers from effectively understanding and using the information obtained.

Business Impact of These Challenges:

- Financial losses due to poor inventory allocation.
- Increased waste from overstocked products.
- Stockouts result in unhappy customers who leave the business, which damages loyalty.
- Operational inefficiencies and reactive decision-making rather than proactive planning.

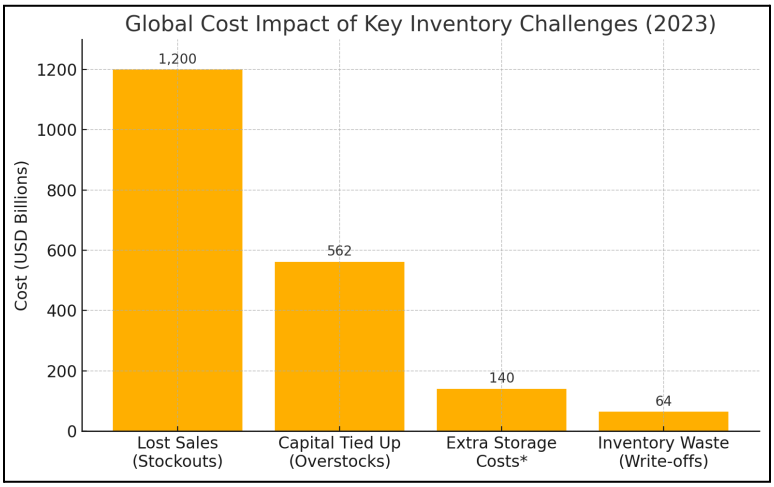

Figure 1. Out-of-stocks ($1.2 T) & overstocks ($562 B): IHL Group, *Retail Inventory Distortion Study* 2023. IHL Group
Holding-cost drag (20-30 % of inventory value): Netstock blog, "Inventory Holding Costs," 2024. Netstock

Inventory waste ($163 B lost annually): Avery Dennison *Integrated Sustainability & Annual Report* 2022 ("The Missing Billions" study). Avery Dennison

As Figure 1 shows, stock-outs alone wiped out roughly $1.2 trillion in 2023, while overstocks locked up another $562 billion. Together, these four cost buckets represent over $2 trillion of avoidable loss, which is exactly the gap our AI system targets. These issues make it clear that a solution needs to deliver both precise demand forecasting together with simple explanations for its outputs.

Solution Overview and Functionalities Chosen

The system we designed uses AI to power inventory forecasting from beginning to end while uniting predictive models with generative AI capabilities to specifically tackle business challenges. The chosen functionalities were chosen based on their capability to balance feasibility with innovation and their ability to generate tangible business value.

Selected Functionalities:

I. Predictive AI for Inventory Forecasting (Narrow AI Component):

- The model predicts product demand through the analysis of sales records combined with weather information, holiday schedules, and periodic patterns.
- Provides the core logic for inventory planning.

II. Generative AI Integration (Core Enhancement):

- Forecast Explanation & Recommendation Generation:
- The system transforms forecasted data into basic descriptions that explain the basis behind inventory recommendations.

Automated Reports & Email Drafts:

- The system provides procurement recommendations in a clear and manager-ready format that saves time while improving consistency.

Conversational Q&A Decision Support (Chatbot):

- Users can ask questions such as "Why is Product X prioritized?" or "What if we reduce the inventory of Product Y?" to which the system will respond in simple language that users can understand easily.

Scenario Simulation via Prompting:

- The system allows users to examine different situations in order to determine possible risks and results prior to procurement choices.

Product Categorization & Trend Detection via Text Analysis:

- The system generates predictions about emerging trends and product classifications through its text analysis capabilities of product descriptions, along with customer reviews and supplier notes to enhance forecasting precision.

Justification for Functionality Selection

The selected functionalities were selected to ensure the system delivers accurate forecasting along with making those forecasts understandable, actionable, and trustworthy for decision-makers. The implementation of generative AI connects sophisticated data science results with real-world business choices to fulfill the mission of enabling SMEs through sophisticated yet user-friendly AI solutions.

The combination of technical feasibility with creativity and business value enables the solution to remain innovative while keeping it feasible for project requirements to provide SMEs with inventory decision-making tools.

AI Solution and Prototype Development

Technical Approach

Our AI-powered inventory forecasting and decision support system is designed as a hybrid solution that combines the strengths of both narrow AI (predictive modeling) and generative AI (language models). The innovation of this solution lies not only in its ability to forecast inventory demand accurately but also in its capacity to transform complex numerical outputs into human-friendly, actionable insights. This feature is particularly vital for small and medium-sized enterprises (SMEs), where decision-makers may not always have the technical background to interpret traditional data science models.

This section explains the technical foundation of the system, the rationale behind the selection of specific AI models, the configuration parameters applied, and the iterative refinement process used to optimize the solution for real-world business use.

A. Predictive Modeling (Narrow AI Component)

I. Sales History Analysis

The forecasting process begins with a thorough analysis of historical sales data across multiple product categories. Understanding past purchasing behaviors is critical for identifying recurring trends and anticipating future demand. To ensure the reliability of the data, an initial cleaning phase is conducted where duplicate entries, missing values, and outliers are removed. The forecasting models receive protection against unusual spikes or dips through interquartile range analysis and z-score standardization during outlier detection.

After cleaning, the data receives seasonal decomposition analysis. Through this analysis, three essential components emerge from the sales data: trend, seasonality, and residuals. The system enables forecasts of different time intervals based on product categories and turnover rates, allowing daily, weekly, and monthly projections. The system provides adaptability to meet different operational needs and business models.

II. External Data Integration

The foundation of historical sales data remains strong, but demand forecasting accuracy requires integration with external variables that shape purchasing behavior. The system brings in three fundamental types of external data sources:

- **Weather Forecasts:** Winter apparel and rain gear, and outdoor equipment show purchasing demand patterns that correlate with weather conditions.
- **Holiday Calendars:** Consumer purchasing behaviors often spike around holidays like Thanksgiving, Black Friday, Christmas, or back-to-school seasons.
- **Seasonal Economic Trends:** Broader economic factors and seasonal cycles (such as tax refund season or vacation periods).

