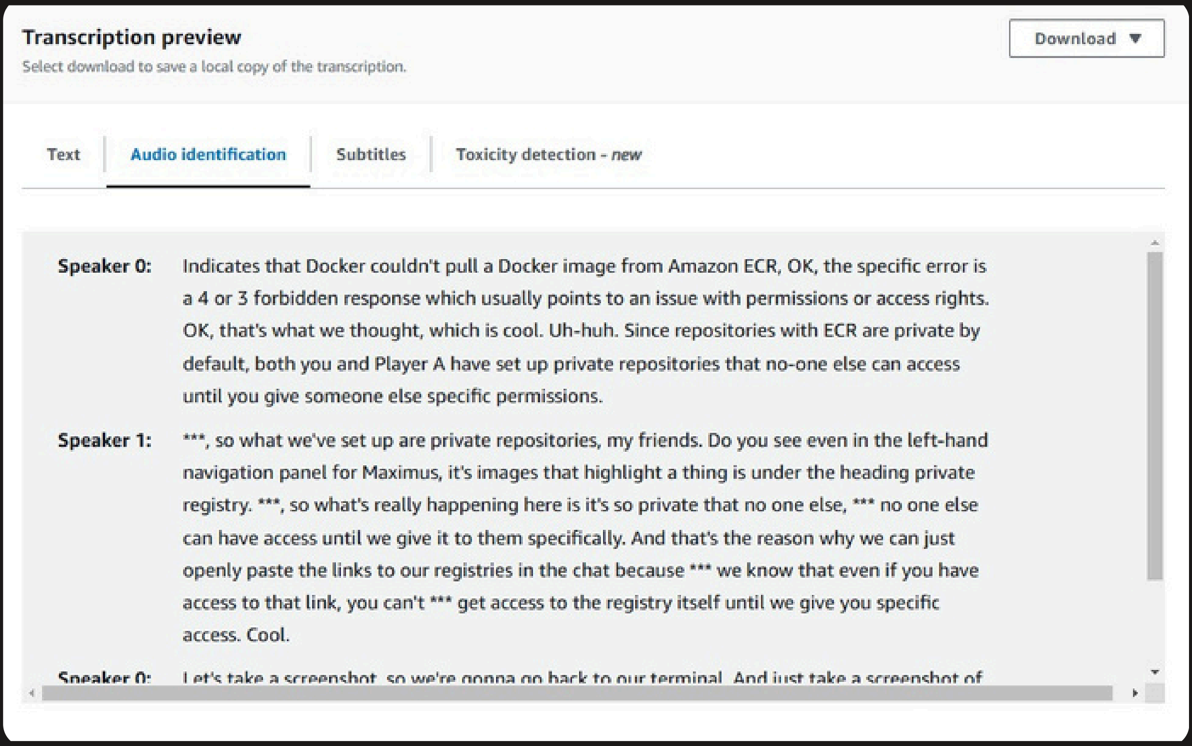
nextwork.org

Transcribe Audio/Video Files with AI

Kanika Mathur
github.com/KanikaGenesis



Transcription preview
Select download to save a local copy of the transcription.
Download ▼
Text
Audio identification
Subtitles
Toxicity detection - *new*

Speaker 0: 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.

Speaker 1: ***, 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, *** 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 *** we know that even if you have access to that link, you can't *** get access to the registry itself until we give you specific access. Cool.

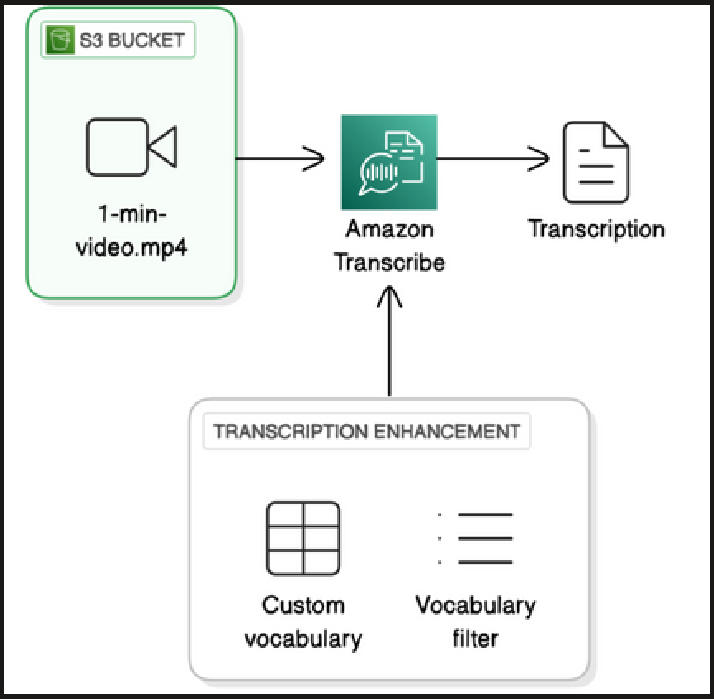
Speaker 0: Let's take a screenshot, so we're gonna go back to our terminal. And just take a screenshot of

Introducing Today's Project!

