# Alexa Privacy and Data Handling Overview

Getting privacy right takes careful attention, and Amazon has built privacy deeply into the Echo hardware and Alexa service, by design.

This paper provides background and details of two areas of interest to Alexa for Business customers and users: when is audio streamed from an Echo device to the Alexa cloud, and how is customer data used while in the Alexa cloud. The device-specific details in this paper are aimed primarily at the three Amazon Echo devices that can be managed with Alexa for Business—Echo Plus, Echo (2nd Generation), and Echo Dot (2nd Generation). Because Echo Show and Echo Spot can be used in conjunction with Alexa for Business, information specific to these two devices is also provided.

Information is presented in three sections. First is an end-to-end overview of the Alexa system. Next is a description of how Echo devices detect their wake word and begin streaming audio to the cloud. Last is a description of data retention and use within the Alexa cloud.

# What is the Alexa System?

We must first understand what the Alexa system is, and how it processes requests. For purposes of this discussion, the Alexa system is comprised of Echo devices listed above—the hardware and software that customers directly interact with and the cloud components—which have the majority of the "smarts": Automatic Speech Recognition, Natural Language Understanding, and Response. Some responses are provided by third party services through "skills." The 3rd-parties that write and publish those skills are responsible for their skill's behavior.

To start, we begin with a very simple tour through the entire system, demonstrated with the example request of "Alexa, what is the weather," so we can see how the request is picked up by an Echo, sent through voice recognition, interpreted, acted upon, and then responded to

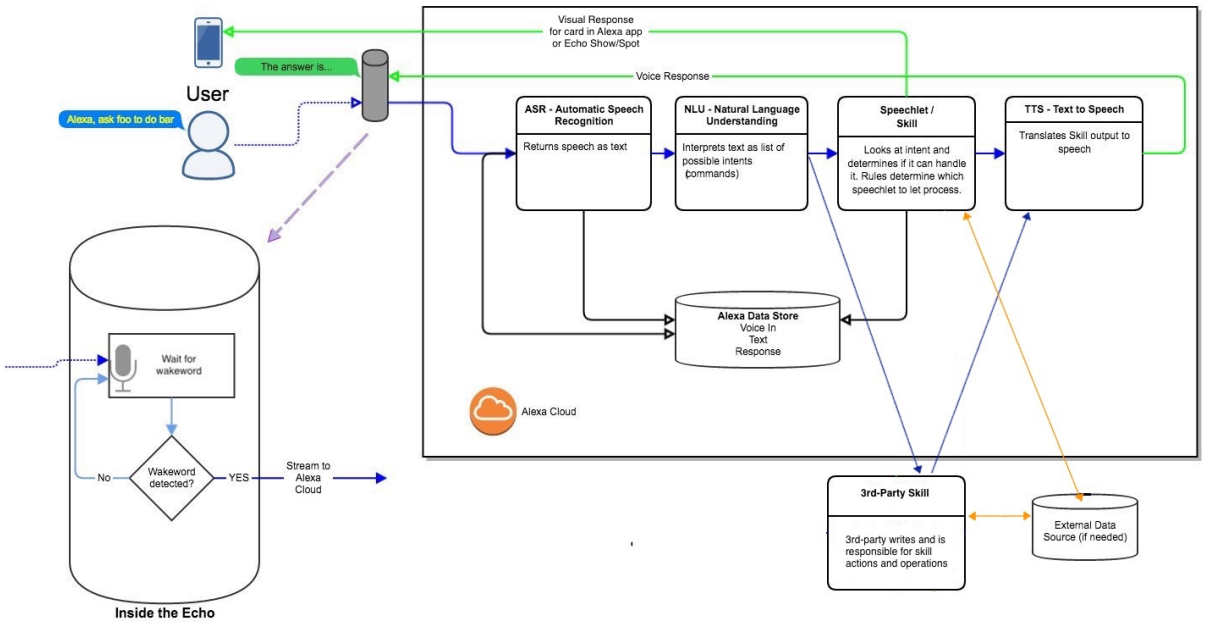


**Figure: Overview of Echo and the Alexa System**

## Echo Devices

Echo devices are the input and output devices for Alexa, and we designed and built them, from the beginning, with multiple layers of security and privacy protections and controls.

