

Tutorial 2 Deep Learning

Source code for Deep Learning Helpers:

<https://github.com/izzajalandoni/Deep-Learning-Helper>

Backpropagation:

https://cs230.stanford.edu/winter2020/section3_exercises.pdf

1.) Try to test existing pretrained models first and see if their predictions work.

An AI assistant is basically made of 4 models:

1) VAD - Voice Activity Detection : this is a very lightweight model that is always on and listening. It has a binary output - presence or absence of human voice. If voice has been detected, it will activate the next block - ASR.

More info about VAD: <https://thegradients.pub/one-voice-detector-to-rule-them-all/>

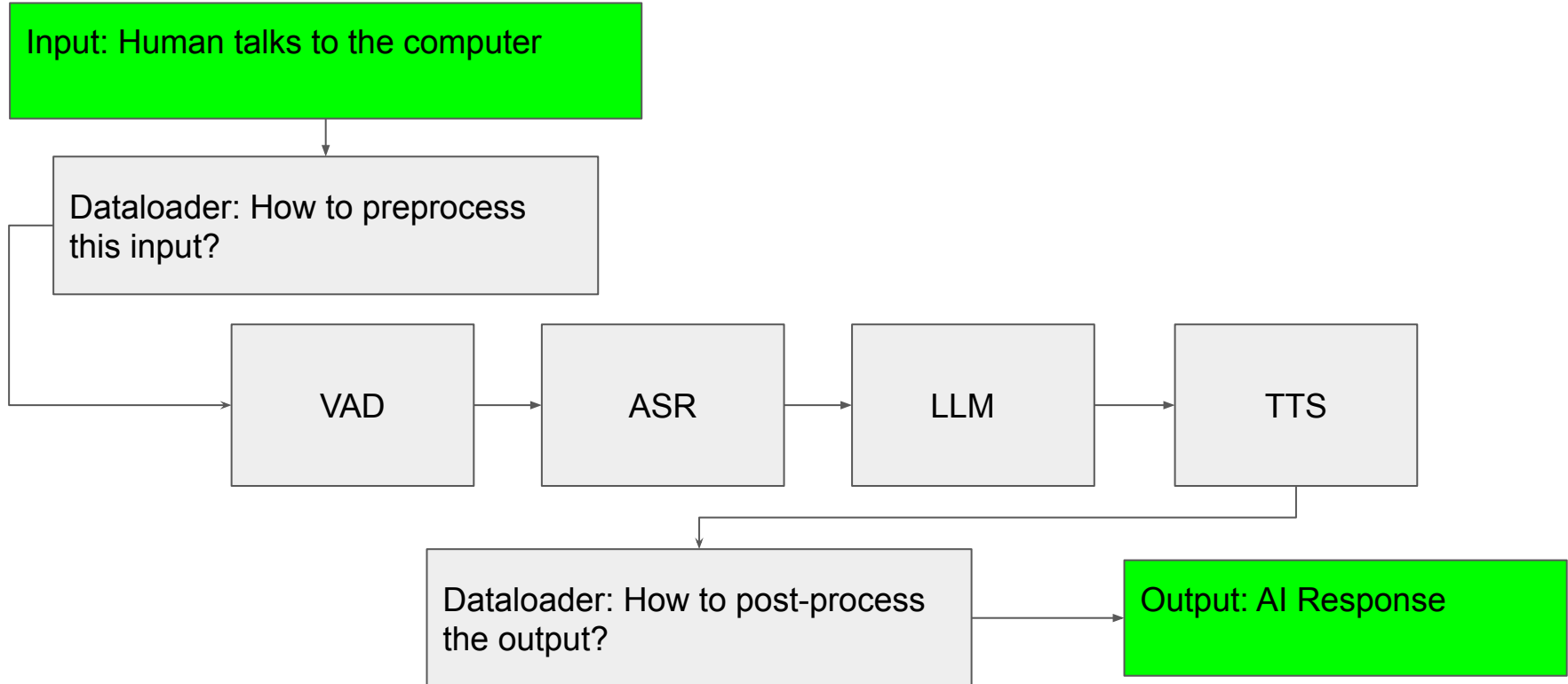
2) Automatic Speech Recognition (ASR) - this model can convert human speech to text. It is the most straightforward way for computers to understand human voice. When the text is already available, it can be passed to an NLP model which is usually an LLM.

More info about ASR and TTS: <https://developer.nvidia.com/blog/an-easy-introduction-to-speech-ai/>

3) Large Language Model (LLM) - after ingesting a huge amount of text scraped from Internet and other digital formats, a model learns fundamental patterns in human language. An LLM can be trained to perform a specific task (eg Question-Answering). You can use an LLM to query an AI for some general information - like "What is the capital of Japan?". The LLM should answer : "Tokyo" or "The capital of Japan is Tokyo". This text can then be passed to the next block - TTS.

4) Text-to-Speech (TTS) - usually trained on a large corpus of text aligned with a corresponding voice waveform. A TTS can produce natural sounding speech. This is the most intuitive interaction with the human. TTS provides an audible sound for humans to make understand the output of LLM in the most intuitive manner.

3.) Create a new dataloader that will manage all the input and output data



1.) Voice Activity Detection Models

<https://thegradient.pub/one-voice-detector-to-rule-them-all/>

2.) Automatic Speech Recognition Models

<https://paperswithcode.com/task/automatic-speech-recognition>

3.) LLM

<https://www.kdnuggets.com/2022/09/john-snow-top-open-source-large-language-models.html>

4.) Text-to-Speech

<https://paperswithcode.com/task/text-to-speech-synthesis>