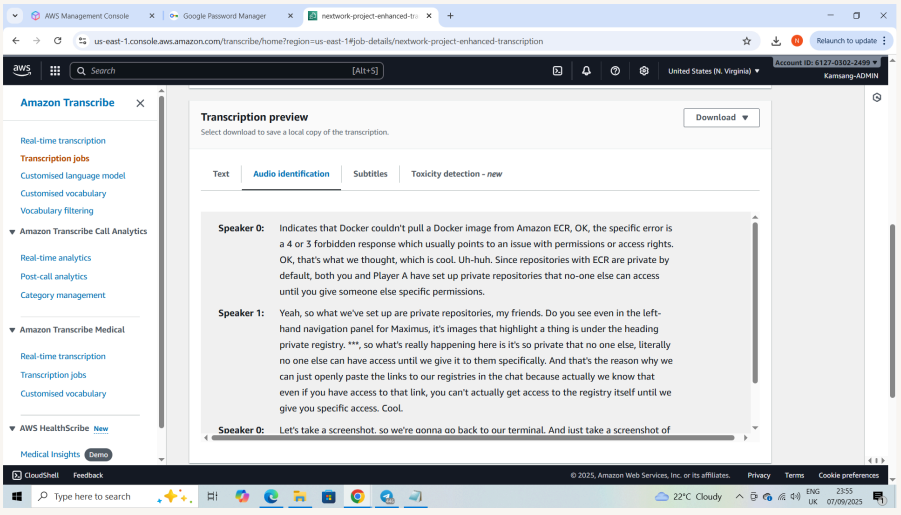
nextwork.org

Transcribe Audio Files with AI

Nchindo Boris



Nchindo Boris
NextWork Student

Introducing Today's Project!

In this project, I will demonstrate how to transcribe videos and audios using AWS services.

Tools and concepts

Services I used were; Amazon Transcribe and S3. Key concepts I learnt include; - Store a video with Amazon S3. - Transcribe your video with Amazon Transcribe. - Write custom vocabularies to improve your transcription's accuracy. - Use custom filters to automatically remove unwanted words. - Experiment with real-time transcription.

Project reflection

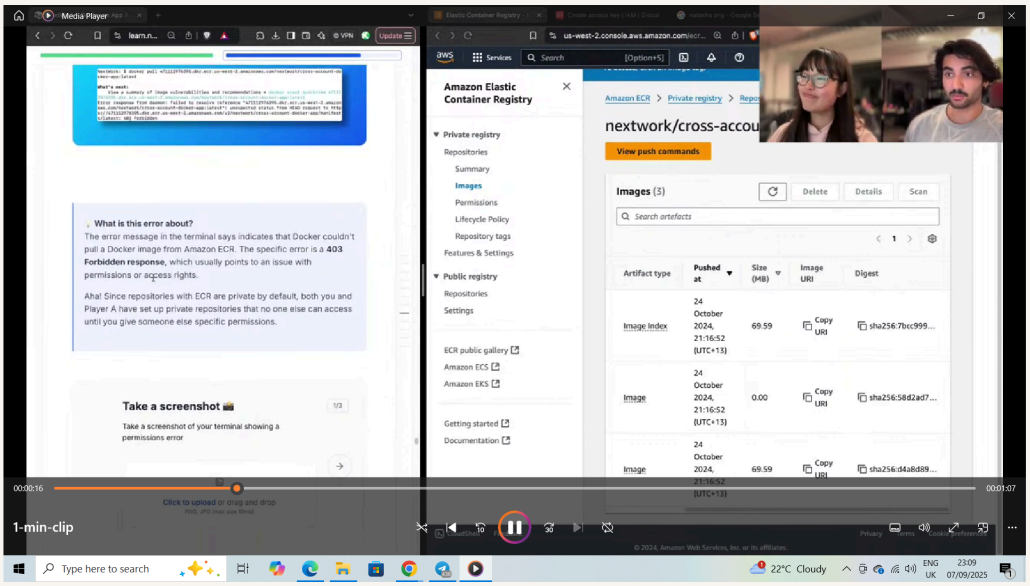
This project took me approximately 40 minutes to complete. The most challenging part was adjusting the custom vocabulary and vocabulary filter to handle tricky terms and remove unwanted filler words effectively. It was most rewarding to see how these customisations significantly improved the transcription quality and to experience live transcription working smoothly

I did this project today because I wanted to learn how AI- powered transcription works, especially using AWS tools like Amazon Transcribe

Nchindo Boris
NextWork Student

S3 and Transcribe

To set up for this project, I'm using an S3 bucket to store the video file I want to transcribe, as Amazon Transcribe requires media files to be stored in S3 to access and process them. The file I'm transcribing is now securely uploaded to the cloud, making it accessible for the transcription service to begin its work.

