amazon

Multi-Agent Design Guide

Table of Contents

VII Mission

In a world with multiple voice services each with different capabilities, we believe customers should have the freedom to choose their preferred services. We aim to deliver this experience through multiple, simultaneous voice services on the same product, each with its own 'wake word' or invocation name—enabling customers to talk to the service of their choice in a secure manner by **simply saying its name.**

VII Pillars

Customer Choice

Building voice-enabled devices that promote customer choice and flexibility through multiple, simultaneous wake words.

Secure Interoperability

Developing voice services that can work alongside others while protecting the privacy and security of customers.

Technology Solutions

Releasing technologies and solutions that make it easier to integrate multiple voice services on a single product.
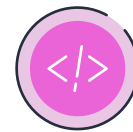
Research and Development

Accelerating machine learning and conversational AI research to improve the breadth, quality and interoperability of voice services.

Introduction

Tip

This guide is not intended to be a comprehensive guide to designing voice agents or the products that use them. Nor will it include specific technical solutions.

Voice agents are a rapidly growing part of the world of consumer products and services. With multiple voice agents available on a single device, each with their own strengths, experiences, and personalities, customers have more options than ever to engage with them.

The increasing number and variety of voice agents does present a challenge to product and agent designers, though. How can we offer customers a set of voice agents whose fundamental interactions and behaviors that allow great customer experiences? Can we simplify requirements and identify design consistencies to reduce development effort?

How to Use this Guide

The VII Design Guide focuses on behavior and interaction design challenges that occur when multiple simultaneous agents exist on a single device. Building on the ideas and experiences of VII members, this Design Guide is intended to:

- Present a common framework and terminology to help facilitate ongoing discussions about multiple simultaneous agent experiences. A Glossary is included at the end of this document.

- Capture device and voice agent design guidance, principles, and best practices needed for multiple voice agents to coexist on a device.

- Empower engineers, designers, and product developers to find innovative ways to achieve even greater benefits for customers.

- Represent a step toward sharing knowledge to improve customer experiences and reduce development effort and cost.

The minimum set of design recommendations for agents to co-exist simultaneously on a device are gathered in Baseline Guidance for each section. In addition, the other recommendations and best practices presented are intended to improve the customer experience.

Customer Choice & Agent Invocation

Customer choice is a bedrock principle of the Voice Interoperability Initiative. All available agents a customer wishes to interact with should be available to them simultaneously, and should be invocable and identifiable by their own unique wake words.

Baseline Guidance

1. A customer should be able to choose from available voice agents for a particular interaction. They should have the option to use multiple simultaneous wake words when more than one agent is registered on a device.

2. Multiple simultaneously registered agents should be available to customers at all times, aside from the following exceptions:

    a. When one agent has been invoked and is actively streaming a customer utterance to the cloud, no other agent's wake word should be detectable. For example, if a customer says "Agent 1, tell me about Agent 2," Agent w2 should not be invoked.

    b. An agent should not be able to invoke any other agent by distributing the wake word via TTS. For example, one agent cannot wake up another agent by speaking its wake word.

3. When an agent is in Speaking state, responding to a customer, the customer should be able to interrupt that agent's response with any other active agent's wake word (barge in).

Customer Choice & Agent Invocation

Agent Preference

Key to the VII pillar of customer choice is the idea that customers can, at any time, use any of the agents registered on their device. Having a selection of agents to choose from, with different capabilities covering a variety of "hero" use cases, offers an individualized experience which customers can explore in their own ways. There may also be situations in which customers do not use a specific wake word to start an interaction, for example when pressing an Action button. In those cases, the product may designate, or allow a customer to choose, which agent will respond.

Invocation

When creating simultaneous multi-agent experiences, it is important to design devices and voice agents which can handle the complexities of invocation, from the variety of wake words, to the availability and possible overloading of buttons. Invoking a specific agent should involve simple, easy-to-remember methods.

Wake Words

For many devices, wake words are going to be the primary method customers use to invoke an agent. The wake word, along with the agent's voice, is also a significant component of persona and brand. If your customers are going to be interacting with your agent using a wake word, keep in mind the following:

- Wake words should be distinct from each other, and easy to remember, both to avoid customer confusion and to minimize the chance that a device will mistake one wake word for another.

