

Design of the H8 Speech Detector

Speech **H8** Detector

Of team H8 Speech Detector, for TechLabs Berlin Winter Semester
2020/2021

By Robin van de Water - UX Design Track

Contents

Figma prototype	1
Introduction	2
What problem do we solve?	3
What value does it provide?	5
User Stories	7
Parents with children using Youtube	7
Youtube viewers	7
Journalists	7
Information Architecture	8
User Flow A	9
User Flow B	10
User Flow C	10
Competitive Analysis	11
Prototype Implementation	12

Figma prototype

The Figma prototype can be found at the following web address:

<https://www.figma.com/file/ppGsy1cSvUC5peDyu3Qyt5/H8-Speech-Detector-Prototype?node-id=101%3A3521>

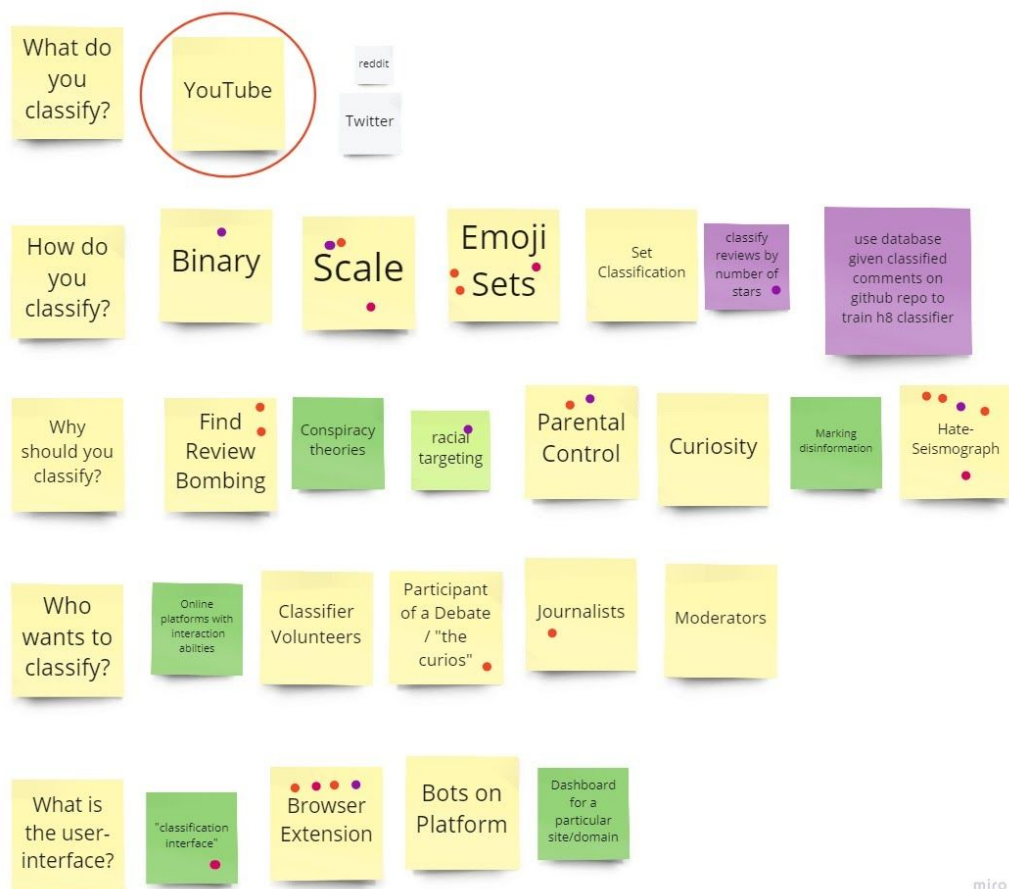
Introduction

Disturbing posts on minorities, political opponents and content creators can be found all over the internet. Some might be discomforting but others may be threatening. There are even structured, organized threats from right-wing trolls and conspiracy theorists, who deliberately torpedo discussions and flood the news feeds and comments of political opponents with hate, insults and death threats. In short: there should be no space for hate speech on the internet. The riots of the Capitol was just one of many events which could possibly be avoided if hate campaigns via social media were detected and the spread blocked. The idea of the H8 Speech Detector was born.

