MICHIGAN STATE UNIVERSITY

Project Plan 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

MSU LiLaC

- MSU Linguistic, Language, and Cultures (LiLaC)
  - Offers degree programs and research in linguistics
- MI Diaries Project done by the Sociolinguistics Lab
  - Aims to chronicle language changes over the course of the pandemic
  - Aims to provide primary sources on the pandemic to historians as well
  - Takes volunteer audio

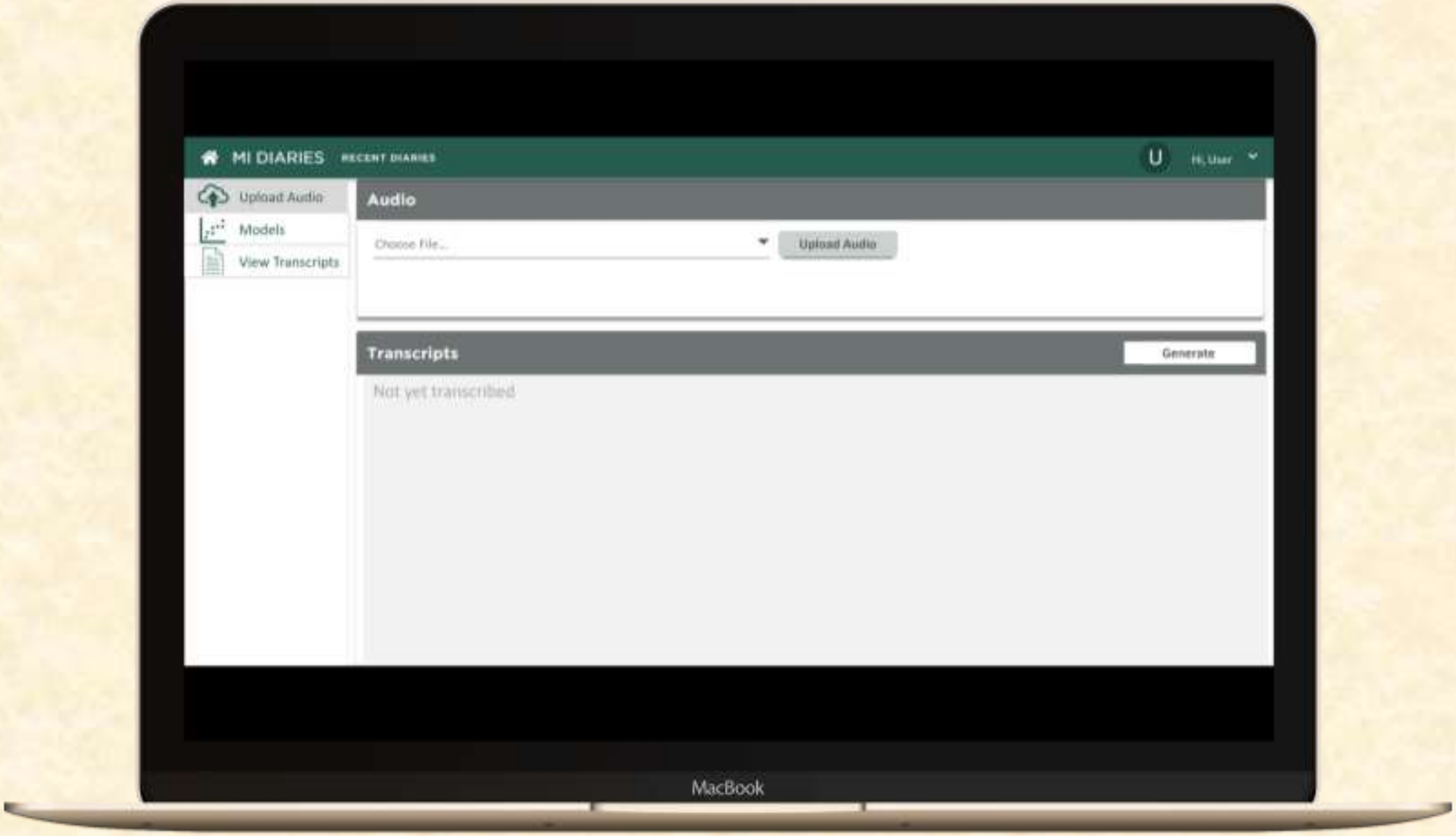
Functional Specifications

- Creating an Automatic Speech Recognition (ASR) to fit into the current pipeline for the MSU Linguistics
- Replacing Google's role in the pipeline
  - Saves money
  - Protects data
  - Improves accuracy with dialectal differences
- ASR
- Speaker Diarization (speaker differentiation)
  - Time-aligned transcript
- Model Retraining
  - Handling inaccuracies

