

## **Beurling**

### **OCELOT AI**

### **Introduction**

This report is a compilation of two larger research projects aimed at trying to better understand Youtube's demonetization bot. The report have the goal and intention to help struggling youtubers to avoid demonetization. Something which has forced countless youtubers to stop making videos on Youtube and for employees to be fired.

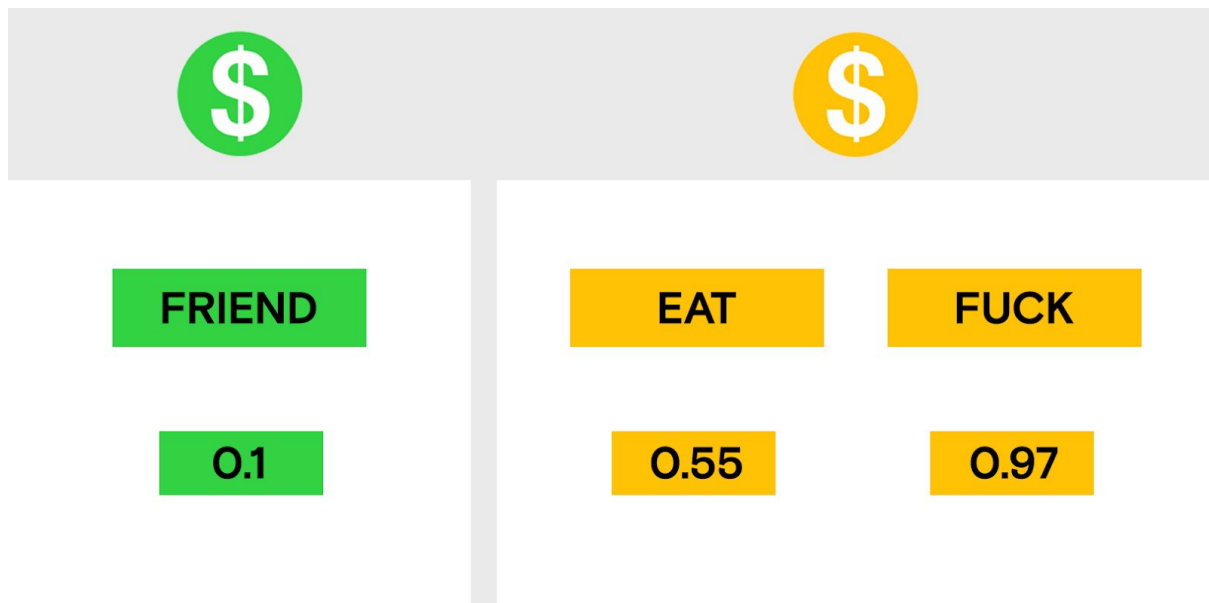
### **Demonetization list project**

The first research project that we will showcase is a project started by the Youtuber Sybreed and later automated by Sealow of Ocelot AI. The project primarily consist of a 15,000 word list of demonetized vs monetized title keywords. The list will serve as a great guideline for which words are at high risk of demonetizing any given video.

You can find the July testing results used for the [report here](#).

Check [Andrew's patreon](#) for updated versions (no payment required)

- We have no affiliation nor will we receive any part of any donations made.
- We do however endorse his future research efforts as it will be of great value to the community.



The list is best interpreted as a list of negatively charged keywords as certain words are deemed to be more severe than others. The demonetization bot classifiers have a hidden confidence level ranging from 0 to 1. If your video is above Youtube's threshold your video will be demonetized and a manual review is required. As a result you can still manage to get some videos monetized with titles containing some of the negatively charged words. You are however putting yourself at a constant disadvantage by doing so. The list is best utilized when a video is demonetized and you need to quickly identify if a word in the title is causing the demonetization.

During our research period Sybreed was able to get several videos monetized by removing totally acceptable words from the title such as "Restaurant" or "house". Our testing list contain endless amounts of similar words that are perfectly fine yet will negatively impact your chance of monetization.

