**Finger Spelling Agent**

**Difficulty:** Level 400

**Business** **Goal:**

The language learning school wants to expand their service offering by providing beginner level learning tools for American Sign Language (ASL) and have asked you to implement a study activity program that will allow students to use their web-cam and practice fingerspelling the english alphabet in ASL.

**Sign Language Considerations:**

ASL is one of many sign languages.

* ASL = American Sign Language
* BASL = Black American Sign Language
  + Dialect of ASL and often uses two hands to sign
* JSL = Japanese Sign Language
* BSL = British Sign Language

Different sign languages use different signs, and have different language features.

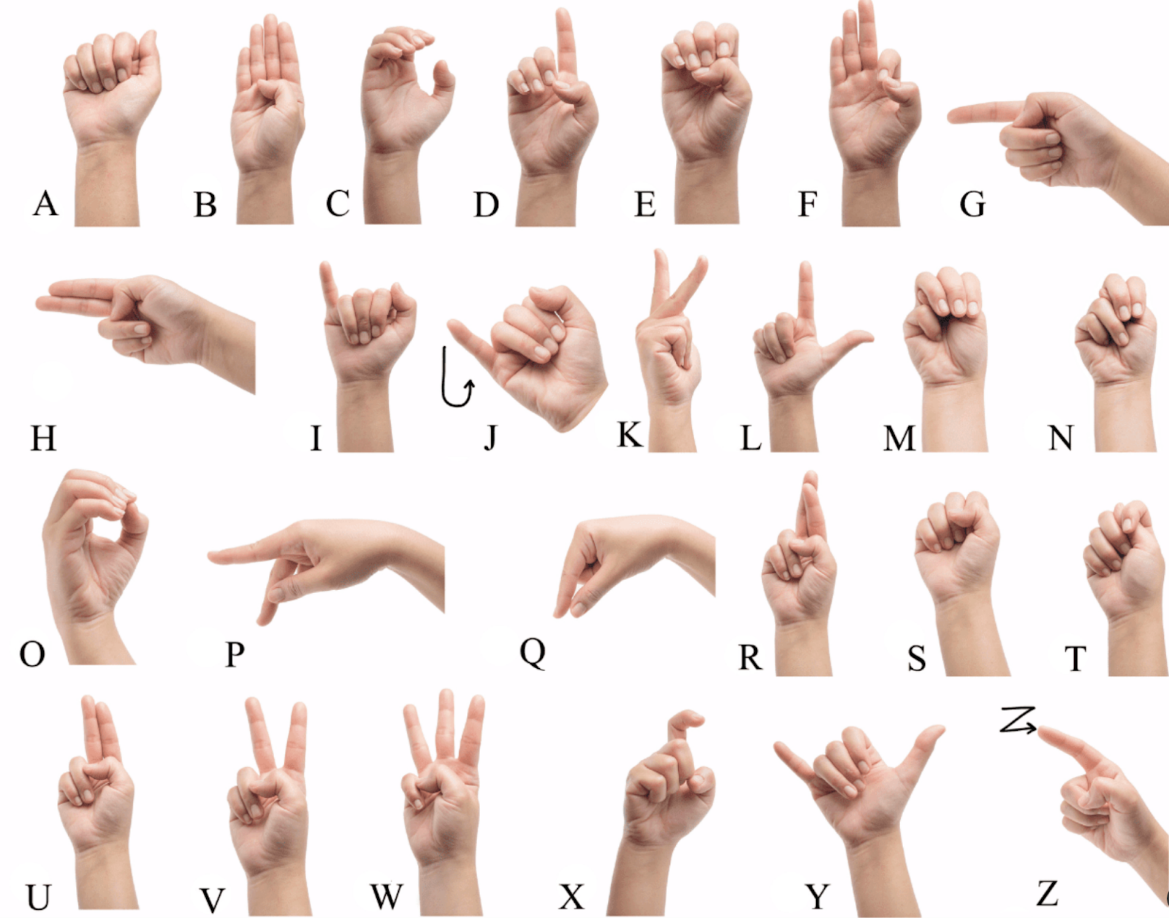
ASL includes more than just hand signs, but one must factor in facial expressions, body movements, posture, mouthing the words, space and directionality.

ASL is its own distinct language with its own unique grammar and syntax

* it's not a word for word signed version of English.

When building accessibility applications for sign language you need to consider all the above and work with the deaf community.

For ASL the english alphabet can be represented using a single handle with a single sign.  
While these images appear static, consider that hand motion is part of identifying the sign.



**Technical Uncertainty:**

* What training data can we use to identify fingerspelling signs?
* How accurate will the singing be if it captures a static image vs video?
  + How would video be represented? 24 frames of videos or checkpoint video frames
* What possible ML Models or architectures exist that have been used for identifying ASL signs?
* What vision strategy should we use to detect signs?
  + Are we detecting hand gestures by drawing a skeleton shape with nodes and lines?
  + Are we detecting the image as a whole against other collection of images?
  + Are we detecting based on motion and direction?
* What is the amount of resources required to run this ML task?
  + Is this for low-inference that can be run on a modern AI PC or mobile phone?
  + Could the vision model utilize NPUs?

**Technical Considerations and Resources**

* [OpenCV](https://opencv.org/) - Computer Vision Library
* [MediaPipe](https://github.com/google-ai-edge/mediapipe) - Live-streaming video
* [ASL Alphabet Datset Kaggle](https://www.kaggle.com/datasets/grassknoted/asl-alphabet)
* [RavenOnur/Sign-Language](https://huggingface.co/RavenOnur/Sign-Language) - hugging face and codelab example of detecting signs
* [aqua1907Gesture-Recognition](https://github.com/aqua1907/Gesture-Recognition) - Gesture Recognition