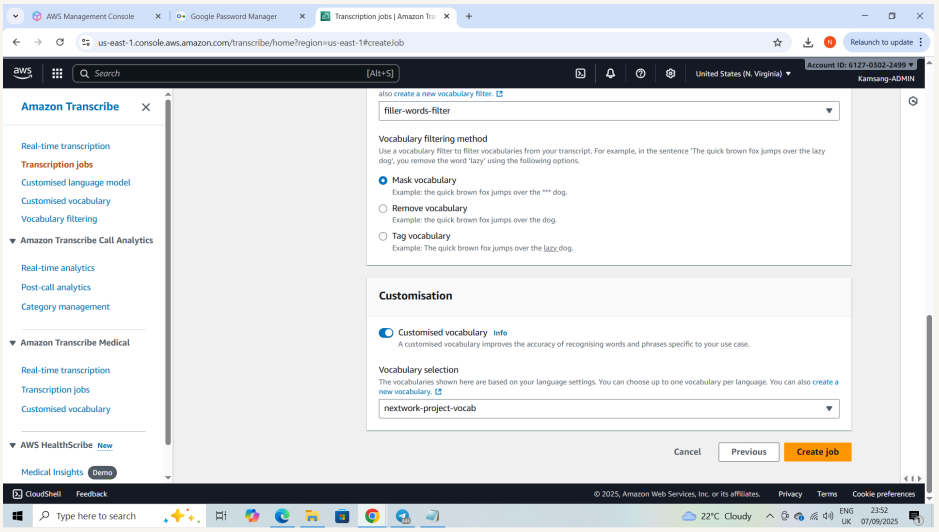


Run A Transcription Job

The steps to run a transcription job include uploading a media file to an S3 bucket, opening the Amazon Transcribe console, creating a new transcription job, selecting the general model type, and starting the job without any additional settings. Overall, this process took me just a few minutes, with the transcription itself taking around 1–2 minutes to complete. It was a straightforward and efficient setup, making it easy to get started with transcribing audio using AWS.

Amazon Transcribe uses model types to optimise transcription accuracy based on the nature of the audio content. These models are trained for specific use cases to better handle unique vocabulary, speech patterns, and background noise. Use cases of model types include the general model for everyday speech, the medical model for healthcare-related content, and the call analytics model for customer service and call centre recordings. Selecting the right model type ensures that the transcription is as accurate and relevant as possible for the given context.

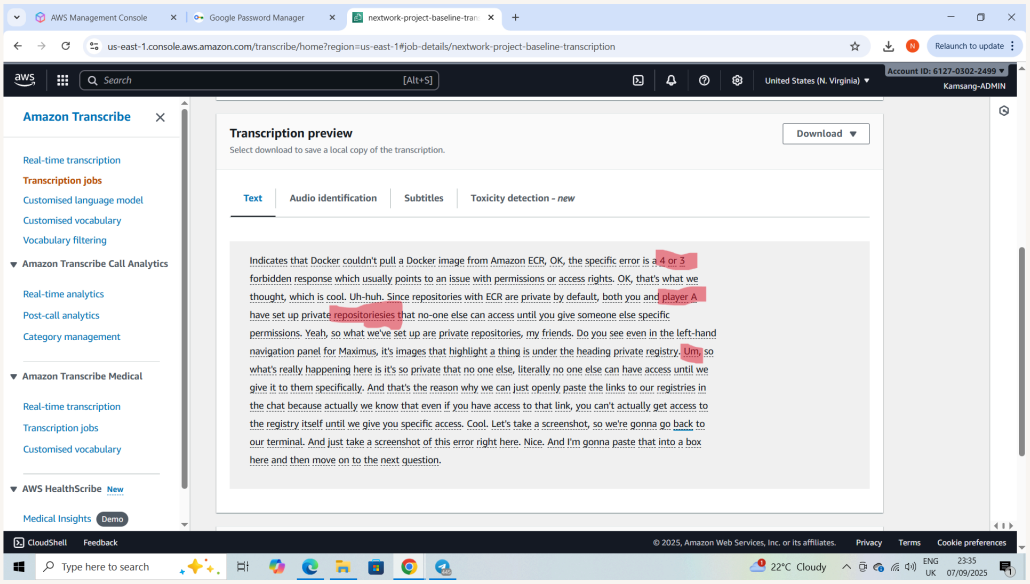
You can customise a transcription further with subtitling, which adds subtitles and speaker partitioning. Subtitles help people who are deaf, hard of hearing, or speak different languages to understand a video or audio. Speaker partitioning helps to label different speakers in an audio recording i.e. who said what.


