

## Project Title / Team OrcaBox

[Placeholder]

### Contact

Name	Alias
Sofia Yang	t-sofiayang
Fredin Vazquez	t-fvzquez
Hadeel Ali	t-hadeelali
William Tom	t-williamtom

### Summary Description

#### Problem Statement

How can we increase the accessibility of visual-based media in browsers for people with visual impairments so they can have equal access to information and experiences on the web?

#### Proposed Feature / Solution

Create a user-centric, in-browser solution, so people with visual impairments can easily access audio-based descriptions of previously inaccessible images, and also develop a feature so they can navigate through video content (**such as keynotes, and presentations?**) and easily access the most relevant/useful information.

### Strategic Rationale

Why is this important? How does it relate to revenue or strategic priorities? What pain points does it solve?

#### Why

##### 1. Web accessibility is a huge problem:

"According to WHO, there are 285 million people worldwide who, due to some disability (i.e. they are suffering with low vision), cannot read all content on a website. 39 million of those people are blind and cannot access any of the content via sight." (<https://www.sitepoint.com/how-many-users-need-accessible-websites/>)

##### 2. Visual-based media is a large blocker for people who have visual impairments

#### 2.1 Images

Many people who have visual impairments use screenreaders to browse websites; However, screenreaders cannot interpret images without '[alt text](#)'.

As of 2023, at least 22% of homepages are missing alt text on images  
(<https://webaim.org/projects/million/#alttext>)

## 2.2 Videos

Videos have a standard rate of 24 image frames per second to communicate large amounts of visual information

However, the visual information would be inaccessible to people with visual impairments, and so they can only obtain video context through sound, and captioned dialogue.

Large production films, etc. have teams to produce 'audio descriptions' to describe visual context, but the vast majority of videos on the internet do not have audio descriptions due to the cost of creating audio tracks manually.

### Pain Points Addressed

1. The issue of 'viewing' an image on a website, and being faced with a non-existent / poor "alt text", and thus being unable to gain visual context.
2. The pain point of missing out on visual context of what is happening in a video / missing out on text presented in a video.
3. The pain point of being unable to easily navigate to the most 'relevant' parts of a video, due to not having enough context (as in 2).

### User Scenarios (WIP)

1. As a user, when I select an image, I wish to hear a comprehensive description of the image with the following attributes. **(P0 Priority?)**
  - o The style of the image (i.e photo/graphic/lineart, grayscale/color )
  - o The entities in the painting (person, places, objects. Commonly identifiable entities such as Eiffel Tower, should be pointed out.)
  - o Any text that is image
  - o Description should describe the overall context (i.e black and white photo of a pigeon flying above the Eiffel tower), rather than just entities (i.e black and white, pigeon, tower)
2. As a user, when I select a video, along with the audio track, I wish to be able to hear a description of what is happening in the video with the following attributes: (This is for the scope of keynotes/presentations/demos) **(P1 Priority?)**
  - o For a graphic/slide is being presented, what is the content? (i.e Figures/Graphs, Graphics, Text, etc). Should all be pointed out.
  - o The 'entities' (person, places, object) and context. (I.e 'A person wearing a grey suit walks onto a stage in the center of the room')
  - o The speech description should not interfere with the experience of hearing the original audio in the video.
3. As a user, when I am navigating through different timestamps in a video, I wish to have an easy experience navigating to what I want to 'watch': **(P2+ Priority)**
  - o I should be able to have an UI to navigate to different shorter clips in a longer video so I can select which part to watch.

**Commented [LZ1]:** How would a user with visual impairment "select" a video? Do you need to check the ways of interactions?

**Commented [SY2R1]:** I believe the user flow of selecting a video is like this for a someone using a screenreader:  
1. Given a website with a list of videos, the user can have the title of each video read out loud to them with the screenreader.  
2. Once the user wants to "select" a video, they use keyboard shortcuts in their own screenreader app to navigate to next/previous element on the webpage until they reach the video they want.  
3. After the user reaches that element, they use a keyboard shortcut to interact with/play the video.

I'm referencing this for the ways a user can interact with a webpage:

<https://dequeuniversity.com/screenreaders/survival-guide>

Thank you for the feedback!

- The navigation should be facilitated with a keyboard for accessibility (tab key navigation)
- As a user, before I navigate to a shorter clip in the longer video, I want to hear a quick summary of what is going on in that portion of the video, so I know if it is important or not.
- For the quick summary, it should be reflected of what is happening in video frames, and what is being spoken in captions.

## Feature Priority