For this reason we decided to tackle hate speech on the biggest video platform on earth: Youtube. Youtube's impact in spreading hate and unwanted content has become undeniable and we were able to find ways in which to extract reaction data from the platform in order to analyze it to produce a so-called "H8 Report". In this report we go into the design decisions that have made this project into what it is.

What problem do we solve?

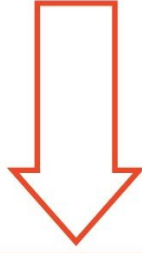
Starting out, in the first meeting, we did some brainstorming that gave shape to the project as it was presented by Urs in the initial presentation. We used Miro to gather answers to questions that Felix, our mentor, posed to us. After voting and revisiting this after another meeting, we decided to make our Hate Speech detector idea work for Youtube, as we found an API that could extract these comments into a format that could be processed by the data scientist. In these same meetings we also decided the why, who and what of the product. We figured that there are people that would like to check content on its toxicity and that a browser extension could be one of the ways to do this. This was later changed into a website, as this gave us more freedom and platform independence (and it fitted in with the web design track). The Miro Board that was created can be seen below, with the little dots as votes for a particular idea.



I then got to work to ask a few fundamental questions to answer together as I learned in my UX Track. The first, most important question can be seen below, answered on the Miro board:

Why are we doing this?

What's the reason this website/app/system needs to exist? What problem does it solve, for you/your client and for users?



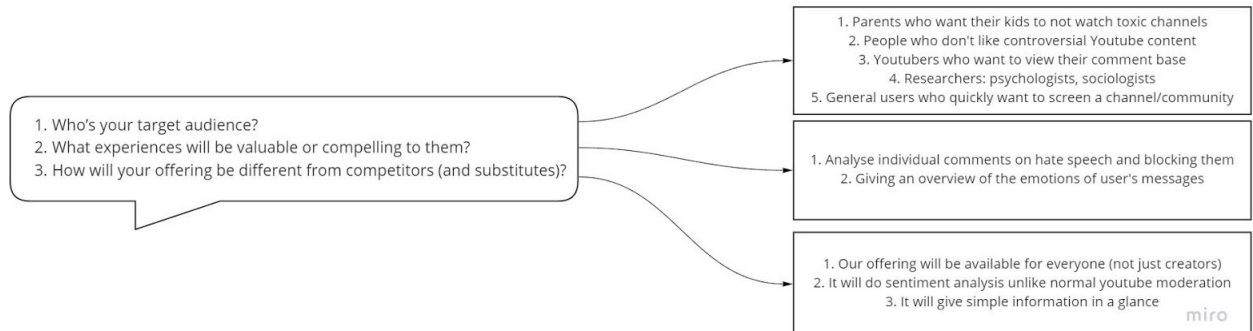
Solving: Unwanted content/communities on Youtube

miro

We mainly wanted to create a non-hostile Youtube for everyone. A world where people can be aware before they “go down the rabbit hole” of violent, hateful content. This gave us motivation and a goal to strive to.

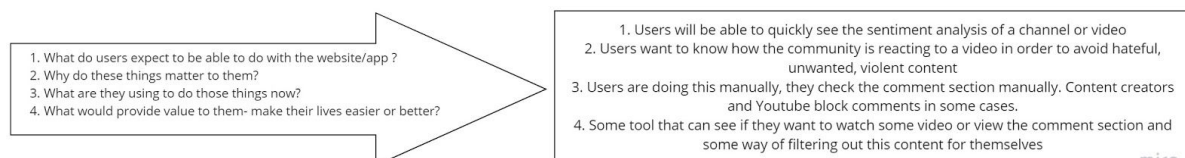
What value does it provide?