In this project, I demonstrate how AI can be used to generate accurate transcriptions for audio and video files. The goal was to explore how transcription technology can enhance app accessibility and support features like voice commands.

Tools and concepts

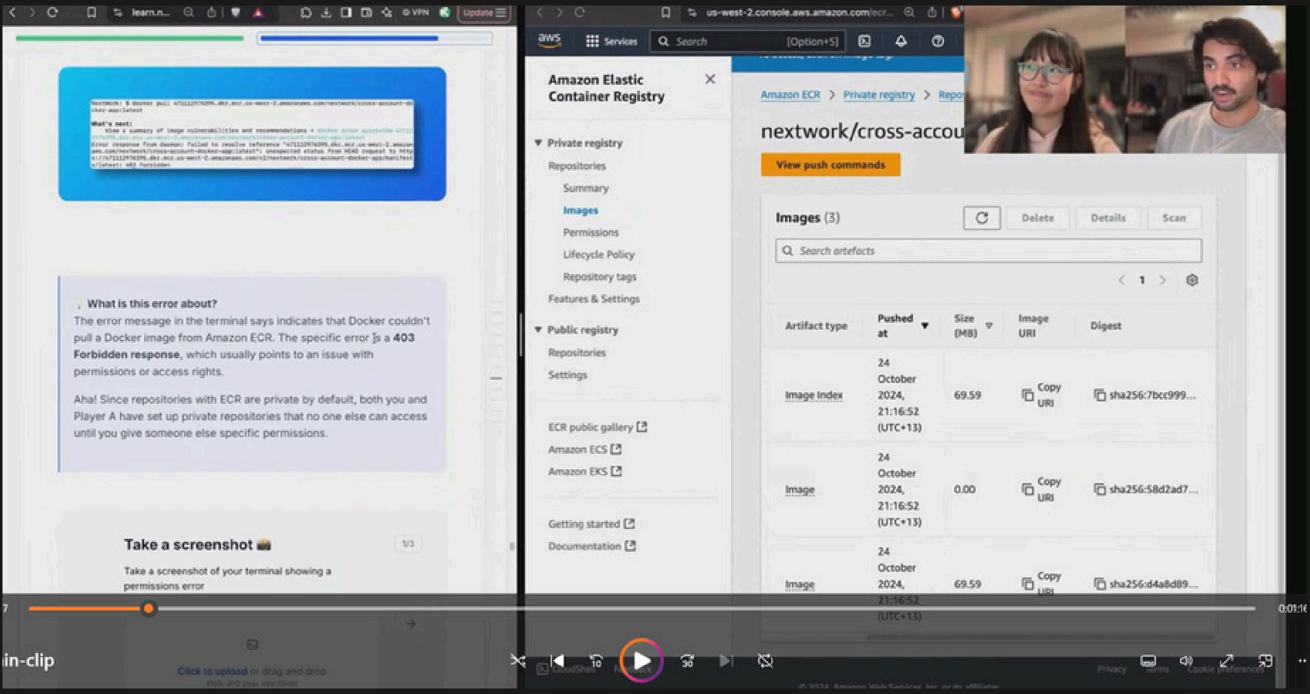
To achieve this, I used Amazon S3 to store my media files and Amazon Transcribe to automatically convert speech into text. Through this project, I learned about key concepts such as vocabulary filters, custom vocabulary creation, and real-time streaming transcription. I also discovered that Amazon Transcribe can automatically detect files for transcription when the S3 bucket name contains the keyword "transcribe," streamlining the workflow between storage and processing.



Kanika Mathur
github.com/KanikaGenesis

S3 and Transcribe

To set up for this project, I’m using an S3 bucket to store a video file because I'm going to transcribe it. The file I’m transcribing is a video clip that's 1 minute and 23 seconds in duration and it shows a demo of a NextWork project on AWS ECR and Docker.



```
NextWork: $ docker pull [illegible].dkr.ecr.us-west-2.amazonaws.com/nextwork/cross-account-docker-app:latest

What's next:
    View a summary of image vulnerabilities and recommendations → docker scout quickview [illegible].dkr.ecr.us-west-2.amazonaws.com/nextwork/cross-account-docker-app:latest
Error response from daemon: failed to resolve reference "[illegible].dkr.ecr.us-west-2.amazonaws.com/nextwork/cross-account-docker-app:latest": unexpected status from HEAD request to https://[illegible].dkr.ecr.us-west-2.amazonaws.com/v2/nextwork/cross-account-docker-app/manifests/latest: 403 Forbidden
```

💡 What is this error about?

The error message in the terminal says indicates that Docker couldn't pull a Docker image from Amazon ECR. The specific error is a **403 Forbidden response**, which usually points to an issue with permissions or access rights.

Aha! 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.