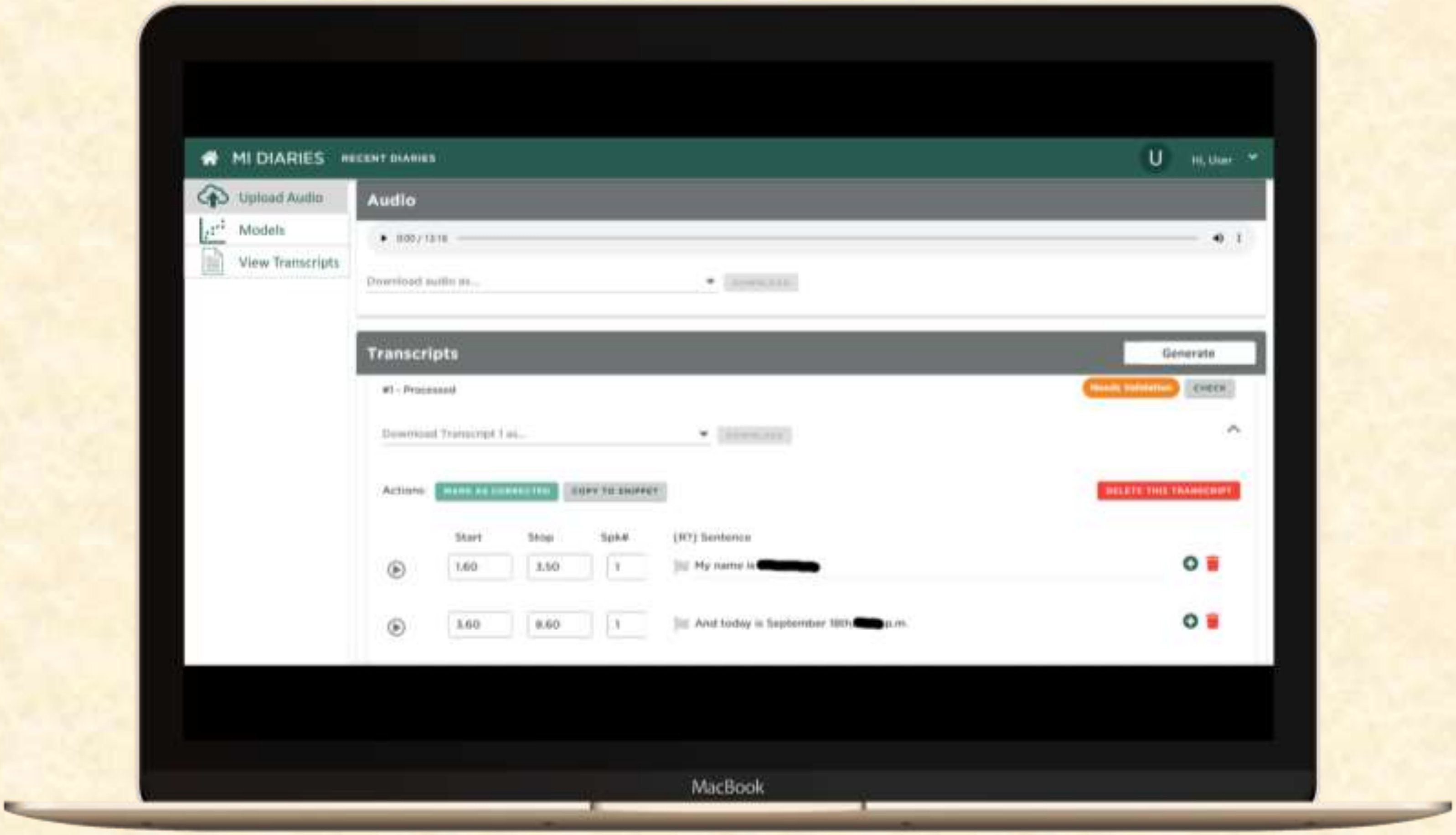
Design Specifications

- Audio Upload
- Hand-Correction Interface
- Sensitive Information Flagging
- Diarization View
- Retraining

