HEALTHCARE AI AGENTS

Amna khan

GHUNCHAS

Generative Answers & Data Grounding

Generative Answers are a key capability of the Healthcare Agent Service. They allow the agent to generate natural language responses while ensuring that the answers are grounded in trusted, verified data sources. This module explains how generative AI is safely used in healthcare scenarios.

Generative Answers Overview

Purpose of Generative Answers

Generative Answers enable the healthcare agent to provide natural, conversational responses to user questions while remaining accurate and reliable.
Instead of returning static or prewritten replies, the agent dynamically generates answers based on relevant data and context.

In healthcare, this allows:

- More flexible and user-friendly conversations
- Clear explanations of medical information
- Faster access to relevant healthcare knowledge

Grounded AI vs Generic AI Responses

Generic AI models generate responses based only on training data, which can lead to hallucinations or incorrect medical information.

Grounded AI, used in Healthcare Agent Service:

- Retrieves information from trusted sources first
- Uses that information as context for generating responses
- Ensures answers are evidence-based and verifiable

This grounding approach is critical in healthcare, where incorrect information can lead to serious risks.

Importance of Trusted Data

Trusted data is the foundation of safe generative AI in healthcare.
The system ensures that all generated responses are based on:

- Verified medical documents
- Approved organizational data
- Trusted public health websites

By grounding responses in trusted data, the agent delivers accurate, consistent, and compliant healthcare information.

Generative Answers on Customer Sources

Generative Answers on Customer Sources allow organizations to use their private and internal data to power healthcare conversations.

Use of Private Organizational Data

Healthcare organizations can use internal data such as:

- Hospital guidelines
- Policy documents
- Patient education PDFs
- Internal medical protocols

This ensures that the agent provides responses aligned with organizational standards and practices.

Azure AI Search (Vector Search)

Azure AI Search is used to index and retrieve customer data efficiently.

Key capabilities include:

- Vector-based semantic search
- High-speed retrieval of relevant documents
- Support for large and complex datasets

Vector search allows the system to understand meaning, not just keywords.

Azure OpenAI Integration

Azure OpenAI is responsible for generating natural language responses.

The process works as follows:

1. Azure AI Search retrieves relevant content
2. The content is passed to Azure OpenAI
3. Azure OpenAI generates a contextual response

This integration ensures high-quality AI responses while maintaining enterprise security.

Secure and Compliant Responses

Security and compliance are enforced at every step:

- Data remains within the organization's Azure environment
- No customer data is used to train AI models
- Responses include disclaimers and evidence references

This makes generative answers suitable for healthcare compliance requirements such as HIPAA and GDPR.

Customer Source Workflow

Uploading Documents (PDFs, Text, Images)

Organizations can upload:

- PDF documents
- Text files
- Image-based documents

These documents form the knowledge base for generative answers.

Vector Embedding Generation

Uploaded documents are converted into vector embeddings, which represent semantic meaning.

This enables:

- Meaning-based search
- Improved relevance of answers
- Better handling of medical terminology

Semantic Retrieval

When a user asks a question:

- The system retrieves the most relevant content using vector similarity
- Results are based on meaning, not exact words

This improves answer accuracy.

Context-Aware Answer Generation

The retrieved content is combined with:

- User query
- Conversation history

Azure OpenAI then generates a clear, contextual answer grounded in the retrieved data.

Generative Answers on Public Sources

Generative Answers on Public Sources allow the agent to safely access approved public medical websites.

Bing Custom Search Integration

Bing Custom Search is used to:

- Search approved medical websites
- Control which domains can be accessed
- Prevent access to unverified sources

This ensures safe and relevant web-based answers.

Domain-Restricted Web Queries

Only pre-approved domains are allowed.

Examples include:

- Government health websites
- Medical research organizations
- Trusted healthcare institutions

This prevents unreliable or misleading sources.

Approved Medical Websites Only

The system blocks:

- General web searches
- Unverified blogs or forums

This guarantees medical accuracy and trust.

Public Source Configuration

1. Custom Search Configuration

Administrators configure:

- Search instance
- Approved websites
- Search scope

This controls how public data is accessed.

2. Allowed Domain Control

Only specified domains are permitted, ensuring content quality.

3. Controlled Web-Based Responses

All responses:

- Include evidence attribution
- Display AI disclaimers
- Provide user feedback options

Scenario Management & User Experience

Scenarios define how users interact with the healthcare agent. This module focuses on managing scenarios and designing effective user experiences.

Built-in Scenarios

Healthcare Agent Service includes built-in scenarios designed for common healthcare needs.

1. Symptom Checker

Guides users through structured questions to assess symptoms and suggest next steps.

2. Medical Condition Information

Provides verified explanations of medical conditions using credible sources.

3.Drug and Side-Effect Information

Delivers accurate information about medications, usage, and side effects.

4.Triage Protocols

Helps determine urgency and directs users to appropriate care levels.

Custom Scenario Creation

Scenario Editor Overview

The Scenario Editor is a visual tool used to design conversation flows.

Visual Flow-Based Design

Developers create scenarios using:

- Nodes
- Conditional paths
- Logical decision points

This makes scenario design intuitive and manageable.

Scenario ID and Metadata

Each scenario includes:

- A unique ID
- Metadata for orchestration and analytics

This helps the orchestrator identify and trigger the correct scenario.

Scenario Elements

Conversational Elements

Includes:

- Statements to provide information
- Prompts to ask user questions

Flow Control Elements

Control conversation logic such as:

- Conditions
- Branching paths

Navigation and Debugging Elements

Help test and troubleshoot scenarios efficiently.

Best Practices for Scenario Design

1. Modular and Reusable Design

Design small, reusable scenarios to simplify maintenance.

2. Performance Considerations

Avoid overly complex flows to maintain responsiveness.

3. Debugging and Testing Strategies

Regular testing ensures:

- Correct flow execution
- Safe responses
- Good user experience