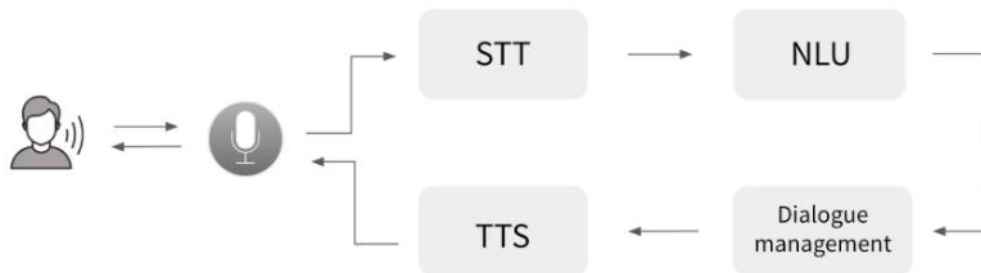


## Using Mozilla DeepSpeech and Mozilla TTS



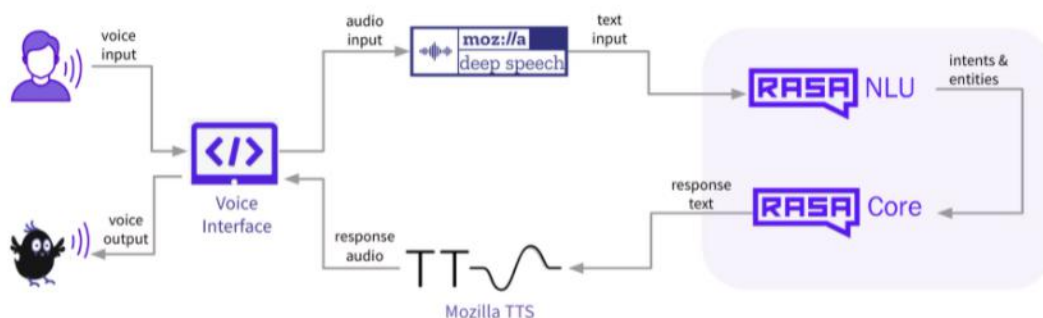
### Why Mozilla tools?

- Mozilla tools come with a set of pre-trained models, but you can also train your own using custom data. This allows you to implement things quickly.
- In comparison to alternatives, Mozilla tools seem to be the most OS agnostic.
- Both tools are written in Python which makes it slightly easier to integrate with Rasa.
- It has a big and active open source community ready to help out with technical questions.

### What is Mozilla DeepSpeech and Mozilla TTS?

- Mozilla DeepSpeech is a speech-to-text framework which takes user input in an audio format and uses machine learning to convert it into a text format which later can be processed by NLU and dialogue system.
- Mozilla TTS takes care of the opposite - it takes the input in a text format and uses machine learning to create an audio representation of it.

To Summarize,



## Steps to be Implemented :

1. The Rasa Assistant
2. Implementing the speech-to-text component
3. Implementing the text-to-speech component
4. Putting it all together

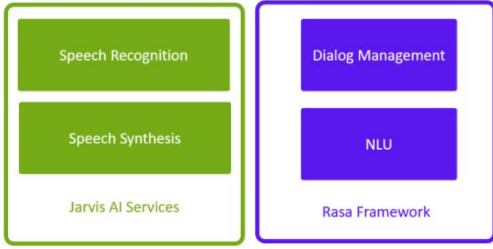
To put all the pieces together and test the voice assistant in action we need two things:

- Voice interface
- A connector to establish the communication between the UI and the backend (Mozilla and Rasa components).

Set up the Rasa voice interface first :

- Install npm and node, following the instructions mentioned in the following link : <https://treehouse.github.io/installation-guides/>
- Clone the Rasa Voice UI repository.

## COMPARISON TABLE :

Parameters	Jarvis(Riva)	Mozilla Tools
Services Provided	Speech Recognition (Speech-To-Text) and Speech Synthesis (text-to-speech).	Mozilla DeepSpeech (speech-to-text) and Mozilla TTS (text-to-speech)
Ingredients of the open source voice assistant	 <p>The diagram illustrates the components of Jarvis AI Services. It consists of two main colored boxes. The left box is green and contains 'Speech Recognition' and 'Speech Synthesis'. The right box is purple and contains 'Dialog Management' and 'NLU'. Below the green box is the label 'Jarvis AI Services' and below the purple box is 'Rasa Framework'.</p>	Rasa, Mozilla DeepSpeech, Mozilla TTS, Rasa Voice Interface
Services	The services needs to have low latency, preferably less than 300 milliseconds. It also requires to be trained.	Mozilla tools come with a set or pre-trained models, but you can also train your own using custom data.
OS	Seems to be less OS agnostic.	In comparison to alternatives, Mozilla tools seem to be the most OS agnostic.
Interfacing requirements	Runs more smoothly on Linux than Windows while interfacing.	Runs smoothly on both Linux, Windows and Android while interfacing.
Packages	Need to install with the NV packages for implementation of a voice assistant chatbot.	No such complications seen.
Requirements	Necessitating a scalable inference framework for NLP tasks on a GPU.	You need good STT and TTS components.
Flexibility	Most flexible than Mozilla Tools.	NLU has to be flexible enough to compensate for the mistakes made by STT.
Issues	Interfacing issues.	Few areas needs improvement, especially at the STT and NLU stage.
Solutions for Improvement	<p>To implement Riva framework NVIDIA packages are needed, For which, GUI is needed, which becomes a tedious job for student members when compared with other Open-Source platforms.</p> <p>Any alternative to make it available without such complications would make it more user friendly.</p>	<ul style="list-style-type: none"> <li>Pre-trained STT models are trained on a quite generic data which makes the model prone to mistakes when used on more specific domains. Building a custom STT model with Mozilla DeepSpeech could lead to a better performance of STT and NLU.</li> <li>Improving the NLU could potentially compensate for some of the mistakes made by STT. A rather simple way to improve the performance of the NLU model is to enhance the training data with more examples for each intent and to add a spellchecker to the Rasa NLU pipeline to correct some smaller STT mistakes.</li> </ul>

Availability	Currently ASR, NLU and TTS models are available in NVIDIA Riva. Trained on thousands of hours of speech data.	No such highly trained models are available. If needed without errors, then need to customize the data in order to avoid mistakes or errors.
Collaboration with Rasa	What makes this collaboration between NVIDIA and Rasa so compelling is that it is the combination of two technological environments who needs each other as much as they compliment each other.	Works almost in the same way.
Sequence of Events	The basic sequence of events shows how the power of Riva NLP and Rasa's NLU capability can be leveraged, especially for longer input.	
Capabilities	The Riva NLP capabilities are astute and the state management can be facilitated within Riva. Integration to existing text base digital assistants will stand Riva in good stead.	Capabilities are not as high as Jarvis, when compared.