Amazon Transcribe FAQs

aws.amazon.com/transcribe/faqs

General

Q: What is Amazon Transcribe?

Amazon Transcribe is an AWS service that makes it easy for customers to convert speech-to-text. Using Automatic Speech Recognition (ASR) technology, customers can choose to use Amazon Transcribe for a variety of business applications, including transcription of voice-based customer service calls, generation of subtitles on audio/video content, and conduct (text based) content analysis on audio/video content.

Q: How does Amazon Transcribe interact with other AWS products?

Amazon Transcribe converts audio input into text, which opens the door for various text analytics applications on voice input. For instance, by using Amazon Comprehend on the converted text data from Amazon Transcribe, customers can perform sentiment analysis or extract entities and key phrases. Similarly, by integrating with Amazon Translate and Amazon Polly, customers can accept voice input in one language, translate it into another and generate voice output, effectively enabling multi-lingual conversations. It is also possible to integrate Amazon Transcribe with Amazon Kendra or Amazon Opensearch (successor to Amazon Elasticsearch Service) to index and perform text based search across audio/video library. To learn more, checkout the MediaSearch or Content Analysis solution.

Q: What else should I know before using Amazon Transcribe service?

Amazon Transcribe service is designed to handle a wide range of speech and acoustic characteristics, including variations in volume, pitch, and speaking rate. The quality and content of the audio signal (including but not limited to factors such as background noise, overlapping speakers, accented speech, or switches between languages within a single audio file) may affect the accuracy of service output. We are constantly updating the service to improve its ability to accommodate additional acoustic variation and content types.

Using Amazon Transcribe

Q: How will developers access Transcribe?

The easiest way to get started with Amazon Transcribe is to submit a job using the console to transcribe an audio file. You can also call the service directly from the AWS Command Line Interface, or use one of the supported SDKs of your choice to integrate with your applications. Either way, you can start using Amazon Transcribe to generate automated transcripts for your audio files with just a few lines of code.

Q: Does Amazon Transcribe support real-time transcriptions?

Yes. Amazon Transcribe enables users to open a bidirectional stream over HTTP2. Users can send an audio stream to the service while receiving a text stream in return in real time.

Q: What encoding does real-time transcription support?

Streaming transcription currently supports 16-bit Linear PCM encoding.

Q: What languages does Amazon Transcribe support?

For information on language support, please refer to this underline documentation page.

Q: What devices does Amazon Transcribe work with?

Amazon Transcribe for the most part is device agnostic. In general, Amazon Transcribe works with any device that includes an on-device microphone such as phones, PCs, tablets, and IoT devices (e.g. car audio systems). Amazon Transcribe API will be able to detect the quality of the audio stream being input at the device (8kHz VS 16kHz) and will appropriately select the acoustic models for converting speech-to-text. Furthermore, developers can call Transcribe API through their applications to access speech-to-text conversion capability.

Q: Are there size restrictions on the audio content that Amazon Transcribe can process?

Amazon Transcribe service calls are limited to 4 hours (or 2GB) per API call for our batch service. The streaming service can accommodate open connections up to 4 hours long.

Q: What programming languages does Amazon Transcribe support?

Amazon Transcribe batch service supports .NET, Go, Java, Javascript, PHP, Python and Ruby.
Amazon Transcribe real-time service supports Java SDK, Ruby SDK, and C++ SDK. Additional SDK support are coming. For more details, visit the Resources page.

Q: My custom vocabulary words are not being recognized! What can I do?

The speech recognition output depends on a number of factors in addition to custom vocabulary entries, so there can be no assurance that if a term is included in the custom vocabulary, it will be correctly recognized.
However, the most frequent reason is that a custom word lacks the correct pronunciation. If you haven't provided a pronunciation for your custom word, please try to create one. If you already have provided one, double-check its correctness, or include other pronunciation variants if necessary. This can be done by creating multiple entries in the custom vocabulary file that differ in the pronunciation field.

Q: Why do I see too many custom words in my output?

Custom vocabularies are optimized for a small list of targeted words; larger vocabularies may lead to over-generation of custom words, especially when they contain words that are pronounced in a similar way. If you have a large list, please try reducing it to rare words and words that are actually expected to occur in your audio files. If you have a large vocabulary covering multiple use cases, split it into separate lists for different use cases. The words that are short and sound similar to many other words, may lead to over-generation (too many custom words appearing in the output). It is preferable to combine these words with surrounding words and list them as hyphen-separated phrases. For example, the custom word "A.D." could be included as part of a phrase such as 'A.D.-converter'.

Q: There are two ways of giving pronunciations, IPA or SoundsLike fields in the custom vocabulary table. Which one is better?