- If you allow customers to choose alternative wake words to invoke an agent, be aware of other agent wake words and alternative options.

Multi-Agent Experiences

This section of the Design Guide addresses the fundamental behaviors that multiple simultaneous agents should employ to provide engaging interactions for customers. These are baseline behaviors, not comprehensive, and are intended to help you to create innovative multi-agent experiences.

Design Foundations

While creating simultaneous multi-agent experiences, use these foundations to help guide your design decisions.

Privacy

When customers interact with multiple agents on a device, it is important to offer them transparency in their experiences. Product makers should ensure that customers know which agent they are interacting with. If a multi-agent experience requires the sharing of any data between agents, it should be made clear to customers that such sharing is taking place, and they should be allowed to give consent.

Predictability

Interacting with multiple agents across different devices may seem more complex to customers. We should strive to avoid confusion, extra cognitive load, and conflicting behaviors. We want to offer customers delightful surprises, not frustrating ones. To help them navigate the complexities, products and voice agents should interact in expected and familiar ways.

Ease of use

Customers should be able to interact intuitively with their agents without having to remember complex rules or confusing guidance. Interactions and commands should feel natural, and customers should have easy access to device controls.

Multi-Agent Experiences

Discovery and Education

Beyond just showing customers how to use a voice agent, a multi-agent product will have to help customers find the agents that are available to them, choose among them, and explore what they are capable of. What is the most effective way to show customers how to get the most out of the voice agents on your product?

Baseline Guidance

1. Customers should easily be able to discover information about the primary uses, benefits, and capabilities of available agents.

2. Customers should be made aware of any multi-agent functionality supported by the device:

   a. Customers should be informed of simultaneously available wake words.

   b. Customers should be informed of Universal Device Command support.

Agent Discovery

Because voice is a largely invisible modality that relies heavily on customer recall, it is important to be clear about how to engage with multiple voice agents, and to get customers excited about the possibilities. The registration and education flow, available to customers either during the product's out-of-box experience (OOBE) or afterward, might be a good place to introduce your customers to available agents. Including additional discovery mechanisms is another great way to encourage customer engagement.

Multi-Agent Experiences

Agent Selection and Registration

- Encourage customers to explore available agents, for example by listing the value or strengths, and including example utterances.

- Be aware of the steps needed for a customer to choose and register agents, and make the process as easy as possible, especially if it occurs during the OOBE process.

- Provide an easily accessible list of registered agents.

Agent Education

- Describe the strengths of each agent. Assigning roles to each agent may help customers understand how best to use multiple agents, and make it easier for them to remember how to interact with them.

- For the example utterances you present, choose utterances with a high value for that particular agent, showcasing unique or valuable use cases or experiences.

Multi-Agent Experiences

Companion App

Much of the customer education and initial settings choices are handled in a product's out-of-box experience (OOBE). The OOBE setup flow can cause cognitive overload for customers due to the many decisions they must consider. For customers who choose not to complete the voice setup during OOBE, or who want to change their settings, offer them an easy way to access those settings at any time, for example in a companion app.

A successful device companion app should allow customers to:

• Set up and manage voice agents at any time.

• Learn how to use their voice agents.

• Access simple controls and preferences for voice agents at any time.

Companion App Best Practices

Locate voice and agent settings in a single, easy-to-find area of the companion app. All products should provide the following settings and controls from within their device's companion app:

• The ability for the customer to register voice agents.

• The ability for the customer to configure key settings for each voice agent (e.g. language selection).

Multi-Agent Experiences

Settings and Preferences

Settings and preferences should be easy for customers to find, understand, and change. They should also include easy to understand terminology and controls for all the agents available for a device.

Device Maker Considerations

The device maker should supply the customer a place to manage their voice agent-specific settings.

- Customers should be able to register voice agents at any time.

- Customers should be able to change agent settings, preferences, and manage voice history (if provided), or should be provided a link to a place to do that.

- Agent-specific settings should be located together in an easy-to-find area.

Agent Maker Considerations

Agent makers should allow customers to change settings and preferences specific to their voice agent.

- Provide to device makers the information necessary to set up the voice agent.