Generative AI Integration

GPT-4o enables our inventory forecasting and decision support system to generate actionable insights through natural language generation, which translates technical data into usable findings. The solution resolves an industry-wide issue that SMEs face regarding how to connect sophisticated analytics with everyday business operations. Generative AI simplifies data interpretation, which produces clear and credible recommendations that non-technical stakeholders can use effectively.

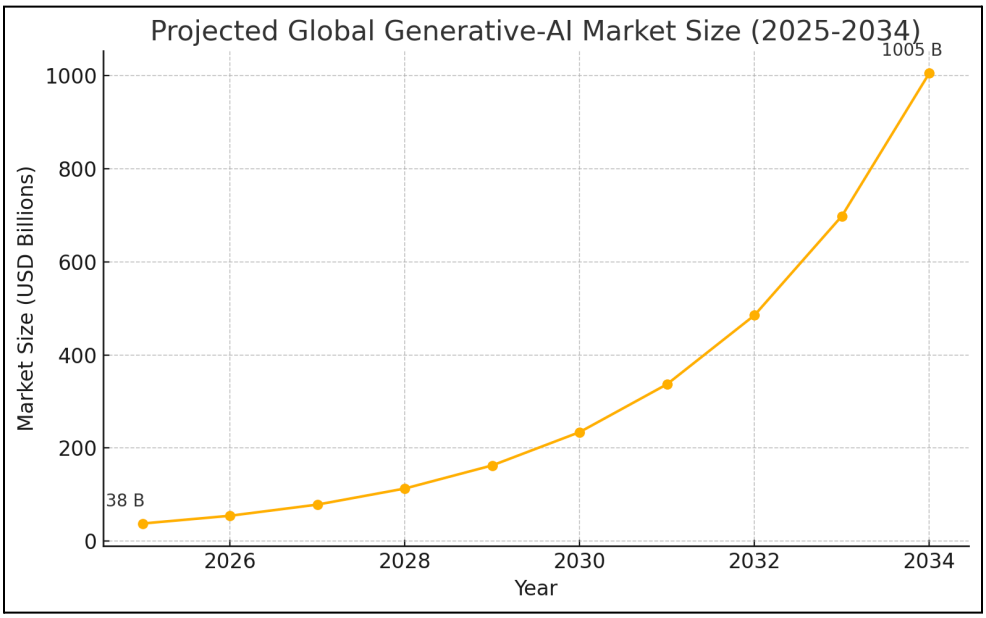

Figure 2. Projected global generative-AI market growth (2025-2034)
Precedence Research, *Generative AI Market Size to Hit USD 1,005 Bn by 2034* (published Jan 2025). Precedence Research

Generative AI continues to gain momentum in all business sectors at an accelerated pace. According to Precedence Research, the global generative AI market is projected to grow from USD 37.89 billion in 2025 to approximately USD 1,005.07 billion by 2034, reflecting a compound annual growth rate (CAGR) of 44.2% (Precedence Research Report). The significant expansion indicates why businesses should include generative AI within their modern solutions.

I. Forecast Explanation & Recommendation Generation

Through this function, the system transforms analytics output into straightforward, business-friendly reports that help procurement managers make decisions based on simple insights. Through key influencers like seasonal trends and holiday effects, and weather patterns, the system explains inventory recommendation logic for better decision confidence.

II. Automated Reports & Email Drafts

Through automated generation, the system creates procurement reports and draft communications, which decreases manual work and provides standardized output. The reports integrate forecast results with recommended inventory adjustments, together with their supporting reasoning, to make internal communication more efficient.

Example Draft:
Our research shows that Product X demand will rise by 30% before the holiday period. The recommended stock increase should prevent any potential stockouts.

III. Conversational Q&A Chatbot

Through its GPT-4o-powered chatbot, users gain interactive access to forecasting system functionalities. The system generates data-based responses to managerial questions about forecasts and recommendations, and their reasoning. The tool enables real-time decision making through its ability to answer queries about Product Y restocking priorities.

IV. Scenario Simulation via Prompting

Decision-makers gain control over hypothetical scenario analysis through "what-if" question functionality, which includes stock adjustment and promotional strategy evaluation. The system presents simulated outcome predictions together with strategic recommendation solutions for proactive risk evaluation and planning purposes.

V. Product Categorization & Trend Detection via Text Analysis

The system offers improved inventory planning through qualitative analysis, which extends beyond basic numeric forecasting methods. The system assesses product descriptions and supplier feedback and customer reviews to group products into "Fast-Moving," "Seasonal," "At-Risk of Overstock", and "Low Demand" categories. The system marks Product C as "At Risk" when multiple reviews demonstrate recurring problems with this item.

The generative AI component enhances the forecasting system through its implemented functionalities, which deliver user-friendly insights for improved operational outcomes and strategic decision-making.

Model Configuration and Prompt Engineering

The success of the generative AI component in our inventory forecasting system relies on well-planned model configuration and efficient prompt engineering. These two elements guarantee that the AI creates dependable and uniform, and user-friendly outputs that satisfy business requirements.

Model Configuration

This system uses GPT-4o (OpenAI) language model because it delivers top-level reasoning abilities and excellent natural language generation. The system implemented these configurations for maximizing model performance across all solution use cases:

Model Selection: GPT-4o because it provides multi-modal functionality along with fast response generation and improved reasoning over GPT-3.5.

Temperature:
For structured fact-based output such as forecast explanations and procurement reports, the temperature is set at 0.3 to generate consistent and deterministic responses.
The model generates diverse yet coherent responses during creative tasks at temperature settings ranging from 0.6 to 0.7.

Top-p (nucleus sampling):
0.9 to maintain quality output while allowing some variability to avoid overly repetitive language.

Max Tokens:
The system adjusts its token maximum according to query type while providing longer limits for scenario simulations and summaries to prevent response truncation.

Frequency Penalty:
Used at a light intensity (0.2–0.3) to prevent word repetition in report and email content.

Presence Penalty:
A moderate setting of 0.4 for Presence Penalty helps the model introduce fresh points when needed, primarily in open-ended Q&A and brainstorming dialogues.