Echo devices use on-device keyword spotting designed to detect when a customer says the wake word. This technology inspects acoustic patterns in the room to detect when the wake word has been spoken using a short, on-device buffer that is continuously overwritten. This on-device buffer exists in temporary memory (RAM); audio is not recorded to any on-device storage. The device does not stream audio to the cloud until the wake word is detected or the action button on the device is pressed. If it does not detect the wake word or if the action button is not pressed, no audio is sent to the cloud. The customer can turn the microphone off at any time by pressing the microphone button on the device. When the button is pressed to turn the microphones off, the microphones are electrically disconnected and a dedicated red LED on the microphone button is illuminated to indicate the microphones are off. As a safeguard, we designed the circuitry of Echo devices so that power can only be provided either to this dedicated red LED or to the device microphones, not to both at the same time. As a result, if the dedicated red LED is illuminated, the microphones are off and cannot stream audio to the cloud.

In the usual case, with the microphone on, when the user says, "Alexa, what is the weather in Menlo Park," the Echo's light ring turns blue to indicate that the device detected the wake word and is streaming audio to the cloud. The system is designed so that communication between the Echo device and the Alexa cloud is protected using TLS 1.2.

All customers have the ability to delete the voice recordings associated with their account; enterprise users by using the Alexa for Business console, and personal users by using the Alexa app or the Amazon website. Customers using personal devices (enrolled users) can also use the Alexa app to view and play back the voice recordings associated with their account.

## Automatic Speech Recognition (ASR)

The first Alexa system to receive data is Automatic Speech Recognition (ASR). It takes the audio stream and turns it into a text string (or set of possible text strings) that are sent to the Natural Language Understanding (NLU) system.

In our case, possible strings could be:

- "what is the weather in menlo park"
- "watt is the weather in menlo park"
- "what is the whether in menlopark"
- "watt is the whether in men lo park"

These strings, and their related confidence scores, are used to improve speech recognition. The string with the highest score (i.e., the one that Alexa acts on) is also stored and is displayed to the user in the Alexa app.

## Natural Language Understanding (NLU)

The "Natural Language Understanding" (NLU) interprets the recognition result and produces an intent. The NLU performs:

- intent classification (determining in this case that the user wishes to get the weather, and returning a Weather intent),
- entity recognition (determining that the user requested the location "Menlo Park"), and
- slot resolution (determining the location identifier for the location "Menlo Park" that can later be used to retrieve the correct weather for that location).

The service now looks at the intent as "Weather" and routes the request to the proper application (Skill) with the slots filled in for location (94025) and time (today).

The data about the chosen intent, and information related to entity recognition and slot resolution, is stored for machine learning purposes.

## Skills

Skills extend what the Alexa system can do, and we've designed our skills program to share only limited information with the third-party developers of those skills. For example, voice recordings are not shared with skills.

In the above example, the skill takes the input (location (Menlo Park) and time (today)) and retrieves the appropriate information from the designated data source, which returns the needed data. The skill then formulates its response, taking the raw data (in this case, temperatures as well as the forecast) and constructing a textual response formatted with SSML (simple speech markup language) which tells the next step, TTS, how to respond. Once the response is generated, it is sent to the response system.

Customer's personal information (e.g. name, address) are not released to the 3rd-party unless specifically requested to be shared by the customer. We also use a permission framework similar to the one used by mobile devices, which requires customers to grant permission to share certain data with skill developers–e.g., Lyft could request permission to access the address the customer has set for their Echo device so Lyft can send a ride to that location, and we would only share that address with Lyft after the customer granted permission. Even when a customer links their Amazon account to a 3rd-party skill account (e.g., when a customer links their account with Lyft, so that Lyft always knows which Lyft account to charge rides to), the 3rd-party doesn't receive the customer's Amazon account identifiers. Instead, they receive a token. Amazon can identify an Amazon customer from this token, but skill developers cannot. Each time a customer talks to a skill, the skill gets the same token for that user. However, different skills get different tokens when they talk to the same user, so even if skill developers share data, they cannot determine that they share common customers based on data that Amazon has sharedEach 3rd-party skill has their own policies concerning storage and retention of skill-specific data. Customers can read the 3rd-party skill developer privacy policies, and see what customer information is being asked for, on the skill detail page on the Amazon website.

## Responses (TTS)

The response system takes the SSML that is produced by the skill, uses text-to-speech (TTS) to generate the audio speech file, and streams the audio to the appropriate device. For many skills, this ends the interaction. Other skills are interactive and will ask follow-on questions that require answers. Echo devices are designed

so that the blue ring on the Echo device is lit when the device is waiting for a response to a question that Alexa has asked.

