

NextWork.org

Transcribe Audio Files with AI



tahirgroot@gmail.com

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 ***. 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: Yeah, *** 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. ***, *** what's really happening here is it's *** 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. ***.

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



TA

tahirgroot@gmail.com
NextWork Student

NextWork.org

Introducing Today's Project!

in this project, I will demonstrate how to use AI to create a transcription for audio and videos. I am doing this project to learn how I can create transcriptions that enable voice commands or improve accessibility for an app.

Tools and concepts

Services I used were Amazon S3 for video storage and Amazon Transcribe for transcription. Key concepts I learnt include live transcribing, custom vocabularies and filters, Transcribe <> S3's connection.

Project reflection

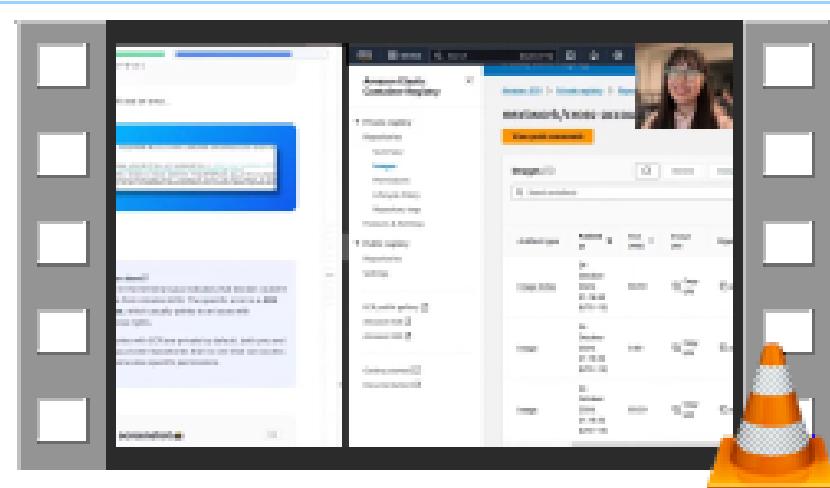
This project took me approximately 1 hour. The most challenging part was creating the custom vocabulary. It was most rewarding to do a live transcript and see the custom vocab + filter at work straight away!

I did this project today to learn about transcription and AI tools like transcribe! This project helped meet my goals by letting us practise transcription + ways to make it even more accurate.

S3 and Transcribe

To set up for this project, I'm using an S3 bucket to store the video file i am transcribing. The file because i am transcribing is a 1 minute clip of an exsisting Nextwork project demo. This is a great opportunity to provide subtitles to my videos!