User testing led to an iterative process of model configuration refinement, which relied on feedback about response clarity and business-focused content, and completeness.

Prompt Engineering Techniques

Business objectives alignment, together with non-technical user understanding, depended heavily on prompt engineering techniques for the generative AI system. The following strategies were implemented:

I. System Role Definition

The implementation of a standardized system created a clear inventory management advisor persona for the AI system. This role guaranteed that every response remained suitable for business use and offered professional language while delivering specific, actionable information.

Example System Prompt:
"You are an inventory management advisor. The main function of your position requires you to help small businesses explain their demand forecasts and make procurement recommendations, as well as address inventory questions through straightforward language.

II. Task-Specific Prompt Templates

Each functionality within the system was supported by custom-designed prompt templates:

Forecast Explanation:
"Analyze the projected demand of Product X throughout the upcoming quarter and provide an explanation of expected changes in market demand. Show the anticipated increase or decline in demand along with references to holidays and seasonal patterns, and historical sales data."

Scenario Simulation:
"Analyze what would happen if Product Y inventory decreased by 15% throughout the next quarter. The evaluation presents potential risks while providing action suggestions through historical sales information and seasonal patterns."

Chatbot Q&A:
"Answer this procurement manager's question about Product Z deprioritization for this month while using data-based explanations and basic terminology."

Trend Detection and Product Categorization:
"Sort these products based on their descriptions and customer reviews into Fast-Moving, Seasonal, At-Risk of Overstock, and Low Demand categories. Each category requires a short explanation that justifies the selection."

III. Multi-Turn Conversation Management

The chatbot conversation maintains context through follow-up prompts that include previous dialogue information. The chatbot developed a memory of the conversation structure through this method, which produced better user experiences during Q&A sessions.

IV. Iterative Refinement and Prompt Testing

The development process for prompts involved multiple evaluation cycles and improvement phases. The assessment of initial outputs evaluated these main factors:

- Clarity and readability.
- The accuracy and pertinence of the information presented.
- The AI should follow its advisory function according to these standards.
- User-friendliness and engagement.

Impact of Configuration and Prompt Engineering

The GPT-4o model achieved enhanced performance and reliability through careful configuration and effective prompt engineering strategies for the inventory forecasting and decision support system. The combination of these two methods was crucial to produce AI outputs that both delivered precise results and maintained a clear understanding and appropriate context, and direct alignment with SME decision-maker requirements.

The model configuration enabled the system to achieve both precise accuracy and flexible creative output. The system needed factual precision for structured procurement reports but allowed variable and diverse expressions for creative scenario simulations. The system's business advisory function received prompt engineering that controlled user-system interactions to maintain consistent responses.

Effectiveness Achieved Through Model Configuration and Prompt Engineering

Accurate and Understandable Forecast Explanations: The AI system produced clear and understandable forecast explanations that explained both demand changes and inventory suggestions.

Professional and Consistent Report Generation: The automated system produced formal reports and email drafts that required no manual editing for their structure or tone.

Interactive Engagement via Chatbot: Through the chatbot interface, managers gained quick access to data-based answers which improved their ability to make timely informed decisions.

Risk-Aware Scenario Simulations: The system generated authentic projections of different "what-if" scenarios which helped organizations implement proactive inventory management and strategic planning.

Effective Trend Detection and Product Categorization: GPT-4o's language abilities enabled precise qualitative data analysis which both detected new problems and properly sorted products.

The well-planned configuration and prompt engineering approach ensured that the solution went beyond technical outputs and fostered user trust, system adoption, and practical application in real business environments.

- **Model:** GPT-4o
- **Temperature:** 0.3 for structured outputs, 0.6–0.7 for creative tasks.
- **Top-p:** 0.9

Prompt Engineering Techniques

The decision-making requirements of non-technical business users heavily depend on prompt engineering, which ensures accurate and usable generative AI outputs. System instructions must be carefully crafted, and task-specific prompts designed while obtaining feedback to ensure consistency and clarity, and effectiveness.

I. Defining the AI Role as an Inventory Management Advisor

The system receives direct instructions to operate as an inventory management advisor, so it knows how to communicate properly. The specific guidance for the AI's role ensures all generated outputs remain both professional and easy to understand for users who do not speak

technical language. Through clear, non-technical language, the AI delivers meaningful data-based information that helps procurement teams make confident decisions.

Example System Prompt:

“Small and medium-sized enterprises can rely on you as their inventory management advisor for professional guidance. Your task is to deliver straightforward demand forecast explanations together with recommended procurement steps and inventory-related answers that maintain a professional tone for non-technical business users.”

II. Development of Task-Specific Prompt Templates

Each system feature receives its own set of prompt templates, which offer specific instructions for forecast explanations, scenario simulations, trend detection, and chatbot Q&A. The templates reduce ambiguity while generating consistent outputs.

- **Forecast Explanation Prompt:**
  "Show the simplified demand forecast explanation for Product X while supporting your recommendation through seasonal patterns and historical sales information."
- **Scenario Simulation Prompt:**
  “Simulate how a 20% reduction of Product Y inventory would affect the company throughout the next month. Analyze past demand patterns to develop alternative actions while describing potential risks that may occur."
- **Chatbot Q&A Prompt:**
  "Please respond to this question from a procurement manager: ‘Why is Product Z deprioritized this month?’ Explain the decision through data evidence in a direct and understandable manner."
- **Trend Detection and Categorization Prompt:**
  "Analyze the product descriptions and customer reviews provided. Use your assessment to divide products into Fast-Moving, Seasonal, At-Risk of Overstock, or Low Demand categories, followed by explanations for your decisions."

The templates maintain the AI's output to stay relevant while delivering insightful responses that match the system's tone and operational purpose.

III. Iterative Refinement Through User Feedback Loops

The system uses tested and improved prompt designs to achieve better output clarity and accuracy through structured cycles of development. This iterative refinement involves:

- **Initial Testing:** The testing phase of prompt templates involves running multiple scenarios to assess readability, together with tone and business alignment.
- **User Feedback Collection:** The system collects user feedback from businesses as well as SME stakeholders and procurement managers to evaluate the clarity and usefulness and trustworthiness of AI-generated content.
- **Prompt Optimization:** Changes are implemented to both prompt structure and phrasing based on feedback which prioritizes output clarity and actionability in all responses.