- Provide material about the agent's "hero" use cases or unique capabilities to be included in the product's education flow.

- If an agent provides a voice history, customers should have access to view and manage their voice history.

Multi-Agent Experiences

Agent Attribution and Branding

In multi-agent experiences, customers should always know what agent they are talking to. Agent attribution may be explicit (e.g. colors, logos), or it may be indicated by differentiated visual and sound cues. Attribution can also include both personality and behavioral characteristics unique to an agent.

Branding

A little bit goes a long way. Be aware of the effect of adding too many distracting branding personas.

The Wake Word

The wake word is a strong brand indicator, and the choice of wake word is an important part of the experience.

Visual attribution

Visual attribution can provide a better customer experience in many cases. Use verbal attribution only when visual cues are not available or when it is of critical importance to know (for example when presenting personal or sensitive information, or when the veracity of information is crucial).

Multi-Agent Experiences

Human Interface Design

Human interface design considerations contribute significantly to a customer's sense of privacy and understanding. These considerations involve the customer interacting with or receiving information from the device itself, not from a voice agent. Products that support more than one simultaneous voice agent have unique challenges both in designing usable controls, such as buttons, and in representing their current state with attention system displays and audio cues.

Physical User Interface

The following best practices are designed to ensure that your customer always knows when a device is active and detecting wake words. These recommendations are vitally important to maintaining customer trust.

Microphone

We strongly suggest that all products which allow hands-free voice activation (not push-to-talk only) have a microphone on/off control.

- Microphone control should be universal, i.e. the microphones are turned on or off for all agents and device functions.

- It is also recommended to provide on/off controls for other device inputs, such as cameras or sensors.

Action

Where appropriate, products should include an "Action" button that functions for each active agent. The Action button should afford the following functions:

- Initiate a new voice interaction

- Interrupt responses and media output from any and all agents

- Stop a sounding Alert.

Volume Adjust

All devices that output sound should include a physical control to adjust the universal volume, affecting all agents.

Multi-Agent Experiences

💡 **Tip**

It is strongly recommended that devices do not overload the microphone on/off button. It is also recommended not to overload the Action button.

Button Interactions

It is recommended that buttons or other controls that interact with agents, such as Play and Pause buttons, do so consistently between agents. Devices should not implement separate sets of similar controls for different agents. Note that this will require that the device be able to direct the button press command to the proper agent. These best practices apply whether the buttons are physical or virtual, and they also inform the decisions about which commands map to Universal Device Commands.

Overloading

Overloaded buttons, or other controls, may have more than one function or be able to invoke more than one agent. They may behave differently based on a mode the device is in, or on the length or pattern of pressing the button. Overloaded buttons are intrinsically difficult to use for customers, who must then remember both the extra functions of the control, as well as more than one method of interaction. With multiple agents on a device, there is a risk of even more complicated interactions. Overloaded buttons should be avoided when possible. If you must overload a button, you should:

- Provide clear and repeatable instructions about the function and use of the button

- Group similar functions to a single control

- Use a label, icon, or some other indication that the button has multiple functions

- Keep the interaction patterns simple and easy to remember, such as a long or short button press.

Multi-Agent Experiences

Universal Device Commands

Universal Device Commands (UDCs) are those commands and controls that a customer may use with any compatible agent to control certain device functions, even if the agent was not used to initiate the experience. UDCs can broadly be classified in two categories:

- Device global commands (e.g. changing the device's volume) that are implemented by each agent separately.

- Cross-agent commands (e.g. stop a sounding timer that was started by another agent) that may require state information to be shared from the device to enable agents to properly interpret the request.

UDCs are a necessary feature for devices with multiple active agents, and aim to satisfy customer expectations and solve for the most common frustration points and address customer expectations. For example, imagine one person sets an alarm and then leaves the room. Then another person enters the room, hears the alarm sound, and wants to turn it off. The interaction should be possible using any compatible agent, and should not result in an agent telling the person that there are no alarms set. Similarly, customers should be able to use any active agent to control the device's volume, much like using the volume control buttons on the device.

Multi-Agent Experiences

Baseline Guidance

- Devices with multiple simultaneous agents should provide access to the device state information that agents need to implement Universal Device Commands, when invoked by the customer.

