# Team Gargantua

## Motivation and goals

- American Sign Language (ASL) is the most widely spoken sign language as well as the third most spoken language in the US (behind English and Spanish). Lack of resources poses a hinderance to effective communication between the deaf/mute and the people around them.
- Our goal is to create a classifier that can identify hand gestures from static images and use it to predict live images from a webcam. Ultimately, we want to deploy a webapp that where the user can control the classifier outputs.

## Training data

- Training data is in the form of a MNIST style dataset, containing pixel values for each image. A total of 24 classes are present, one for each alphabet (except J and Z which require motion).
- This data has been converted to images and upscaled so that the training data is comparable to the webcam images. To remove training bias due to handedness we performed lateral flip on the training images to account for both left and right handed signs. Additionally, we normalized the images to improve performance across different skin tones.

## Pipeline

- The upscaled and transformed training images were then fed into a pre-trained ResNet model, implemented using fastai library. The framework was run on Google Collab and eventually deployed on Google Cloud Platform.
- Among the three ResNet models we tried, ResNet34 turned out to be the model with optimal trade-off between accuracy and runtime. We were able to reach an accuracy of 99% on the validation set.

## Deployment

- The model deployment can be seen at https://asl.elder-rabbit.com/fuzzy-octo-guacamole/asm-webcam.html. We would like to acknowledge Jannu Sudhakar for his help with the server side setup.
- The app allows the users to capture the hand gesture using their webcams and provides real-time predictions of the ASL alphabets. For reference the ASL alphabets are provided on the app.

## Limitations and Future Work

- Processing times are slow because of our budget constraints, since hosting with a GPU on Google Cloud Platform is expensive.
- The accuracy on webcam images is lower than that on the validation set due to higher resolution of webcam images.
- In the future, we would like to employ object detection tools to automatically detect hands of users as well as dynamic gestures.