



CASE STUDY

# Ensignt Improves Insurance Proposal Design Quality

—  
With End to End LLMOps

A conversation-first workflow accelerates resolution time

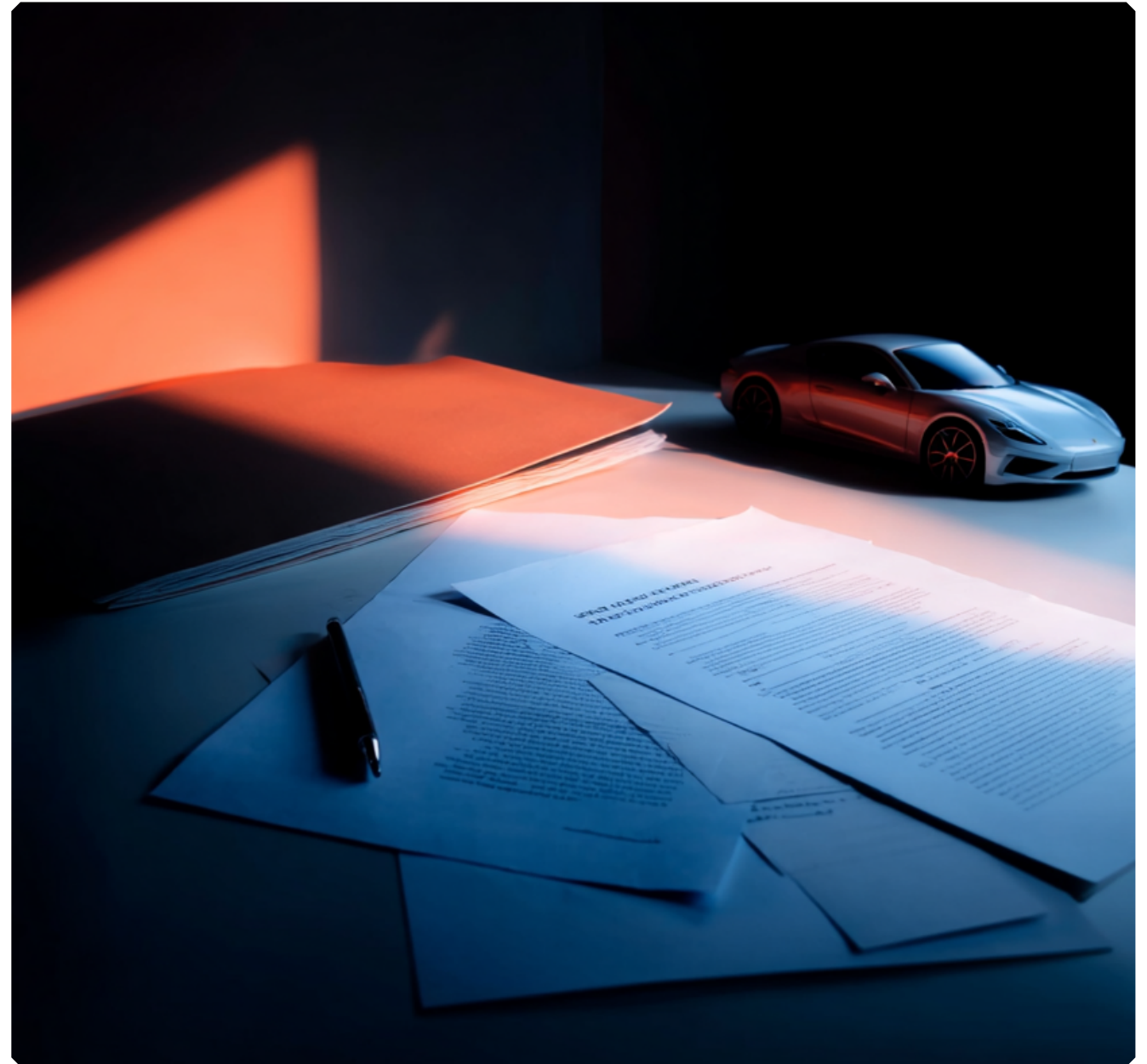




Ensignt supports financial professionals and insurance carriers with solutions that streamline sales for life insurance, long-term care, and annuities. As part of this effort, Ensignt developed ENSI, a GenAI-powered chatbot that serves as a virtual case designer for advisors and those who assist them creating sales proposals. The assistant supports intake, product clarification, and quote preparation through a single conversational interface that remains available outside standard business hours.

Advisors can paste client information directly into ENSI or speak to ENSI rather than completing structured forms. The assistant parses the input, identifies missing attributes, and requests only the information required to proceed. This interaction reduces email exchanges, keeps cases moving, and supports reliable and timely sales proposals.

As ENSI adoption increased, Ensignt focused on strengthening reliability, quality measurement, and operational visibility. The goal was to support continuous improvement while maintaining consistent performance as usage expanded. To meet this need, Ensignt worked with SoftServe to develop the full GenAI chatbot solution and the supporting LLMOps foundation.