- The data sent by the device to an invoked agent about ongoing activity states on the device should be minimal and specific to actions that UDCs allow the agent to take. For example, if a customer uses one agent to begin a timer but then invokes a second agent to stop that timer when it rings, the only information the second agent should receive about the timer is that there is stoppable sounding timer on the device (and not, for example, details about the duration of the timer or which agent originally set it).

- Agents invoked to take action on an ongoing activity should not use the device state information provided for any other purpose than to fulfill the UDC request.

Recommended Universal Device Commands

A recommended set of UDCs is listed below. Your product may include other UDCs depending upon the experience and agent capabilities. When considering implementing additional commands, keep in mind:

- Customers may initiate long-running activities that will be stopped later, and the customer may not remember which agent to use (e.g. media playback).

- Unprompted activities may begin that require customer interaction and they may not know which agent to use in order to respond (e.g. sounding timer).

- Customers may want any agent to be able to control global device settings (e.g. volume).

- A command to one agent should not bypass authentication or other security requirements for any other agent.

Multi-Agent Experiences

In this version of the Design Guide, we include the following categories of Universal Device Commands that multi-agent devices should consistently support:

Device Activity	Feature	Use Case
Sounding Alert	Stop Timer	As a user, I want to invoke any compatible agent on device and say 'Stop' (and variants like 'Cancel' or the action button) to stop a sounding (i.e. completed) timer, regardless of which agent created the timer.
	Dismiss Alarm	As a user, I want to invoke any compatible agent on device and say 'Stop' (and variants like 'Cancel', 'Quiet' or the action button) to dismiss a sounding alarm, regardless of which agent set the alarm.
	Dismiss Reminder	As a user, I want to invoke any compatible agent on device and say 'Stop' (and variants like 'Cancel', 'Dismiss') to dismiss a sounding reminder, regardless of which agent set the reminder.
Playing Media	Stop Media	As a user, I want to invoke any compatible agent on device and say 'Stop' to stop any active media playback including music, radio, long-form audio (ebooks, podcasts, news, etc.) and videos.
Playing Camera Feed	Stop Camera Feed	As a user, I want to invoke any compatible agent on device and say 'Stop' (and variants like 'End') to stop any active streaming smart home camera feeds.
Incoming Call	Reject Calls	As a user, I want to invoke any compatible agent on device and say 'Reject' (and variants like 'Stop', 'Cancel') to reject an incoming phone, audio, or video call, regardless of which agent provides the service.

Multi-Agent Experiences

In this version of the Design Guide, we include the following categories of Universal Device Commands that multi-agent devices should consistently support:

Device Activity	Feature	Use Case
Agent Speaking	Stop Speech	As a user, I want to invoke any compatible agent on device and say 'Stop' (and supported variants like 'End', 'Cancel', etc.) to stop any ongoing agent speech activity, regardless of which agent is speaking.
(varies)	Global Foreground Stop	As a user, I want to invoke any compatible agent on device and say 'Stop' (and supported variants like 'End', 'Shut up', etc.) to stop the intended foreground activity when there is more than one active session (e.g. Timer over Music, Weather TTS over Music), regardless of which agent initiated the activities.
N/A	Volume Control (up/down)	As a user, I want to invoke any compatible agent on device and say 'Set volume up/down' (and variants like 'turn it up/down') to change the global volume setting, regardless of which agent set it last.
	Volume Control (to level N)	As a user, I want to invoke any compatible agent on device and say 'Set volume to N' (where N = values from 0 to 10) to change the global volume setting, regardless of which agent set it last.
	Volume Mute	As a user, I want to invoke any compatible agent on device and say 'Mute' to mute the global volume setting, regardless of which agent set it last.

Multi-Agent Experiences

Agent Transfer

When multiple agents co-exist on a device, customers perceive agents as separate from each other, each with their own capabilities, invocations, and experiences. Sometimes, though, customers might make a request to an agent that it cannot directly fulfill. One way to help the customer in that situation is to implement an interoperability pattern called Agent Transfer.

