HEALTHCARE AI AGENTS

GHUNCHAS

Core Architecture & Components

Healthcare Agent Service is designed specifically for the healthcare domain, where accuracy, safety, trust, and compliance are critical. Its architecture combines advanced AI capabilities with strong governance, ensuring that AI-powered agents can be safely used in real clinical and healthcare environments.

Core Components Overview

Healthcare Agent Service consists of several core components. Each component has a clearly defined role and together they form a complete, secure, and intelligent healthcare AI system.

Healthcare Orchestrator

The Healthcare Orchestrator is the central control layer of the Healthcare Agent Service. It manages how the agent understands user requests, selects data sources, triggers workflows, and generates responses.

Unlike traditional chatbots that follow fixed scripts, the orchestrator uses intelligent reasoning to dynamically decide:

- What the user is asking for
- Which scenario should be activated
- Which plugin or knowledge source should be used
- How to apply safeguards before responding

This orchestration layer ensures that every response follows healthcare safety standards and organizational rules.

Healthcare Intelligence

Healthcare Intelligence refers to the AI reasoning capability of the system that understands medical language and healthcare-specific contexts.

This intelligence enables the agent to:

- Interpret medical terms and symptoms
- Understand patient and clinician questions
- Respond in a medically appropriate tone

It is powered by Large Language Models (LLMs), but unlike general-purpose AI, this intelligence is constrained and guided to avoid unsafe or misleading medical responses.

Credible Sources

Credible Sources are trusted, verified sources of medical information used by the agent to ground its responses.

These sources ensure that the AI does not rely on guesses or unsupported knowledge. Examples include:

- Government health organizations (CDC, FDA)
- Medical research libraries (US National Library of Medicine)
- Hospital-approved documents and internal PDFs

By grounding answers in credible sources, the system reduces hallucinations and increases medical accuracy.

Health Safeguards

Health Safeguards are a critical part of Healthcare Agent Service. They are built-in mechanisms that ensure responsible and ethical AI use in healthcare.

These safeguards include:

- AI disclaimers to clarify that responses are not a substitute for professional medical advice
- Evidence attribution so users can see where information comes from
- Clinical checks to avoid unsafe recommendations
- Feedback options so users can report issues

Health safeguards operate automatically and cannot be bypassed, ensuring consistent safety.

Scenarios & Scenario Editor

Scenarios are structured conversation flows designed to handle specific healthcare tasks.

Examples of scenarios include:

- Symptom checking and triage
- Patient education
- Appointment guidance

The Scenario Editor allows developers and healthcare organizations to:

- Design step-by-step conversation flows

- Control how questions and answers are presented
- Combine AI responses with rules and workflows

This ensures consistent, predictable, and safe agent behaviour.

Plugins (Built-in, Customer, OpenAPI)

Plugins allow the healthcare agent to connect with external systems and services.

Types of plugins include:

- **Built-in plugins**: Provided by Microsoft (e.g., healthcare knowledge services)
- **Customer plugins**: Organization-specific systems like hospital databases
- **OpenAPI plugins**: Third-party APIs

Plugins enable the agent to perform real-world actions such as retrieving records, searching documents, or triggering workflows.

Healthcare-Oriented System Design

Healthcare Agent Service is not a generic AI platform. It is specifically designed to meet the unique requirements of healthcare environments.

Healthcare-Adapted Orchestration Logic

The orchestration logic is adapted to healthcare needs, where safety and compliance are more important than creativity.

This logic ensures:

- Conservative and responsible responses
- Avoidance of diagnosis or treatment claims
- Escalation to professionals when needed

The system prioritizes patient safety over speed or flexibility.

LLM-Driven Decision Making

Large Language Models (LLMs) play a key role in understanding natural language and generating responses.

However, in Healthcare Agent Service:

- LLMs are guided by orchestration rules
- Outputs are validated through safeguards

- Responses are grounded in trusted sources

This controlled use of LLMs ensures reliable and safe AI behaviour.

Integration with Azure OpenAI

Healthcare Agent Service integrates with Azure OpenAI, which provides enterprise-grade AI models.

Azure OpenAI ensures:

- Secure data handling
- No use of customer data for model training
- Compliance with healthcare regulations

This makes it suitable for sensitive healthcare data and conversations.

Extensible and Modular Architecture

The architecture is modular, meaning components are loosely coupled and easily extendable.

This allows organizations to:

- Add new data sources
- Integrate additional plugins
- Scale agents across departments

The modular design supports long-term growth and customization.

Healthcare Orchestrator & Conversation Flow

Module 3 focuses on how the Healthcare Orchestrator manages conversations, routes queries, and ensures safe responses.

Healthcare Orchestrator

Role of the Orchestrator

The orchestrator acts as the decision-making engine of the healthcare agent.

It ensures that:

- User intent is correctly understood
- Appropriate workflows are triggered
- Safety policies are enforced

Without the orchestrator, AI responses would be uncontrolled and unsafe.

Intent Detection and Routing

When a user sends a query, the orchestrator analyses the intent behind the message.

For example, it identifies whether the user is:

- Asking for medical information
- Seeking appointment help
- Reporting symptoms

Based on this intent, the query is routed to the correct scenario or plugin.

Plugin Selection Mechanism

The orchestrator automatically selects plugins based on:

- Instruction titles
- Plugin descriptions
- Context of the conversation

This allows seamless integration of multiple systems without manual intervention.

Disabling Conflicting Language Models

To avoid inconsistent or unsafe outputs, conflicting or unnecessary language models are disabled.

This ensures:

- Consistent response behavior
- Clear control over AI output
- Alignment with healthcare safety policies

Orchestrator Workflow

User Query Analysis

The workflow starts when a user submits a question. The orchestrator analyses:

- Language
- Intent
- Urgency

This analysis sets the direction for the entire response process.

Context and History Evaluation

The orchestrator evaluates:

- Previous messages
- Conversation context
- User-provided information

This helps generate relevant and coherent responses.

Intelligent Source Selection

The orchestrator selects the best source to answer the query:

- Credible medical databases
- Customer documents
- Approved public sources

This selection ensures accuracy and trust.

Response Generation with Safeguards

The final response is generated using AI and then passed through health safeguards.

Each response includes:

- A disclaimer
- Evidence references
- Feedback options

This guarantees transparency and safety.

Orchestrator Fallback Mechanisms

Generative Credible Fallback

If the primary source cannot answer the query, the orchestrator uses generative answers grounded in credible sources.

This ensures continuity without sacrificing accuracy.

Fallback to Customer Sources

If public credible sources are insufficient, the system falls back to organization-specific documents such as internal guidelines or PDFs.

Fallback to Built-In Sources

As a final option, built-in Microsoft-approved sources are used to ensure the user still receives a helpful response.

Control and Configuration Options

Administrators can configure:

- Fallback order
- Allowed data sources
- Safety thresholds

This gives full control over agent behaviour.