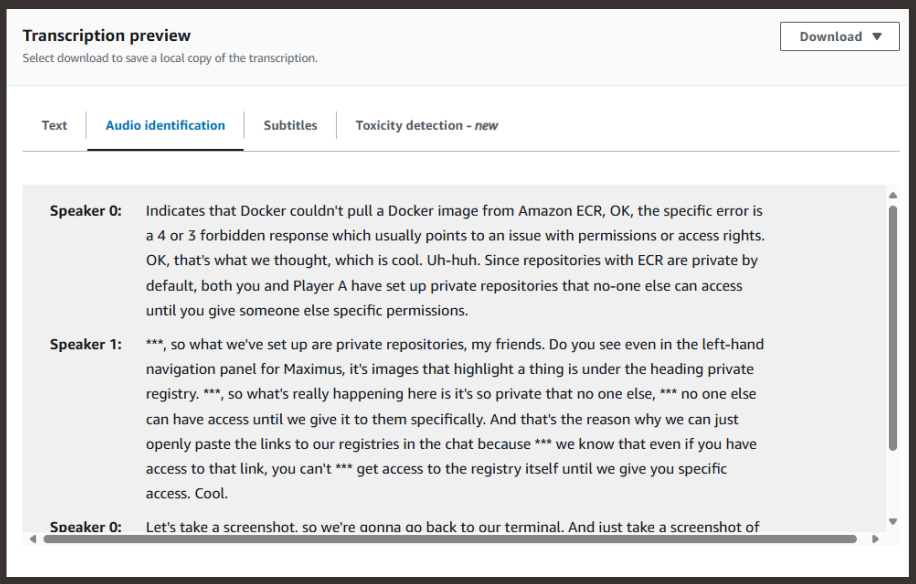
nextwork.org

Transcribe Audio Files with AI

ME Melvin J Bonner



ME
Melvin J Bonner
NextWork Student

Introducing Today's Project!

In this project, I will demonstrate how to transcribe a video with Amazon Transcribe and write custom vocabularies to improve my transcription's accuracy. I'm doing this project to learn Amazon Transcribe.

Tools and concepts

Services I used were Amazon Transcribe and Amazon S3. Key concepts I learned include real-time transcription, custom vocabulary, and vocabulary filtering.

Project reflection

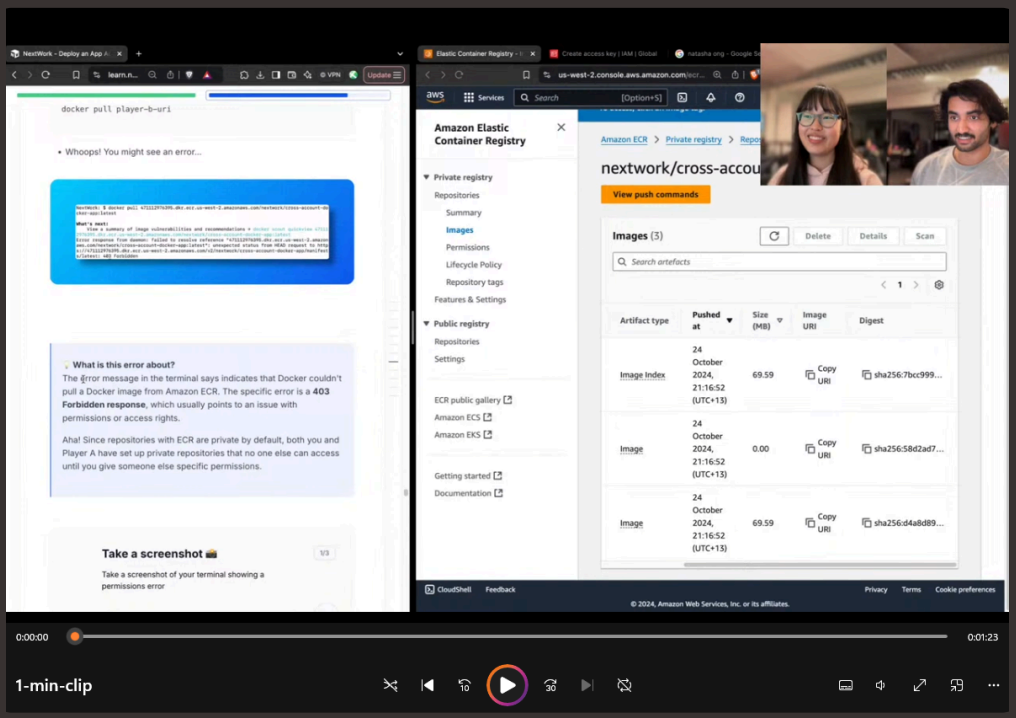
This project took me approximatley 3 hours to complete. The most challenging part was doing the real-time transcription. It was rewarding to complete this project and get another AWS skill under my belt.

