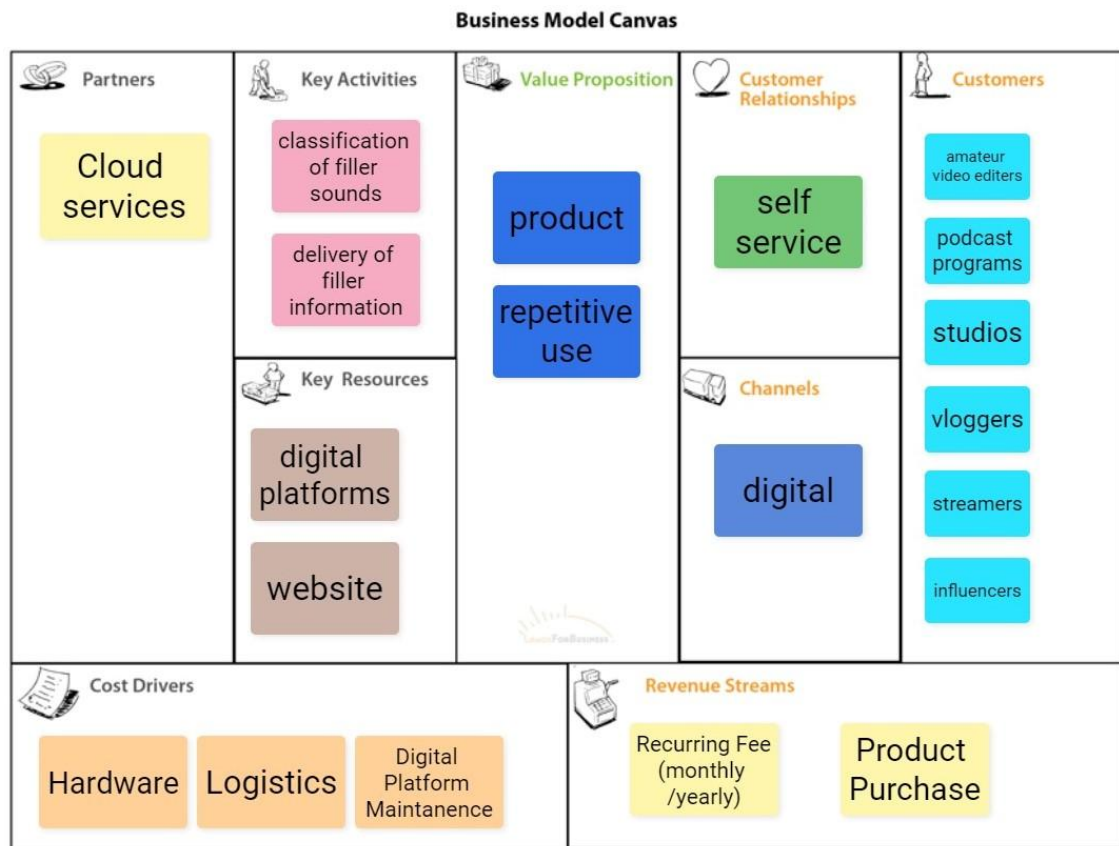


OUR MODEL



We are Qompose.AI, a multidisciplinary team that united during the Womanium Quantum computing and entrepreneurship hackathon and initiated our first AI focused project. Our business model is currently in a very short form since our project requires a high amount of time for technical realization and debugging therefore we only could have time for a short proposal but we are a hard working and dedicated team and hope our product already proved its usefulness.

Target audience Qompose.AI :

We aim to develop various innovative AI solutions in the long run particularly focused on audiovisual projects therefore we target a wide variety of audience depending on the particular product.

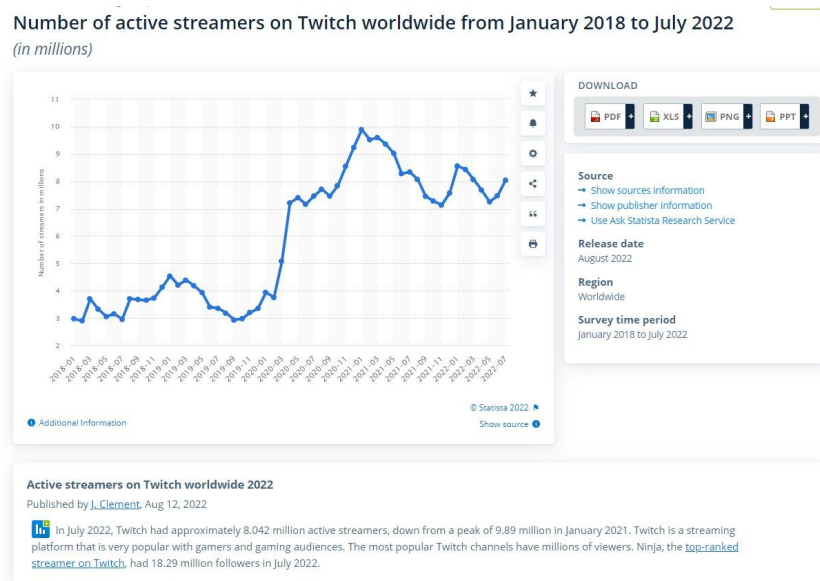
Target Audience Speechfire:

Our product Speechfire is a web based solution addressing anyone that takes over the extremely time consuming video editing task. We target a wide spectrum of audience

ranging from professionals such as influencers, podcast providers, video editing professionals to amateur users. Simply, anyone who wishes to have a video file with no speech errors. The influencers are especially concerned about social media algorithms, they aim for the best results, while there are also just regular people who want to have videos with better audio.

