# **Industry Capstone Project**

# **Spotify**

## Problem definition

*[50-100 word description of the problem which the candidates need to solve]*

“Hi Google” doesn’t work with Google Assistant because Google Assistant is listening for two separate words. This phrase is often called a **wake word**. Wake words are critical to the design of low-power machine learning process data to wake up the device for full processing. Words such as “Ok google” are oftentimes used to wake up AI assistant devices before they begin processing speech.

While typical virtual assistant systems tend to classify wake words as speech commands, this process can generate false alarms from the limited number of negative training samples in the dataset as shown in the latest work <https://www.aclweb.org/anthology/2020.nlposs-1.9.pdf>

For this project we want to design a machine learning model that can process a custom wake word. For example, if you were to speak “Hey FourthBrain, when does class begin” the model should be able to respond with the class time or verify if it heard the question accurately or not.

For this work we will utilize public domain data sets such as:

1. Mozilla common Voice: <https://commonvoice.mozilla.org/en/datasets>
2. Google speech commands dataset:

<https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html>

## 

## Key Research Questions/ Technological constraints that the Project will Answer

1. Can a machine model detect a wake word?

2. Is it possible to wake a machine learning model up with a custom wake word while enhancing the classification of voice commands?

3. Can the machine model identify any type of wake word or is it just a specific wake word?

4. Can we improve the Alexa / Google Assistant experience with wake words that are easily understandable?

5. Can we define what are the qualities that make a *good wake word* and what are the qualities that make a *bad wake word*

## Final deliverables at the end of the project

*[Please list the desired technical deliverables from the project team in as much detail as possible]*

1. End-to-end pipeline that processes audio signals, detects wake words and classifies voice commands thereafter.

2. Quantitative and qualitative performance of custom wake word detectors.

3. Deployed ML model

4. Technical report on System setup and performances.

## Key activities/ technologies the project team may be expected to undertake/ work with

*[E.g. What kind of technology stack you prefer they work with, the datasets they may need to work on, what kind of analysis they may be expected to undertake, etc.]*

Python

Tensorflow or pytorch

Flask or FastAPI

Deployment on AWS or Google cloud.

## Expected learning outcomes

*[What do you expect the candidates to learn from the project? Please mention the technical skills they will grow over the project.]*

1. Building end-to-end ML pipelines

2. ML Model development for audio processing, wake work detection

3. Communication and technical reporting

| Desired Team Size (if any): | 2-3 |
| --- | --- |