IPA allows for more precise pronunciations. You should provide IPA pronunciations if you are able to generate IPA (e.g., from a lexicon that has IPA pronunciations or an online converter tool).

Q: I'd like to use IPA but I'm not a linguistic expert. Is there an online tool I can use?

Several standard dictionaries, such as the Oxford English Dictionary or the Cambridge Dictionary (including their online versions) provide pronunciations in IPA. There are also online converters (e.g. easypronunciation.com or tophonetics.com for English) — however, note that in most cases these tools are based on underlying dictionaries and may not generate correct IPA for some words, such as proper names. Amazon Transcribe does not endorse any third-party tools.

Q: Do I need to use different IPA standards that are specific to a different accents of the same language? (e.g. US English versus British English)?

You should use the IPA standard that is appropriate for the audio files you will be processing — e.g., if you are expecting to process audio from British English speakers, use the British English pronunciation standard. The set of allowed IPA symbols may differ for the different languages and dialects supported by Amazon Transcribe; please make sure that your pronunciations contain only the allowed characters. Details on the IPA character sets can be found in the documentation: https://docs.aws.amazon.com/transcribe/latest/dg/how-vocabulary.html#charsets

Q: How can I provide the pronunciation using SoundsLike field in the custom vocabulary table?

You can break a word or phrase down into smaller pieces and provide a pronunciation for each piece using the standard orthography of the language to mimic the way that the word sounds. For example, in English you can provide pronunciation hints for the phrase *Los-Angeles* like this: *loss-ann-gel-es*. The hint for the word Etienne would look like this: *eh-tee-en*. You separate each part of the hint with a hyphen (-). You can use any of the allowed characters for the input language.

Q: How do two different ways of providing acronyms (with periods and without periods but with pronunciations) work?

If you use an acronym containing periods, the spelling pronunciation will be generated internally. If you do not use periods, please provide the pronunciation in the pronunciation field. For some acronyms, it is not obvious whether they have a spelling pronunciation or a word-like pronunciation (e.g., NATO is often pronounced 'n eɪ t oʊ' (*nay-toh*) rather than 'ɛn eɪ ti oʊ' (N. A. T. O.)).

Q: Where can I find examples of how to use custom pronunciations?

You can find sample input formats and examples in the documentation: https://docs.aws.amazon.com/transcribe/latest/dg/how-vocabulary.html.

Q: What happens if I use the wrong IPA? If I am uncertain, am I better off not inputting any IPA?

The system will use the pronunciation you provide; this should increase the likelihood of the word being recognized correctly if the pronunciation is correct and matches what was spoken. If you are not certain you are generating correct IPA, please run a comparison by processing your audio files with a vocabulary that contains your IPA pronunciations, and with a vocabulary that only contains the words (and, optionally, display-as forms). If you do not provide any pronunciations the service will use an approximation, which may or may not work better than your input.

Q: When using DisplayAs forms, can I display character sets unrelated to the original language being transcribed? (e.g. output "Street" as "街道").

Yes. While phrases may only use a restricted set of characters for the specific language, UTF-8 characters apart from \t (TAB) are permitted in the DisplayAs column.

Q: Is Automatic content redaction or personally identifiable information (PII) redaction available with both batch and streaming APIs for Transcribe?

Yes, Amazon Transcribe supports automatic content redaction or PII redaction for both batch and streaming APIs.

Q: What languages are supported for Automatic content redaction / PII identification and redaction?

Please refer to the Amazon Transcribe documentation for information on the language availability of automatic content redaction / PII redaction.

Q: Does Automatic content redaction also redact sensitive personal information from the source audio?

No, this feature does not remove sensitive personal information from the source audio. You can however redact personal information from the source audio yourself using the start and end timestamps that are provided in the redacted transcripts for each instance

of an identified PII utterance.

Q: Can I use Automatic content redaction for redacting personal information from the existing text transcripts?

No, Automatic content redaction only works on audio as an input.

Q: What else should I know before using Automatic content redaction?

Automatic content redaction is designed to identify and remove personally identifiable information (PII), but due to the predictive nature of machine learning, it may not identify and remove all instances of PII in a transcript generated by the service. You should review any output provided by Automatic content redaction to ensure it meets your needs.

Q: Are there any differences between Automatic content redaction for streaming and batch APIs?

Yes there are two additional capabilities supported by Automatic content redaction for the streaming API which are not supported by the batch API. You can decide to only identify PII and not redact when using content redaction with streaming API. Also you have the ability to identify or redact specific PII types with streaming API. For example, you can redact just the social security number and credit card information and keep other PII like names and email addresses.