I did this project today to get a better understanding of Amazon Transcribe. this project met my goals.

ME
Melvin J Bonner
NextWork Student

S3 and Transcribe

To set up for this project, I'm using an S3 bucket to to store the files because Amazon Transcribe, doesn't store the files. The file I'm transcribing is a 1-minute excerpt of our project demo on containers and ECR.



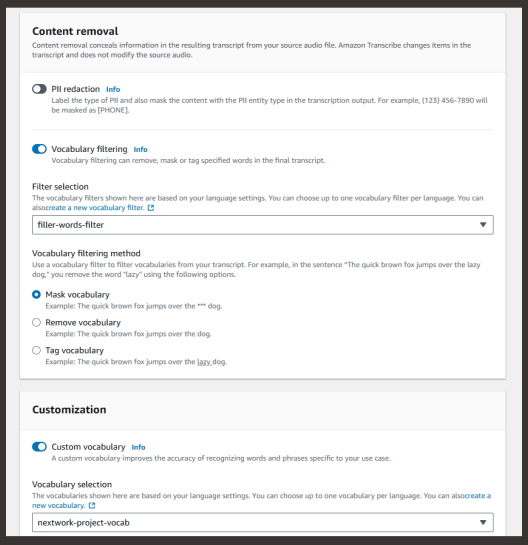
ME
Melvin J Bonner
NextWork Student

Run A Transcription Job

The steps to run a transcription job include selecting the language settings, selecting the model type, selecting the input data, and selecting the output data.

Amazon Transcribe uses model types to set instructions that tells Amazon Transcribe how to convert speech to text. Use cases for model types include a general one for everyday speech or a custom one for technical subjects.

You can customise a transcription further with subtitling, which adds a display of words spoken in a video and speaker partitioning which helps to label different speakers in an audio recording.

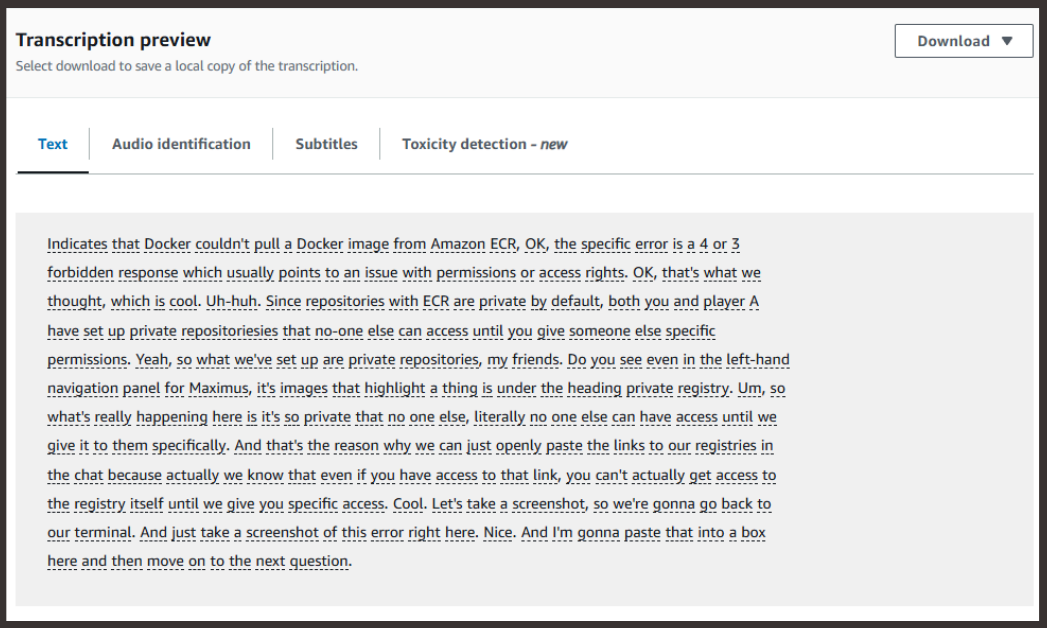


ME
Melvin J Bonner
NextWork Student
nextwork.org

Baseline Transcript Review

To start using Amazon Transcribe, I first ran a baseline transcription job, which is a practice to do the first run without using any custom settings. This is because I want to see how the service performs with the raw audio before I start making any changes. My baseline transcription gives me a starting point to compare to later when I add custom settings.