To determine the value of a product, one can ask a few questions. This has undeniable importance as one must identify if it has value before actually making something that is already there, outdated, or just not useful in general. These questions I laid out and answered, once again, on a Miro board.



We mainly targeted parents as these have a controlling role for their children. Children might not be able to, themselves, identify what is wrong or right and it is important that they don't base beliefs on content that is hateful, racist or overly vulgar. We then realized that selective Youtube viewers, Youtubers themselves, Researchers, or any Youtube viewer could benefit from this solution. We decided, given the time-limit we have, to make a sort of hate report that could be generated by inserting a reference to a video.

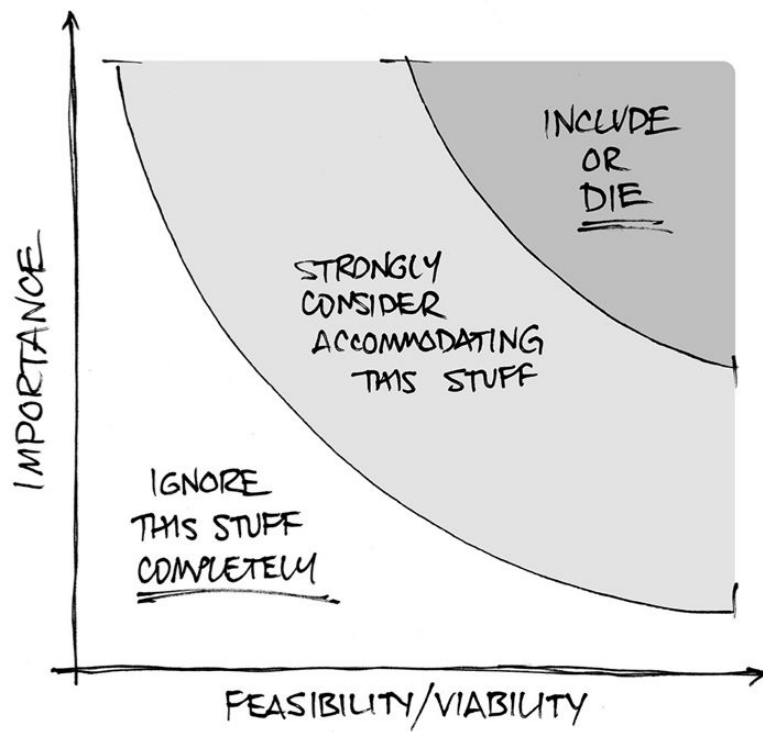
We then asked further questions based upon these target audiences:



We really think that this could make the lives of people better when they know that a certain Youtuber has an audience that is "hateful". We can, before someone checks the specific comments themselves, already give the user an indication of the type of things that are reacted to a certain video.

When looking at the features we had to implement, we decided to first research for the libraries and available technology that could realize our main goal. This is why we did not do a thorough analysis of features: we knew that we wanted to have a working classifier and interface that could give the user the basic functionality as a proof of concept.

Retrospectively, this seemed like a wise decision as we only had 5 people, and one UX-Designer. We ended up with only 3 people of which no web-developer. So Urs and me decided to take this role upon us, leaving little time for other UX-design focus in the project. We discussed this by analyzing the figure below.



User Stories

The user story core is made up of the following template:

- *As a [type of user], I want to [perform some task] so that I can [achieve some goal].*
- *Given that [some context], when [some action is carried out], then [a set of observable outcomes should occur].*

Parents with children using Youtube

As a parent, I want to make sure my children watch the right type of content so that I can be assured that they won't be exposed to hate speech, conspiracies, and threats.

Given that I see my children watch something, when they are on Youtube, then I should be able to see what kind of discussion this content provokes and what my children can see.

Youtube viewers

As a youtube viewer, I want to check if a video has hateful reactions so that I can avoid watching it and getting more negative content recommended to me.