Q: Which AWS regions is automatic content redaction or PII redaction available?

Please refer to the Transcribe documentation for information on the availability of automatic content redaction and PII redaction for batch and streaming APIs in the AWS regions.

Q: Which APIs support Automatic language identification?

Automatic language Identification is currently supported for batch and streaming APIs.

Q: What languages can Amazon Transcribe automatically identify?

Amazon Transcribe can identify any of the languages supported by the batch and streaming APIs.

Q: Does Amazon Transcribe identify multiple languages in the same audio file?

Amazon Transcribe only identifies the dominant language in an audio file.

Q: Is there any way to restrict the list of languages to choose from for Automatic language identification?

Yes, you can specify a list of languages that might be present in your media library. When you provide a list of languages, the identified language will be chosen from that list. If no languages are specified, the system will process the audio file against all the languages supported by Amazon Transcribe and select the most probable one. The accuracy of language identification is better when a select list of languages is provided.

Pricing & Availability

Q: What does it cost?

Refer to the Amazon Transcribe Pricing page to learn more.

Q: What AWS regions are available for Amazon Transcribe?

Please refer to the AWS Global Infrastructure Region Table.

Data Privacy

Q. Are voice inputs processed by Amazon Transcribe stored, and how are they used by AWS?

Amazon Transcribe 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 Amazon Transcribe and other Amazon machine-learning/artificial-intelligence technologies. Use of your content is important for continuous improvement of your Amazon Transcribe 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. Please see https://aws.amazon.com/compliance/data-privacy-faq/ for more information. You may opt out of having your content used to improve and develop the quality of Amazon Transcribe and other Amazon machine-learning/artificial-intelligence technologies by using an AWS Organizations opt-out policy. For information about how to opt out, see Managing AI services opt-out policy.

Q: Can I delete data and artifacts associated with transcription jobs stored by Amazon Transcribe?

Yes. You can use available Delete APIs to delete data and other artifacts associated with transcription jobs. If you have issues doing so, contact AWS support.

Q: Who has access to my content that is processed and stored by Amazon Transcribe?

Only authorized employees will have access to your content that is processed by Amazon Transcribe. 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. Please see https://aws.amazon.com/compliance/data-privacy-faq/ for more information.

Q: Do I still own my content that is processed and stored by Amazon Transcribe?

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

Q. What happens to my data used in training custom language models? Will I still own it?

When submitting text data that is used to train a dedicated model, the customer has ownership of both the original text data as well as the generated custom model. The text data will neither be stored, nor used to improve our general speech recognition engine. Models produced by using CLM are self-contained, and accessible only by the customer.

Q. Since the service will not be retaining my training data, are there any drawbacks or degradation to the transcription quality or overall service experience?

There will be no transcription quality degradation resulting from our service not storing your training data. Once the training data is used to actually produce a custom language model, the model itself becomes available for repeated use at your discretion. The original training set you uploaded is expunged from our systems. The only drawback is if you require technical support. Because we do not retain your original training data, we would not have convenient access to those assets or related intermediate artifacts, should you require support team to investigate potential service issues. Support would still be available, but not as expedient because we may need to ask for additional information from you.

Q. How can I reuse the data for future model updates or improvements?

Since training data is not stored, the same data set and any additional data will have to be uploaded again to train new models. When there is an update to the base model provided by Amazon Transcribe, you will be notified. To take advantage of the latest base model, you should submit your data to train a new model. You will then have both the original custom model that you previously generated and also the new version to use.

Q. How I delete a model?

Users are able to delete any customer language model that they generated at their discretion.

Q: Is the content processed by Amazon Transcribe moved outside the AWS region where I am using Amazon Transcribe?

Any content processed by Amazon Transcribe is encrypted and stored at rest in the AWS region where you are using Amazon Transcribe. Some portion of content processed by Amazon Transcribe may be stored in another AWS region solely in connection with the continuous improvement and development of your Amazon Transcribe customer experience and other Amazon machine-learning/artificial-intelligence technologies. If you opt out of having your content used to develop the quality of Amazon Transcribe and other Amazon machine-learning/artificial-intelligence technologies by contacting AWS Support, your content will not be stored in another AWS region. You can request deletion of voice inputs associated with your account by contacting AWS Support. 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. Please see https://aws.amazon.com/compliance/data-privacy-faq/ for more information.

Q: Can I use Amazon Transcribe in connection with websites, programs or other applications that are directed or targeted to children under age 13 and subject to the Children's Online Privacy Protection Act (COPPA)?

