AI CONCIERGE AGENT HELPS HR CONNECT WITH EMPLOYEES

What is a concierge agent

A concierge agent is an AI-enabled, natural language interface that provides information and services to users. More than just a chatbot, a concierge agent allows users to interact with organizational data and systems by using generative AI to assist with common tasks and dynamically synthesize and contextualize responses.

What the AI concierge agent does

Catalyte's AI concierge agent uses Workday and internal documentation to handle basic HR inquiries and improve the flow of knowledge. Employees can use the agent 24/7 to ask questions, request time off and perform other HR-related tasks.

How it does it

The concierge agent leverages the Azure AI platform. The user submits requests for HR information to an orchestration layer which "decides" how requests should be satisfied, either via search results of indexed corporate documentation or through API calls to Workday. It then presents the most relevant information to the user.

Benefits of a concierge agent

The concierge agent provides 24/7 support, without limiting employees to office hours or physical locations - ideal for Catalyte's 100% remote workforce. It also provides a more convenient and personalized employee experience, increasing satisfaction and engagement. Because the agent automates routine HR tasks, it reduces the HR team's administrative burden and operational costs.

Solution architecture

The concierge agent is a combination of time-tested software design principles and novel AI service offerings.

Isolating solutions components within appropriate Microsoft Azure resources makes it scalable, secure and flexible.
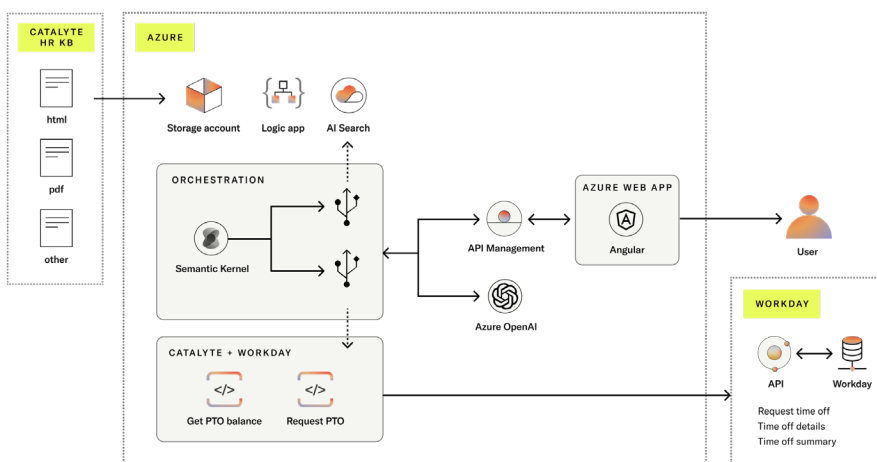
1. User interface

User inquiries for HR information originate in the solution's front end, a conversational concierge. It was created with the Angular framework and hosted in an Azure Web App. This allows for:

+ Dynamic data rendering without the need for page reloads

+ Seamless deployment with potential for different development environments

+ Simple, user-friendly design aligned with Catalyte's branding

2. Messaging gateway

User requests from the front end are passed to a centralized messaging gateway. Here, Azure API Management securely and efficiently makes the solution's inner workings available. This approach allows for:

+ Enforcement of security policies and rate limits

+ Protection of downstream components by managing authentication, authorization and other security measures

+ Analytics and logging capabilities to monitor the performance and usage

3. Orchestration

The orchestration layer is responsible for accurately answering user questions based only on configured resources. Inquiries are classified as those that can be satisfied by searching indexed HR content and/or interacting with Workday. Hosted as an Azure Function App, Microsoft's Semantic Kernel performs this function. All requests and responses which pass through the agent benefit from Azure OpenAI's advanced natural language processing (NLP) capabilities.