The text of the response is stored by the Alexa system so that users of personal devices can review past answers using the Alexa app. Access to this data is not available to Alexa for Business users or administrators for devices managed by Alexa for Business. In addition, the response can be used by the Amazon team who built the specific skill to ensure that Alexa is providing relevant answers to queries and that the TTS system is properly translating the text to speech.

A technical overview of ASR, NLU, and Skill development is presented in *Just ASK: Building an Architecture for Extensible Self-Service Spoken Language Understanding*, Proceedings of the Neural Information Processing Systems 2017 Conference, March 2018 (https://arxiv.org/pdf/1711.00549.pdf)

# How Does an Echo Detect its Wake Word and Send Audio to the Alexa Cloud?

Amazon Echo devices are designed to use on-device keyword spotting to detect the wake word and only the wake word. Unless the microphone is turned off (discussed below), this technology inspects acoustic patterns in the room to detect when the wake word has been spoken using a short, on-device buffer that is continuously overwritten. There are multiple algorithms running on the Echo device looking for the specified wake word. At this point no audio is sent to the Alexa cloud.

If the algorithms do not detect the wake word, then the Echo device continues to wait for the wake word, continuously overwriting the contents of the small internal audio buffer. Importantly, Echo devices do not keep local records of audio; they keep only a small amount of audio to detect the wake word. This on-device buffer exists in temporary memory (RAM); audio is not recorded to any on-device storage.

When the wake word is detected or the action button is pressed, a connection to the cloud is opened up. The Echo device turns on the blue ring and starts streaming the audio, starting with a fraction of a second of audio prior to the wake word and continuing until the Alexa system in the cloud turns off the audio stream. Echo devices use a signal processing technique called *beam forming* to emphasize the user's speech from the desired direction while suppressing audio interference (like conversations outside the room) from other directions. Customers see beamforming take place on an Echo device with a visual cue—the lightest blue color on the light indicator points towards the source of the audio that is being recorded.

If Alexa is activated using the wake word, the first step that occurs when the stream reaches the cloud is that the audio is reanalyzed using the more powerful processing capabilities of the cloud to verify the wake word was spoken. These additional algorithms are in the cloud, and not on the device, for reasons including requiring more processing power than the Echo device has available or using machine-learning derived models based on recent learnings. The on-device algorithms are automatically updated on a regular basis. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or if Alexa was activated using the action button), our ASR and NLU systems process the customer's request so Alexa can respond appropriately. As our speech recognition system analyzes the audio stream, the system continually attempts to determine when the customer's request to Alexa has ended and then immediately ends the audio stream.

The light ring then typically flashes blue/light blue until the response is ready for playback. It then sends the response (the blue ring pulses while Alexa is speaking), and the Echo device returns to monitoring for its wake word.

**TO SUMMARIZE HOW AN ECHO DEVICE WORKS:**

1. Unless the mic mute button is lit red, the Echo device is analyzing audio to detect its wake word and only its wake word. No audio is sent to the Alexa cloud while this is happening.
2. Wake word detection is done on the Echo device.
3. Only when the wake word is detected or the action button on the device is pressed does audio stream to Alexa in the cloud.
4. When audio is streaming to Alexa, the blue light is turned on.
5. Alexa is designed to stop processing the audio if it determines the speech is not intended for Alexa.
6. The audio stream closes immediately once our ASR system determines the customer has stopped speaking the request

## Preventing an Echo Device from Responding

We make it very easy for anyone using the device to turn the microphones off.

There may be certain times that users may not want the Echo device to respond to the wake word. All Echo devices have a microphone on/off button. When this button is pressed, the microphones are electrically disconnected and a dedicated red LED on the microphone button is illuminated to indicate the microphones are off. As an additional safeguard, we designed the circuitry of Echo devices so that power can only be provided either to this dedicated red LED *or* to the device microphones, not to both at the same time. As a result, if the dedicated red LED is illuminated, the microphones are off and cannot stream audio to the cloud. The system is designed so when in this state, the microphone cannot be turned back on by software or by voice (and if turned off prior to removing power from the Echo device, will still be off when the power is restored).

