## Real Time Internal Sales Agent

## **TABLE OF CONTENTS**

1. Executive Summary
2. Introduction & Background
3. Project Vision & Goals
4. System Overview & Architecture
   * 4.1 High-Level Architecture Components
   * 4.2 Real-Time Voice Communication for Near Zero-Latency, Multi-Lingual Conversations
5. Key Modules & Phases of Development
6. Detailed Classification of Conversational Factors
   * 6.1 Query Type Classification
   * 6.2 Query Content & Specifics
   * 6.3 Customer Intent Classification
   * 6.4 Query Language & Tone
   * 6.5 Time & Urgency Classification
   * 6.6 Preferred Communication Mode
   * 6.7 Transactional Details & Payment Terms
   * 6.8 Documentation & Certificate Requests
   * 6.9 Delivery & Logistics
   * 6.10 Negotiation & Pricing Sensitivity
   * 6.11 Customer Type & History
   * 6.12 Internal Action & Clarification (Agent Side)
   * 6.13 Edge Cases & Special Considerations
7. Query Processing Criteria, Prioritization, Groups, and Stages
8. Dynamic Chatbot Interactions & Flow
9. Front-End Dashboard (No-Code/Low-Code Platform) for Model Training & Configuration
10. Integration with Sales ERP & Other Systems
11. Security, Compliance & Data Governance
12. 25 Example Scenarios (Edge Cases) [**NEW FORMAT**]
13. Technical Feasibility & Architecture Confirmation
14. Implementation Sprints
15. Conclusion & Next Steps

**1. EXECUTIVE SUMMARY**

This expanded document describes a **Real-Time Conversational AI Sales & Fulfillment Agent** specifically designed for a **B2B Raw Material Company** in the construction industry—though it can be applied to **any vertical** by swapping out the underlying Product Information Management (PIM) data. Key features:

* **Highly scalable** design, leveraging dynamic PIM integration for product details.
* **Front-end no-code/low-code platform** to manage AI model configurations, conversation flows, discount thresholds, and more.
* **Real-time voice** capabilities with near zero-latency STT (Speech-to-Text) and TTS (Text-to-Speech) pipelines, plus advanced multi-lingual mixing (Hinglish/Marathi-English).
* **ERP** integration for inventory, pricing, invoice management, and billing checks.
* **Human-like** interactions that never start with a sales pitch—**the agent waits** for the customer to speak first.

By unifying these elements, the solution addresses **complex B2B sales**, handles negotiations, collects specification details, arranges deliveries, and escalates edge cases with minimal friction.

**2. INTRODUCTION & BACKGROUND**

**2.1 Industry Context**

Within the **construction** sector, raw materials like steel, cement, aggregates, and chemicals are purchased in large volumes. Purchases often include:

* **Technical specs** (grades, thicknesses, finishes).
* **Complex negotiation** around price, credit, or partial deliveries.
* **Documentation** (invoices, MTC, MTR).

Current manual processes are time-consuming and error-prone. A real-time conversational AI agent brings **speed, consistency,** and **scalability** to meet growing demand or new markets.

**2.2 The Need for an AI Sales Agent**

1. **24/7 Engagement**: Capture inquiries or urgent calls at any hour.
2. **Reduced Errors**: Automatic quotes from ERP/PIM without manual mistakes.
3. **Flexibility & Scalability**: Adapt quickly to new product lines, geographies, or industries.
4. **Integrated Data**: Conversation context merged with ERP and CRM for a single source of truth.
5. **Future-Proof**: By adopting robust architecture, new features—like deeper analytics—can be added with ease.

**2.3 Project Mandate**

* **Design** a robust, real-time voice/text-based agent that respects B2B raw material complexities.
* **Implement** zero-latency streaming solutions for STT/TTS to achieve a near human-like conversation flow.
* **Empower** sales teams with a no-code dashboard for model training, discount adjustments, and conversation logic updates.
* **Extend** the architecture to function in any industrial domain by hooking up a new or extended PIM schema.