Given that I want to watch a video, when I see a video in my recommended feed, then I should be able to screen this video beforehand.

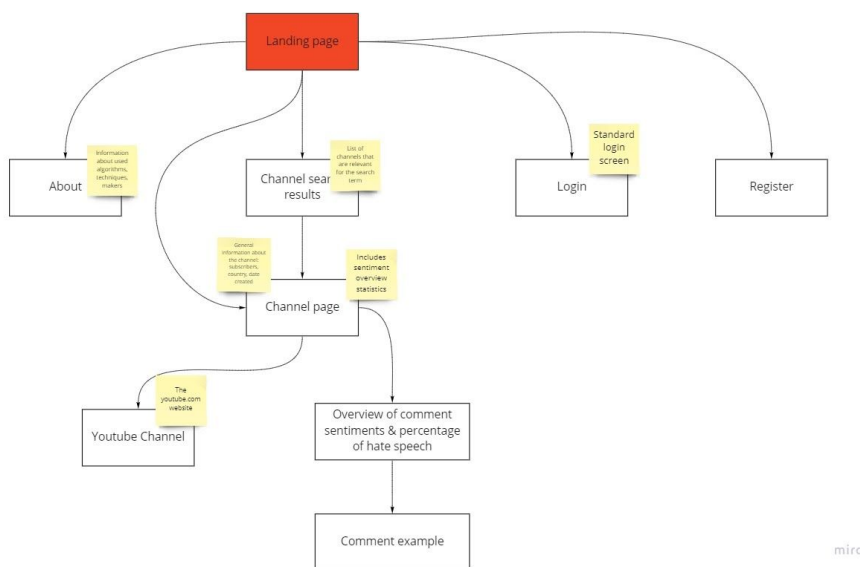
Journalists

As a Journalist, I want to check if a community is hateful so that I can write about it or research this. Given that I got a lead that some community is toxic, when researching a social debate, then I can easily see what the sentiment of this youtube audience is about in this topic.

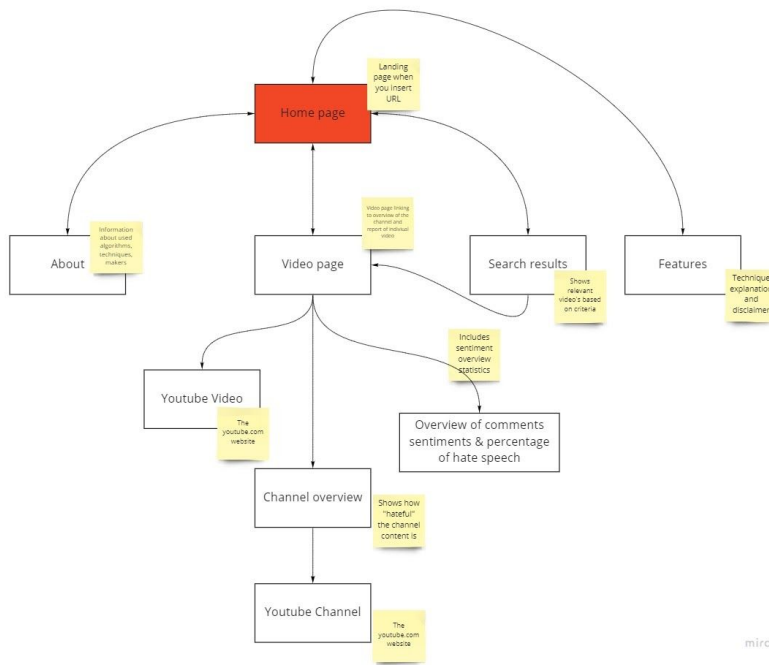
Information Architecture

The information architecture is very important for the eventual experience of the user using our website. The information architecture was changed and adoption after discovering the likely direction and capabilities of the hate detection algorithm. Given the limited time and knowledge, we could expect that this were to happen due to certain development choices that are better than others.