While reviewing the baseline transcript, I noticed repositories is misspelled as repositoriesies. Speech fillers like um can be excluded for a better reading experience. 403 Forbidden, a technical term, is mis-transcribed to 4 or 3 forbidden player, A is not an inaccuracy, but it should be capitalized to Player A as it's a title. These happened because transcription errors can happen for a bunch of reasons, like background noise, someone speaking unclearly, or just the tricky nature of language itself. Even though tools like Amazon Transcribe are pretty smart, they can still get tripped up by accents, local slang, or special terms that aren't in their usual training setup too. This might lead to some mistakes or missing parts in what gets written down. Also, simple typos can pop up if the model mishears words or bumps into terms it doesn't recognize often.



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.

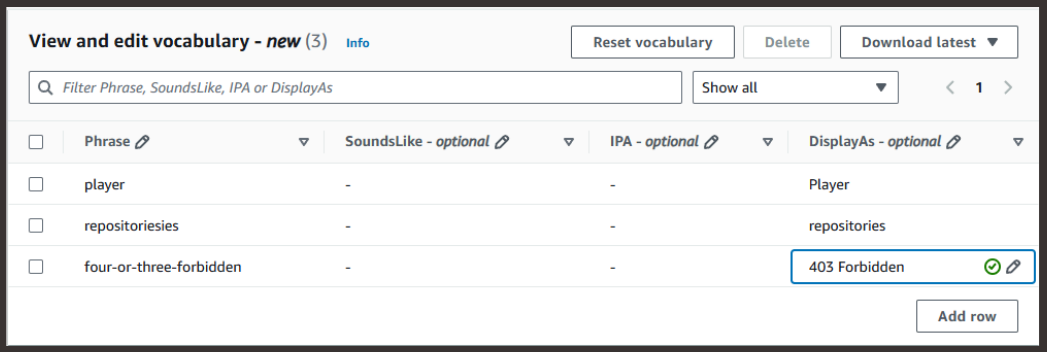
ME
Melvin J Bonner
NextWork Student

Custom Vocabulary

I can resolve transcription inaccuracies using a custom vocabulary, which is a list of words or phrases that you want Amazon Transcribe to recognize. A custom vocabulary improves transcription accuracy by helping Amazon Transcribe understand words that it might not normally recognize.

To create an item in a custom vocabulary, you need to define two values. They are Phrase and DisplayAs. The Phrase column is where you enter the specific words or phrases you want Transcribe to recognize. The DisplayAs column is where you tell Transcribe how these phrases should appear in the transcription, so they're displayed in a standardized or preferred format.