## How long does the Echo device stream audio to the cloud?

As Alexa analyzes the audio stream, the service continually attempts to determine when the customer's request to Alexa has ended and then immediately ends the audio stream. In some circumstances, in response to customer commands, the stream will open again for a customer to follow up, including if the customer's request involves multiple interactions. For example, if you say, "Alexa, set the timer," Alexa will respond with "Timer for how long?" and will open the audio stream to wait for your response. Similarly, an interactive skill like "20 Questions" will ask questions and Alexa will open the audio stream for your response.

Customers may also elect to enable the Follow-Up mode setting (on a device-by-device basis). Follow-Up mode allows Alexa to respond to a series of requests in rapid succession without the customer needing to repeat the wake word for each request, but only after being woken by an initial request with the wake word.

In all cases, the blue light will be illuminated to indicate to customers that the Echo device is streaming audio to the cloud. Customers can also enable an audible tone that plays at the start and end of each request.

## Alexa Calling

When using Alexa calling, the light ring (or bar) will glow green, to indicate that the microphone is on and audio is streaming. In this case, the audio is not being streamed to the Alexa ASR and NLU systems. Instead, it is being routed to either another Alexa-enabled product (for Alexa-to-Alexa calling) or to the phone system (if placing a telephone call). Calls are not recorded by Alexa. During the call, the Echo device is still monitoring for its wake word, so that you can say things like "Alexa, end call". Even during a call, you will

note that the light ring on the Echo device turns blue when its wake word is detected, to show that the audio directly after the wake word is going to Alexa.

#### ECHO SPOT AND ECHO SHOW

Echo Show and Echo Spot are devices that are also equipped with a camera. Just like other Echo devices, there is a button on the device to turn off the microphones, which also turns off the camera. In addition, it is worth noting that the camera only sends video or pictures to the cloud when a customer requests this be done. While an Echo Show or Echo Spot is waiting for its wake word, it does not send any images or videos to the cloud (just as it doesn't send any audio).

There are two other differences with the Echo Show and Echo Spot versus the Echo, Echo Dot, and Echo Plus. They both have a red LED to indicate the microphone is off—either on the button itself or on the front of the device—but they do not have a light ring. Instead, they are designed to display either a red bar (Echo Show) or red ring (Echo Spot) on the screen when the microphone/camera button is pressed. They also display a symbol (⌀) on the screen to indicate the microphone and camera are off. Similarly, the blue "audio streaming to the Cloud" indicator shows as a blue ring or line on the screen.

### Summary of Wake Word Detection and Audio Streaming to Alexa

To wrap up this section, we have addressed the following:

1. Echo devices are the input/output devices for the Alexa system, which is in the cloud.
2. Audio is sent from an Echo device to Alexa in the cloud *only* when the wake word is detected or the action button is pressed.
3. The blue light ring is on when audio is being sent from an Echo device to Alexa
4. Alexa double-checks, with cloud-side verification, to ensure that the wake word was really spoken.
5. Alexa ends the audio stream when it determines that the customer is no longer talking to Alexa.
6. The microphone button is under the user's control and prevents the Echo device from detecting the wake word.
7. The circuitry of the red LED on the microphone button is electronically connected to the microphones. If the red LED on the microphone button is lit, the microphones are off.
8. During an interaction with Alexa, the blue light ring will turn on if Alexa is waiting for input.

## Data Retention and Use in Alexa

Alexa is designed to get smarter every day—this is accomplished through the power of machine learning and the cloud**.** The more a customer uses Alexa, the more it adapts to their speech patterns, vocabulary, and personal preferences. When the customer says the wake word, their subsequent phrases are processed and stored in the cloud to respond to the customer's request and to improve the customer's experience and our services, including training our speech recognition and natural language understanding systems so Alexa can better understand customers' requests.

### Data Storage

Different types of data are used and stored by the Alexa system to provide the Alexa service. Configuration parameters are set by the user either on the device or using the Alexa app. These parameters include such

things as the device location (set by the administrator or user), preferred time zone and unit measures, volume level, and other preferences.

Audio and text inputs are the core piece of Alexa data. As described above, voice recordings are processed through speech-to-text algorithms and then through natural language processing algorithms to extract the user's intent and the parameters of the Alexa query. These systems use machine learning techniques to continuously improve themselves with each input.

## Data Retention

Data is stored in multiple forms and for multiple purposes in various Amazon services, such as S3 and DynamoDB (under the control of the Alexa service). Each data type has an associated retention policy and access policy. We minimize the data we retain while still allowing Amazon to provide the service to customers (including allowing enrolled users to review and play back their voice recordings) and to build, test, debug, and improve our systems.