**3. PROJECT VISION & GOALS**

1. **Streamlined Automation**: Move the typical phone/email negotiations, quoting, and order confirmations into a single AI-driven conversation.
2. **Human-Like Interactions**: The agent should wait silently for the user’s input and only ask minimal clarifications—primarily requesting the user to send further specs on WhatsApp or email.
3. **Seamless ERP & PIM Integration**: Ensure real-time data for inventory checks, dynamic pricing, and invoice generation.
4. **Scalability**: Designed from the ground up so that hooking different product lines or industries is straightforward.
5. **No-Code/Low-Code Management**: Give business stakeholders direct control to re-train or alter conversation flows without heavy developer intervention.

**4. SYSTEM OVERVIEW & ARCHITECTURE**

**4.1 High-Level Architecture Components**

1. **Voice/Chat Interface Layer**
   * Could be Twilio or Exotel for phone calls.
   * Web Chat, WhatsApp Business API, or Telegram for text-based channels.
2. **NLP & STT/TTS Pipeline**
   * **Speech-to-Text** (STT) and **Text-to-Speech** (TTS) must be near real-time.
   * A **Conversational AI Engine** (GPT-based or Dialogflow) processes queries and decides next steps.
3. **Integration & Orchestration Layer**
   * **ERP Integration**: Inventory, order management, invoice creation.
   * **PIM**: Live product specs and pricing.
   * **CRM**: Payment history, customer classification (VIP, new, frequent defaulter).
4. **No-Code Front-End Dashboard**
   * Drag-and-drop conversation flow editor.
   * Configurable discount thresholds, fallback scenarios.
   * Model training interface (upload domain texts, re-train, switch model versions).
5. **Database & Logs**
   * Store conversation transcripts, fallback triggers, usage analytics.
   * Maintain system event logs for debugging and compliance.

**4.2 Real-Time Voice Communication for Near Zero-Latency, Multi-Lingual Conversations**

In **B2B** contexts, voice calls remain a major interaction channel. Achieving **human-like** conversation requires:

1. **Ultra-Low Latency STT**
   * **Streaming STT**: Tools like Google Cloud Streaming, Deepgram, or Whisper with real-time streaming support.
   * As the user speaks, partial transcripts are generated within **<300 ms** segments.
2. **Incremental NLP Processing**
   * **Token-by-token streaming** to the LLM (if the model supports partial ingestion of input).
   * Minimizes “thinking gaps.”
3. **Low Latency TTS**
   * Provide chunk-based or **streaming TTS**. The agent can start speaking the partial response while generating the remainder.
   * High-quality TTS engines (e.g., ElevenLabs, Amazon Polly) with multi-lingual or code-switched voices.
4. **Multi-Lingual Mixing** (Hinglish, Marathi-English)
   * The STT engine must identify language shifts on-the-fly.
   * The TTS engine must be able to produce natural-sounding text for code-mixed phrases.
   * The LLM internally keeps track of the conversation’s language context.
5. **Network Architecture for Zero-Latency**
   * Deploy the **NLP model** on a high-throughput, GPU-based instance with auto-scaling.
   * Use **edge locations** or **regional servers** close to major customer bases to reduce round-trip times.
   * Optimize concurrency so that multiple calls or chats do not throttle performance.
6. **Workflow & Voice Channels**
   * **When a user calls**, audio streams to the STT service. Partial transcripts feed the LLM.
   * The LLM streams back text, which is **immediately** converted to speech and played to the user.
   * Any PIM or ERP lookups happen asynchronously in parallel, with results fed back into the conversation state.

By carefully orchestrating these pipelines, the system can produce an experience that feels almost like a **live human** on the other end, with minimal audible delays—even in multi-lingual code-switched contexts.

**5. KEY MODULES & PHASES OF DEVELOPMENT**

**5.1 Phase 1: Requirements & Prototype**

* Gather real or representative data from your B2B raw material environment.
* Create a **minimal prototype** with text-based queries hooking up to a dummy PIM.

**5.2 Phase 2: Core NLP & Integration**

* Integrate **advanced STT** and **TTS** for real-time voice calls with multi-lingual support.
* Fine-tune GPT-based models on domain-specific terms (steel grades, typical negotiations, credit processes).
* Connect with your **ERP** for real inventory checks, pricing tables, order creation.

**5.3 Phase 3: Front-End No-Code Dashboard**

* Develop the **drag-and-drop** flow builder for conversation logic.
* Provide an easy interface to manage **model training** (upload domain texts or conversation logs).
* Offer a user-friendly approach to set discount or credit thresholds, fallback responses, user role management, etc.