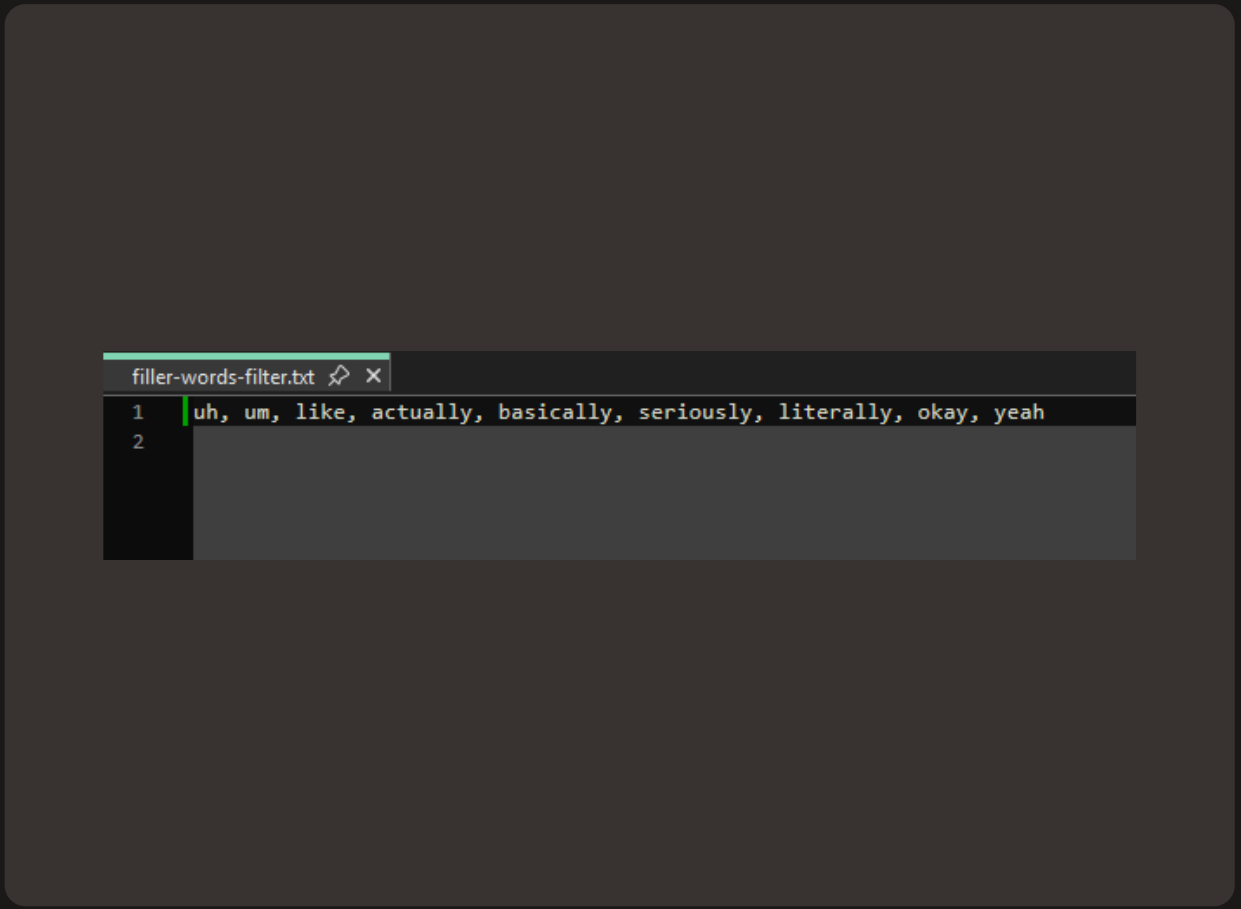
My custom vocabulary defines specific terms that are industry jargon with repositories and 403 Forbidden.


View and edit vocabulary - new (3)
Info
Reset vocabulary
Delete
Download latest
Filter Phrase, SoundsLike, IPA or DisplayAs
Show all
1
Phrase
SoundsLike - optional
IPA - optional
DisplayAs - optional
player
-
-
Player
repositoriesies
-
-
repositories
four-or-three-forbidden
-
-
403 Forbidden
Add row

Vocabulary Filters

Another feature in Transcribe is vocabulary filtering, which block out words you don't want to appear in your transcripts. It's different from custom vocabularies because custom vocabulary helps Amazon Transcribe understand specific words or phrases that it might not normally pick up.

My vocabulary filter removes filler words, sensitive information, or anything else I don't want in my transcript. To set up this filter, I first created a text file that contained a list of filler words.