During an Agent Transfer, the customer makes a request of an agent (Agent 1) who cannot directly fulfill their request (e.g. "I can't do that"). However, if Agent 1 is aware of another agent (Agent 2) on the device which can likely fulfill that request, Agent 1 can summon the other agent to assist the customer. No data or context is passed between agents during a transfer, and the customer repeats their request directly to Agent 2 without needing to say the wake word. This pattern reduces some of the friction customers may face in completing their tasks.

Tip

It is important to use appropriate landmarking and sound cues, as shown in the example, to make it clear to the customer that a transfer is taking place.

Example

Customer: "<Agent 1>, play music."
(Agent 1 can't play music, but it knows Agent 2 can likely fulfill that request)

Agent 1: "Hmm, This sounds like something Agent 2 can help you with."
(Transfer sound cue)

Agent 2: "Hi, how can I help?"
(Customer does not need to say the wake word but must repeat the request)

Customer: "Play music."

Agent 2: "Ok, playing top 80s rock."
(Music plays on the device via Agent 2)

Privacy & Security

Strong privacy and security guidelines are also key aspects to earning and maintaining customer trust. This section describes some of the basic building blocks involved in designing simultaneous multi-agent experiences to help earn and maintain customer trust.

Customer Privacy

Devices should provide transparent, easily predictable and expected behaviors and experiences to customers when operating a device with multiple simultaneously available voice agents.

Baseline Guidance

The presence and use of multiple agents should not compromise a customer's privacy.

- Device makers should ensure that a customer's voice recording (or "utterance") is sent only to the agent that the customer intends to invoke (i.e. the agent whose wake word the customer uses).

- Devices or agents should implement an attention system (eg. LEDs or voice chrome) to ensure customers know that an agent is collecting a voice recording.

- Customers should be able to easily understand when any voice recording is shared between agents, and have the ability to provide consent for experiences that require sharing recordings or other types of data.

- Each voice agent should provide customers transparency by enabling them to see and understand which voice recordings were handled by that agent. If an agent provides a voice history, customers should be able to delete it.

Privacy & Security

Attention States and Attention System

The attention system on a device is an important factor in building and maintaining customer trust in your device. Just as it is for single agent products, multi-agent products should clearly communicate the current attention state to customers. Customers should easily be able to understand what state the device is in, or any active agent on the device, as well as when that state changes. This section describes recommendations for attention system behaviors in multi-agent interactions.

All coexisting agents should convey at least the 3 core attention states:

Listening

An agent has been invoked, either by voice or touch, and is recording a customer utterance.

Thinking

The agent or device is processing a request or waiting for a reply from the voice service. (This may not apply when, for example, local agents have no perceived latency between Listening and Speaking.)

Speaking

Playing a voice reply, or otherwise delivering a response to the customer. (Optional for non-voice responses or for devices replying using visuals on a screen).

Visual and sound cues for the 3 core attention states should be clear and easy to understand for all active agents, even if some are unique to an agent or device.

Privacy & Security

Baseline Guidance

1. All agents and devices should convey to customers the core attention states: **Listening, Thinking** (when applicable), and **Speaking** (eg. displayed on the device, listed in Settings, or indicated in a companion app).

2. Agents should not use attention state colors and sound cues which conflict in meaning. For example, the same color should not be used as Listening for one agent and Mic Off for another.

3. It is very important for a product to convey a device's **Microphone On/Off** state.

Agent and Device Security

Securing a device with multiple simultaneous voice agents requires a multifaceted approach in each step of the development process and beyond. Device and agent makers should evaluate potential threat scenarios by performing threat modeling for all features and use cases for their device. The following list represents general security guidelines.

Baseline Guidance

The presence and use of multiple agents should never compromise the security of the device or the customer's data.

- A device should not store any data related to personal customer information. Any required storage of personal data should be minimized and encrypted.

- All customer data in the cloud should be handled in a secure manner (eg. access control, automatic logging, encryption, multi-factor authentication).

- A device should have hardware and software security capabilities that include secure boot, a trusted compute boundary, an anti-roll-back mechanism, and should support hardware-based cryptographic engines.

Privacy & Security

- A device should implement sufficient hardening and access control techniques to limit system access to authorized users, processes, or applications.

- A device should implement adequate authorization, authentication, and input sanitization mechanisms.