Only those who have an approved need to access certain data to accomplish their job are given access to that data—access is granted via specific, audited permissions and access to customer data requires review and approval by the responsible managers. Additionally, the permissions to access this data are reviewed and positively confirmed by management at least quarterly and access is audited.

Sensitive customer data in the Alexa system (such as voice recordings) is stored in databases and encrypted at rest and in transit, using Amazon's internal key management systems.

Some system level data is also stored in log files, for either service troubleshooting purposes, or security incident resolution. Troubleshooting logs contain information necessary for developers to troubleshoot the Alexa system, but do not contain customer voice recordings or data derived from customer voice recordings, such as slot values or the TTS response. Access to these logs is restricted to teams needing access to this data to perform their business functions. Troubleshooting logs are encrypted and their access is audited.

Security logs are retained for purposes of audits and are restricted to those operating in security incident roles. They contain data that describe (1) when systems or users authenticated themselves to the system and (2) which systems and users accessed which data, and when. Again, these logs are encrypted and the data in them is used to ensure that system use complies with applicable policies.

We apply retention policies to data to minimize the data we retain. Data is retained when it serves a business purpose (including providing the service to customers and improving our systems) or as necessary to comply with law.

Metrics are stored in databases. Metrics are used for internal business processes, to direct system improvements, for systems performance analysis and reporting, and for customer reports. Access to metrics is restricted to the teams and individuals that need this data to perform their work. As with other data access, these permissions are reviewed and approved at least quarterly.

The speech recognition and natural language understanding in the Alexa system are based on machine learning (ML) algorithms. Data sets from real use cases are fed into the various ML systems to build new algorithms and improve existing algorithms. Again, access to speech and derived data in Amazon's ML systems is strictly controlled and audited. Third-party skill developers store data they receive through customers' use of their skills according to their own privacy policies. Customers can review the privacy policies of 3rd-party skill developers by visiting the skill's detail page on the Amazon website.

## Data Use

Data in the Alexa system is used to deliver and improve Amazon's services and as described in Amazon's privacy notice. This includes responding to customers' requests, allowing customers to review and listen to their voice recordings, and training Alexa to better understand and interpret customers' requests. As discussed above, if you use a third-party Alexa skill, Alexa will provide the content of your requests (but not the voice recordings) to the skill so the skill can respond accordingly.

Consistent with the AWS user agreement, we do not disclose your content to any government except in each case as necessary to comply with the law or a binding order of a governmental body. Unless it would violate the law or a binding order of a governmental body, we will give you notice of any legal requirement or order with respect to your content under the AWS user agreement.

# Appendix: Privacy & Security FAQs from AWS Website

https://aws.amazon.com/alexaforbusiness/faqs/

### How do Amazon Echo devices recognize the wake word?

Amazon Echo devices use on-device keyword spotting to detect the wake word. When these devices detect the wake word, they stream audio to the cloud, including a fraction of a second of audio before the wake word.

### Can I turn off the microphone on Echo devices?

Yes, you can turn off the microphone by pushing the microphone on/off button on the top of your device. When the microphone on/off button turns red (on the Echo Show there is a red LED), the microphone is off. The device will not respond to the wake word until you reactivate the microphone by pushing the microphone on/off button again. An organization cannot turn on a device's microphones via the Alexa for Business Console if the device's microphones have been turned off.

### How do I know when an Echo device is streaming my voice to the Cloud?

When an Echo device detects the wake word the light ring around the top of your device turns blue, to indicate that the device is streaming audio to the Cloud (for Echo Show and Echo Spot, you will see a blue bar or ring on the screen). When you use the wake word, the audio stream includes a fraction of a second of audio before the wake word. The audio stream closes once your question or request has been processed.

For personal devices you can enable a 'start of request sound,' a short audible tone that plays after the wake word is recognized to indicate that the device is streaming audio. You can also enable an 'end of request sound' that will play a short audible tone at the end of your request, to indicate that the connection has closed and the device is no longer streaming audio. This is available within the Sounds settings in the Alexa App (Settings > [Your Device Name] > Sounds).

### What can an organization tell their users about the user's information when using a corporate skill on an enrolled account or using a device managed by the organization?