Amazon Transcribe
Real-time transcription
Transcription jobs
Customised language model
Customised vocabulary
Vocabulary filtering
Amazon Transcribe Call Analytics
Real-time analytics
Post-call analytics
Category management
Amazon Transcribe Medical
Real-time transcription
Transcription jobs
Customised vocabulary
AWS HealthScribe New
Medical Insights Demo
filler-words-filter
Vocabulary filtering method
Use a vocabulary filter to filter vocabularies from your transcript. For example, in the sentence 'The quick brown fox jumps over the lazy dog', you remove the word 'lazy' using the following options.
Mask vocabulary
Example: the quick brown fox jumps over the *** dog.
Remove vocabulary
Example: the quick brown fox jumps over the dog.
Tag vocabulary
Example: The quick brown fox jumps over the lazy dog.
Customisation
Customised vocabulary Info
A customised vocabulary improves the accuracy of recognising words and phrases specific to your use case.
Vocabulary selection
nextwork-project-vocab
Cancel
Previous
Create job
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
22°C Cloudy

Baseline Transcript Review

To start using Amazon Transcribe, I first ran a baseline transcription job, which means I transcribed the audio without applying any custom settings or enhancements. This is because it's important to see how well the service performs on its own, using just the raw audio. The baseline transcription acts as a reference point, helping me measure how much improvements like custom vocabularies or filters actually enhance the final transcription later on.

While reviewing the baseline transcript, I noticed several inaccuracies. For example, “repositories” was misspelt as “repositoriesies”, and technical terms like “403 Forbidden” were mis-transcribed as “4 or 3 forbidden”. Speech fillers like “um” were included, which can be distracting in the final text. Additionally, “player A” should have been capitalised to “Player A” to reflect its proper use as a title in the context of the video. These happened because transcription tools can struggle with unclear pronunciation, background noise, specialised vocabulary, or words not commonly found in their training data.