- A device should implement secure transmission of data between a device and the cloud, such as use of latest TLS, certificate validation of cloud endpoints.

- A device should implement a secure software update process to apply all security patches.

Baseline Guidance Summary

The following is a summary of the design guidelines included in the previous sections.

Customer Choice and Agent Invocation

1. A customer should be able to choose from available voice agents for a particular interaction. They should have the option to use multiple simultaneous wake words when more than one agent is registered on a device.

2. Multiple simultaneously registered agents should be available to customers at all times, aside from the following exceptions:

   a. When one agent has been invoked and is actively streaming a customer utterance to the cloud, no other agent's wake word should be detectable. For example, if a customer says "Agent 1, tell me about Agent 2," Agent 2 should not be invoked.

   b. An agent should not be able to invoke any other agent by distributing the wake word via TTS. For example, one agent cannot wake up another agent by speaking its wake word.

3. When an agent is in Speaking state, responding to a customer, the customer should be able to interrupt that agent's response with any other active agent's wake word (barge in).

Baseline Guidance Summary

Discovery & Education

1. 1. Customers should easily be able to discover information about the primary uses, benefits, and capabilities of available agents.

2. 2. Customers should be made aware of any multi-agent functionality supported by the device:

    a. Customers should be informed of simultaneously available wake words.

    b. Customers should be informed of Universal Device Command support.

Universal Device Commands

1. Each agent which is simultaneously available on a device should be capable of completing, at a customer's request, the minimum set of Universal Device Commands listed below.

2. If a customer invokes one agent to take an action on an ongoing activity that was started by another agent, no data should be shared about the activity beyond what is necessary to complete the request. For example, if a customer sets an alarm with Agent 1, the alarm sounds, and then the customer requests that Agent 2 turn off the alarm, the only data that Agent 2 should be provided is that there is a stoppable activity on the device.

3. Agents invoked to take action on an ongoing activity should not use the device state provided for any other purpose than to fulfill the UDC request.

Baseline Guidance Summary

Privacy

The presence and use of multiple agents should not compromise a customer's privacy.

1. Device makers should ensure that a customer's voice recording (or "utterance") is sent only to the agent that the customer intends to invoke (i.e. the agent whose wake word the customer uses).

2. Devices or agents should implement an attention system (eg. LEDs or voice chrome) to ensure customers know that an agent is collecting a voice recording.

3. Customers should be able to easily understand when any voice recording is shared between agents, and have the ability to provide consent for experiences that require sharing recordings or other types of data.

4. Each voice agent should provide customers transparency by enabling them to see and understand which voice recordings were handled by that agent. If an agent provides a voice history, customers should be able to delete it.

Attention States and Attention System

1. All agents and devices should convey to customers the core attention states: **Listening, Thinking** (when applicable), and **Speaking** (eg. displayed on the device, listed in Settings, or indicated in a companion app).

2. Agents should not use attention state colors and sound cues which conflict in meaning. For example, the same color should not be used as Listening for one agent and Mic Off for another.

3. It is very important for a product to convey a device's **Microphone On/Off** state.

Baseline Guidance Summary

Security

The presence and use of multiple agents should never compromise the security of the device or the customer's data.

1. A device should not store any data related to personal customer information. Any required storage of personal data should be minimized and encrypted.

2. All customer data in the cloud should be handled in a secure manner (eg. access control, automatic logging, encryption, multi-factor authentication).

3. A device should have hardware and software security capabilities that include secure boot, a trusted compute boundary, an anti-roll-back mechanism, and should support hardware-based cryptographic engines.

4. A device should implement sufficient hardening and access control techniques to limit system access to authorized users, processes, or applications.

5. A device should implement adequate authorization, authentication, and input sanitization mechanisms.

6. A device should implement secure transmission of data between a device and the cloud, such as use of latest TLS, certificate validation of cloud endpoints.

7. A device should implement a secure software update process to apply all security patches.

Glossary

In order to advance clear discussions of multi-agent product design, it is important to use a consistent shared vocabulary. The list included here is not meant to be comprehensive, but to promote a widely adopted set of standard terms.

A

Active Agent

Any voice agent currently capable of responding to a customer invocation.