## Barriers to a smooth advisor experience

ENSI already offered a promising path toward faster proposal design. To deliver the full value, Enight wanted to address the following work pattern challenges advisors faced day to day:



Routine questions slowed advisors down and reduced time for client conversations.



Traditional web forms interrupted workflow and made client profiling slower.



Email exchanges delayed finalization of proposals and often pushed work into the next day.



Clarifying product information across carriers required additional follow-up.



Model behavior became harder to track as usage expanded.



Manual testing could not cover the full range of real user paths.

Enight needed a structured way to observe every interaction and evaluate quality at scale.

## A Gen AI chatbot built around real advisor workflows

The team developed a complete GenAI chatbot solution that supports case design from intake through quote preparation. The solution focused on clear interaction flow, predictable behavior, and traceable outcomes.

The assistant includes:



### Conversational proposal design

ENSI guides advisors through proposal creation in a single conversational flow. The assistant identifies missing information, asks only for what is required, and keeps progress moving without unnecessary steps.



### Paste-first data capture

Advisors drop client information directly into the conversation. ENSI parses the content, detects gaps, and gathers only the details needed to generate accurate quotes.



### 24x7 advisory support

Advisors can progress cases outside business hours. Clients receive answers sooner, and conversations continue without waiting for the next day.



### Embedded product knowledge

ENSI draws on detailed carrier product information and helps advisors compare options without additional outreach. It also makes this information more accessible to those that prepare proposals for their advisors.

This solution established the functional foundation of ENSI. To support quality at scale, Enight added a complete LLMOps layer.



# LLMOps foundation supporting reliability and continuous improvement

Enight implemented an LLMOps foundation that governs how the GenAI chatbot is evaluated, tested, observed, and improved over time. This foundation connects every conversation to measurable quality signals.