- **Multi-Turn Conversation Checks:** The prompts receive improvements to maintain context retention during chatbot interactions, which allows the system to handle complex queries through multiple exchanges.

Introduction to the Iterative Refinement Approach

The system achieves better performance through ongoing evaluation and refinement, which leads to delivering natural, effective interactions that assist real-world inventory decision-making.

The main objective of developing the AI-Powered Inventory and Forecasting System for SMEs involved creating forecasts that were both accurate and accessible to non-technical business users. A structured iterative refinement process became essential to reach this goal because system prompts and AI parameters, and outputs needed continuous testing and evaluation for systematic feedback-based improvements. The following section demonstrates the actual methods together with specific examples and materials that show the development of our AI solution across multiple cycles.

System Prompts and Role Definition

The effectiveness of any generative AI system starts with defining its role and purpose clearly, especially when used in business applications. The system generated generic or incomplete outputs during the early stages because users provided unclear instructions to the system. The AI-generated inventory recommendations without providing enough context or explanation became a major problem during this phase.

We created a specific system prompt that defined the AI's function as an inventory management advisor to solve this issue. The role-based approach directed the AI to produce responses that supported business objectives through straightforward communication and organized recommendations while using non-technical language.

Use-Case	User Prompt	V1 System Prompt (Improved)	Final System Prompt (Refined)
Forecast Explanation	Explain the demand forecast for Product X over the next quarter. Describe the key drivers (seasonality, holidays, weather, historical trend) and end with specific procurement actions in plain English.	You are a business analyst who explains inventory forecasts to operations managers. Summarize the main demand drivers in everyday language and finish with 2-3 procurement suggestions.	You are an inventory-forecast advisor for SME retailers. When asked to explain a forecast, produce **three sections**: 1 *One-paragraph overview* of expected demand change. 2 *Bullet list* of drivers (seasonality, holiday calendar, weather outlook, historical trend) ➜ cite key figures or % changes. 3 *Procurement actions* with quantities or reorder timing. Keep wording

			non-technical, ≤180 words total.
Scenario Simulation	Simulate how a 20 % reduction of Product Y inventory over the next month will affect stock-outs, holding cost, and revenue. Use past demand patterns for context and suggest alternative actions while describing potential risks.	You are an inventory-planning assistant. Run a quick simulation of the requested scenario, mention likely effects on stock-outs, holding cost, revenue, and give at least one mitigation idea.	You are an inventory-scenario analyst for SME supply chains. For any "what-if" request, output **four blocks**: 1 *Metrics Table* (Baseline vs 20 % cut): stock-out %, holding-cost $, revenue $. 2 *Risk Analysis* (3 bullets). 3 *Alternative Actions* (2–3 options, e.g., staggered replenishment). 4 *Recommendation* (max 2 sentences). Use plain English, no jargon.
Chatbot Q&A	A procurement manager asks: "Why is Product Z deprioritized this month?" Explain the decision using data evidence in plain English.	You are an AI that answers procurement managers' inventory questions with concise data-based explanations and a brief suggestion.	You are an inventory-management advisor for SMEs. Reply in **two parts**: • *Data Evidence* – 3 short bullets (e.g., "Forecast ↓12 % vs avg", "Gross margin 5 pp below target", "Supplier delay 8 days"). • *Recommendation* – one sentence. Keep total length ≤120 words, avoid jargon.
Trend Detection & Categorization	Here are product descriptions and customer reviews. Categorize each product as Fast-Moving, Seasonal, At-Risk of Overstock, or Low Demand, and give a one-sentence justification for every assignment.	You are a merchandising assistant. Read the text, assign each product to a demand category, and give a brief reason.	You are a merchandising analyst who segments products for SME retailers. Return a **Markdown table** with columns: *SKU/Name · Category · Reason (≤25 words)*. Allowed categories: Fast-Moving, Seasonal, At-Risk of Overstock, Low Demand. Base every reason on evidence from the supplied descriptions or reviews. No extra commentary outside the table.

This change alone helped significantly reduce misunderstandings between system outputs and user expectations, especially for procurement managers who required clear, step-by-step explanations.

AI Parameter Configurations

Choosing appropriate AI parameters was central to achieving a balance between factual accuracy and natural-sounding language. Throughout the development process, several combinations of temperature, top-p, frequency penalty, and presence penalty were tested. The goal was to ensure that the AI could remain consistent when providing structured reports while still offering flexibility and creativity during scenario simulations or open-ended Q&A sessions.

Parameter Testing Summary:

Parameter	Tested Values	Final Choice	Purpose
Temperature	0.2, 0.3, 0.5, 0.7	0.3 for reports, 0.6–0.7 for Q&A and scenarios	Low temperature ensured accuracy for structured outputs; moderate temperature allowed creative variation where needed
Top-p (nucleus sampling)	0.8, 0.9, 1.0	0.9	Balanced creativity with coherence
Frequency Penalty	0.0, 0.2, 0.3	0.2–0.3	Reduced repetition in procurement reports.
Presence Penalty	0.3, 0.4, 0.5	0.4	Encouraged diverse reasoning in answers
Max Tokens	400–3000 (based on task type)	Dynamic allocation	Ensured long-form responses for scenario simulations and summaries

Key Insight from Testing:

Early testing at temperatures above 0.7 generated creative responses but sometimes produced answers that lacked focus. The combination of a 0.3 temperature setting for factual outputs with a 0.9 top-p produced reliable and structured content.

Prompt Variations and Refinements

The development of the prompt design required multiple stages because reviewers assessed output quality based on clarity and actionability, and alignment with the AI's defined role.

User Feedback and Iterative Adjustments

A key part of refinement was to collect structured feedback from real users, including business students, procurement interns, and SME managers. These users reviewed sample outputs and provided feedback on clarity, relevance, and ease of use.

