

## **Problem Statement**

### **Extracting audio from vernacular videos and converting it to text using Natural Language Processing (NLP) technologies**

## **Proposed Solution**

Speech Recognition is a popular topic under machine learning concepts. Speech Recognition is getting used more in many fields. For example, the subtitles that we see on Netflix shows or YouTube videos are created mostly by machines using Artificial Intelligence. Other great examples of speech recognizers are personal voice assistants such as Google's Home Mini, Amazon Alexa, Apple's Siri.

Entity extraction, also known as named entity extraction (NER), enables machines to automatically identify or extract entities, like product name, event, and location. It's used by search engines to understand queries, chatbots to interact with humans, and teams to automate tedious tasks like data entry.

Entity extraction is a text analysis technique that uses Natural Language Processing (NLP) to automatically pull-out specific data from text, and may classify it according to predefined categories. . These are named entities, the words, or phrases. This includes proper names but also numerical expressions of time or quantity, such as phone numbers, monetary values, or dates.

## **The project goals**

Zee Media has a rich library of vernacular videos (mostly Hindi & English). The goal of the model is to create an audio file from video and take the input audio and extract the entities. Using NLP techniques application must process these videos and generate audio files & text transcripts from the audio file and the related entities. Process this text to extract entities and enrich videos with relevant tags. This information will be used to improve the content recommendations.

The end goal of the project is to achieve at least one or all the following use cases

Use Case1: Convert video to audio -> radio/podcast model

Use Case1a: Audio to audio summarization -> Audio briefs

Use Case1b: Convert Audio to text -> CC for text below the video

Use Case 1c: Convert text to entity extraction -> improves video searchability.

Use Case1d: Summarization of text extracted.

Use Case 2: Convert text to audio -> audio generation of text stories

**DISCLAIMER** This document is strictly private, confidential and personal to its recipients and should not be copied, distributed or reproduced in whole or in part, nor passed to any third party.



DISCLAIMER This document is strictly private, confidential and personal to its recipients and should not be copied, distributed or reproduced in whole or in part, nor passed to any third party.