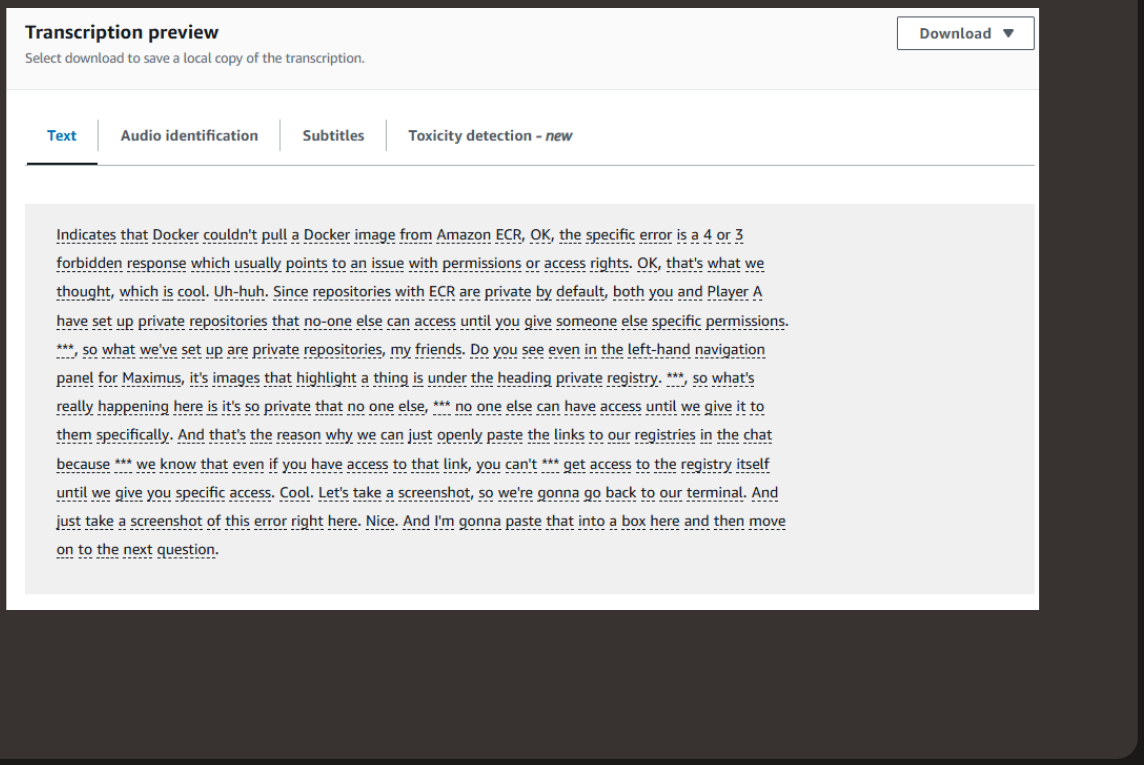


ME
Melvin J Bonner
NextWork Student

Enhanced Transcription

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

The enhanced transcription is better than the baseline because most of the errors have been removed. The onl yerror that remains is the "403 Forbidden" error.

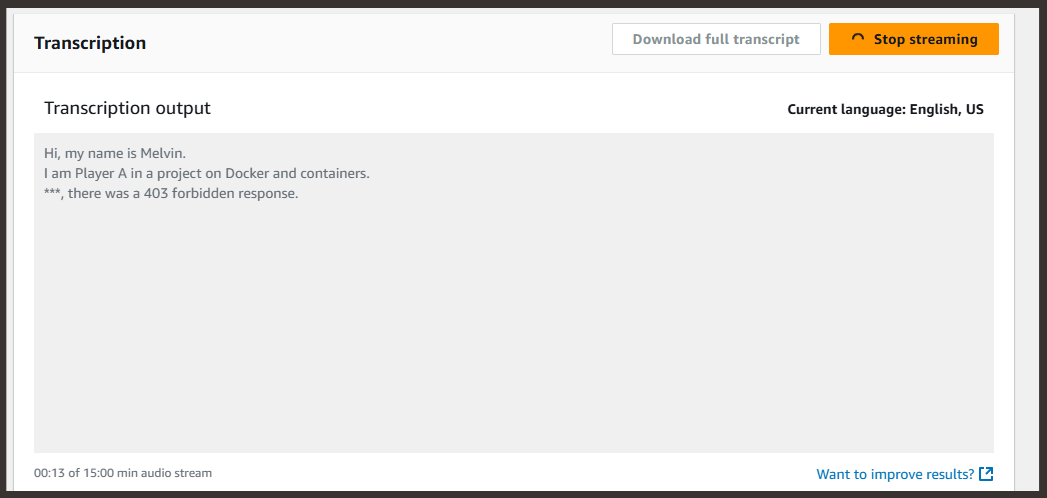


ME
Melvin J Bonner
NextWork Student

Real Time Transcription

For my project extension, I experimented with real-time transcription, which converts speech to text instantly as the speaker is still talking.

Even during real-time transcription, I could use features like printscreen function. Overall, compared to a transcription job, real time transcription was more of a useful feature.



nextwork.org

The place to learn & showcase your skills

Check out nextwork.org for more projects