I conceived a architecture where you would end up on a landing page that would have channel search results, then a channel page for that specific channel where links would be to the channel and a detailed overview for that channel's video. The first design can be seen below:



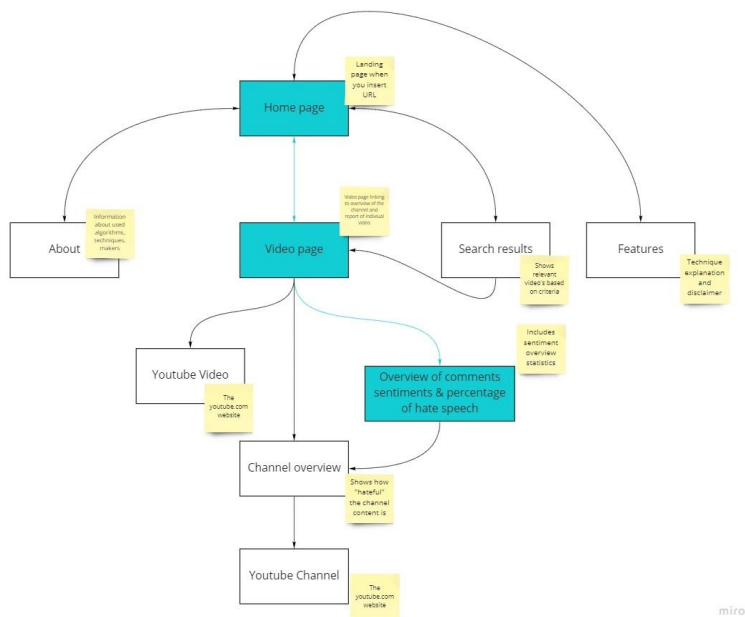
After deliberation we ended up not implementing an account system as it would make the product more complicated and it would cost implementation resources. We also decided to specifically focus on video's as this was an easier starting point than the channel architecture that I designed before:



Several User Flows through the project in this iteration, are detailed below.

User Flow A

User Flow A describes the standard user flow that we expect: the user inserts the Youtube URL and gets directly to the video page where he/she clicks on the overview of the video and leaves our webpage after having been informed.



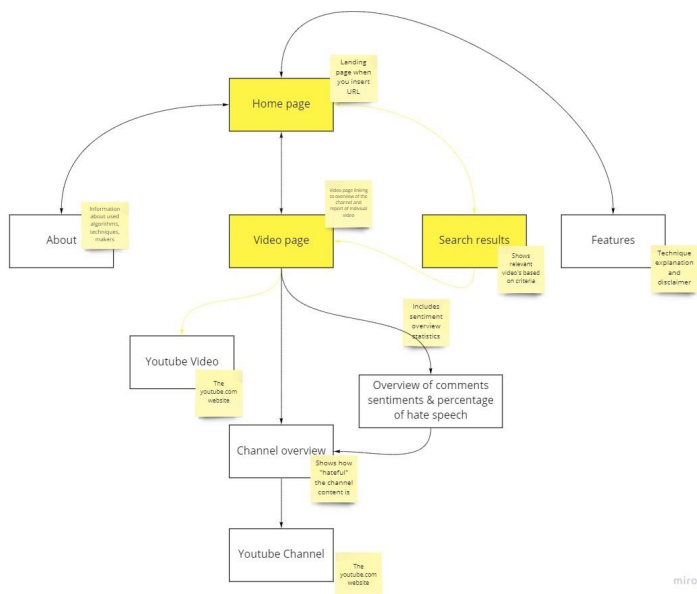
User Flow B

Another user flow is shown below, the user inserts a video title and ends up on the search results page where they pick the desired video after finding it in the list. After this they end up on the analysis page again and they want some more information about the hateful content of the channel in general so they navigate there. After this, they navigate to the youtube channel itself.