**5.4 Phase 4: Advanced Features & Edge-Case Handling**

* Implement a more sophisticated negotiation approach, possibly involving **AI-based** price recommendations if historical data is available.
* Add **credit checks**: If the ERP flags the customer as overdue, the agent automatically references that.
* Implement **partial shipments**, multi-stage deliveries, or specialized compliance documents.

**5.5 Phase 5: Scalability & Multi-Industry Deployment**

* Build a **universal** approach for hooking up new or different PIM data sets (e.g., for chemicals, automotive parts).
* Prepare for language expansions if the business moves to other regions (Tamil-English, Telugu-English, etc.).
* Optimize concurrency and performance to handle **high volumes** of voice calls or chat sessions in parallel.

**6. DETAILED CLASSIFICATION OF CONVERSATIONAL FACTORS**

**6.1 Query Type Classification**

1. **Sales Inquiry**
2. **Price Request**
3. **Product Specification Inquiry**
4. **Negotiation Attempt**
5. **Delivery Status Follow-Up**
6. **Billing/Invoice Inquiry**
7. **Documentation Request**
8. **Payment-Related Query**
9. **Complaint/Issue Reporting**
10. **General/Miscellaneous**

**6.2 Query Content & Specifics**

* Material type (TMT bars, HR sheets, CR coils, or other raw materials).
* Specifications like grade (Fe500, Fe550D), thickness, length, finish.
* Quantity in tons or kg.
* Delivery location & timeline.
* Payment terms or credit conditions.

**6.3 Customer Intent Classification**

* **Immediate Purchase**
* **Future Inquiry**
* **Competitor Comparison**
* **Negotiation**
* **Complaint or Issue Resolution**

**6.4 Query Language & Tone**

* Language: English, Hinglish, Marathi-English (avoid mixing Hindi & Marathi in same conversation).
* Tone: polite, urgent, frustrated, aggressively negotiating, etc.

**6.5 Time & Urgency Classification**

* Immediate (same-day or within hours).
* Urgent (1-2 days).
* Normal (2-3 days or more).
* Low Priority (future reference).

**6.6 Preferred Communication Mode**

* Voice (phone call).
* WhatsApp.
* Email.
* SMS or internal chat.

**6.7 Transactional Details & Payment Terms**

* Payment Terms: net 7, net 15, net 30, partial advance.
* Overdue Payment checks in ERP.
* Payment method: NEFT, RTGS, cheque, digital gateway.

**6.8 Documentation & Certificate Requests**

* Material Test Certificate (MTC).
* Mill Test Report (MTR).
* GST invoice, duplicate invoice.
* Delivery challan or compliance docs.

**6.9 Delivery & Logistics**

* Delivery timeline feasibility (immediate, scheduled).
* Logistics type (dedicated truck, part-load).
* Delivery address or project site.
* Tracking updates & follow-ups.

**6.10 Negotiation & Pricing Sensitivity**

* Negotiation level: mild, moderate, aggressive.
* Competitor references.
* Bulk discount possibility.
* Manager approval needed?

**6.11 Customer Type & History**

* New or returning.
* Frequent buyer or “never finalizes.”
* High-value or low-value.
* Past payment behavior (timely, delayed).

**6.12 Internal Action & Clarification (Agent Side)**

* Stock check from ERP.
* Price fetch from PIM or dynamic pricing table.
* Management escalation for special discount.
* Payment verification or block if overdue.

**6.13 Edge Cases & Special Considerations**

* Overdue amounts but wanting new orders.
* Impossible deliveries (huge quantity, minimal time).
* Persistent discount demands.
* PO mismatch (quoted specs vs. official PO).

**7. QUERY PROCESSING CRITERIA, PRIORITIZATION, GROUPS, AND STAGES**

1. **Immediate vs. Deferred**: Immediate handle if standard specs, deferred if special approval needed.
2. **Priority Groups**:
   * Group A: VIP or urgent large-volume orders.
   * Group B: Standard inquiry, normal timeline.
   * Group C: Price-checkers or info-only.
3. **Stages**:
   * Intake -> Verification (ERP/PIM) -> Negotiation -> Commitment -> Follow-up.