We also conducted an experiment using 100 titles taken from top search results for "gay" and "lesbian" to approximate how many LGBTQ video titles are unfairly affected by the well documented LGBTQ demonetization. We then retested the titles that were demonetized by replacing the LGBTQ terminology with "friend" and "happy". To test whether the same titles would be okay for monetization had they not featured "gay" or "lesbian". This is also a great way to test if Youtube's classifiers are in fact capable of recognizing context to an acceptable level.

## **Demonetization list project**

### **Part 1 - Methodology**

The demonetization list project testing was conducted during a 2 month period. The first testing period included 1,500 words that were manually tested by Sybreed

between the 2nd of June to the 5th of July, 2019. The second period included 14,000 words and was automated using the Youtube Data API by Sealow from the 6th of July to the 21st of July. All videos tested were 1-2 seconds long and featured no visual or audio content that could trigger demonetization. All videos had a waiting period of around 2 hours as most title changes triggered a rescan within 10-30 minutes regardless of how many title changes were made at once. The total amount of videos never exceeded 300 and we also kept track of which videos the words were scanned on. The latter being to later verify that weird outliers did in fact update from green to yellow and vice versa. All future versions including the list for September will be available here and will be tested by Sybreed.

## Demonetization list project

### Part 2 - Outliers

During our testing we found a significant amount of weird outliers. We compiled 300 of these outliers in to a [google spreadsheet](#). The list contain words such as “Restaurant, ebooks, you, program, sunglasses, then, there, that, these, stair, sql, database, images, photos, cameras, hotel, hotels, expedia, gay, lesbian, from, farms, farming, kilometer, profit, say, says, saving, tickets and tourism” just to name a few.

It should be made clear that terms such as “woman”, “gay”, “lesbian”, “democrat” and “liberal” are likely negatively charged due to their use in political commentary that is often deemed to be non advertiser friendly. That is not to say that this doesn’t have a negative impact on minorities and will lead to discriminatory demonetization of perfectly acceptable videos regardless of context. The classifiers that Youtube use are probabilistic and if certain words are overrepresented for manually confirmed demonetized videos then the demonetization classifiers will have an increasingly harder and harder time differentiating between what is and isn’t acceptable. Resulting in more and more perfectly acceptable videos being demonetized and requiring manual review.



A very good example of this is “Animals” which in July lead to videos being demonetized due to the large amount of animal fighting videos on Youtube. Despite the titles making no mention of fighting or abuse.

That is not to say that most of these outliers are understandable however. As the list of outliers contain words such as “Restaurant”, “Sunglasses”, “Xerox”, “Paypal” and “Ebooks”.

A few weird trends among these outliers are other companies and products “Paypal, Xerox, Expedia, Hotels, Zorb, Zumiez, Wordpress, Dell and Shrek” were all demonetized during our testing. Likewise payment related terms such as “profit, billing, revenues, saving, savings, selling, sells, referral, referrals, payroll, euro” were all demonetized during our testing. Another trend is programming terms such as “Database, SQL, program, http, xml, xhtml, username, password” which should be perfectly fine for all types of monetization unless explicitly stated alongside hacking terms or feature frequent swearing. Neither of which were present during our testing.

## **Demonetization list project**

### **Part 3 - LGBTQ extended testing**

In order to estimate how widespread unfair demonetization of terms such as “gay” and “lesbian” are we conducted a small experiment featuring 100 titles from the top search results for “gay”, “lesbian” and “LGBTQ”. The demonetized titles were then modified by replacing LGBTQ terminology with “friend” and “happy” to see if the same videos would now be monetized. The [full spreadsheet can be found here](#).