Yes, subject to your compliance with the Amazon Transcribe Service Terms, including your obligation to provide any required notices and obtain any required verifiable parental consent under COPPA, you may use Amazon Transcribe in connection with websites, programs, or other applications that are directed or targeted, in whole or in part, to children under age 13.

Q: How do I determine whether my website, program, or application is subject to COPPA?

For information about the requirements of COPPA and guidance for determining whether your website, program, or other application is subject to COPPA, please refer directly to the resources provided and maintained by the United States Federal Trade Commission. This site also contains information regarding how to determine whether a service is directed or targeted, in whole or in part, to children under age 13.

Amazon Transcribe Call Analytics

Q. What is Amazon Transcribe Call Analytics?

Amazon Transcribe Call Analytics is an AI-powered API that provides rich call transcripts and actionable conversation insights that you can add into their call applications to improve customer experience and agent productivity. It combines powerful speech-to-text and custom natural language processing (NLP) models that are trained specifically to

understand customer care and outbound sales calls. As a part of AWS Contact Center Intelligence (CCI) solutions, this API is contact center agnostic and makes it easy for customers and ISVs to add call analytics capabilities into their applications.

Q. What can I do with Amazon Transcribe Call Analytics?

With Call Analytics, you can quickly add valuable intelligence such as customer and agent sentiment scores, call drivers, and call categories directly as an API output to any inbound or outbound call application. Supervisors can use those applications to better understand customer and agent interactions, identify trending issues, and track performance metrics across different categories.

Q. How do I get started with Amazon Transcribe Call Analytics?

You can use Transcribe Call Analytics through the APIs and AWS Management Console. Analytics jobs can be created, monitored, and deleted through the API or the console. In the console, you will see a list of analytics jobs, and a job detail page with input parameters and a JSON output preview. In addition to this, you will be able to create, edit, and delete categories through the APIs or Console for the automated contact categorization feature.

Q. Which languages does Amazon Transcribe Call Analytics support?

Amazon Transcribe Call Analytics currently supports Arabic (Gulf), Mandarin Chinese (Mainland China), English (United States), English (Great Britain), English (Ireland), English (Scotland), English (Wales), English (Australia), English (India), French (Canada), French (France), German (Germany), German (Switzerland), Hindi (India), Italian (Italy), Japanese (Japan), Korean (Korea), Portuguese (Portugal), Portuguese (Brazil), Spanish (United States), and Spanish (Spain).

Q. In which AWS regions is Amazon Transcribe Call Analytics available?

Amazon Transcribe Call Analytics is currently available in US West (Oregon), US East (N. Virginia), Canada (Central), Europe (London), Europe (Frankfurt), Asia Pacific (Mumbai), Asia Pacific (Seoul), Asia Pacific (Singapore), Asia Pacific (Sydney), and Asia Pacific (Tokyo) regions.

Q. How does the pricing work for Amazon Transcribe Call Analytics?

Amazon Transcribe Call Analytics API will be separately priced from existing Transcribe APIs. Please see the Transcribe pricing page for more details.

Q. Is Amazon Transcribe Call Analytics HIPAA eligible?

Yes. Amazon Transcribe Call Analytics is HIPAA eligible.

Amazon Transcribe Medical

Q. What is Amazon Transcribe Medical?

Amazon Transcribe Medical is an automatic speech recognition (ASR) service that makes it easy for developers to add medical speech-to-text capabilities to their applications. Using Amazon Transcribe Medical, you can quickly and accurately transcribe medical dictation and conversational speech into text for a variety of purposes, such as recording physician notes or processing in downstream text analytics to extract meaningful insights.

Q. What can I do with Amazon Transcribe Medical?

Amazon Transcribe Medical uses advanced machine learning models to accurately transcribe medical speech into text. Transcribe Medical can general text transcripts that can be used to support a variety of use cases, spanning clinical documentation workflow and drug safety monitoring (pharmacovigilance) to subtitling for telemedicine and even contact center analytics in the healthcare and life sciences domains.

Q. Do I need to be an expert in automatic speech recognition (ASR) to use Amazon Transcribe Medical?

No, you don't need any ASR or machine learning expertise to use Amazon Transcribe Medical. You only need to call Transcribe Medical's API, and the service will handle the required machine learning in the backend to transcribe medical speech to text.

Q. How do I get started with Amazon Transcribe Medical?

You can get started with Amazon Transcribe Medical from the AWS Management console or by using the SDK. Please refer to this technical documentation page for details.

Amazon Transcribe Medical provides a free tier so that you can test the service. Refer to this pricing page for more information.

Q. Which languages does Amazon Transcribe Medical support?

Amazon Transcribe Medical currently supports medical transcription in US English.