Screen Mockup: Transcription

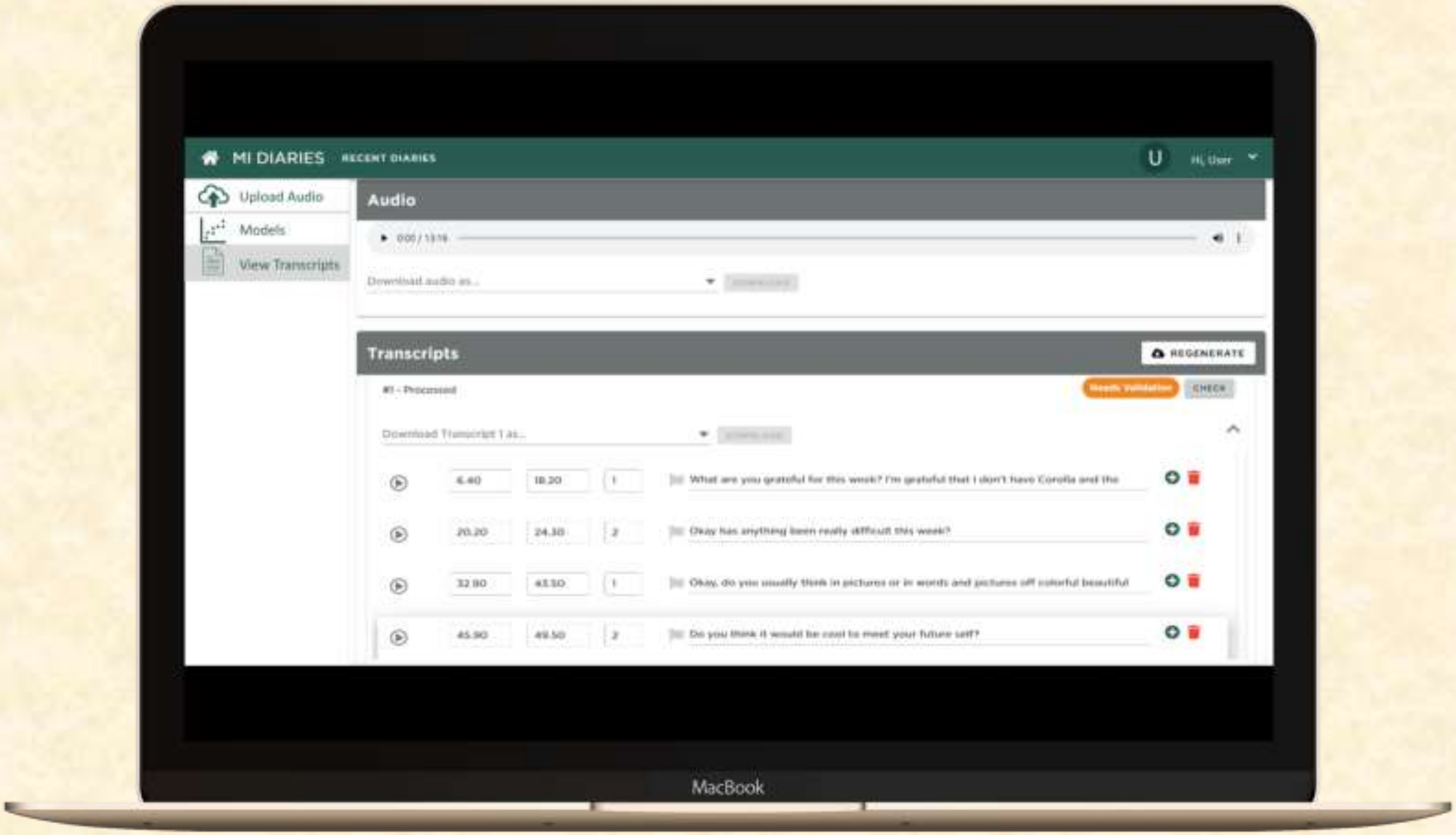


Screen Mockup: Transcription (2)



Start	Stop	Spk#	(R?) Sentence
1.60	3.50	1	My name is [redacted]
3.60	8.60	1	And today is September 18th [redacted] p.m.