You can tell them that the organization has no access to the information it receives about how they use a personal device, outside of when they interact with corporate skills. The organization may receive engagement metrics (device and skill usage metrics) for shared devices. In either case, the organization has no access to any voice recordings.

Voice recordings from shared devices being managed by Alexa for Business can be deleted from the Alexa for Business management console. If a user has enrolled their personal account, they can view and delete individual voice recordings associated with their account using the Alexa companion app, or all recordings by visiting Manage Your Content and Devices

### When an organization manages shared devices using Alexa for Business, what information does that organization have access to?

The organization can see and control which skills are enabled on a shared device, the room where it's assigned, and the settings applied to the device.

### When an organization manages shared devices using Alexa for Business, does the organization have access to voice recordings made by users of the shared device?

No, unlike with a personal Alexa-enabled device where a user can review their voice recordings in the Alexa companion app, Alexa for Business organizations cannot access any voice recordings or text transcripts of what a user said. In addition, the organization doesn't see Alexa's responses to users' queries.

### What data do skill developers for Alexa for Business have access to?

Skill developers receive the information about their skill and its usage that is made available to skill developers in the Alexa Skills Kit developer portal. They also have access certain information about shared devices via the Alexa for Business API.

### What controls do organizations have over personal accounts that they let enroll and join their Alexa for Business account?

Organizations can control which of their users can enroll and join their personal account to the organization's Alexa for Business account. In addition, they can require a user create a voice profile to access corporate resources like calendars.

### What information does an organization receive about its users' Amazon accounts when users enroll their personal account with the organization's Alexa for Business account?

The organization does not have any access to the user's personal Amazon account. The organization does not receive the name or email that the personal account uses. As with shared devices, the organization has no access to the voice recordings on a personal device, including deleting voice recordings.

### Are voice inputs processed by Alexa for Business stored, and how are they used by Alexa for Business?

Alexa for Business may store and use voice inputs processed by the service solely to provide and maintain the service and to improve and develop the quality of Alexa for Business and other Amazon machine learning and artificial intelligence services. Use of your content is necessary for continuous improvement of your Alexa for Business customer experience, including the development and training of related technologies. We do not use any personally identifiable information that may be contained in your content to target products, services, or marketing to you or your end users. Your trust, privacy, and the security of your content are our highest priority and we implement appropriate and sophisticated technical and physical controls, including encryption at rest and in transit, designed to prevent unauthorized access to, or disclosure of, your content and ensure that our use complies with our commitments to you.

### How can voice recordings be deleted?

The organization can delete voice recordings for shared devices they manage via the Alexa for Business console. The organization does not have any access to these voice recordings, other than the ability to delete them. Personal device users can can view and delete specific voice recordings associated with their accounts by going to History in Settings in the Alexa app, drilling down for a specific entry, and then tapping on the delete button. Or, personal device users can delete all voice recordings associated with their accounts for each of their Alexa-enabled products by selecting the applicable product at Manage Your Content and Devices.

Deleting voice recordings may degrade your Alexa for Business experience.

### Who has access to my content that is processed and stored by Alexa for Business?

Only authorized employees will have access to your content that is processed by Alexa for Business. Your trust, privacy, and the security of your content are our highest priority and we implement appropriate and sophisticated technical and physical controls, including encryption at rest and in transit, designed to prevent unauthorized access to, or disclosure of, your content and ensure that our use complies with our commitments to you.

Alexa for Business provides access to non-AWS services provided by third parties and other Amazon entities (for example, skills and features that provide movie showtimes and traffic information). If you or your users use those non-AWS services, we may exchange information related to those requests with the parties providing the applicable services and that information is subject to the privacy and security practices of those parties.

### Do I still own my content that is processed and stored by Alexa for Business?

You always retain ownership of your content and we will only use your content with your consent.

### Is the content processed by Alexa for Business moved outside the AWS region where I am using Alexa for Business?

Any content processed by Alexa for Business is encrypted and stored at rest in the AWS region where you are using Alexa for Business. Some portion of content processed by Alexa for Business may be stored in another AWS region solely in connection with the continuous improvement and development of your Alexa for Business customer experience and other Amazon machine learning and artificial intelligence services. Your trust, privacy, and the security of your content are our highest priority and we implement appropriate and sophisticated technical and physical controls, including encryption at rest and in transit, designed to prevent unauthorized access to, or disclosure of, your content and ensure that our use complies with our commitments to you.