

MICHIGAN STATE

UNIVERSITY

Alpha Presentation

On-Premises ASR Pipeline for Michigan English

The Capstone Experience

Team Michigan State University Linguistics

Eden Seo

Jacob Caurdy

Maria Irimie

Kyle Reinhart

Yichen Ding

Department of Computer Science and Engineering

Michigan State University

Spring 2022



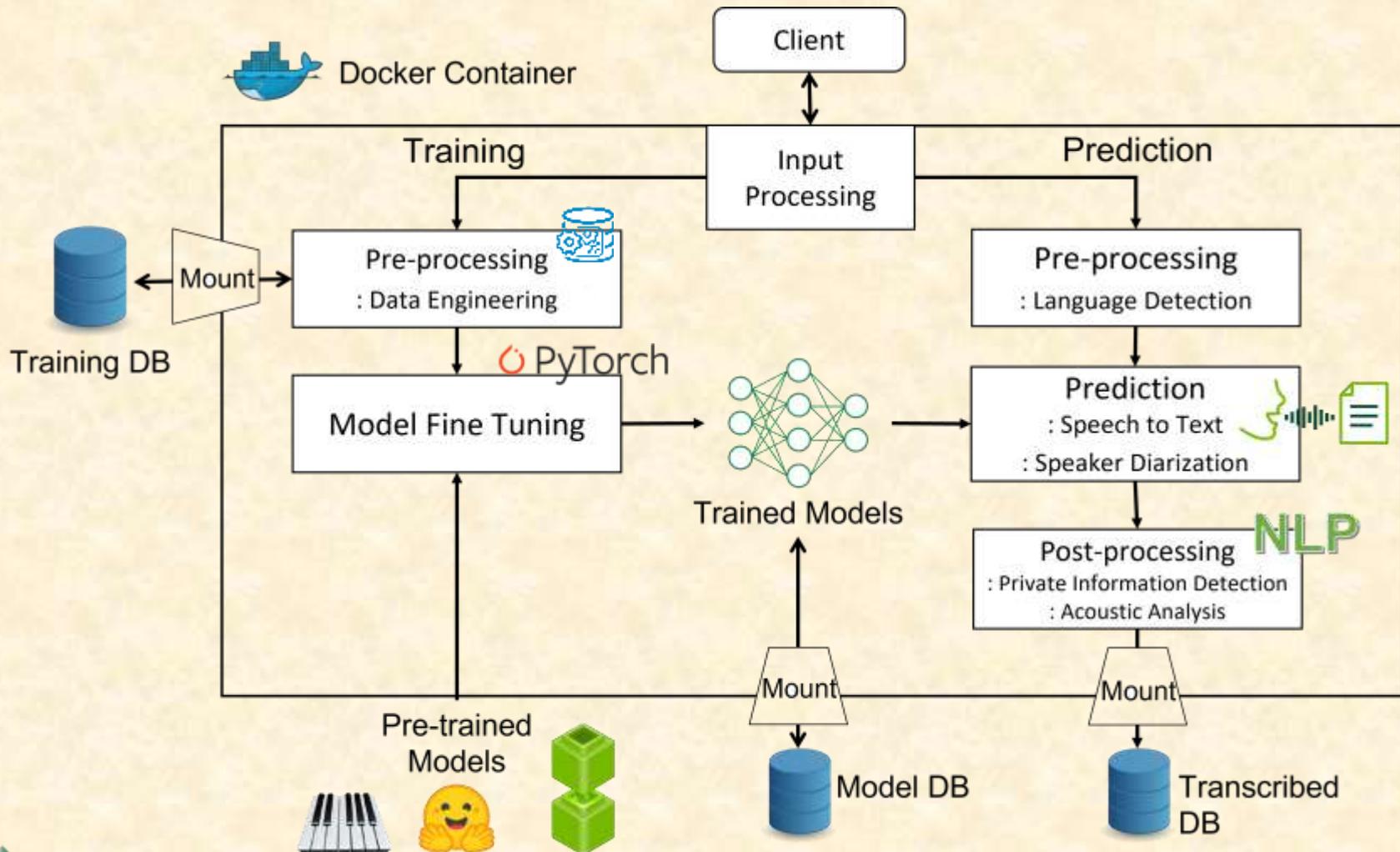
*From Students...
...to Professionals*

Project Overview

- Problem: The Department of Linguistics, Languages and Cultures' MI Diaries project needs to replace their outsourced speech-to-text software from Google to an in-house package
- Motivation: Privacy and monetary concerns
- Solution: All-in-one speech-to-text machine learning pipeline which takes in audio files and outputs time-aligned transcripts