Take a screenshot 📸 1/3

Take a screenshot of your terminal showing a permissions error

Click to upload or drag and drop

Amazon Elastic Container Registry

Amazon ECR > Private registry > Repo

nextwork/cross-accou

Images (3)

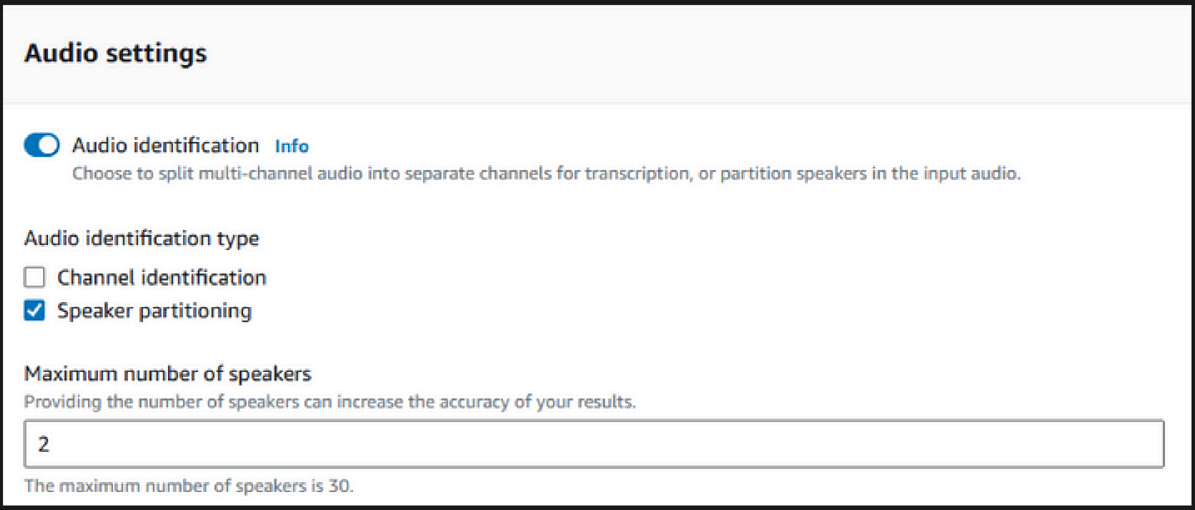
Artifact type	Pushed at	Size (MB)	Image URI	Digest
Image Index	24 October 2024, 21:16:52 (UTC+13)	69.59	Copy URI	sha256:7bcc999...
Image	24 October 2024, 21:16:52 (UTC+13)	0.00	Copy URI	sha256:58d2ad7...
Image	24 October 2024, 21:16:52 (UTC+13)	69.59	Copy URI	sha256:d4a8d89...

Run A Transcription Job

The steps to run a transcription job include language selection and the model type. I specified the input data, which is where the video file is. I also provided the output data, which is where the results of the transcription job will be stored. Overall, this process took me 5 minutes to set up.

Amazon Transcribe uses model types to learn how to translate speech to text. A model type is like a guide that Transcribe will use to understand how to translate the language; there are many types of models used in conversations. For example, there are model types that are customized to focus on specific fields like medicine, law, and others.

We can customize a transcription further with subtitling, which adds subtitles to a video (great for accessibility or translations) and speaker partitioning, which helps to identify multiple speakers in an audio file.



Audio settings

● Audio identification Info
Choose to split multi-channel audio into separate channels for transcription, or partition speakers in the input audio.