Desktop > aws projects > Transcribe Audio Files with AI



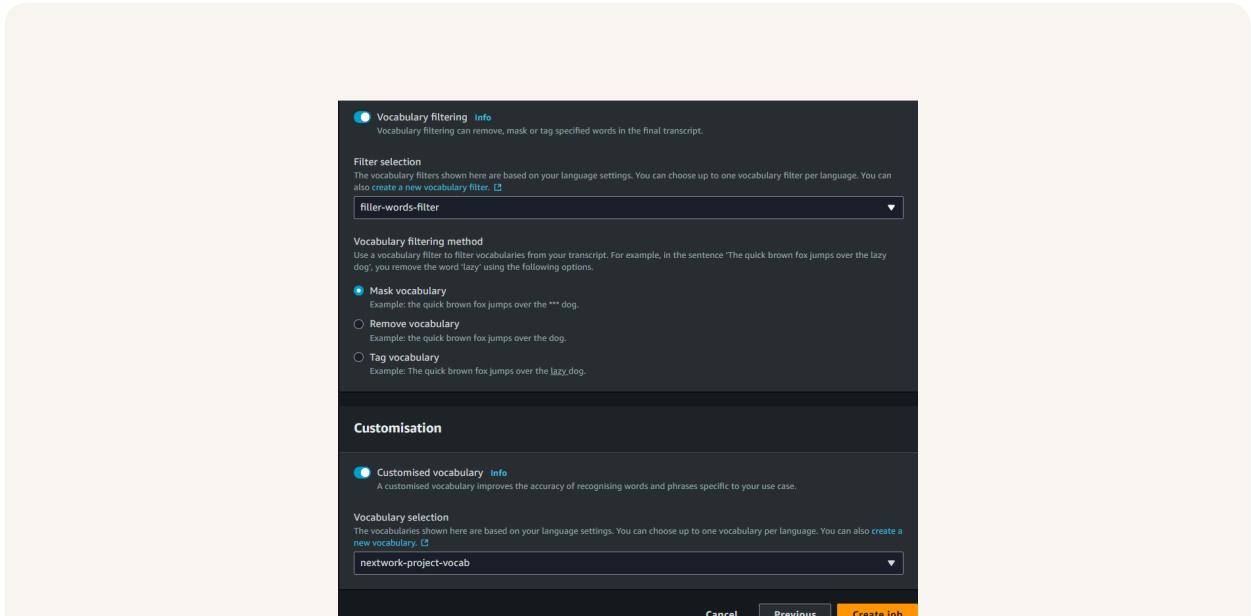
1-min-clip.mp4

Run A Transcription Job

The steps to run a transcription job include selecting a language and model type; selecting input and output data(i.e. where the video/audio is from, and where the transcription should be stored). Overall, this process took me about 5 minutes.

Amazon Transcribe uses model types to learn how to translate speech to text. You can think of a model type as in instruction or guide that transcribe uses. Use cases of model types include creating custom models for medical/legal conversations.

You can customise 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 a audio file.



Baseline Transcript Review

To start using Amazon Transcribe, I first ran a baseline transcription job, which means a transcription that doesn't use any of the additional features or tools provided by Amazon Transcribe. This lets me compare it with the results of any changes.

While reviewing the baseline transcript, I noticed a few inaccuracies, including typos (e.g. 'repositories' spelt as 'repositoriesies'), jargon (e.g. '403 Forbidden' transcribed as '4 or 3 forbidden'), filler words (e.g. 'um') and lack of context.

The screenshot shows the 'Transcription preview' section of the Amazon Transcribe console. At the top, there's a 'Download' button with a dropdown arrow. Below it, a message says 'Select download to save a local copy of the transcription.' A navigation bar at the bottom includes tabs for 'Text' (which is selected), 'Audio identification', 'Subtitles', and 'Toxicity detection - new'. The main content area displays a block of transcribed text:

```
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. 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 defined list of technical or misspelt words that I want Transcribe to process accurately. A custom vocabulary improves accuracy by defining the correct transcription.

To create an item in a custom vocabulary, you need to define two values. They are 'Phrase' which means the phrase we want Transcribe to detect; and 'DisplayAs', which is the way we want Transcribe to display the word in its transcriptions.

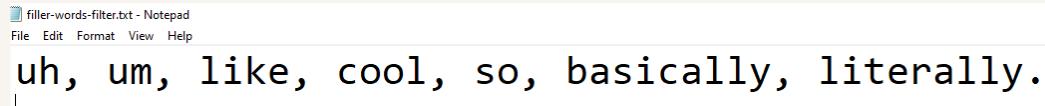
My custom vocabulary defines the proper spelling/treatment of three inaccuracies in my baseline transcription. First, 'player' -> 'Player'. Second, '4 or 3 forbidden' -> '403 Forbidden'. Third, 'repositoriesies' -> 'repositories'.

View and edit vocabulary - new (3)					Info	Reset vocabulary	Delete	Download latest ▾
<input type="text"/> Filter phrase, SoundsLike, IPA or DisplayAs					Show all	< 1 >		
	Phrase ⚠	SoundsLike - optional ⚠	IPA - optional ⚠	DisplayAs - optional ⚠				
<input type="checkbox"/>	player	-	-	Player				
<input type="checkbox"/>	repositoriesies	-	-	repositories				
<input type="checkbox"/>	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 I DO want to transcribe (accurately), whereas filtered words isn't wanted.

My vocabulary filter removes unwanted text, like filler words('um'). To set up this filter, I first created a filter file, which is a comma separated list of words



A screenshot of a Windows Notepad window titled "filler-words-filter.txt - Notepad". The window shows a menu bar with File, Edit, Format, View, and Help. The main text area contains the following text:

```
uh, um, like, cool, so, basically, literally.
```

Enhanced Transcription

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

The enhanced transcription is better than the baseline because it does not have the same typos; its now capitalised 'player' to turn into a title, and filler words are correctly identified and masked!. I'll need more training to fix '403' error

Transcription preview
Select download to save a local copy of the transcription.

[Download ▾](#)

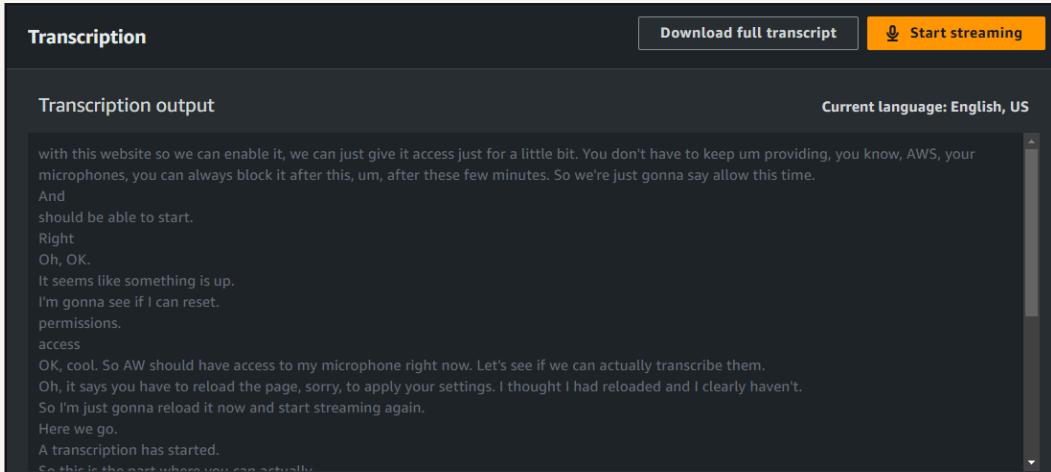
[Text](#) [Audio identification](#) [Subtitles](#) [Toxicity detection - new](#)

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 **. 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 , ** 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. **, ** what's really happening here is it's ** 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. **. Let's take a screenshot, ** 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.

Real Time Transcription

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

Even during real-time transcription, I could use features like vocabulary filtering and custom vocabulary. Compared to a transcription job, real-time transcription was still just as accurate, although '403 forbidden' still has errors.



The screenshot shows a transcription interface with a dark header bar. On the left, it says "Transcription". On the right, there are two buttons: "Download full transcript" and "Start streaming" (which is highlighted in yellow). Below the header is a section titled "Transcription output" on the left and "Current language: English, US" on the right. The main area contains a scrollable text box displaying a transcribed conversation:

```
with this website so we can enable it, we can just give it access just for a little bit. You don't have to keep um providing, you know, AWS, your
microphones, you can always block it after this, um, after these few minutes. So we're just gonna say allow this time.
And
should be able to start.
Right
Oh, OK.
It seems like something is up.
I'm gonna see if I can reset.
permissions.
access
OK, cool. So AW should have access to my microphone right now. Let's see if we can actually transcribe them.
Oh, it says you have to reload the page, sorry, to apply your settings. I thought I had reloaded and I clearly haven't.
So I'm just gonna reload it now and start streaming again.
Here we go.
A transcription has started.
So this is the part where you can actually...
```



NextWork.org

Everyone should be in a job they love.

Check out nextwork.org for
more projects