System Architecture



Uploaded File Outputs Transcript

The screenshot shows a web browser window with the address bar displaying '127.0.0.1:8050'. The page title is 'MI Diaries'. On the left sidebar, there is a microphone icon with 'MI' inside, and buttons for 'Upload' and 'Train'. Below these is the text 'Transcriptions'. The main content area features a dashed box with the text 'Drag and Drop or Select Files'. Below this, a file named '0a15bb21-3993-4496-8672-f3be45769356.wav' is shown with an audio player interface displaying '0:00 / 15:00'. Below the audio player is a table with the following data:

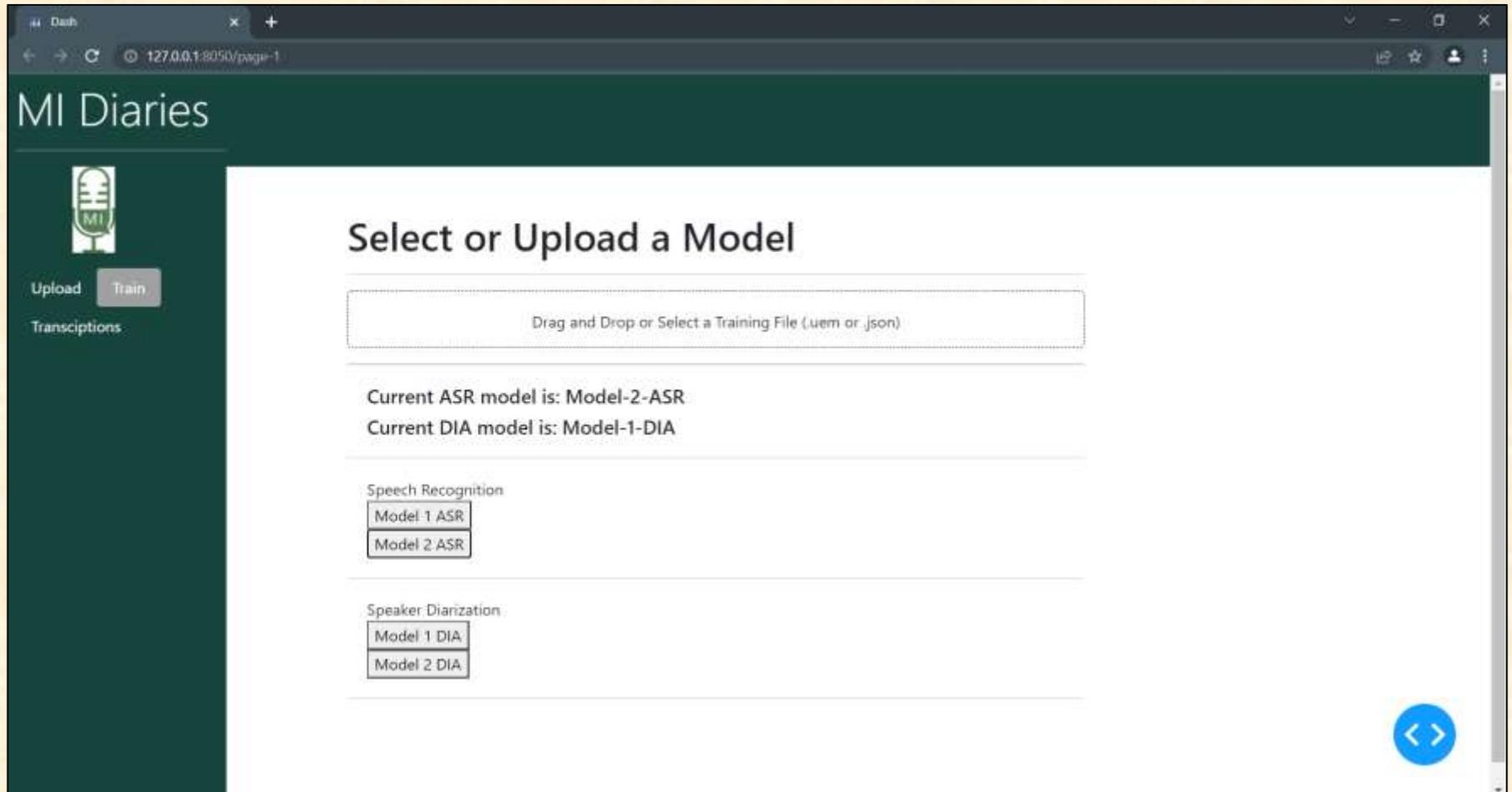
start	end	speaker	transcript
01:003	02:035	1	Hey, there! Did you know you can save 15% or more on car insurance?
02:045	03:058	2	What! That's an incredible amount of savings. How do I get in on it?
03:092	04:024	1	All you have to do is download the Geico mobile app and walkthrough the set up.
04:057	05:012	1	Once you are done, you will on your way to endless savings.
05:092	06:053	2	Wow. Thank you. Now I'll finally be able to afford that doodad for my girl.