Audio identification type

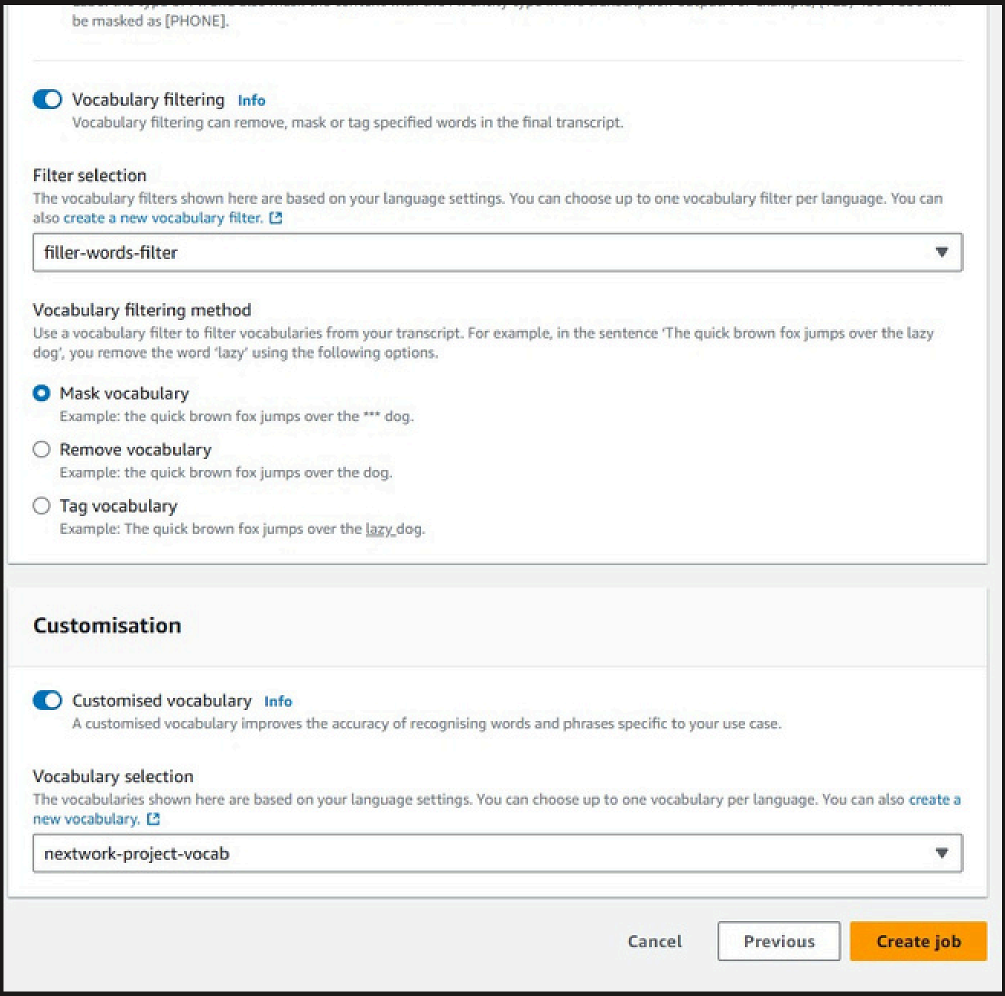
□ Channel identification

☑ Speaker partitioning

Maximum number of speakers
Providing the number of speakers can increase the accuracy of your results.

2

The maximum number of speakers is 30.



be masked as [PHONE].

● Vocabulary filtering Info
Vocabulary filtering can remove, mask or tag specified words in the final transcript.

Filter selection
The vocabulary filters shown here are based on your language settings. You can choose up to one vocabulary filter per language. You can also create a new vocabulary filter.

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.

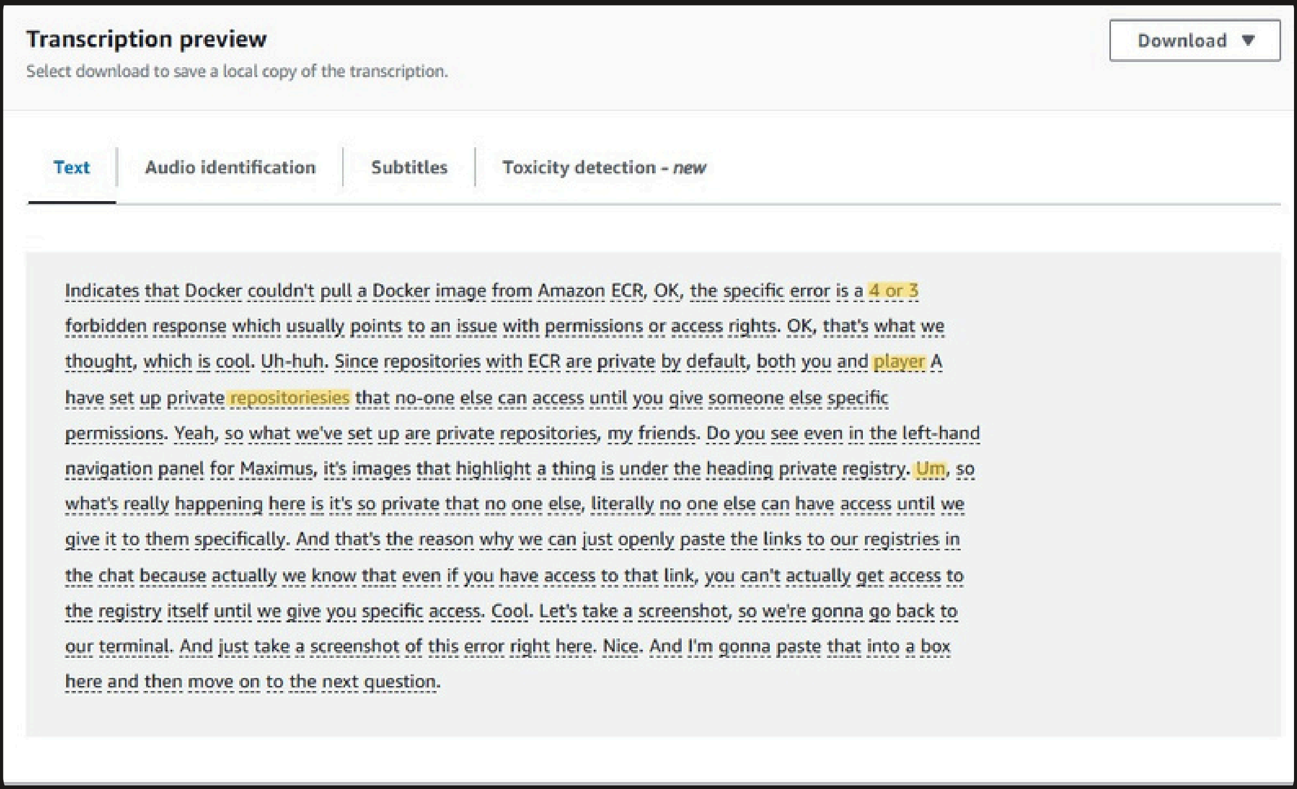
Vocabulary selection
The vocabularies shown here are based on your language settings. You can choose up to one vocabulary per language. You can also create a new vocabulary.

nextwork-project-vocab ▼

Baseline Transcript Review

To start using Amazon Transcribe, I first ran a baseline transcription job, which means that my transcription won't use any additional settings or tools that are offered by the Transcribe service. This will allow me to monitor the outcomes every time I make a change in the original transcription.

