

Time: 23/11/2023 – 15/12/2023

- We have completed our MoSCoW Requirement List:

Must:

- Use appropriate music (intro, outro)
- Generate script for the episode.
- Create podcast Structure (intros, greetings, meaningful filtered information)
- Convert text to speech (generate text based on a topic or description and then use text to speech tool, eg ElevenLabs)
- Create a functioning and not overloading UI (which is easy to use and navigate)
- Implement a scheme where a recent video (or article) can be transcribed and put into chatgpt to create a new episode on the topic of that video. That is because chatgpt doesn't use the current data

Should:

- Add detailed episode descriptions
- Create user suggestions on topics (users could type a topic that they would like an episode to cover). We'll have a box where users can enter their preferences and the future episodes will be about those.
- Create a help page (explaining how to navigate the tool)
- Create an episodes search (Based on keyword or types).
- Clone voices
- Have a diverse range of guests and topics
- Simulate interviews with historical figures
- Feature saved episodes option

Could:

- Add comedic elements in the episodes (the user can select the style of the episode whether its comedic, formal, entertaining etc).
- Add music based on sentiment
- Generate Images for the video
- Use relevant stats or graphics (eg if the episode covers a recent sport match, there can be a relevant graph to show some interesting data) (remove)
- Add immersive 3d audio (remove)
- Have a Light/Dark theme
- Allow the episodes download
- Add share options.

Won't have:

- Implement augmented reality
- Have a website translation (not really a priority as the podcast is solely in english).

There might be some adjustments to it as we discuss it once more with our client.

We have also looked at some more models for different generations involving text-to-speech, text generation and text-to-image. We also looked at text-to-video, but the

implementation of it is unlikely as the models are very expensive and we are not prioritising the video aspect of our podcast, focusing on text and audio.

The models we are planning to use are:

1. OpenAI – script generation
2. ElevenLabs – text-to-speech conversion
3. MidJourney – text-to-image generation

The plan for next few weeks as we go on holiday is to:

- Assign parts of programming to team members.
- Start programming (website and script, audio, video generation).
- Create github repository.
- Continue to research.