Below the table is a section titled 'Raw Content' with the following text:

```
data:audio/wav;base64,Uk1GRkaotwFXQVZFZm10IBAAAAAABAAEAgD4AAAB9AAACABAATElTVBoAAAABJKZPSVNGVA4AAABM  
YXZmNTcuODMuMTAwAGRhdGEaQlCBAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAA...
```

A blue circular button with a double arrow icon is located in the bottom right corner of the application area.

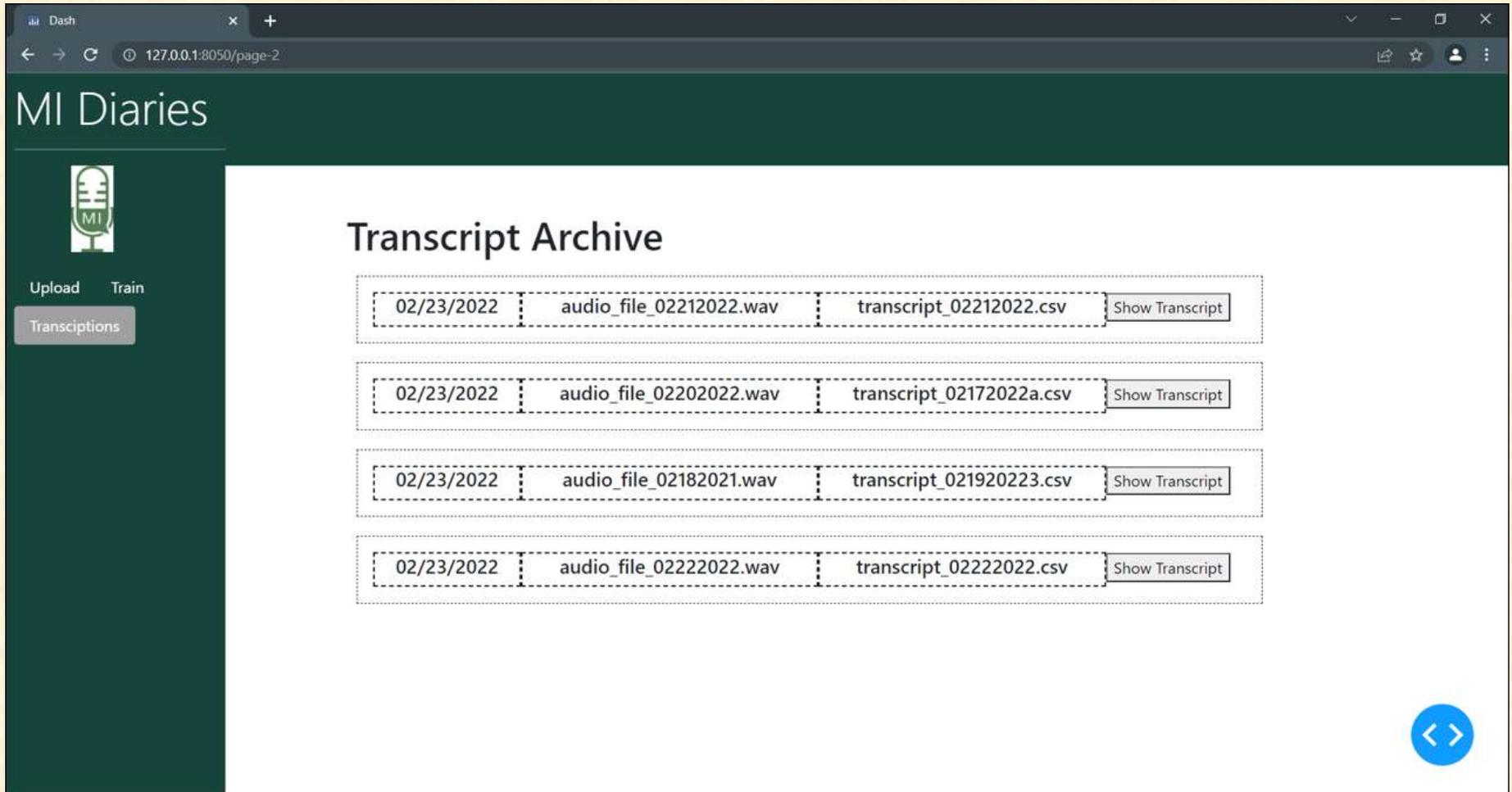
Selecting a Model



The screenshot shows a web browser window with the address bar displaying "127.0.0.1:8050/page-1". The page title is "MI Diaries". On the left sidebar, there is a microphone icon with "MI" inside, and buttons for "Upload" and "Train". Below the sidebar, the text "Transcriptions" is visible. The main content area is titled "Select or Upload a Model". It features a large dashed border box with the text "Drag and Drop or Select a Training File (.uem or .json)". Below this, it displays the current models: "Current ASR model is: Model-2-ASR" and "Current DIA model is: Model-1-DIA". Under the heading "Speech Recognition", there are two buttons: "Model 1 ASR" and "Model 2 ASR". Under the heading "Speaker Diarization", there are two buttons: "Model 1 DIA" and "Model 2 DIA". A blue circular button with left and right arrows is located in the bottom right corner of the page.