While reviewing the baseline transcript, I found a few inaccuracies, including typos, jargon, filler words, and lack of context. For example, I found three mistakes. One mistake was a typo on 'repositoriesies', which was meant to be 'repositories'. The '4 or 3 forbidden' is a jargon error since it should be a '403 Forbidden error'. It also shouldn't use a filler word like 'um,' which is not a proper word.


Transcription preview
Select download to save a local copy of the transcription.
Download ▼
Text
Audio identification
Subtitles
Toxicity detection - *new*

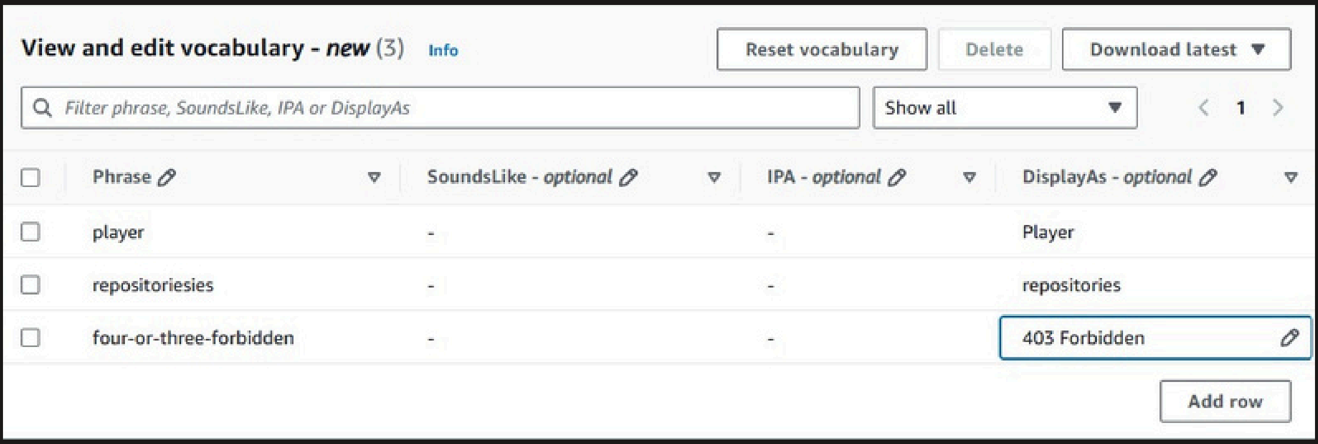
Indicates that Docker couldn't pull a Docker image from Amazon ECR, OK, the specific error is a 4 or 3 forbidden response which usually points to an issue with permissions or access rights. OK, that's what we thought, which is cool. Uh-huh. Since repositories with ECR are private by default, both you 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 list of jargon and misspelt words that I'll need Transcribe to process and correct. A custom vocabulary improves the accuracy of the transcription by providing the correct version of the transcription.

To create an item in a custom vocabulary, we need to define two values. They are Phrase and DisplayAs; the 'Phrase' refers to the phrase or word we want Transcribe to recognize. The second one is 'DisplayAs', which refers to how we want Transcribe to show the phrase in the transcription.

My custom vocabulary defines proper spelling/ treatment of the three inaccuracies found in the baseline transcription. The first one is 'player,' and it will be displayed as 'Player.' The second one is '4 or 3 forbidden', which will be corrected to display as '403 Forbidden'. The third one is a spelling error, which was 'repositoriesies'. It will be displayed as 'repositories'.