It takes raw video speech recordings as input, identifies, labels and generates markers for the undesired sections that require to be removed during editing. It has a well defined task that anyone that wants to edit a video in speed is part of our target.

Our target audience is expected to grow increasingly as it is already possible to see in this chart that shows the increasing numbers of only twitch streamers that would highly benefit this product:



The problem that we are offering a solution for:

Producing a professional and consumer oriented end piece is not an easy job. It requires a tremendous amount of energy and long working hours from the users, considering the fact that most of them are nonprofessionals.

We are able to offer an automated solution which will have the highest accuracy rate thanks to our own AI model. With our website initially we will be giving the exact time_stamps corresponding to disfluencies which will be an easy guide for users to detect and delete sections of their videos. As the next step we will also provide video editing options where the AI is a great option for better video transition.

Our goal is to simplify the editing task by making use of an advanced Deep Learning technique which reduces the amount of time required for this process into mere seconds.

Our method is based on acoustic image recognition which provides much better results than its text-to-speech counterparts and the Deep Learning model that we use consists of a multilayered architecture containing multiple Res-NET blocks each aiming a different category using squeeze-and-excite neural networks and other specific Attention models. As a final touch we collect the Data that we get as output from the AI, convert it into a form that is compatible with video-editing software and generate a preview on the website with markings. Currently we are using a simplified version of the neural network architecture for demo purposes which is trained and gives best results with the English language. The demo model offers 30%+ accuracy which we expect to increase to 91%+ based on proven research as we soon upgrade to our advanced model. Unfortunately none of the computing solutions that we currently own were powerful enough to run this heavy deep learning model which requires a very powerful GPU. However we based it on a previously proven research with the mentioned accuracy rate. We are totally capable of training our DL model with better hardware and more time for fine tunings and even reach higher accuracy by providing a larger variety of datasets and further model development while our company plan is to include an R&D division beyond implementing ready made research.

Our current accuracy rate with the simple CNN model, the advanced FLuentNet Research is to be found in research folder in our github page:

```
Model: "sequential"
Layer (type)                Output Shape                Param #
-----
conv2d (Conv2D)              (None, 198, 99, 32)        320
max_pooling2d (MaxPooling2D) (None, 99, 50, 32)         0
batch_normalization (Batch Normalization) (None, 99, 50, 32)         128
conv2d_1 (Conv2D)            (None, 97, 48, 32)         9248
max_pooling2d_1 (MaxPooling2D) (None, 49, 24, 32)         0
batch_normalization_1 (Batch Normalization) (None, 49, 24, 32)         128
conv2d_2 (Conv2D)            (None, 48, 23, 32)         4128
max_pooling2d_2 (MaxPooling2D) (None, 24, 12, 32)         0
batch_normalization_2 (Batch Normalization) (None, 24, 12, 32)         128
flatten (Flatten)            (None, 9216)                0
dense (Dense)                (None, 64)                  589888
dropout (Dropout)            (None, 64)                  0
dense_1 (Dense)              (None, 6)                   390
Total params: 604,358
Trainable params: 604,166
Non-trainable params: 192

Epoch 1/2
4/4 [=====] - ETA: 0s - loss: 5.7759 - accuracy: 0.2400
Epoch 00001: val_loss improved from inf to 1.81036, saving model to SE-ResNet_for_disfluency.hdf5
4/4 [=====] - 1s 267ms/step - loss: 5.7759 - accuracy: 0.2400 - val_loss: 1.8104 - val_accuracy: 0.2900
Epoch 2/2
4/4 [=====] - ETA: 0s - loss: 3.8025 - accuracy: 0.3100
Epoch 00002: val_loss improved from 1.81036 to 1.80585, saving model to SE-ResNet_for_disfluency.hdf5
4/4 [=====] - 0s 118ms/step - loss: 3.8025 - accuracy: 0.3100 - val_loss: 1.8058 - val_accuracy: 0.2400
```

We are also working on solutions that will work seamlessly with popular audio and video editing programs such as plugins and widgets. Among which are DaVinci Resolve, Adobe Premiere, youtube, Protools and many more..

We aim to collaborate with key partners

Initially there are lots of options for partnership but what immediately comes to our mind is that when we implement our core AI solution to an official plug-in with a video-edit software our partnership will be beneficial to both sides.

Monetization solutions.

The easiest and the most profitable monetization is recurring fee via subscription that could be based on monthly, yearly usage. There would be different types of subscriptions including one in which a client can only export the markers, another being where client also get access to our plug in of platform of their own choice that would trim the video for them in the software using automated script provided by us, and last one being where they get access to all (export the markers, access the plugins, and the ability to trim the video on our platform). This long term plan would not take place right away, however, in order to market and get recognition in the industry, we would keep it easy access for everyone.