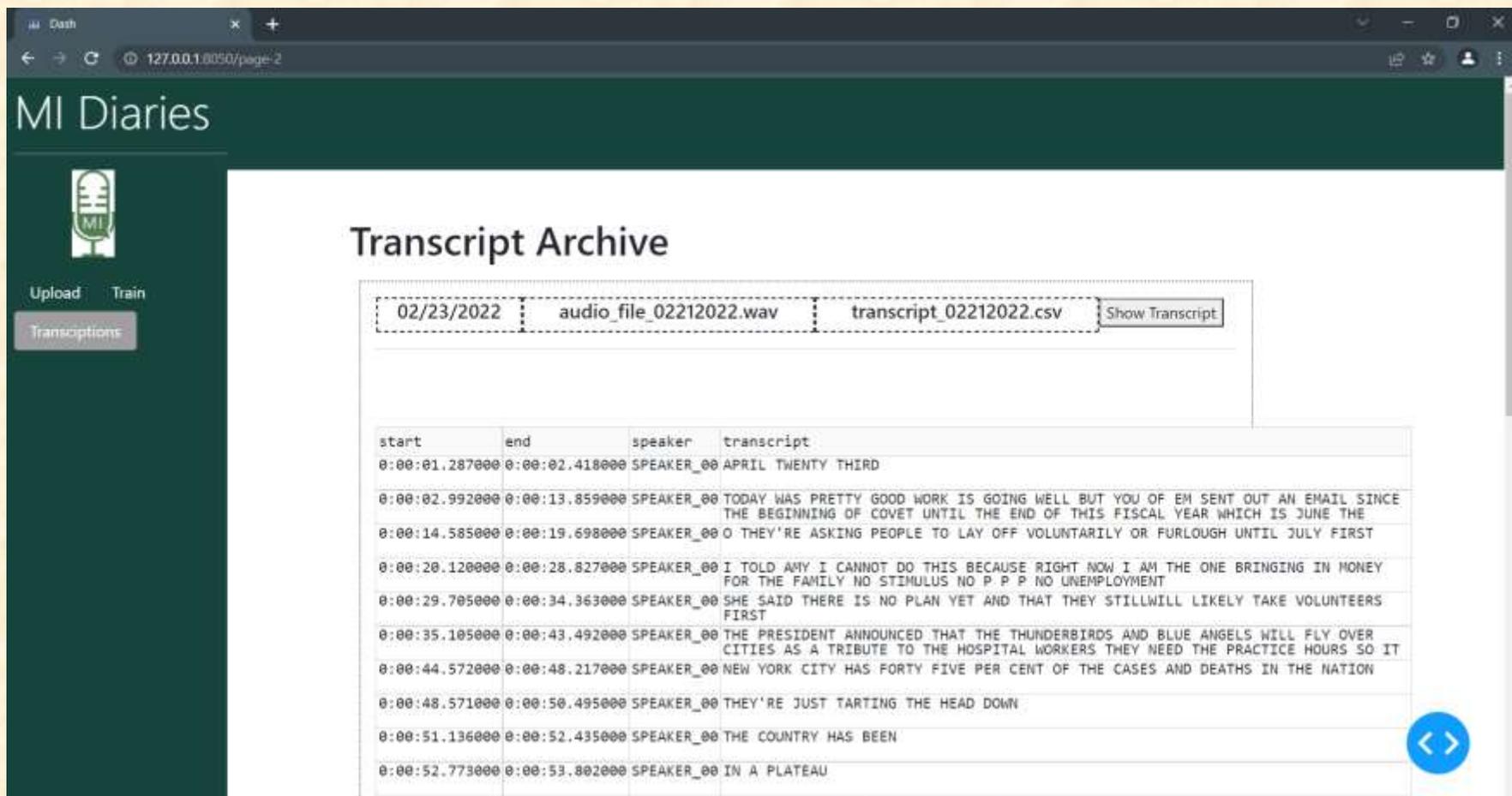
Archive of all submitted



The screenshot shows a web browser window with the address bar displaying '127.0.0.1:8050/page-2'. The page title is 'MI Diaries'. On the left sidebar, there is a microphone icon with 'MI' inside, and buttons for 'Upload', 'Train', and 'Transcriptions'. The main content area is titled 'Transcript Archive' and contains a table of four entries, each with a date, audio file name, transcript file name, and a 'Show Transcript' button.

Date	Audio File	Transcript File	Action
02/23/2022	audio_file_02212022.wav	transcript_02212022.csv	Show Transcript
02/23/2022	audio_file_02202022.wav	transcript_02172022a.csv	Show Transcript
02/23/2022	audio_file_02182021.wav	transcript_021920223.csv	Show Transcript
02/23/2022	audio_file_02222022.wav	transcript_02222022.csv	Show Transcript

Displaying Archived Transcription



The screenshot shows a web browser window displaying the 'MI Diaries' application. The page title is 'MI Diaries' and the URL is '127.0.0.1:8050/page-2'. The application interface includes a sidebar with a microphone icon and the text 'MI', and buttons for 'Upload', 'Train', and 'Transcriptions'. The main content area is titled 'Transcript Archive' and features a table with columns for date, audio file name, and transcript file name. A 'Show Transcript' button is visible next to the transcript file name. Below the table, a list of transcript entries is displayed, each with a start and end time, a speaker identifier, and the transcript text.