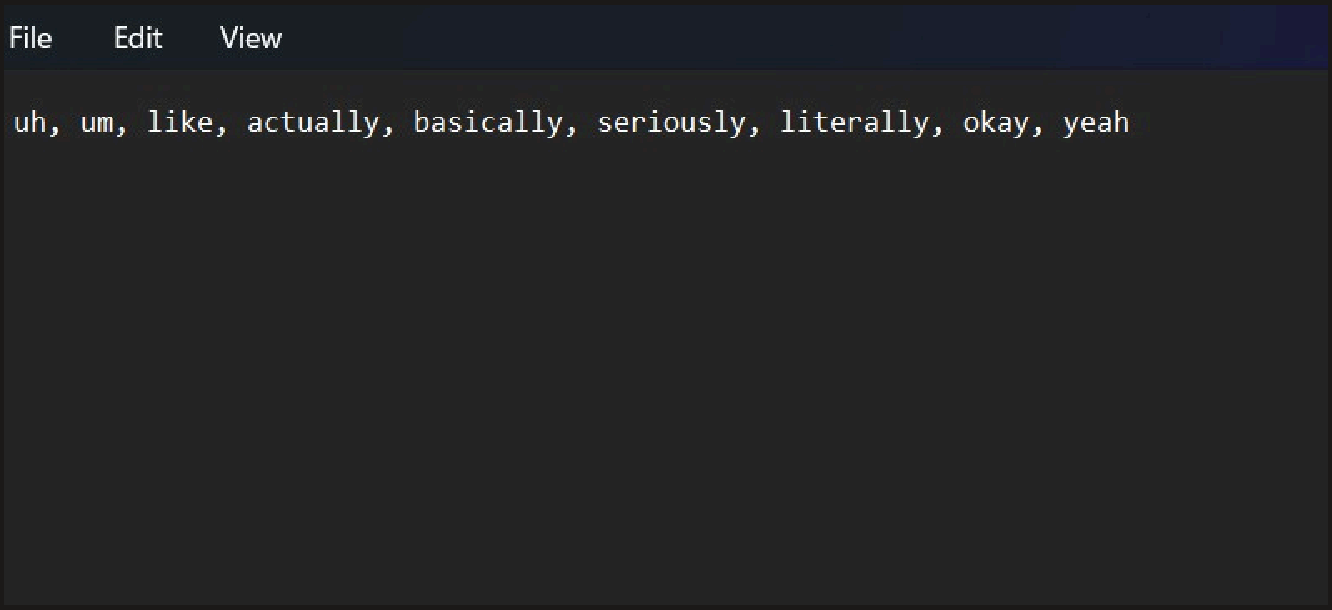

View and edit vocabulary - *new* (3)
Info
Reset vocabulary
Delete
Download latest ▼
Filter phrase, SoundsLike, IPA or DisplayAs
Show all ▼
< 1 >
Add row

☐	Phrase	SoundsLike - *optional*	IPA - *optional*	DisplayAs - *optional*
☐	player	-	-	Player
☐	repositoriesies	-	-	repositories
☐	four-or-three-forbidden	-	-	403 Forbidden

Vocabulary Filters

Another feature in Transcribe is vocabulary filtering, which is a tool for removing unwanted words. It's different from custom vocabularies - vocabularies are used for words we DO want to transcribe (accurately), whereas filtered words aren't wanted.

My vocabulary filter will remove unnecessary text; there are filler words that will get removed, like 'um.' To set up this filter, I first created a filter file; this file contains a list of words, and each of them is separated by a comma. These words will get removed when Transcribe processes the transcription with this filter.

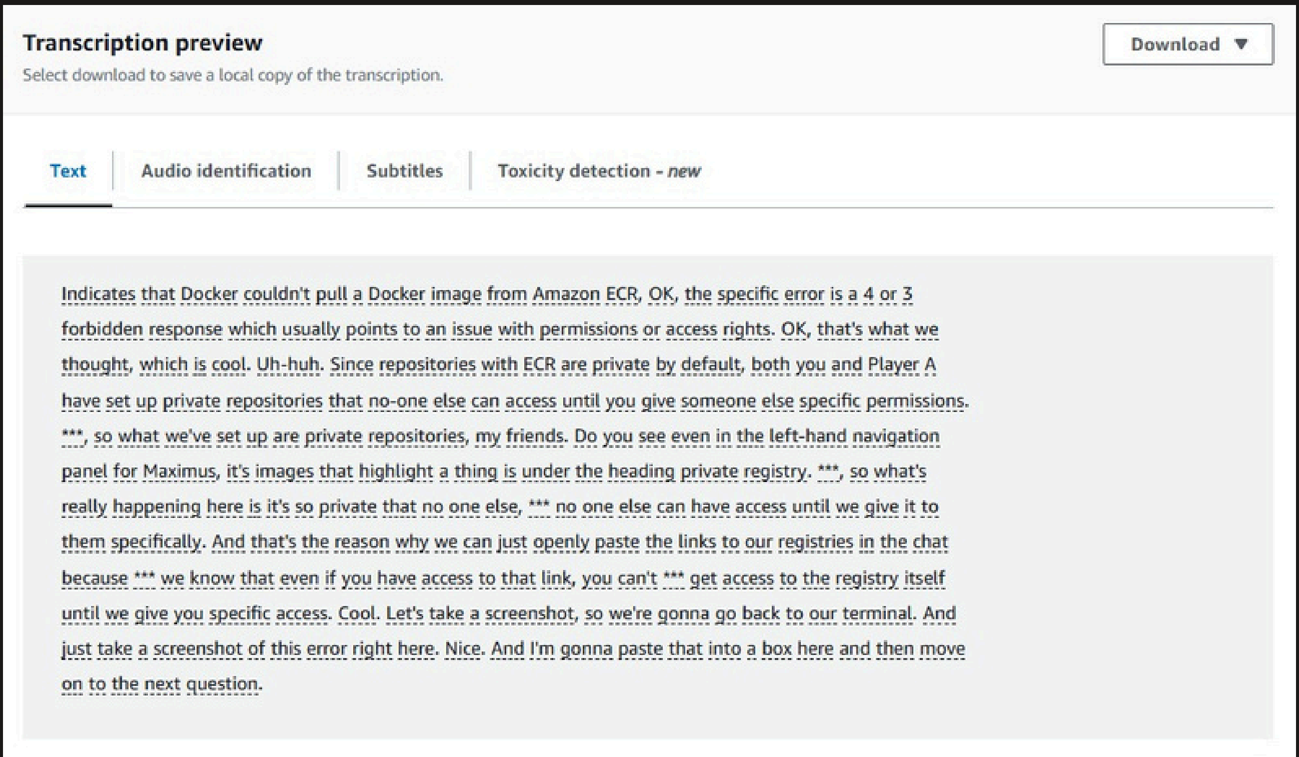


```
uh, um, like, actually, basically, seriously, literally, okay, yeah
```