**8. DYNAMIC CHATBOT INTERACTIONS & FLOW**

1. **Initial Greeting**: The AI remains silent until the customer speaks.
2. **Minimal Clarification**: “Please WhatsApp your spec details for clarity.”
3. **Data Retrieval**: Real-time from PIM (product specs) and ERP (stock, invoice).
4. **Negotiation**: Possibly apply discount rule sets or escalate.
5. **Order Confirmation**: Write to ERP, share invoice.
6. **Edge Case Escalation**: If overdue or mismatch, politely mention it.
7. **Closing**: “Thank you, we appreciate your business. Let me know if anything else is required.”

**9. FRONT-END DASHBOARD (NO-CODE/LOW-CODE PLATFORM) FOR MODEL TRAINING & CONFIGURATION**

**9.1 Objectives**

* Provide a **visual**, drag-and-drop environment to define conversation flows.
* Let business users **configure** discount thresholds, fallback responses, etc.
* Manage **model training**: upload domain text, schedule fine-tuning, track version changes.
* Real-time **monitoring**: watch conversation logs, see fallback rates, success metrics.

**9.2 Key Features**

* **Flow Editor**: Graphical blocks for greetings, stock checks, discount logic.
* **Model Panel**: Switch GPT versions or update existing ones with new data.
* **Integration Settings**: Connect to ERP, CRM, or new PIM endpoints.
* **User & Role Management**: Admins, managers, operators each with distinct permissions.

**9.3 Technology Stack for No-Code Platform**

* **Front-End**: React or Vue with a robust flowchart engine (ReactFlow, JointJS).
* **Backend**: Node.js, Python, or Java microservices controlling model endpoints.
* **Storage**: Relational DB (PostgreSQL) or NoSQL for conversation flows, logs.

**9.4 Benefits**

* Business agility (rapid modifications).
* Non-technical staff can oversee day-to-day changes.
* Faster iteration on conversation flows and discount logic.

**10. INTEGRATION WITH SALES ERP & OTHER SYSTEMS**

**10.1 ERP Integration**

* Auto-check inventory levels, confirm open balances.
* Create orders in the ERP upon user confirmation.
* Issue invoice or delivery note from within the conversation flow.

**10.2 PIM Integration**

* Access up-to-date product specifications.
* Retrieve real-time or date-based pricing references.
* Reflect new product lines or specs in the no-code dashboard for immediate rollout.

**10.3 CRM Integration**

* Identify the customer’s past interactions or outstanding payments.
* Personalize conversation based on loyalty or negotiation patterns.
* Create leads or track new opportunities automatically.

**10.4 Payment Gateway**

* Direct link for invoice settlement if needed.
* Validate payment confirmations in real time.

**11. SECURITY, COMPLIANCE & DATA GOVERNANCE**

**11.1 Authentication & Authorization**

* Role-based access to the no-code platform.
* Multi-factor authentication for admin or finance roles.

**11.2 Encryption**

* Enforce TLS on all external communications.
* Encrypt conversation logs if storing any sensitive data (like payment details).

**11.3 Compliance**

* Follow local data protection laws (GDPR, PDP Bill in India, etc.).
* Potential domain standards: BIS, ISO if relevant for certain materials.

**11.4 Audit Trails & Data Retention**

* Keep logs of conversation flow changes in the no-code editor.
* Retain transcripts securely for a defined period, then purge or anonymize.

**12. 25 EXAMPLE SCENARIOS (EDGE CASES) [NEW FORMAT]**

In each example, we follow the format:

**Customer (Hinglish):**  
*English Translation:*  
**Agent:**  
*English Translation:*  
**Modules Triggered:\_** (Which system modules or steps from the architecture are engaged?)