Feedback Theme	Observed Issue	Adjustments Made
Clarity of Explanations	Some outputs were still too technical.	Simplified prompt language and removed jargon.
Actionable Recommendations	Recommendations lacked clear next	Required the AI to provide step-by-step

	steps.	guidance in prompts.
Consistency Across Outputs	Tone varied between chatbot responses and reports.	Standardized tone in system instructions and templates.
Scenario Depth	Simulations occasionally lacked risk assessment.	Refined prompts to explicitly require risk analysis.

Conclusion

The AI-powered inventory and Forecasting System developed for SMEs successfully addresses the key challenges of inaccurate demand forecasting, lack of advanced tools, and the difficulty of interpreting complex data models. By integrating predictive analytics with generative AI, the solution delivers both precise inventory predictions and clear, understandable recommendations, bridging the gap between sophisticated data science and practical business decision making. This hybrid approach ensures that SMEs, which often lack the resources of larger corporations, can make informed procurement decisions, reduce risks associated with overstocking or stockouts, and improve operational efficiency through data-backed insights.

Through rigorous iterative refinement, including prompt engineering, AI parameter optimization, and user feedback loops, the system evolved into a reliable and user-friendly solution. The continuous adjustments and testing cycles enhanced the clarity, accuracy, and business alignment of the AI outputs, ensuring trust and usability among non-technical decision makers. The project demonstrates how well-designed AI solutions, when combined with thoughtful user-centric design, can democratize access to advanced analytics and empower SMEs with the decision support tools necessary for competitive success.

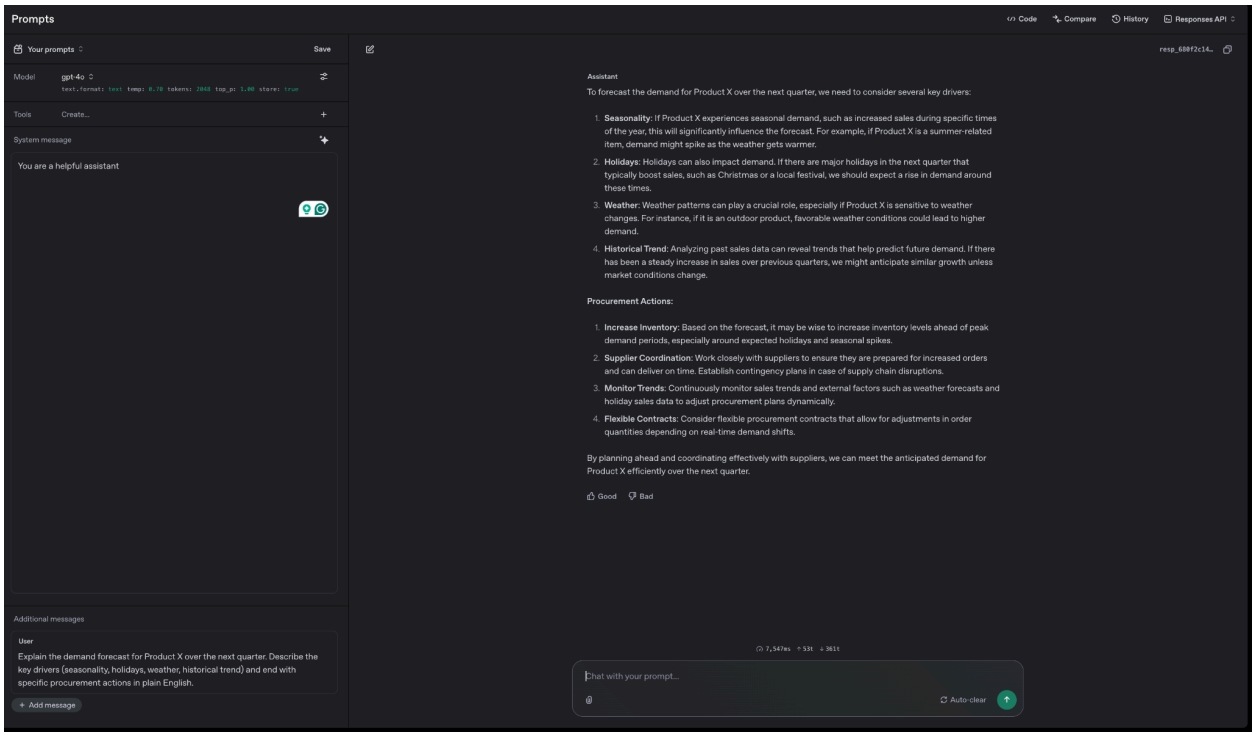

Figure 3: User Prompt Example (Initial Forecast Explanation Request) This figure shows the original user prompt and the basic system response explaining demand drivers and procurement actions without structured formatting.

