Case Study

Guided by GenAI

Transforming travel inquiry handling with
generative AI and Sonata's agent assist solution

Summary

Sonata delivered a generative AI-powered agent assist solution that transformed the client's customer
inquiry response process. By implementing natural language processing (NLP) with context filtering
and retrieval-augmented generation (RAG) technology, the solution reduced average handling time by
50-60% while automatically handling 20-25% of customer inquiries, eliminating the need for resource
scaling during peak seasons.

Client Overview

A leading travel and tourism enterprise specializing in leisure
travel services, offering a comprehensive portfolio that includes
flights, hotel bookings, and vacation packages. With annual
revenues in the multi-billion-dollar range, the company
operates across numerous international markets and manages
a diverse suite of digital platforms that serve millions of
travelers globally.

Headquarter	Revenue
Hanover, Germany	**€23.167 B**

Pressure Points

The client's customer service operations were overwhelmed by long handling times and seasonal
query surges. Agents spent significant time searching through disparate document repositories for
brand- and destination-specific information, especially during peak seasons.

10–12 minutes average handling time per customer inquiry	25–30% of inquiries require brand- or market-specific responses	Inability to scale operations during 2X seasonal surge in query volumes

Solutions

Sonata implemented a comprehensive generative AI-powered agent assist solution that streamlined
the entire customer inquiry response process. The system utilized advanced NLP with sophisticated
context filtering to understand specific brand, source market, country, and booking segment details for
each question. Through a RAG approach, the solution rapidly located, prioritized, and generated
relevant answers from a tailored knowledge base, ensuring accurate, context-sensitive responses while
reducing agent dependency.

Solution architecture

EventBridge scheduler triggers automated article processing every set hour, triggering the "kb-articlesprocessor" function	Lambda "kb-articles-processor" sends GET request to Mindtouch to get a unsorted list of articles	Lambda gets articles from the response and creates individual kb-article-work-items that are put on the queue
Lambda "kb-articles-processor" upserts record in mindtouch-extraction-service-log with "pageid" as key	Lambda "kb-article-work-item-processor" gets messages from queue (each work item is processed individually)	

Technology used

Harmoni.AI features

- Sanitation of user PI information
- User consent & awareness
- Content moderation
- Data security – user authentication
- Dashboard telemetry

Results that Speak Volumes

Enhanced customer satisfaction through faster, accurate responses	Greater operational agility during peak periods	Improved consistency in brand- and region-specific query handling	Stronger compliance with user data protection and moderation standards

Results that Speak Volumes

50-60%* Reduction in Average Handling Time (AHT)	**20–25%** of inquiries managed directly by the agent assist system	Eliminated need for seasonal resource ramp-up