1. **Scenario 1: Overdue Payment & New Order Request**  
   **Customer (Hinglish):** “Bhaiya, mujhe 10 ton aur bhej do. Payment next week kar dunga.”  
   *English Translation:* “Brother, send me 10 more tons. I will pay next week.”  
   **Agent:** “Sir, aapke upar 2 lakh pending hai. Pehle clear kar dijiye; main details WhatsApp par share kar raha hoon.”  
   *English Translation:* “Sir, you have 200k overdue. Please clear it first; I’m sending details on WhatsApp.”  
   **Modules Triggered:** CRM check (customer’s overdue), ERP for outstanding invoice, conversation logic for payment block.
2. **Scenario 2: Repeated Negotiation at Unrealistic Prices**  
   **Customer (English):** “Your competitor is 15% cheaper; match the price or I’m out.”  
   **Agent:** “Sir, that price is below our cost. Can you WhatsApp their quote? I’ll verify internally.”  
   *English Translation:* (N/A—English scenario)  
   **Modules Triggered:** Price negotiation logic, discount threshold check, potential management escalation.
3. **Scenario 3: Impossibly Urgent Delivery**  
   **Customer (Hinglish):** “Aaj hi 50 ton deliver kar sakte ho?”  
   *English Translation:* “Can you deliver 50 tons today?”  
   **Agent:** “Sir, itna urgent mushkil hai. Main logistics se check karke 30 minute mein aapko WhatsApp pe update deta hoon.”  
   *English Translation:* “Sir, that’s difficult. I’ll check with logistics and update you on WhatsApp in 30 minutes.”  
   **Modules Triggered:** Logistics scheduling module in ERP, voice-based conversation flow, partial fallback to WhatsApp for confirmation.
4. **Scenario 4: Mismatch in PO and Quotation**  
   **Customer (English):** “I’ve emailed you the PO. Please process ASAP.”  
   **Agent:** “Sir, your PO states 10 mm thickness, but we quoted 8 mm. Could you confirm on WhatsApp?”  
   *English Translation:* (N/A—English scenario)  
   **Modules Triggered:** PO mismatch detection, ERP check, conversation logic for re-verification.
5. **Scenario 5: Language Switch (Marathi-English)**  
   **Customer (Marathi-English):** “Mala 20 ton HR sheets lagtil, rate sangal ka?”  
   *English Translation:* “I need 20 tons of HR sheets, can you give me the rate?”  
   **Agent:** “Sure sir! Please WhatsApp var tumchi exact specification sanga, me turant rate pathavto.”  
   *English Translation:* “Sure sir! Please share your exact specification on WhatsApp, I’ll send the rate immediately.”  
   **Modules Triggered:** STT/TTS multi-lingual pipeline, PIM integration for HR sheets, conversation logic for minimal clarifications.
6. **Scenario 6: Customer Only Inquires for Price, Never Buys**  
   **Customer (English):** “What’s the current rate for Fe500, 16mm TMT?”  
   **Agent:** “We’ve shared quotes multiple times, sir. Is there something blocking you from placing an order?”  
   **Modules Triggered:** CRM retrieval of negotiation history, possible “frequent enquirer” flow.
7. **Scenario 7: Request for Material Test Certificate**  
   **Customer (English):** “Send me the MTC for last month’s shipment.”  
   **Agent:** “Certainly, please confirm your order number on WhatsApp. I’ll email or WhatsApp the MTC.”  
   **Modules Triggered:** Documentation module in ERP, conversation flow for doc retrieval.
8. **Scenario 8: Bulk Order with Complex Specs**  
   **Customer (English):** “I need 100 tons TMT Fe550, 12 mm, plus 8 mm coiled rods, delivered in 3 parts.”  
   **Agent:** “Sir, that’s detailed. Could you send the full specs and schedule on WhatsApp? I’ll finalize the price.”  
   **Modules Triggered:** PIM dynamic retrieval for multiple SKUs, multi-stage delivery logic in ERP.
9. **Scenario 9: Customer Asks for Bill after Payment Delay**  
   **Customer (Hinglish):** “Bill bhej do, main abhi pay kar dunga.”  
   *English Translation:* “Send me the bill, I’ll pay right away.”  
   **Agent:** “Zaroor sir, aapka pending invoice abhi WhatsApp kar raha hoon.”  
   *English Translation:* “Sure sir, I’m sending your pending invoice on WhatsApp right now.”  
   **Modules Triggered:** ERP invoice retrieval, conversation logic for sending docs.
10. **Scenario 10: Ambiguous Inquiry**  
    **Customer (English):** “I need some steel.”  
    **Agent:** “Could you specify the type, grade, thickness, or quantity? If it’s more convenient, WhatsApp me the details.”  
    **Modules Triggered:** Minimal query classification, fallback for incomplete info.
11. **Scenario 11: Over-Enthusiastic Delivery Promise by Customer**  
    **Customer (English):** “I’ll pick up 10 tons in 2 hours.”  
    **Agent:** “Sir, loading alone may take up to 4 hours. Let me confirm earliest pickup time in ERP and revert.”  
    **Modules Triggered:** Logistics scheduling check, conversation logic.
12. **Scenario 12: Customer Changing Specs Mid-Conversation**  
    **Customer (English):** “Actually, I need CR coils, not HR.”  
    **Agent:** “Sure, let me revise the pricing. Please finalize your specs on WhatsApp so we avoid confusion.”  
    **Modules Triggered:** PIM re-fetch, flow update, ephemeral quoting.