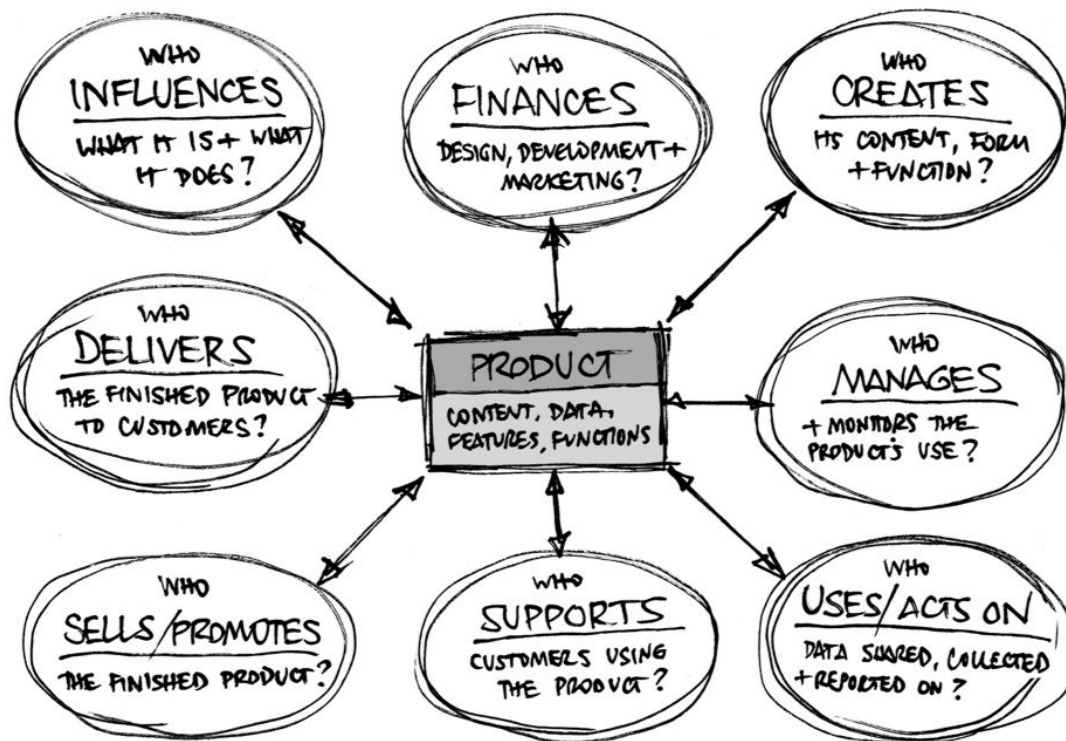
User Flow C

The user in the following user flow visits our webpage and inputs a video title from their memory. They click on a video they were interested in but realise they haven't watched it yet. The user leaves the site by clicking on the video and watches the video itself.



Competitive Analysis


When starting a business, a competitive analysis is essential for surveying the landscape of existing products and businesses. When looking at our product, the limited time, resources, and the intent of the project, relatively little time was put into this. I initially proposed this approach to the team, and they agreed as it has to be a learning experience for the practical roles of implementation. We therefore focused on making a Minimum Viable Product firstly. If we would like to make a viable business model in the future, however, we have a couple of pointers, such as the image below, to get started on this.



Prototype Implementation

For the implementation we were missing any web developer and we were only with 3 people. Urs and I decided to take care of the back-end, respectively front-end. This meant that it would have to be a very simple implementation with only a few of the pages and functionality realized. Luckily, I used the bootstrap framework in my prototype design in Figma, which made the visual elements relatively easy to reproduce. Part of the result can be seen below:

H8 [Home](#) [Features](#) [About](#)



H8 Speech
Detector

Check your Youtube Content

[Generate Report](#)

We check your youtube videos for hateful content and deem if it is appropriate to watch for children. Just enter any Youtube adress here and let us generate our reports. Enter the URL of the Youtube video or, if you don't know the URL, we search for it on the Youtube website.