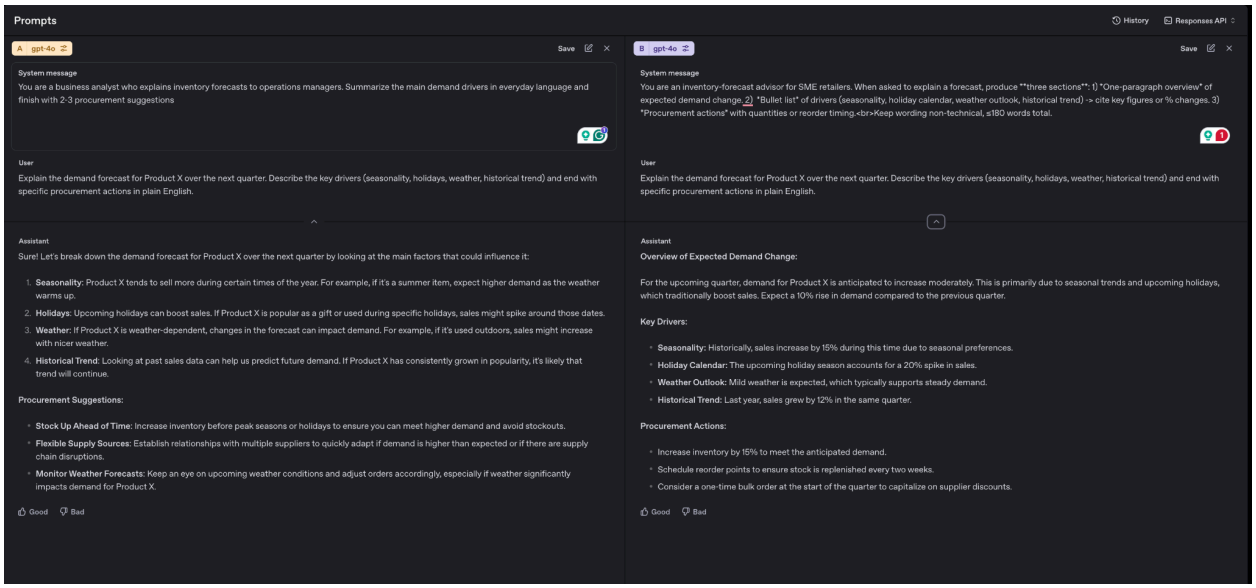

Figure 4 (Left Side): Improved V1 Prompt Output (Business Analyst Tone) The left side presents a more organized forecast with key drivers and procurement suggestions, but lacks specific data points. **Figure 4 (Right Side): Final Refined Prompt Output (Structured SME Advisor Response)** The right side shows the final refined output with a clear overview, bullet-pointed key drivers with data, and actionable procurement steps for SME retailers.

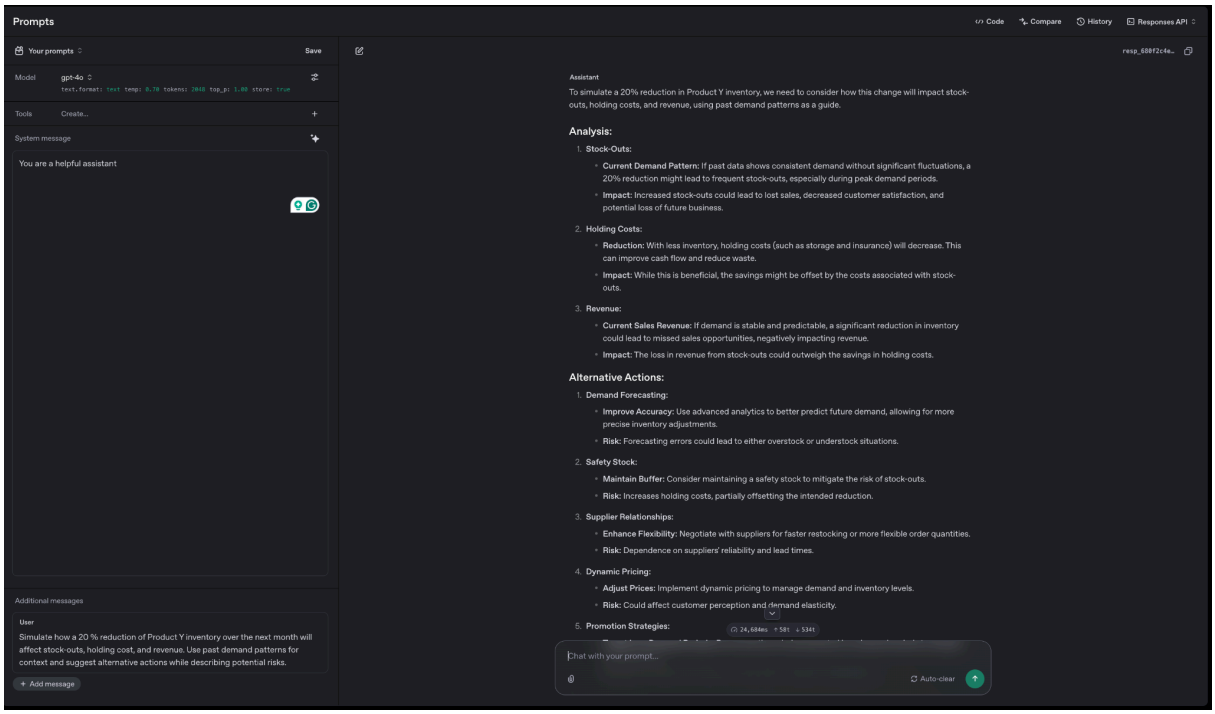

Figure 5: Scenario Simulation Prompt Example (Initial Forecast Explanation Request) This figure shows the original user prompt and the basic system response explaining demand drivers and procurement actions without structured formatting.