Agent (Voice Agent)

The digital "person" that the customer interacts with through conversation (turn taking). An agent has its own brand (voice, personality), method of invocation (custom wake word, Action button), and one or more unique capabilities.

Agent Arbitration

The process of determining which voice agent will participate in an interaction.

Agent Attribution

The action of clearly ascribing a response to the Agent responsible for providing it. (See also Content Attribution.)

Agent Transfer

When customers ask something of an agent who cannot directly fulfill their request, the agent can summon a second agent to assist. No data or context is passed between agents during a transfer and the user repeats their request directly to the second agent, but doesn't have to invoke the second agent using its wake word.

Glossary

Assessed Agent Arbitration

A method of Agent Arbitration whereby a service or mechanism selects which agent will participate in an interaction based on an assessment of all relevant interaction factors.

Attention States

The stages of a voice agent's interaction with a customer. Minimally comprised of the Listening, Thinking, and Speaking states, the attention states can also include states such as Do Not Disturb or Notifications Pending. (See also Attention System.)

Attention System

The combination of all visual and audible cues presented to a customer to communicate a voice agent's attention state. The attention system is typically displayed through animated patterns of lights or colors, along with synchronized sound cues. (See also Attention States.)

C

Cloud AI

Server-resident infrastructure supporting an agent's ASR, NLU, NLG, TTS, NN and ML based interactions. In a multi-agent scenario, multiple agents may use a single Cloud AI.

Content Attribution

Attribution informs the customer of the source of the information or content that they are getting from an agent. From a customer perspective, this enhances the clarity and credibility of some information. For brands, this gives them recognition for the services they are providing. Attribution can be given either verbally ("According to…") or visually. (See also Agent Attribution.)

Glossary

D

Disambiguation

When the recognition system hypothesizes two or more possible resolutions to a user utterance, it may ask the user to choose between the various interpretations to decide which was meant by the user.

H

Household

Group of one or more customers that have agreed to share some aspect of their agent experience

I

Intent

The specific action a user wishes to perform. The specific command that is derived from the range of natural language utterances users may speak to convey their intention. The capability needed to respond to specific intents may determine which agent in a multi-agent scenario will respond to the customer.

Invocation

The method whereby an agent or capability is initiated. This could be a spoken wake word or button press from the customer, or a contextual event such as a timer, geofence or other circumstantial event. Each voice agent will generally have its own unique wake word or other invocation method.

Glossary

L

Landmarking

The practice of prepending the response with a sound to orient a customer where they are in the experience. This can be spoken by either the sending or receiving agent (e.g. "Alexa can help with that" or "its Alexa..."). Landmarking provides attribution to the agent handling the request and clarifies for the customer which agent is handling the request.

Locations

A person may use an agent across multiple stationary locations or on-the-go.

M

Multi-Agent Product or Device

A product designed to support multiple voice agents.

Multiple Simultaneous Wake Word (MSWW)

When two or more wake words are able to invoke voice agents on the same device at all times.

Multi-turn Interaction

An interaction between a customer and an agent that includes more than one utterance or response. It is often used by an agent or domain to ask for additional information from the customer, or to continue an experience. It is characterized by not requiring the customer to invoke the agent beyond the initial start of the interaction.

Glossary

O

Agent Orchestration

See Agent Arbitration.

P

Persona

The characteristics, or personality, of an agent including its name, wake word, voice, accent, and visual appearance. Each agents has its own persona, and a single agent may also offer a range of identifiable or selectable personas.

Push-to-talk

An umbrella term that covers invocation of an agent by means of a physical or on-screen affordance such as a button. Push-to-talk includes both tap-to-talk and hold-to-talk implementations. Different agents may be invoked by a single, "overloaded" push-to-talk affordance.

R

Registered Agent

A voice agent which is available on a device and authenticated with customer credentials

U

Universal Device Commands

Commands or controls recognized across a range of voice agents on a device.

Glossary

V

Voice Agent

See Agent.

W

Wake Word

A phrase or a word in a purposeful human-initiated utterance that can be detected to allow an associated agent to start acting on the utterance following the wake word. Otherwise known as "named invocation." A multi-agent device will recognize different wake words for different agents or personas.