Q. Which medical specialties does Amazon Transcribe Medical support?

Amazon Transcribe Medical supports transcription for an expanding list of primary care and specialty care specialties. Visit our documentation for a full list of supported medical specialties.

Q. In which AWS regions is Amazon Transcribe Medical available?

Amazon Transcribe Medical is currently available in US-East (N. Virginia), US East (Ohio), US West (Oregon), Canada (Central), EU (Ireland), and Asia Pacific (Sydney).

Q. How is Amazon Transcribe Medical priced?

Refer to the Amazon Transcribe Medical pricing page to learn more about pricing details.

Q. Is Amazon Transcribe Medical HIPAA eligible?

Yes.

Q. Is the content processed by Amazon Transcribe Medical used for any purpose other than to provide the service?

Amazon Transcribe Medical does not use content processed by the service for any reason other than to provide and maintain the service. Content processed by the service is not used to develop or improve the quality for Amazon Transcribe Medical or any other Amazon machine-learning/artificial-intelligence technologies.

Q. Does Amazon Transcribe Medical learn over time?

Yes, Amazon Transcribe Medical uses machine learning and is continuously being trained to make it better for customer use cases. Amazon Transcribe medical does not store or use customer data used with the service to train the models

Q. What else should I know before using the Amazon Transcribe Medical service?

Amazon Transcribe Medical is not a substitute for professional medical advice, diagnosis, or treatment. You and your end users are responsible for exercising your and their own discretion, experience, and judgment in determining the correctness, completeness, timeliness, and suitability of any information provided by Amazon Transcribe Medical. You and your end users are solely responsible for any decisions, advice, actions, and/or inactions based on the use of Amazon Transcribe Medical.

Amazon Transcribe Medical may not accurately identify protected health information in all circumstances, and does not meet the requirements for de-identification of protected health information in accordance with HIPAA. You are responsible for reviewing any output provided by Amazon Transcribe Medical to ensure it meets your needs.

Custom Language Models

Q: What functionality does custom language models provide today?

You can use custom language models (CLM) to train and develop language models that are domain-specific. CLM currently supports Australian English, British English, Hindi, US English and US Spanish for batch transcriptions and US English for streaming transcriptions. CLM supports the simultaneous use of custom vocabulary for batch transcriptions.

Q: How much and what type of training data do I need? How do I obtain the data? Does the data need to have a specific format?

The text data should be relevant to the audio that will be transcribed using the custom model; it should contain as many of the domain-specific words, phrases, and word combinations as possible. We recommend using at least 100k and at most 10M words of

running text. Text data resources can be obtained from any in-house or public sources (e.g., using text from customers' websites). We recommend each plain text file contain 200,000 words or more, but not exceeding 1GB in overall file size. The text should be in UTF-8, and use one sentence per line. Each sentence should contain punctuation. Users are responsible for spell-checking, removing formatting characters, and validating the encoding.

Q: How do I use custom language models (CLM)?

To train a custom language model, customers simply supply the text data in an Amazon S3 bucket. Users can then use the Amazon Transcribe service console to load and process the data to train a custom language model. Training is fully automated and requires minimal intervention from the user. When the final custom model is ready, it is made available in the customer's AWS account for transcribing domain-specific audio files. Moreover, customers can train multiple custom models to use for a variety of different use cases.

Q: Are improvements guaranteed? Is it worth expending the effort of collecting text data?

Improvements are not guaranteed – the change in performance will depend on how closely the text data matches the audio, and on the amount of data provided. More data is generally better, but most importantly, the data should cover words and word sequences that are expected to occur in the audio files you intend to transcribe. Improvements to transcription accuracy will depend on the quality of the training data as well as the use case. In some scenarios, general benchmarking indicates as much as 10% to 15% relative accuracy improvement.

Q: How long does model training take? When will I be able to use it?

Model training usually takes between 6 and 10 hours. The length of training time depends on how large the data set is. The custom model will be available directly after training has been completed.

Q: How will I be able to use the model? How will I know whether it works better than the generic model provided by Amazon Transcribe?

The model will be made available in a customer's account under a model ID assigned by the customer prior to the training process. In order to use the model, a flag with the model ID needs to be added to the transcription request. Customers should test the model on their audio files and compare the output against results obtained from the generic engine.

Q: How many custom language models can I train? Can I have multiple models enabled concurrently for my account?

You may concurrently train up to 5 different models at any given time per AWS account. For each account, you can store a maximum of 10 models by default. If more are required, service limit increases can be made here.

Q: Are custom acoustic models supported?

No. Custom acoustic models are not supported. Custom language models are built off of text data that is relevant to your use case or domain.

Learn more about Amazon Transcribe pricing

Visit the pricing page