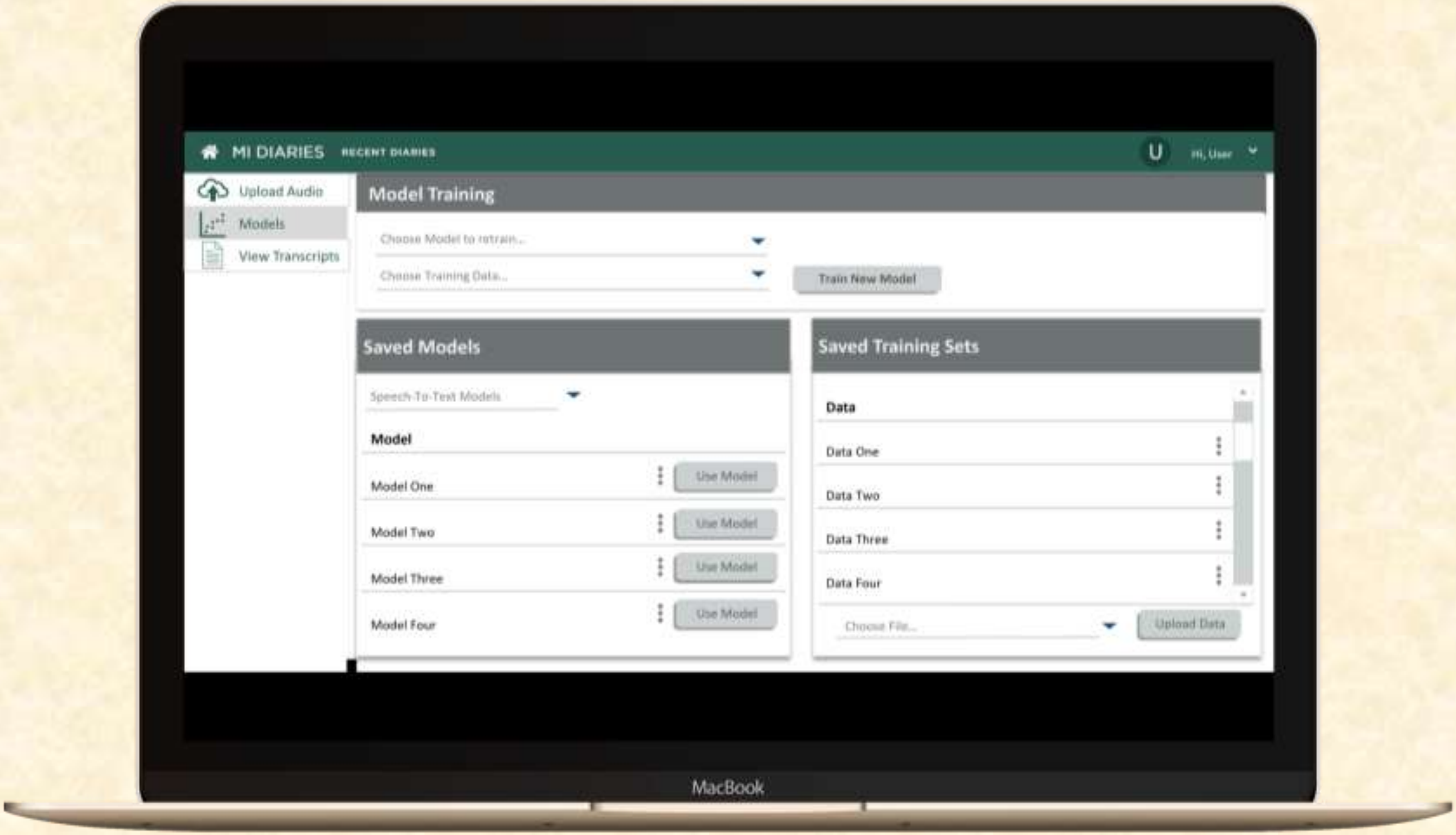
Screen Mockup: Diarization View



#1 - Processed

6.40	18.20	1	What are you grateful for this week? I'm grateful that I don't have Corolla and the
20.20	24.30	2	Okay has anything been really difficult this week?
32.80	43.50	1	Okay, do you usually think in pictures or in words and pictures off colorful beautiful
45.90	49.50	2	Do you think it would be cool to meet your future self?

Screen Mockup: Retraining



Model Training

Choose Model to retrain...

Choose Training Data...

