Assignment 1:

Assume you have a system that takes audio from the speaker containing the following problems:

- Consistent unwanted back-ground sound across all wav files.
- Long silences mid speech.
- Sudden increases and decreases in volume across a single audio file and across all wav files.
- Inconsistent sound wav lengths

The required pipeline does the following:

1. Take in speech recorded by the user.
2. Detects whether someone is speaking or not.
3. If someone is speaking, the system detects the gender of the speaker and shows it to the user.

You are given the following:

- The following pre-trained voice activity detection model **https://pytorch.org/hub/snakers4_silero-vad_vad/**
- A pre-trained voice-encoder model of your choice
- The following data-set:
    - https://drive.google.com/file/d/1HRbWocxwCIGy9Fj1MQeugpR4vOaL9ebO/view

Your task is to implement the required pipeline system.

The ONLY allowed pre-trained models are the ones provided to you and/or the voice-encoder model of your choosing. Everything else should be implemented from scratch by you. You are allowed to use python packages such as librosa, soundfile, pytorch, etc.