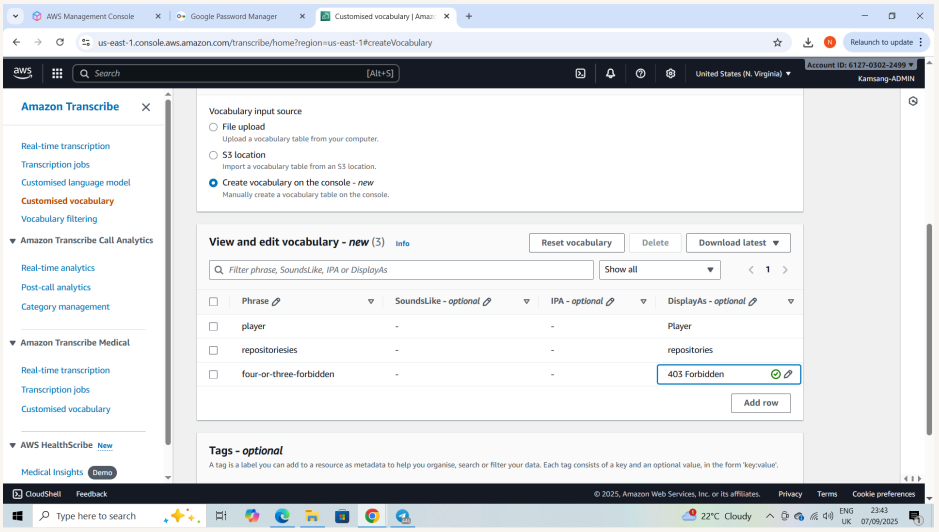

Amazon Transcribe
Real-time transcription
Transcription jobs
Customised language model
Customised vocabulary
Vocabulary filtering
Amazon Transcribe Call Analytics
Real-time analytics
Post-call analytics
Category management
Amazon Transcribe Medical
Real-time transcription
Transcription jobs
Customised vocabulary
AWS HealthScribe New
Medical Insights Demo
Transcription preview
Select download to save a local copy of the transcription.
Download
Text
Audio identification
Subtitles
Toxicity detection - new
Indicates that Docker couldn't pull a Docker image from Amazon ECR, OK, the specific error is a 4 or 3 forbidden response which usually points to an issue with permissions or access rights. OK, that's what we thought, which is cool. Uh-huh. Since repositories with ECR are private by default, both you and player A have set up private repositoriesies that no-one else can access until you give someone else specific permissions. Yeah, so what we've set up are private repositories, my friends. Do you see even in the left-hand navigation panel for Maximus, it's images that highlight a thing is under the heading private registry. Um, so what's really happening here is it's so private that no one else, literally no one else can have access until we give it to them specifically. And that's the reason why we can just openly paste the links to our registries in the chat because actually we know that even if you have access to that link, you can't actually get access to the registry itself until we give you specific access. Cool. Let's take a screenshot, so we're gonna go back to our terminal. And just take a screenshot of this error right here. Nice. And I'm gonna paste that into a box here and then move on to the next question.

Custom Vocabulary

I can resolve transcription inaccuracies using a custom vocabulary, which is a personalised list of words and phrases that I want Amazon Transcribe to recognise correctly. A custom vocabulary improves transcription accuracy by helping the service identify specialised terms, technical jargon, acronyms, or unique phrases that aren't part of the default language model, ensuring they are transcribed properly and consistently.

To create an item in a custom vocabulary, you need to define two values. They are the Phrase, which is the exact word or term you want Amazon Transcribe to recognise, and DisplayAs, which is how you want that phrase to appear in the transcription. This helps ensure that words are both correctly identified and consistently formatted in the final transcript

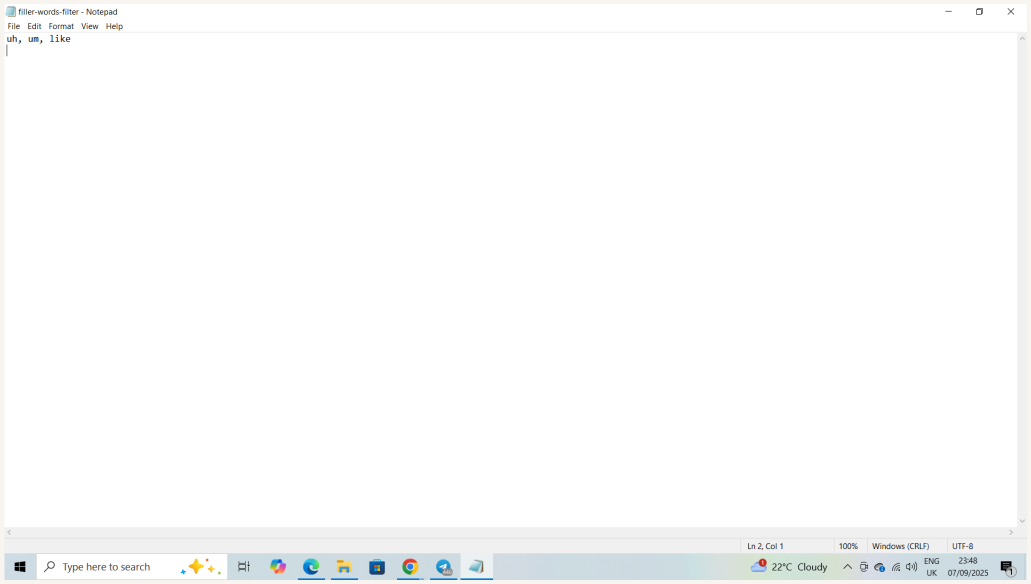
My custom vocabulary defines specific terms and corrections to improve transcription accuracy, including fixing the misspelt "repositoriesies" to "repositories", capitalising "player" to "Player", and correcting "four-or-three-forbidden" to properly represent "403 Forbidden" without spaces or numbers.


Amazon Transcribe
Real-time transcription
Transcription jobs
Customised language model
Customised vocabulary
Vocabulary filtering
Amazon Transcribe Call Analytics
Real-time analytics
Post-call analytics
Category management
Amazon Transcribe Medical
Real-time transcription
Transcription jobs
Customised vocabulary
AWS HealthScribe New
Medical Insights Demo
Vocabulary input source
File upload
Upload a vocabulary table from your computer.
S3 location
Import a vocabulary table from an S3 location.
Create vocabulary on the console - new
Manually create a vocabulary table on the console.
View and edit vocabulary - new (3) Info
Reset vocabulary
Delete
Download latest
Filter phrase, SoundsLike, IPA or DisplayAs
Show all
Phrase
SoundsLike - optional
IPA - optional
DisplayAs - optional
player
repositoriesies
four-or-three-forbidden
403 Forbidden
Add row
Tags - optional
A tag is a label you can add to a resource as metadata to help you organise, search or filter your data. Each tag consists of a key and an optional value, in the form 'key:value'.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Type here to search
22°C Cloudy

Vocabulary Filters

Another feature in Transcribe is vocabulary filtering, which automatically removes or masks unwanted words or phrases from transcripts, such as sensitive information. It's different from custom vocabularies because instead of teaching Transcribe to recognise and correctly display specific terms, vocabulary filters block or clean up words I don't want to appear in the transcription, making it a more efficient way to manage unwanted content.