start	end	speaker	transcript
0:00:01.287000	0:00:02.418000	SPEAKER_00	APRIL TWENTY THIRD
0:00:02.992000	0:00:13.859000	SPEAKER_00	TODAY WAS PRETTY GOOD WORK IS GOING WELL BUT YOU OF EM SENT OUT AN EMAIL SINCE THE BEGINNING OF COVET UNTIL THE END OF THIS FISCAL YEAR WHICH IS JUNE THE
0:00:14.585000	0:00:19.698000	SPEAKER_00	O THEY'RE ASKING PEOPLE TO LAY OFF VOLUNTARILY OR FURLOUGH UNTIL JULY FIRST
0:00:20.120000	0:00:28.827000	SPEAKER_00	I TOLD AMY I CANNOT DO THIS BECAUSE RIGHT NOW I AM THE ONE BRINGING IN MONEY FOR THE FAMILY NO STIMULUS NO P P P NO UNEMPLOYMENT
0:00:29.705000	0:00:34.363000	SPEAKER_00	SHE SAID THERE IS NO PLAN YET AND THAT THEY STILLWILL LIKELY TAKE VOLUNTEERS FIRST
0:00:35.105000	0:00:43.492000	SPEAKER_00	THE PRESIDENT ANNOUNCED THAT THE THUNDERBIRDS AND BLUE ANGELS WILL FLY OVER CITIES AS A TRIBUTE TO THE HOSPITAL WORKERS THEY NEED THE PRACTICE HOURS SO IT
0:00:44.572000	0:00:48.217000	SPEAKER_00	NEW YORK CITY HAS FORTY FIVE PER CENT OF THE CASES AND DEATHS IN THE NATION
0:00:48.571000	0:00:50.495000	SPEAKER_00	THEY'RE JUST TARTING THE HEAD DOWN
0:00:51.136000	0:00:52.435000	SPEAKER_00	THE COUNTRY HAS BEEN
0:00:52.773000	0:00:53.802000	SPEAKER_00	IN A PLATEAU



What's left to do?

- Core Features
 - Switch training process to NVIDIA NeMo
 - Overlap detection
 - Docker
 - Integrate into MI Diaries App
- Additional Features
 - FAVE acoustic analysis
 - Sensitive Information Extraction
 - More models for dialects and languages



Questions?

?

?

?

?

?

?

?

?

?