13. **Scenario 13: Threat to Move to Competitor if No Discount**  
    **Customer (English):** “Reduce 5% or I’ll switch to another supplier.”  
    **Agent:** “We value your business, sir. Could you share your final quantity on WhatsApp? I’ll try my best internally.”  
    **Modules Triggered:** Discount logic, potential manager escalation, conversation flow for competitor reference.
14. **Scenario 14: Payment Mode Confusion**  
    **Customer (Hinglish):** “Check se payment chalega kya?”  
    *English Translation:* “Can I pay by cheque?”  
    **Agent:** “Sir, hum abhi digital payment prefer karte hain. Bank details main WhatsApp par bhej deta hoon.”  
    *English Translation:* “Sir, we currently prefer digital payment. I’ll WhatsApp you the bank details.”  
    **Modules Triggered:** Payment preferences, ERP settlement checks.
15. **Scenario 15: Request for Duplicate Invoice (Past Order)**  
    **Customer (English):** “I lost my old invoice from 6 months ago.”  
    **Agent:** “Sure, please share approximate date or order reference. I’ll retrieve and WhatsApp or email it.”  
    **Modules Triggered:** ERP historic invoice retrieval, doc sending flow.
16. **Scenario 16: Partial Delivery Request**  
    **Customer (English):** “I need 10 tons now, 10 next month.”  
    **Agent:** “Yes, we can schedule in two shipments. Please confirm the split on WhatsApp.”  
    **Modules Triggered:** ERP partial order logic, scheduling.
17. **Scenario 17: Material Shortage / Out of Stock**  
    **Customer (Hinglish):** “Fe500 16 mm hai kya?”  
    *English Translation:* “Do you have Fe500 16 mm?”  
    **Agent:** “Filhaal wo out of stock hai. Fe500D hai, chalega?”  
    *English Translation:* “Currently it’s out of stock. We have Fe500D. Will that do?”  
    **Modules Triggered:** Stock check, alternative suggestion flow.
18. **Scenario 18: Customer Complains About Quality**  
    **Customer (English):** “The batch you sent is off-spec.”  
    **Agent:** “Sorry to hear that, sir. Could you share the batch label or a photo on WhatsApp? I’ll escalate to QC.”  
    **Modules Triggered:** Complaint flow, possible QC escalation.
19. **Scenario 19: Customer Email vs. WhatsApp Preference**  
    **Customer (English):** “Email me the invoice, please. I don’t use WhatsApp.”  
    **Agent:** “Sure, I’ll email the PDF. For quicker clarifications, I’m available by phone if needed.”  
    **Modules Triggered:** Document sending via email, fallback conversation flow.
20. **Scenario 20: High-Value Customer Requiring Priority Dispatch**  
    **Customer (Hinglish):** “Bhai, main regular leta hoon, is order ko priority de do.”  
    *English Translation:* “Brother, I’m a regular buyer, please prioritize this order.”  
    **Agent:** “Ji sir, main abhi isko priority dispatch mark kar deta hoon. Aapko schedule details 1 ghante mein WhatsApp pe mil jaayenge.”  
    *English Translation:* “Yes sir, I’ll mark it for priority dispatch. You’ll receive the schedule details on WhatsApp in 1 hour.”  
    **Modules Triggered:** ERP flag for priority shipping, VIP flow.
21. **Scenario 21: Repeated Request for Different Sizes**  
    **Customer (English):** “Check 8mm, 10mm, and 12mm TMT bars. Also, what about 6mm coils?”  
    **Agent:** “Let’s avoid confusion. Please list all your required sizes on WhatsApp, and I’ll compile one final quote.”  
    **Modules Triggered:** PIM repeated SKU fetch, conversation fallback for big specs.
22. **Scenario 22: Payment Link or Gateway Request**  
    **Customer (English):** “Send me a payment link, I’ll pay online.”  
    **Agent:** “Certainly, let me generate a link right away. Please confirm once payment is done.”  
    **Modules Triggered:** Payment gateway, invoice reference check.
23. **Scenario 23: Very Polite but Non-buying Customer**  
    **Customer (English):** “Could you tell me your best rate for next quarter?”  
    **Agent:** “We’ve shared multiple quotes before. Any reason you haven’t placed an order? We’d love to assist further.”  
    **Modules Triggered:** CRM note (frequent enquirer), conversation logic for polite follow-up.
24. **Scenario 24: Bulk Cross-Industry Query**  
    **Customer (English):** “Do you supply chemicals or just steel?”  
    **Agent:** “Currently, we focus on steel, but we can adapt our PIM for new product lines. Let me check feasibility.”  
    **Modules Triggered:** Potential cross-industry expansion, PIM reconfiguration logic.
25. **Scenario 25: Negative Sentiment or Angry Customer (Delivery Delay)**  
    **Customer (Hinglish):** “Mera order abhi tak kyun nahi aaya? Aap log bahut slow ho!”  
    *English Translation:* “Why hasn’t my order arrived yet? You people are very slow!”  
    **Agent:** “Maaf kijiye sir, main abhi logistics se check karta hoon aur turant aapko actual status WhatsApp kar deta hoon.”  
    *English Translation:* “I’m sorry, sir. I’ll check with logistics immediately and WhatsApp you the actual status.”  
    **Modules Triggered:** Delivery tracking in ERP, complaint flow, real-time update to user.