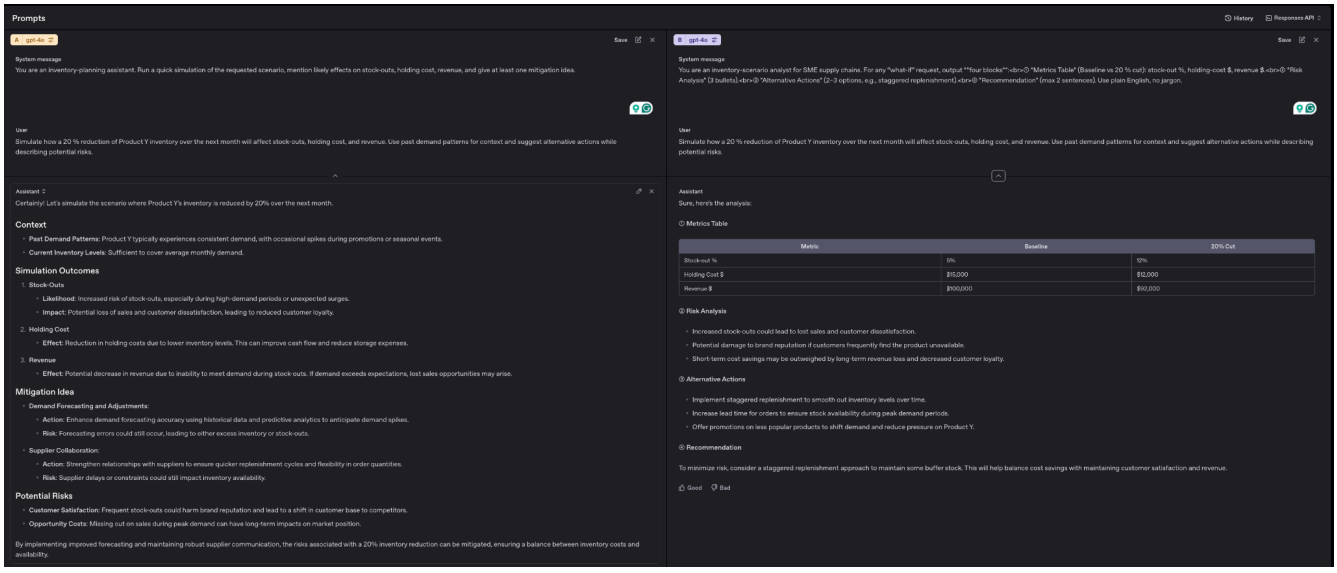

Figure 6 (Left Side): Improved V1 Prompt Output (Initial Forecast Explanation Request – Scenario Simulation) The left side presents a structured scenario simulation explaining the impacts of a 20% inventory reduction with context, outcomes, and mitigation ideas, but without clear quantified comparisons. **Figure 6 (Right Side): Final Refined Prompt Output (Initial Forecast Explanation Request – Scenario Simulation)** The right side shows the final refined output with a clear **Metrics Table**, focused **Risk Analysis**, suggested **Alternative Actions**, and a concise **Recommendation**, making the scenario results easy to understand and apply

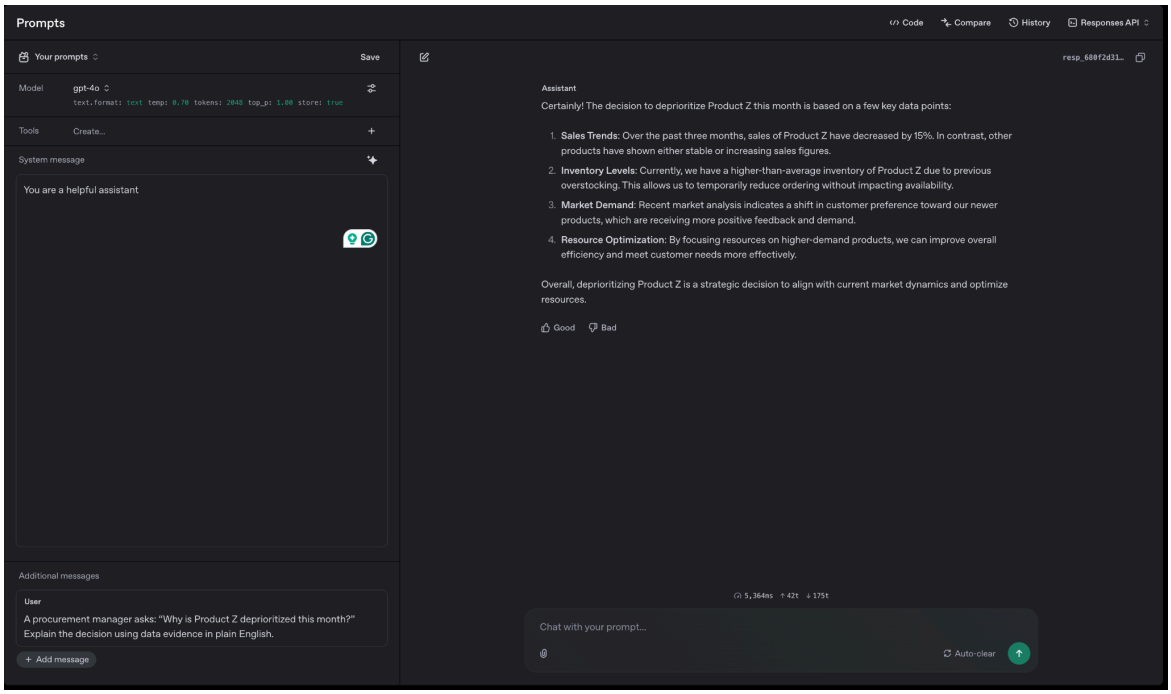

Figure 7: User Prompt Example (Initial Forecast Explanation Request – Chatbot Q&A) This figure shows the original user prompt asking the chatbot to explain why **Product Z is deprioritized** using clear data evidence. The system responds with key factors like sales trends, inventory levels, and market demand, providing a simple, data-backed explanation for the procurement manager.

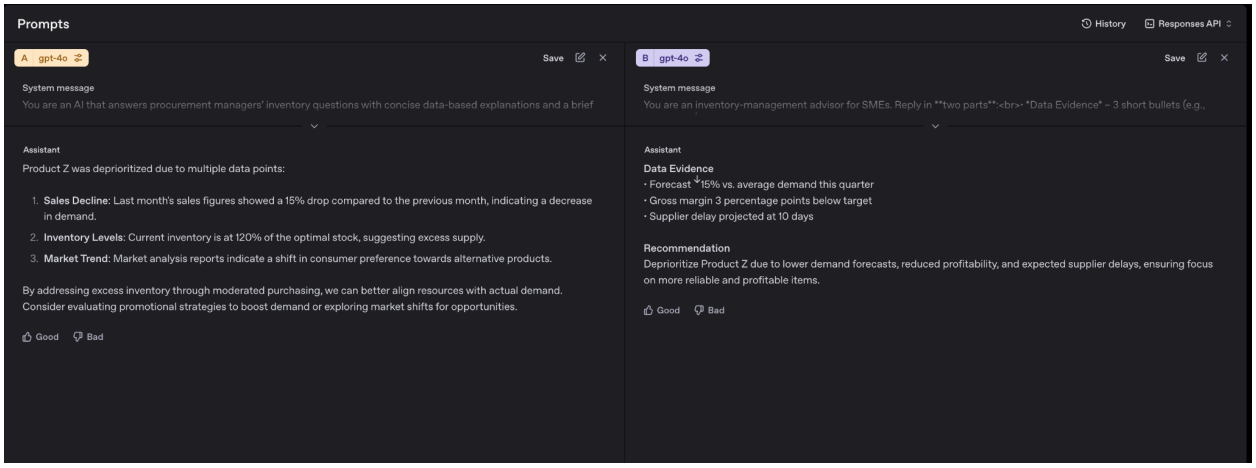

Figure 8 (Left Side): Improved V1 Prompt Output (Initial Forecast Explanation Request – Chatbot Q&A) The left side shows the **V1 output**, where the system provides a general explanation with supporting data points like sales trends and inventory levels but without a structured format or concise bullet style. **Figure 8 (Right Side): Final Refined Prompt Output (Initial Forecast Explanation Request – Chatbot Q&A)** The right side presents the final refined output, organized into Data Evidence (three concise bullets) and a one-sentence Recommendation, making it clear, actionable, and easy for procurement managers to understand.