4. Search

In building the concierge agent, HR documents were persisted to Azure Blob Storage, where the content was indexed and vectorized. Requests satisfied by searching HR content are dynamically routed to Azure AI Search. This allows for:

+ Text and images to be searchable

+ Integrated optical character recognition to convert images to text for better indexing and search accuracy

+ Semantic search and improved retrieval of relevant information by applying vectorization techniques to text content

5. Workday

Requests pertaining to information residing in Workday are dynamically routed to Catalyte's internal APIs. Developed in C#, .NET and hosted as an Azure Function App, these custom components manage authentication and communicate with Workday's REST APIs. This approach provides a secure connection to the Workday cloud service for retrieving employee information while also providing interfaces to enable the concierge agent to act on an employee's behalf.

CATALYTE

Additional design considerations

For example, Microsoft's Azure OpenAI Service provides access to the capabilities offered by large language models (LLM). This allows for the intuitive use of human language to interact with technology, also known as natural language processing (NLP). Retrieval augmented generation (RAG) further refines the capabilities offered by natural language processing (NLP). Semantic Kernel governs the runtime decision-making and determines the services needed to satisfy user requests. Lastly, the solution's foundation rests on offerings provided by Microsoft's Platform-as-a-Service (PaaS).

1. Azure OpenAI

Azure OpenAI provides access to the solution's underlying LLM (GPT 4o) by processing all incoming requests and outgoing responses. Leveraging an LLM that has been trained on vast amounts of diverse text data enables advanced capabilities for understanding, generating and interacting with human language. By integrating LLMs, businesses can create more sophisticated and intuitive applications that understand nuanced language, provide contextually relevant responses and enhance user experiences.

2. Azure AI Search

Azure AI Search was used to index and vectorize content. This involves transforming textual data into numerical vectors that represent the semantic meaning of content. Azure AI Search uses advanced NLP and machine learning techniques to analyze and index vast amounts of data, creating a searchable vector space. By vectorizing content, Azure AI Search enhances search relevance and enables advanced features such as semantic search, where results are ranked based on meaning rather than just keyword matching. This improves the user experience and the accuracy of information discovery.

3. Retrieval-augmented generation

Retrieval-augmented generation integrates line-of-business (LOB) systems, like Workday, using NLP to generate queries. This bridges the divide between unstructured human language and structured data sources. Users of the concierge agent can interact with complex enterprise systems using natural language.

The NLP engine interprets the intent and dynamically generates precise queries to retrieve relevant data from LOB systems. This simplifies access to critical business information and allows for more intuitive interactions.

4. Semantic Kernel

Semantic Kernel is a lightweight, open-source development kit. With it, developers can build AI agents and integrate the latest AI models into C#, Python and Java codebases. It enables developers to combine NLP, contextual understanding, and machine learning with code, allowing for the coordination of various services, tasks and API calls. By leveraging Semantic Kernel, the concierge agent autonomously manages more complex operations, adapts to evolving inputs and enhances user interactions with contextually relevant responses and actions.

Semantic Kernel's native points of extensibility, called plugins, play a particularly important role in the solution's adaptive capabilities. Plugins allow the solution to intelligently select and activate different tools, models, or functionalities depending on the specific needs of a user's input. This results in more flexible and context-aware responses, as the AI can identify the most appropriate resource to handle a given task. With the ability to dynamically switch between plugins, the solution can provide more accurate, efficient and relevant outcomes that meet the user's original intent.

5. Service orientation

Service-based architectural principles involve designing systems where individual services are modular, loosely coupled and focused on specific business functions. Each service operates independently, whether responsible for user authentication, NLP or integrating with external APIs. This approach encourages flexibility and scalability while enabling orchestration of various services.

6. Azure platform-as-a-service

Microsoft's PaaS capabilities allow organizations to build, deploy and manage applications with greater efficiency and scalability, without the need to manage the underlying infrastructure. Azure PaaS provides a comprehensive environment that includes pre-configured services for computing, storage, networking and databases. It also provides integrated tools for development, testing, and deployment. By utilizing Azure PaaS, businesses can accelerate time-to-market, reduce operational overhead and focus more on innovation and application development. This approach supports rapid scaling, ensures high availability and the ability to integrate with other Azure services.

CATALYTE