**13. TECHNICAL FEASIBILITY & ARCHITECTURE CONFIRMATION**

**13.1 Technical Feasibility**

1. **Cloud Infrastructure**: AWS, Azure, or GCP can provide robust environment for real-time streaming.
2. **LLM Hosting**: GPT-4 or GPT-like open-source solutions can be deployed with GPU auto-scaling.
3. **Zero-Latency Goal**: Achieved with streaming STT/TTS and partial token-by-token LLM inference.
4. **ERP & PIM**: Must have stable REST/GraphQL APIs for real-time data retrieval.

**13.2 Architecture Confirmation**

* **Microservices** or serverless event-driven approach recommended for each component (STT, TTS, conversation logic).
* **Load Balancing** across multiple NLP containers to handle concurrency.
* **Data Layer**: A secure DB (PostgreSQL, MongoDB, or Elasticsearch) for logging conversations.

With these measures in place, near real-time voice interactions, accurate quoting, and dynamic negotiation flows become **technically feasible**.

**14. IMPLEMENTATION SPRINTS**

1. **Sprint 1**: Set up cloud environment, minimal text-based prototype, mock PIM.
2. **Sprint 2**: Add real-time voice integration, connect to actual ERP, enable partial STT streaming.
3. **Sprint 3**: Build out no-code dashboard for conversation flows, discount thresholds, user roles.
4. **Sprint 4**: Expand advanced negotiation logic, partial deliveries, MTC auto-sharing.
5. **Sprint 5**: Thorough testing, user acceptance, refine multi-lingual TTS.
6. **Sprint 6**: Secure deployment, performance optimization, finalize production launch.

**15. CONCLUSION & NEXT STEPS**

This **mega-detailed** document merges **all** prior information with further elaboration on:

* **Zero-latency real-time voice** capabilities for multi-lingual code-switching.
* A standardized **example scenario format** listing 25 edge cases with translations and modules triggered.
* A **no-code front-end** approach for easy configuration.
* **Technical** feasibility, architecture layout, and phased deployment plan.

**Immediate Action Items**

1. **Review** this extended plan with stakeholders.
2. **Initiate** environment setup on AWS/Azure, deciding STT/TTS providers.
3. **Assemble** an agile team to address sprints in chronological order.
4. **Implement** the proof-of-concept, gather user feedback, and refine for final go-live.