Enhanced Transcription

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

The enhanced transcription is better than the baseline because the inaccuracies were corrected. The misspelled 'repositoriesies' was spelled correctly to 'repositories'. The 'player A' phrase was capitalized to become 'Player A'. The vocabulary filter worked, and the filler words were masked and replaced with '***.' We'll need more training to fix 403 error.

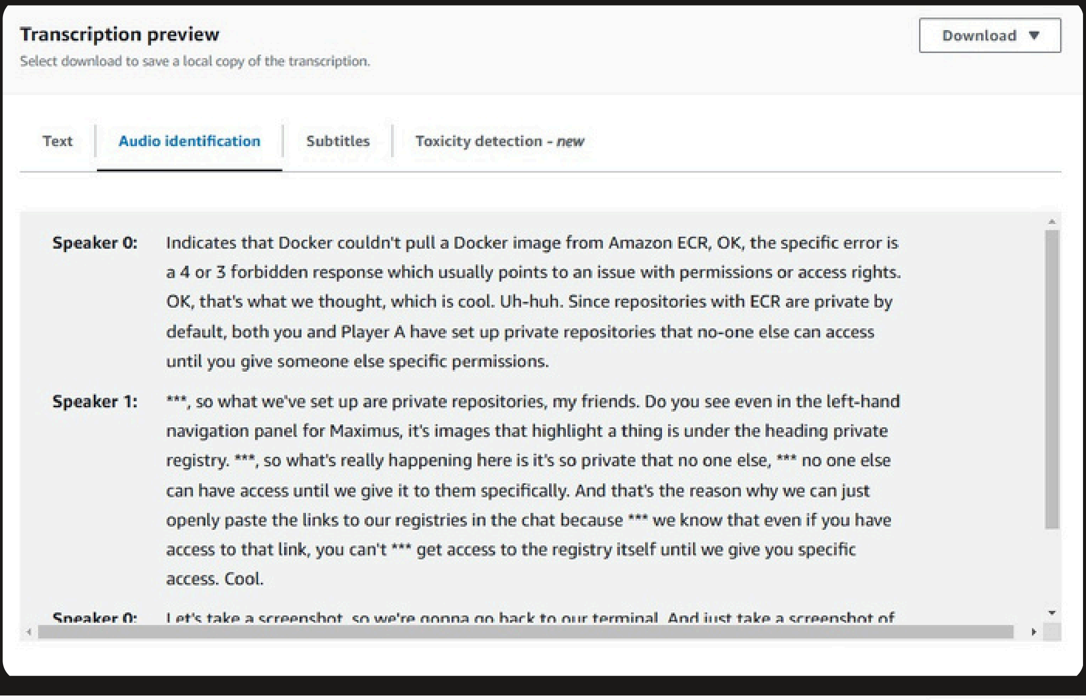


Transcription preview

Select download to save a local copy of the transcription.

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. ***, 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, *** 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 *** we know that even if you have access to that link, you can't *** 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.

One of the features I explored was speaker identification, where the service can automatically label different speakers, such as Speaker 0 and Speaker 1 which makes it easy to understand conversations involving multiple people. I also learned how Transcribe provides timestamps with the text, which can be used to create subtitles (in formats like SRT) or captions for videos allowing users to follow along visually.