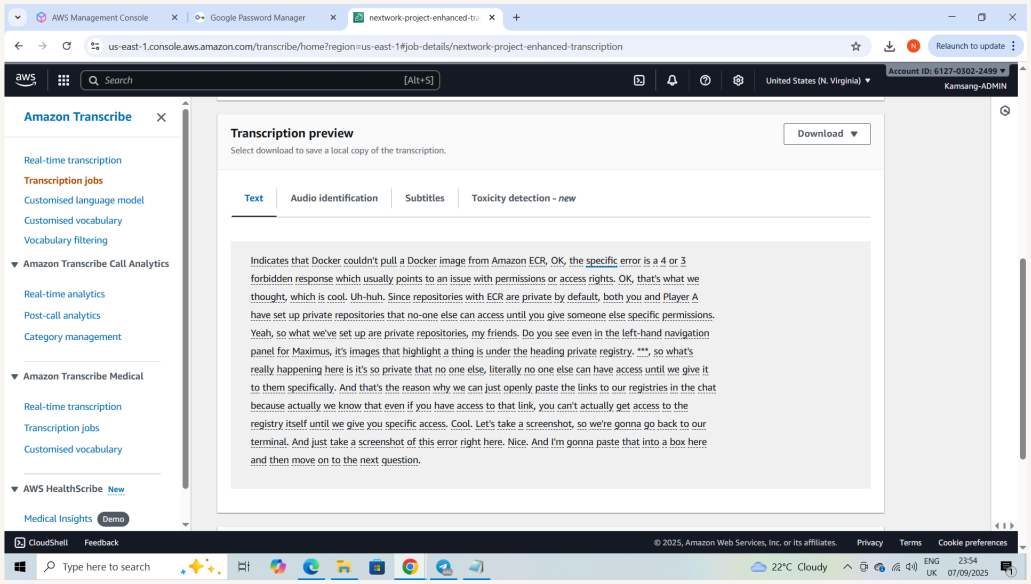
My vocabulary filter removes unwated text, i.e. common filler words like "um," "uh," "like," and other unnecessary speech fillers that clutter the transcription. To set up this filter, I first created a plain text file listing these unwanted words, separated by commas, and then uploaded it to Amazon Transcribe to apply the filter during transcription.


filler-words-filter - Notepad
File Edit Format View Help
uh, um, like
Ln 2, Col 1
100%
Windows (CRLF)
UTF-8
Type here to search
22°C Cloudy
ENG
UK
23:48
07/09/2025

Enhanced Transcription

I ran a new transcription with my custom vocabulary and filtering settings

The enhanced transcription is better than the baseline because it correctly displays key terms from my custom vocabulary, such as properly capitalising Player A and fixing the spelling of repositories. It also removes filler words like “um” using the vocabulary filter, making the transcript cleaner and easier to read. Additionally, speaker partitioning helps separate who is speaking, which improves clarity in conversations. Although “403 Forbidden” is still incorrect, this shows that further customisation or training may be needed for highly technical terms. Overall, the transcription is noticeably more accurate than the baseline version.