Saved Models

Speech-To-Text Models

Model
Model One
Model Two
Model Three
Model Four

Saved Training Sets

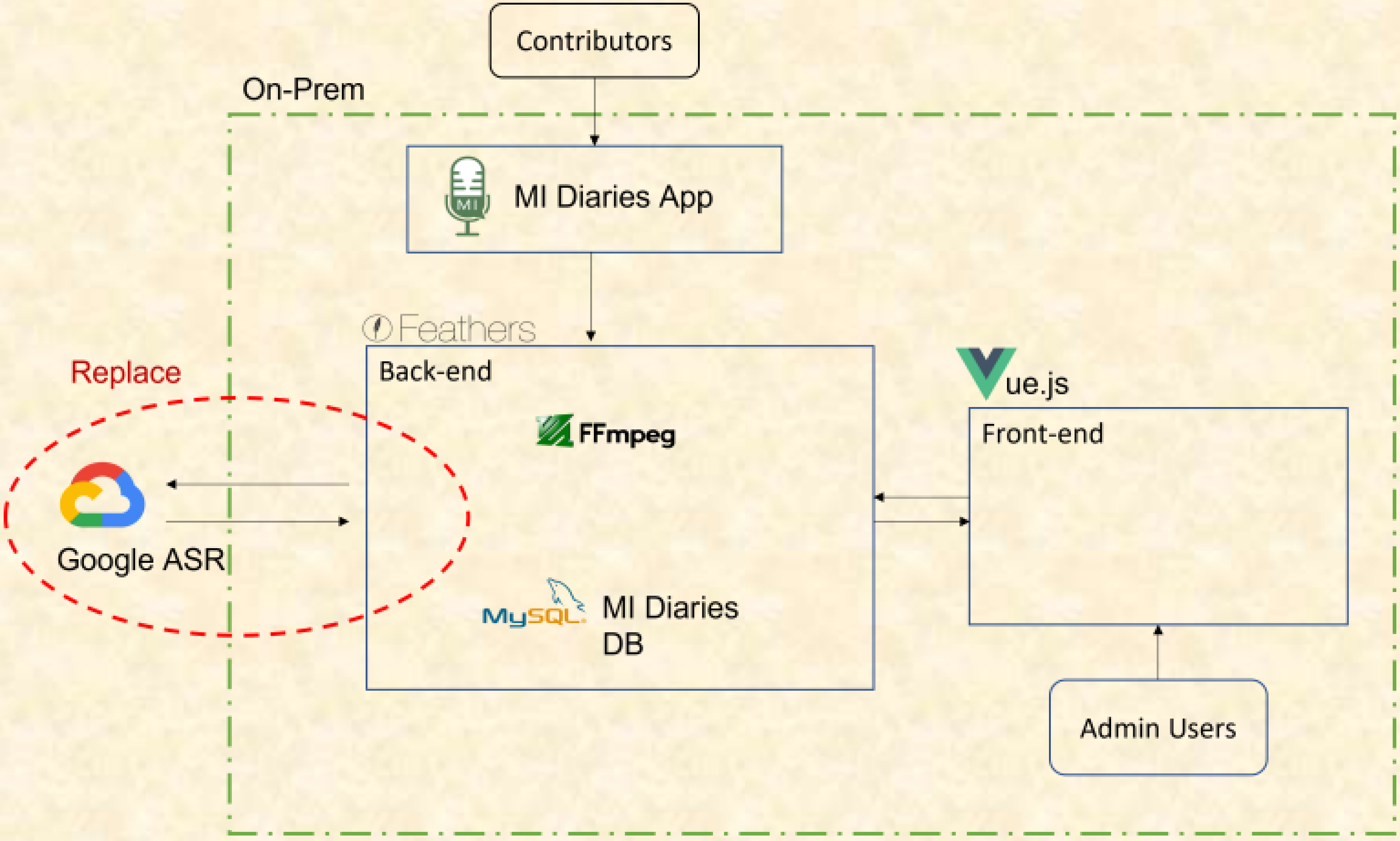
Data
Data One
Data Two
Data Three
Data Four

Choose File...