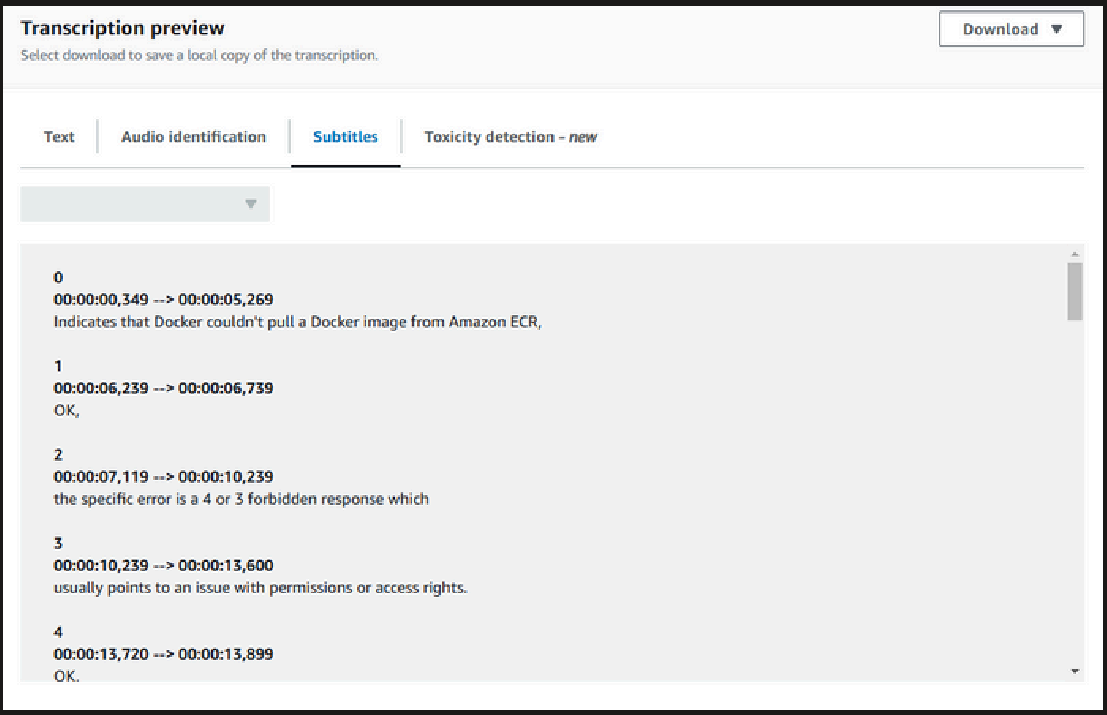
Transcription preview

Select download to save a local copy of the transcription.

Speaker 0: 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.

Speaker 1: ***, 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, *** 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 *** we know that even if you have access to that link, you can't *** get access to the registry itself until we give you specific access. Cool.

Speaker 0: Let's take a screenshot, so we're gonna go back to our terminal. And just take a screenshot of



Transcription preview

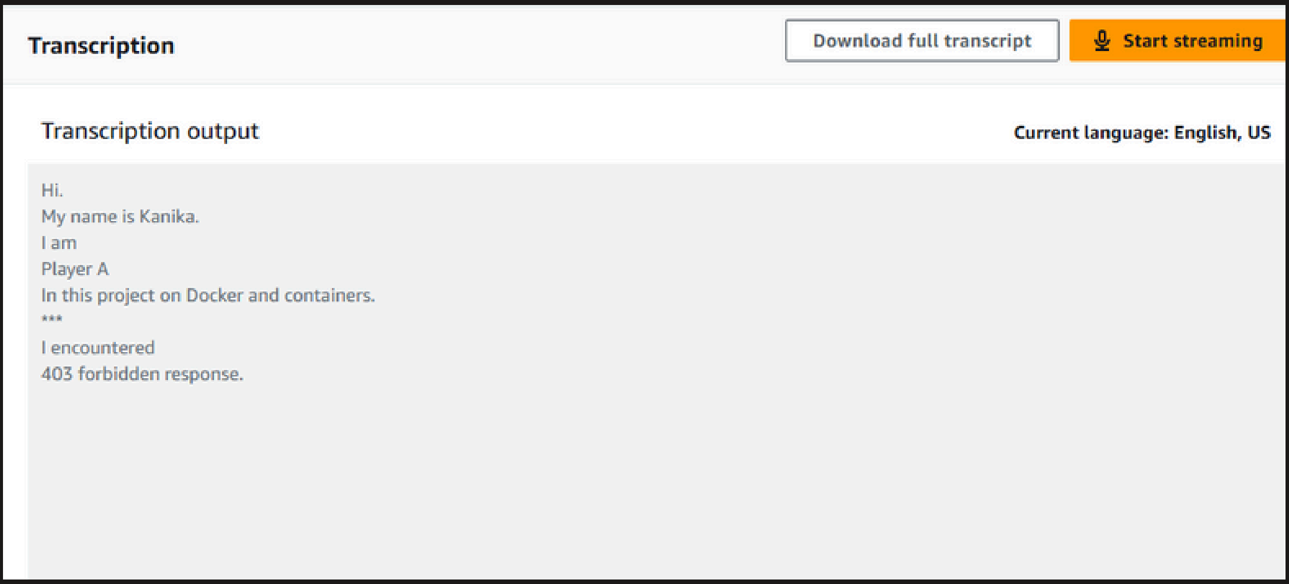
Select download to save a local copy of the transcription.

```
0
00:00:00,349 --> 00:00:05,269
Indicates that Docker couldn't pull a Docker image from Amazon ECR,

1
00:00:06,239 --> 00:00:06,739
OK,

2
00:00:07,119 --> 00:00:10,239
the specific error is a 4 or 3 forbidden response which

3
00:00:10,239 --> 00:00:13,600
usually points to an issue with permissions or access rights.

4
00:00:13,720 --> 00:00:13,899
OK.
```

Real Time Transcription

For my project extension, I experimented with real-time transcription, which is a transcription that happens live while the speaker is still talking. This is helpful for apps that want to deliver live captioning or voice commands.

Even during real-time transcription, I could use features like customized vocabulary and vocabulary filtering. Overall, compared to a transcription job, real- time transcription was still just as accurate. Although, the '403 Forbidden' still has errors.



Transcription

Transcription output

Current language: English, US

Hi.
My name is Kanika.
I am
Player A
In this project on Docker and containers.
***
I encountered
403 forbidden response.

NextWork.org

Everyone should be in a job they love.

Check out nextwork.org for more projects