## CHATBOT ARCHITECTURE AND TECHNOLOGIES

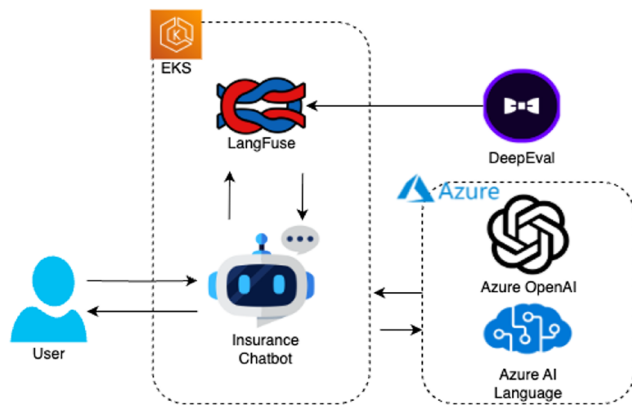


Figure 1. Tracing and evaluation solution

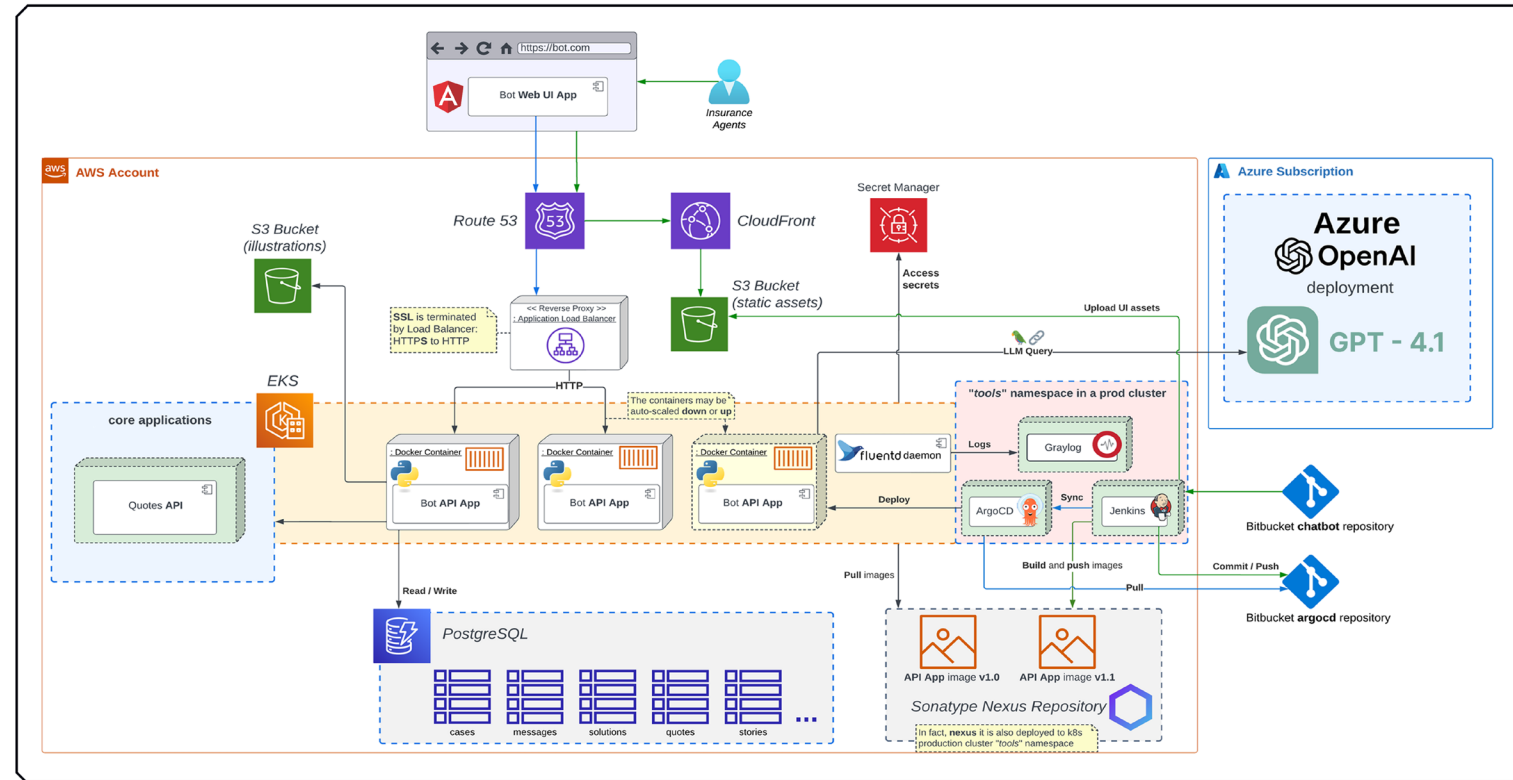


Figure 2. Chatbot architecture



## CONVERSATION TRACING



Every ENSI session is recorded from start to finish. Tracing includes routing steps, model calls, and intermediate decisions, giving Enight full visibility into how conversations unfold. Session-level and message-level cost tracking provides clarity into the operational footprint of each interaction.

## CENTRALIZED PROMPT MANAGEMENT



Prompts are stored in a single organized library with full version histories. This structure helps the team track how prompt changes influence real conversations and compare performance across iterations.

## USER FEEDBACK LOOP



A built-in rating tool allows advisors to provide quick feedback. Each rating links directly to the relevant conversation trace, enabling the team to understand the context behind positive and negative experiences.

## METRICS AND AUTOMATED EVALUATION



ENSI uses automated scoring for every completed or idle conversation. DeepEval evaluates interactions against five metrics based on real advisor needs. These metrics reveal where conversations stall and where improvements are needed:

Agent Assignment Gate

Required Attributes Completeness

Disbursement Zero Validation

Proactive Guidance

Bot Repetition Detection

## AUTOMATED CONVERSATION SIMULATION



A virtual insurance agent runs predefined scenarios through ENSI using three behavioral styles. These simulations validate parsing accuracy, confirm that required information is collected, and help Enight detect regressions before they reach production.

## ANONYMIZATION VALIDATION



Enight evaluated a PII-redaction approach to protect sensitive information in traces. The method showed strong detection accuracy. Future improvements to asynchronous processing will support production deployment.



# Measurable results driven by LLMOps best practices

The combination of a complete GenAI chatbot solution and a structured LLMOps foundation produced clear operational results:

**57**

automated test cases and growing completed across three user styles

**16x**

faster regression testing - completed in 2 hours instead of four business days

Earlier detection of parsing and attribute-collection issues

Repeatable, consistent quality control across all scenarios

Evaluation, testing, and feedback loops allowed Enight to refine ENSI with confidence and maintain stable performance as usage expanded.



## Ready to strengthen your AI operations?

SoftServe helps teams design, operate, and improve GenAI chatbot solutions through structured LLMOps practices. These practices support evaluation, feedback, testing, and visibility required for long-term production use.

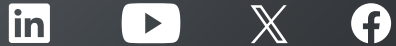
Connect with SoftServe to explore how an end-to-end LLMOps foundation supports dependable conversational AI at scale.



## About SoftServe

SoftServe is a premier IT consulting and digital services provider. We expand the horizon of modern technologies to solve today's complex business challenges and achieve meaningful outcomes for our clients.

## Social Links



[info@softserveinc.com](mailto:info@softserveinc.com)  
[www.softserveinc.com](http://www.softserveinc.com)

## Contact

### **NORTH AMERICAN HQ**

201 W 5th Street, Suite 1550  
Austin, TX 78701  
+1 866 687 3588 (USA)  
+1 647 948 7638 (Canada)

### **EUROPEAN**

30 Cannon Street  
London EC4 6XH  
United Kingdom  
+44 333 006 4341