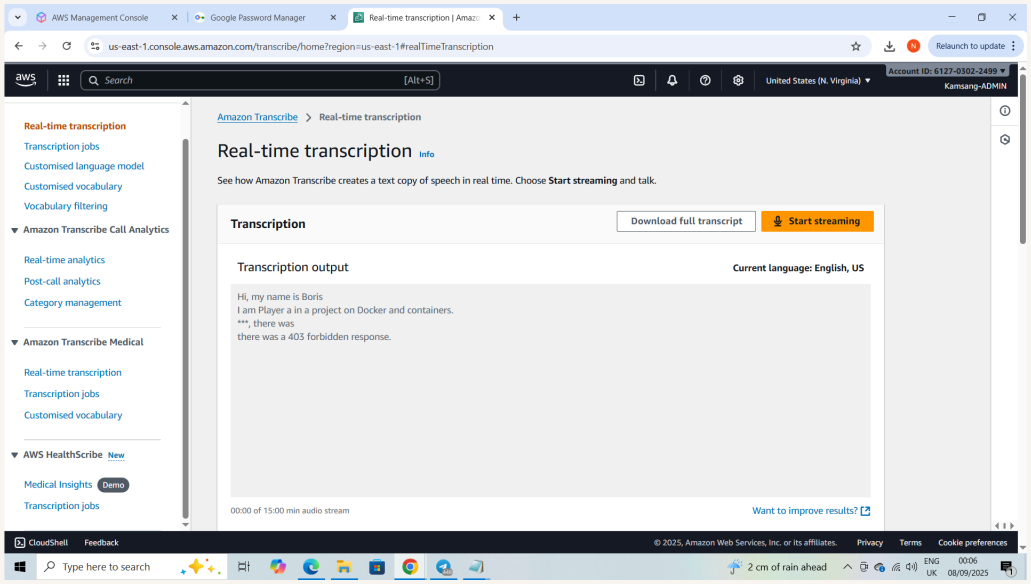

Amazon Transcribe
Real-time transcription
Transcription jobs
Customised language model
Customised vocabulary
Vocabulary filtering
Amazon Transcribe Call Analytics
Real-time analytics
Post-call analytics
Category management
Amazon Transcribe Medical
Real-time transcription
Transcription jobs
Customised vocabulary
AWS HealthScribe New
Medical Insights Demo
Transcription preview
Select download to save a local copy of the transcription.
Download
Text
Audio identification
Subtitles
Toxicity detection - new
Indicates that Docker couldn't pull a Docker image from Amazon ECR, OK, the specific error is a 4 or 3 forbidden response which usually points to an issue with permissions or access rights. OK, that's what we thought, which is cool. Uh-huh. Since repositories with ECR are private by default, both you and Player A have set up private repositories that no-one else can access until you give someone else specific permissions. Yeah, so what we've set up are private repositories, my friends. Do you see even in the left-hand navigation panel for Maximus, it's images that highlight a thing is under the heading private registry. ***, so what's really happening here is it's so private that no one else, literally no one else can have access until we give it to them specifically. And that's the reason why we can just openly paste the links to our registries in the chat because actually we know that even if you have access to that link, you can't actually get access to the registry itself until we give you specific access. Cool. Let's take a screenshot, so we're gonna go back to our terminal. And just take a screenshot of this error right here. Nice. And I'm gonna paste that into a box here and then move on to the next question.

Nchindo Boris
NextWork Student

Real Time Transcription

For my project extension, I experimented with real-time transcription, which converts speech to text instantly i.e. as the speaker is still talking. This very useful feature for apps that want to offer live captioning and voice commands!

Even during real-time transcription, I was able to apply custom vocabulary and vocabulary filtering, which helped improve the recognition of specific terms and remove unnecessary filler words. Overall, compared to a standard transcription job, real-time transcription was faster and surprisingly effective, though there were still occasional inaccuracies with complex terms like “403 Forbidden”. Despite that, it performed well in recognising phrases from my custom setup, making it a powerful option for use cases like live captioning, voice assistants, or interactive applications.


AWS Management Console
Google Password Manager
Real-time transcription | Amazo
us-east-1.console.aws.amazon.com/transcribe/home?region=us-east-1#realTimeTranscription
Relaunch to update
aws
Search
[Alt+S]
United States (N. Virginia)
Account ID: 6127-0302-2499
Kamsang-ADMIN
Real-time transcription
Transcription jobs
Customised language model
Customised vocabulary
Vocabulary filtering
Amazon Transcribe Call Analytics
Real-time analytics
Post-call analytics
Category management
Amazon Transcribe Medical
Real-time transcription
Transcription jobs
Customised vocabulary
AWS HealthScribe New
Medical Insights Demo
Transcription jobs
Amazon Transcribe > Real-time transcription
Real-time transcription Info
See how Amazon Transcribe creates a text copy of speech in real time. Choose Start streaming and talk.
Transcription
Download full transcript
Start streaming
Transcription output
Current language: English, US
Hi, my name is Boris
I am Player a in a project on Docker and containers.
***, there was
there was a 403 forbidden response.
00:00 of 15:00 min audio stream
Want to improve results?
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Type here to search
2 cm of rain ahead
ENG
UK
00:06
08/09/2025

The place to learn & showcase your skills

Check out nextwork.org for more projects