

Promotion Call Service Structure (Optimized Version)

1. Core Concept

STT (Speech-To-Text): Only records customer speech in real-time, storing it as text in the database. Human Review Process: A human reviews STT records and creates the most ideal response as an audio recording. TTS/Recording Cache: Stores pre-recorded audio for immediate playback by the chatbot. Chatbot Training: When a similar question is detected in the future, the chatbot automatically plays the pre-recorded audio.

2. Overall Flow

[Promotion Call Start] → [Customer Speech → STT Recording] → [Human Review → Ideal Answer Recording] → [Recording Stored + Chatbot Training] → [Immediate Playback for Future Similar Questions]

3. First-Time Customer Question Flow (Initial Data Collection Stage)

Goal: Accurately record customer speech and later create the most ideal response for chatbot use. Step 1: Promotion Call Start - Call bot initiates the call and plays a pre-recorded opening message. Step 2: Customer Speech (New Question) - The customer's speech is converted to text via STT in real-time. - Text and timestamps are stored in the database. - The chatbot does NOT respond yet; a generic closing message is played: "Dear customer, I will send you detailed information via text message." Step 3: Human Review of STT Text - A human reviewer examines the stored STT text to understand the customer intent. Step 4: Create Ideal Answer Recording - The reviewer writes the perfect answer script and records it using a human voice (not Fish-Speech). Step 5: Register in Chatbot Training - The recorded audio is saved in the tts_cache and mapped to the original customer question. Example: "What is the commission percentage?" → audio/commission_001.mp3 Step 6: Send Text Message - After the call, the system automatically sends a detailed text message to the customer.

4. Second-Time or Later Similar Question Flow (Automated Response Stage)

Goal: Allow the chatbot to instantly play a pre-recorded response without real-time TTS. Step 1: Promotion Call Start - Call bot initiates the call as usual. Step 2: Customer Speech (Repeated Question) - The speech is converted to text via STT and stored. - The chatbot searches the database for similar past questions using similarity detection (e.g., cosine similarity ≥ 0.85). Step 3: Instant Pre-Recorded Audio Playback - If a match is found, the chatbot plays the corresponding pre-recorded audio file immediately. - Response is provided with zero delay. Step 4: Continuous STT Recording - Even during repeat interactions, customer speech is continuously recorded. - If a new, unique question is detected, it is sent back to the human review process.

5. Data Feedback Loop

[Customer Speech STT Recording] → [Human Review & Answer Design] → [Recording Creation by Human Voice] → [Chatbot Training Registration] → [Future Customer Calls: Instant Playback] → [New Questions Sent Back to Human Review]

6. Technical Architecture

[Promotion Call Start] → [Customer Speech] → (STT Conversion Only, No Immediate AI Response) → [Text Storage + Recording Update] → [Human Review & Ideal Answer Recording] → [Pre-Recorded Audio + Chatbot Training] → [Real-Time Playback for Similar Future Questions]

7. Database Structure

call_session: Basic call info (call_id, phone, timestamps, recording URL). call_audio_chunk: Real-time HLS recording chunk management. call_utterance: STT text storage. tts_cache: Pre-recorded audio storage and mapping.

8. Expected Benefits

Speed: Zero delay for repeated questions, no real-time TTS required. Quality: Human review ensures only the best answers are trained into the chatbot. Data Accumulation: Continuous collection of speech and recordings improves service. Cost Savings: Minimized Fish-Speech real-time usage, reducing GPU costs. Scalability: As repeated questions grow, automation ratio increases while human involvement decreases.

9. Summary

First Question: STT records customer speech only, human reviews and creates ideal answer later. Second and Later Questions: Chatbot instantly responds using pre-recorded answers. Continuous Loop: New questions are always sent back to the human review process, gradually expanding chatbot coverage and automation. Result: Initial quality is guaranteed by human review, while speed and efficiency improve through increasing automation over time.