Dan Shannon, Derek Charron, David Saie
CIS 735 Machine Learning for Security
Deliverable 2
October 15th, 2023

Voice Recognition Authentication
We want to explore the use of voice detection as a biometric authentication model and its attack surface. Our plan is to use a dataset of open source recorded voices saying phrases from the Harvard Sentences. Then record ourselves saying Harvard Sentences and gain a voice profile. Then we will use our voice to authenticate based on the model built from the recorded voices.

What type of identifying parameters do we look for in voice data? What are the attack surfaces? Does the language being spoken affect the model's effectiveness? Can the model tell the difference between a recorded voice and a spoken voice, or a spoken voice and an AI generated voice?

Mozilla Common Voice Dataset: https://commonvoice.mozilla.org/en/datasets
Harvard Sentences: https://www.cs.columbia.edu/~hgs/audio/harvard.html