Taking products such as Recut and Timebolt as reference which are in the market around 90 dollars range that do a simple task a 2 line of python code could do. We expect a high revenue with this product.

The hackathon

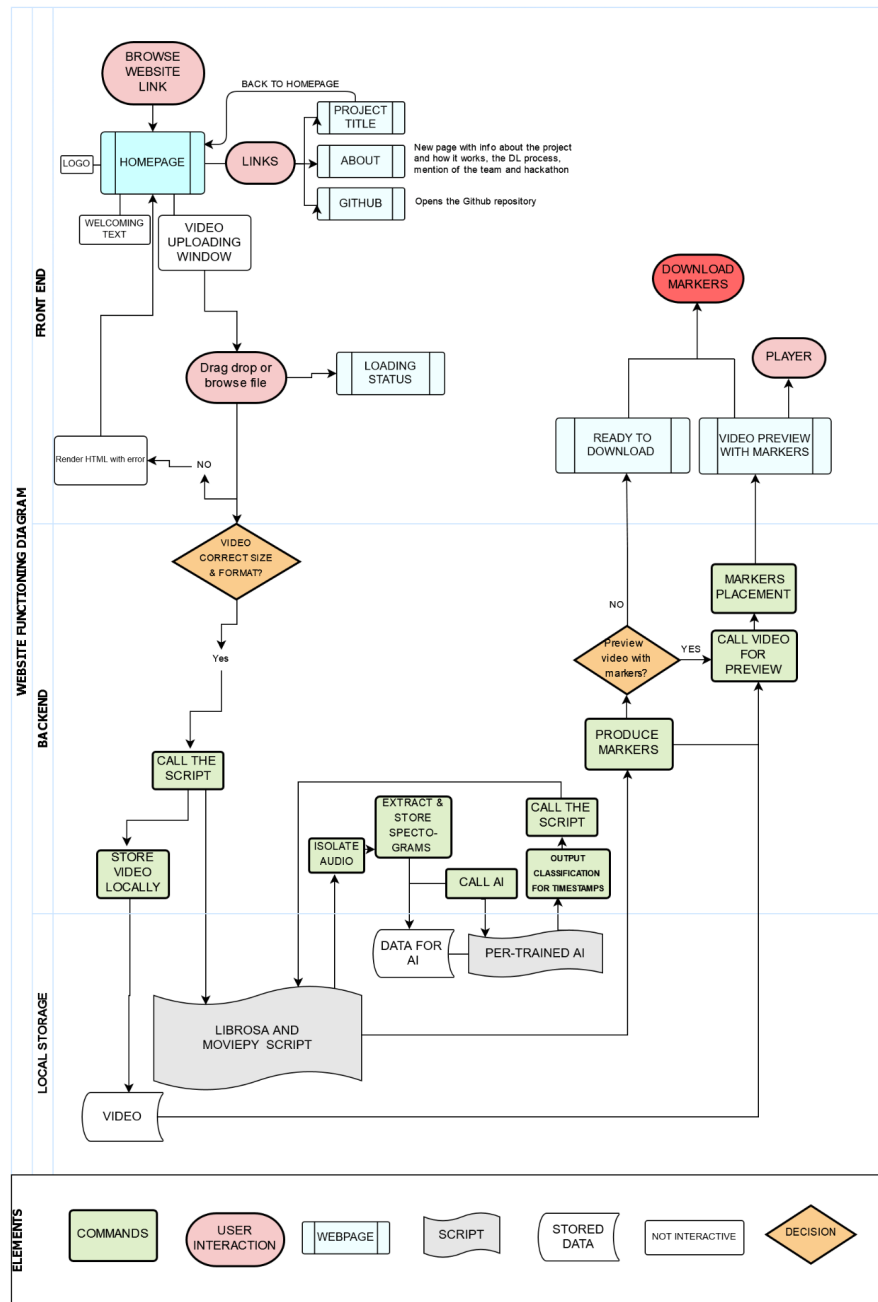
Most of the steps that we took are open to see on our Notion page where we tried to document as much as we could. We included the notion page link

<https://www.notion.so/yap-/Womanium-Video-AI-Project-d3a2f8ed80644bdcaa2c72a5af71d8ce>

which is also in github including a folder of all the research we did and the main references we consulted.

Current Website features:

The website is temporarily hosted at <http://gomposeai.herokuapp.com/> and it runs on flask project. There are 4 pages, first one being home page, which takes a video of only certain formats. Save the video, extract the audio and save that in local folder. It also is connected with libraries like dropzoneJS, Bootstrap, moviepy to perform these actions. After the AI operations, the player web page is rendered while lays out the video, spectrogram, and trimming markers for user to analyze. It also gives an option to save the markers. The layout of the website is as follows:



Future of the website:

The video stored locally will be stored on a cloud to easy access, the client will be provided with login feature and access to old videos, video trimming feature where it will be trimming the parts marked by AI. The client will be able to export the markers in more than one format.