### 1. Image captioning

Image captioning involves the task of generating language descriptions of visual content, like image or videos. Task to understand objects in images, relate to one another and translating it all into natural-sounding language.

Accurately describing a complex scene requires a deeper representation of what’s going on in the scene, capturing how the various objects relate to one another and translating it all into natural-sounding language.

Accurate image captioning is a challenging task that requires advancing the state of the art of both computer vision and natural language processing.

This kind of system could eventually helps in:

* visually impaired people understand pictures,
* Provide alternate text for images in parts of the world where mobile connections are slow, and make it easier for everyone to search on Google for images.
* Semantic image search
* Bering visual intelligence to chat board
* Help visually impaired people to see world around them

Reference for datasets:

<http://cocodataset.org/#explore>

<https://www.flickr.com/photos/tags/dataset/>

<http://academictorrents.com/details/9dea07ba660a722ae1008c4c8afdd303b6f6e53b>

### 2. Toxicity classification

To understand toxic conversational attributes, in comments text form. Each comment in Train dataset has a toxicity label (target), and models should predict the target toxicity for the Test data.

This attribute (and all others) are fractional values which represent the fraction of human raters who believed the attribute applied to the given comment.

For evaluation, test set examples with target >= 0.5 will be considered to be in the positive class (toxic).

The data also has several additional toxicity subtype attributes. Models do not need to predict these attributes for the competition; they are included as an additional avenue for research. Subtype attributes are: severe\_toxicity , obscene, threat, insult, identity\_attack, sexual\_explicit

Reference for dataset:

<https://www.kaggle.com/c/jigsaw-unintended-bias-in-toxicity-classification/data>

### 3. Speech recognition

Voice-to-text is a type of speech recognition program that converts spoken to written language.

This type of speech recognition software is extremely valuable to anyone who needs to generate a lot of written content without a lot of manual typing.

It is also useful for people with disabilities, like the hearing impaired. That makes it difficult for them to use a keyboard.

e.g Medical Speech, Transcription, and Intent

This data contains thousands of audio utterances for common medical symptoms like “knee pain” or “headache,” totalling more than 8 hours in aggregate. Each utterance was created by individual human contributors based on a given symptom. These audio snippets can be used to train conversational agents in the medical field.

This Figure Eight dataset was created via a multi-job workflow. The first involved contributors writing text phrases to describe symptoms given. For example, for “headache,” a contributor might write “I need help with my migraines.” Subsequent jobs captured audio utterances for accepted text strings.

Note that some of the labels are incorrect and some of the audio files have poor quality. I would recommend cleaning the dataset before training any machine learning models.

This dataset contains both the audio utterances and corresponding transcriptions.

Reference for dataset:

<https://www.kaggle.com/paultimothymooney/medical-speech-transcription-and-intent>

<https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/data>