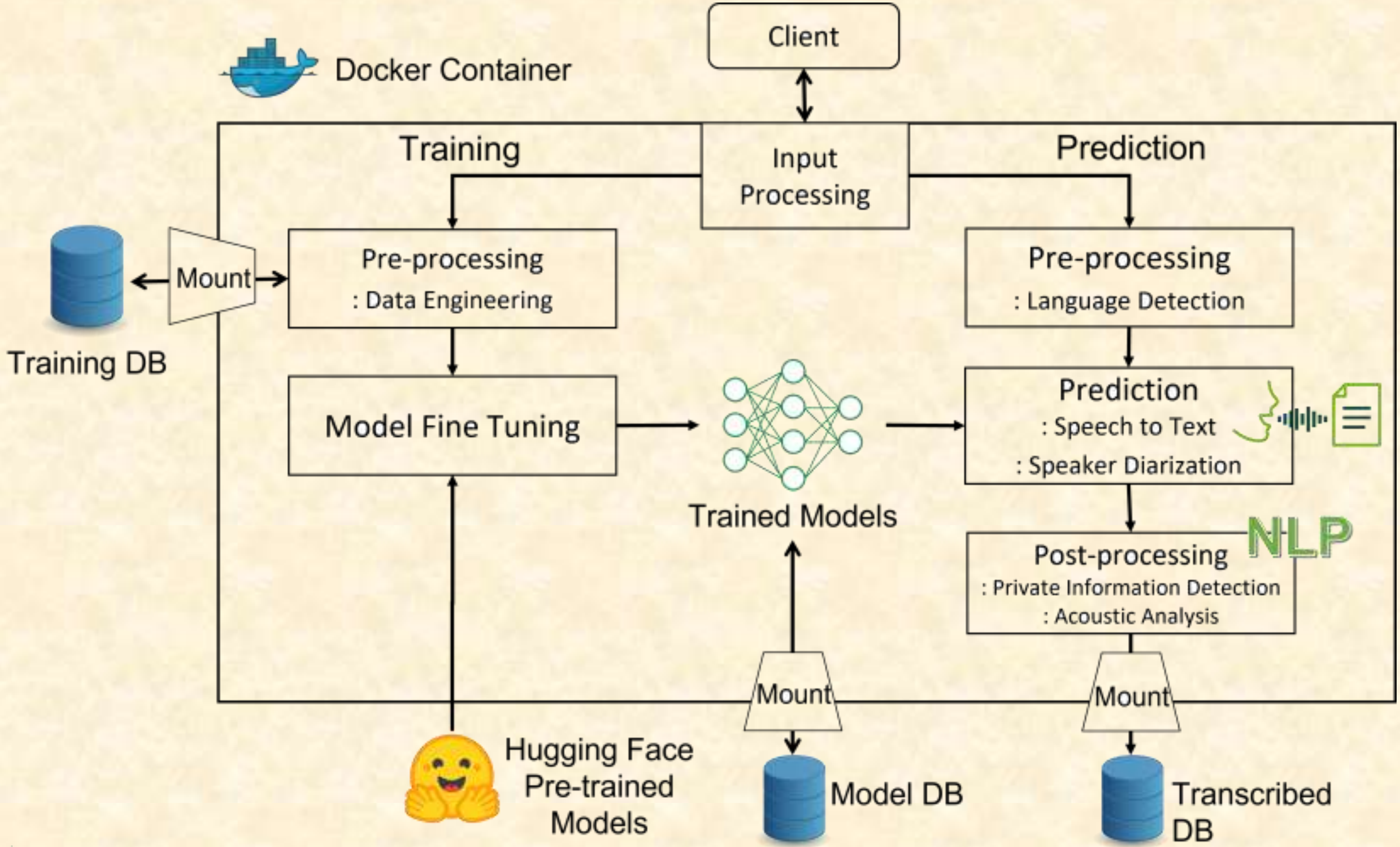
Technical Specifications

- ASR Pipeline consists of two processes
  - Training and Prediction
- Training
- Prediction
  - Pre-processing, prediction, and post-processing
  - Combines both models
- Docker
  - Portability
  - Potential GPU Acceleration

System Architecture



System Architecture



System Components

- Hardware Platforms
  - iMacs
  - MSU EGR GPU Computes
- Software Platforms / Technologies
  - Machine Learning
    - HuggingFace
      - PyTorch
      - Wav2Vec2, WavLM
    - Natural Language Processing
      - NLTK, spaCy, Gensim
  - Docker
  - GitHub
  - Python and PyCharm

Risks

- **Inadequate Data for Speaker Diarization**
  - Currently not enough labeled data for supervised learning in speaker diarization
  - Mitigation: Self-supervised models or using publicly available labeled datasets
- **Training and Predictions Times without GPUs**
  - Speed for training and prediction models is bottlenecked by the model architecture
  - Mitigation: Possible to use a smaller model which may be less accurate if speed is critical.
- **Extending to different dialects and languages**
  - Extension to different dialects is dependent on data we don't have
  - Mitigation: We will create features for users to upload their own training datasets which can fine-tune ASR models provided by Hugging Face.

Questions?

? ? ? ?

? ?

? ? ?