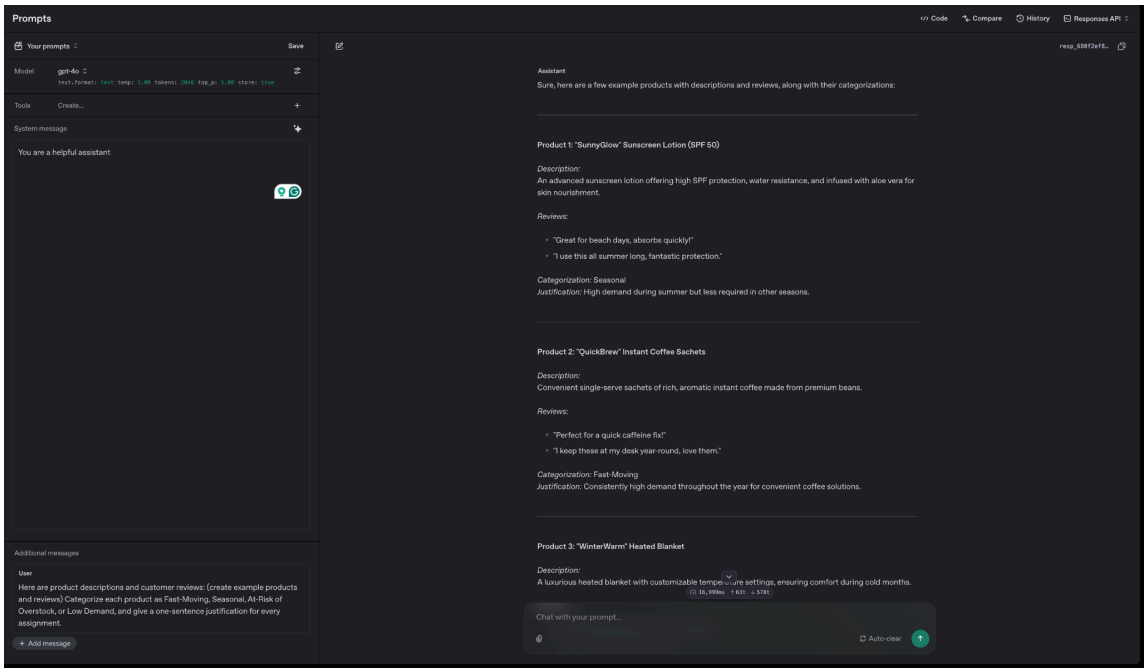

Figure 9: User Prompt Example (Initial Trend Detection & Categorization Request) This figure shows the original prompt asking to categorize products with justifications, using basic unstructured responses.

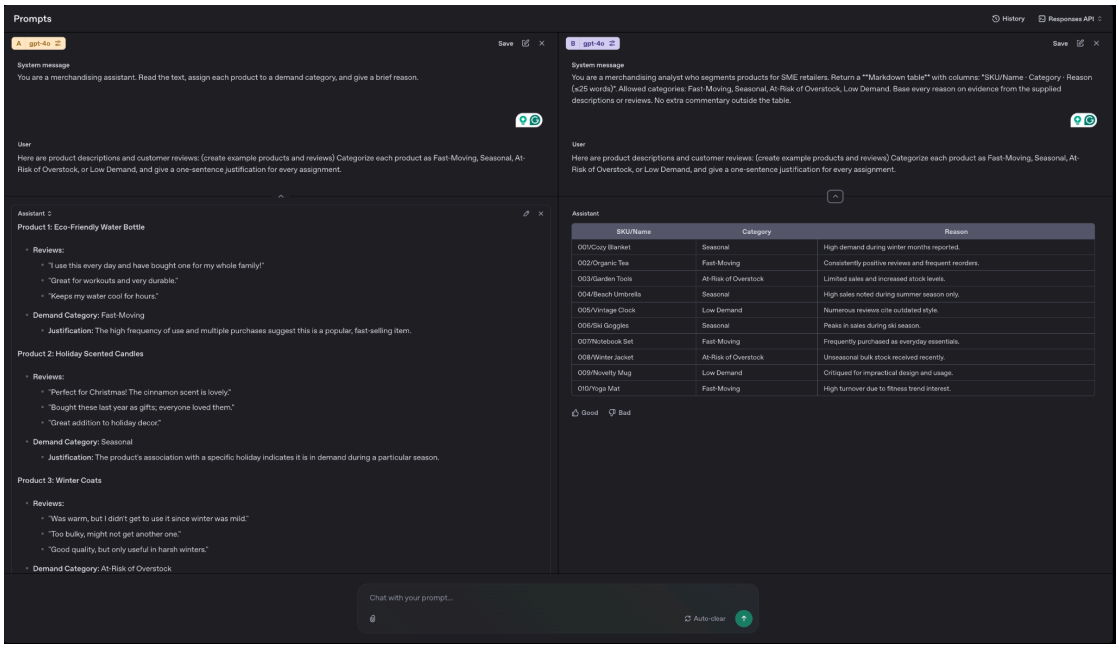

Figure 10 (Left Side): Improved V1 Prompt Output (Initial Trend Detection & Categorization Request) The left side shows the V1 output, where products are categorized with reasons based on reviews, but the response lacks structured formatting and consistency across entries. **Figure 10 (Right Side): Final Refined Prompt Output (Initial Trend Detection & Categorization Request)** The right side presents the refined output as a clear markdown table with SKU/Name, Category, and concise Reason, making the categorization easy to read, compare, and justify at a glance.