33 out of the 100 titles tested that we deemed fit for monetization were demonetized despite being perfectly fine by all standards. The list of demonetized titles include titles such as:

“Gay and Lesbian Guide to Vienna - VIENNA/NOW”

“LGBT Tik Tok Compilation in Honor of Pride Month”

“Top 10 Lesbian Couples in Hollywood Who Got Married”

“Lesbian Princess”

“Lesbian daughters with mom”

What is more shocking is the fact that when we re-tested these titles after replacing the LGBTQ terminology with “friend” and “happy”. Every single video that we tested was now monetized. Which more than explain the LGBTQ community’s outrage over being demonetized for LGBTQ terminology. As perfectly acceptable context is still

demonetized by the bot where the exact same videos are monetized without the LGBTQ terminology.

This is not a matter of LGBTQ personalities being demonetized for something that everyone else would also be demonetized for. This is LGBTQ terminology like “gay” and “lesbian” being the sole reason for a video being demonetized.

“Happy and friend Guide to Vienna - VIENNA/NOW”

“Happy Tik Tok Compilation in Honor of friend Month”

“Top 10 happy Couples in Hollywood Who Got Married”

“Happy Princess”

“Happy daughters with mom”

It is crucial to realize that this arguably discriminatory pattern of the demonetization bot is still present in 33% of the titles tested despite being a well known issue for almost 2 years now. This issue is made worse given comments by Youtube’s VP of all products Neal Mohan who in an interview with the Mexican Youtuber “Luisito Comunica” stated the following after being asked about why LGBTQ terminology lead to videos getting demonetized.

*“While we might have made mistakes in the past. The nature of our algorithms is that they continue to learn from those mistakes. Continue to get better”*

Mohan then go on to mention that Youtube and Google have programs in place that relate to “Algorithm fairness”. He also state that Google strive to be an industry leader in “Machine learning fairness”.

Given these comments it’s made clear that Mohan highly insinuate that LGBTQ demonetization is no longer an issue and that it’s a thing of the past. Given our testing results it’s made clear that these comments are not accurate. While the current situation may certainly be better than 2 years ago it is far from what most LGBTQ creators would deem as acceptable performance. Hearing Youtube’s VP of all products calling this issue resolved when the reality is far different is a worrying statement. As it would imply that Youtube are significantly less likely to allocate resources to try and fix the issue if they no longer believe that it’s an actual issue.

Youtube’s ceo Susan Wojcicki however had a much more reasonable response in a [recent interview with the Youtuber Alfie Deyes](#).

*“We work incredibly hard to make sure that our systems are fair. We have a ML fairness initiative, ML stands for machine learning. To make sure that our algorithms, the way that our machines work are fair. There is no policy, of course”*

*“We work incredibly hard to make sure that when our machines learn something. Because a lot of our decisions are made algorithmically that our decisions are fair. And we have a whole committee and a whole process to make sure that we are managing fairness of how our algorithms work”*

In regards to the questions whether “there’s no like flagged words from specifically that community that is like, more likely to be demonetized?” Wojcicki replied that “there shouldn’t be, no”. Our testing result is however proof of the contrary as every demonetized title was deemed to be advertiser friendly the moment that we replaced the LGBTQ terminology.

That is however not to take away from the fact that it’s clear that Wojcicki genuinely does care about the LGBTQ community. Because while the demonetization bot do inhabit several arguably discriminatory patterns it should be made clear that this is not due to Youtube policies nor a lack of programs in place to mitigate algorithmic discrimination. It’s simply the result of the probabilistic nature of the machine learning classifiers used by the demonetization bot.

Because while it might be easy to believe that the classifiers will eventually be able to fully recognize the context given enough training data. The reality is often that the data is full of noise such as several LGBTQ videos that are talking about sex and several anti LGBTQ videos that have all been confirmed by manual review. As a result the classifier will have a significantly harder time differentiating between what is and isn’t acceptable which have resulted in several LGBTQ videos being unfairly demonetized.

It should also be made clear that another anti LGBTQ shooting or similar tragedy in the future could have massive implications on the future demonetization of LGBTQ terms. Even though such an unfortunate development would certainly be understandable from Youtube’s end. It does however showcase the downfall of using probabilistic machine learning classifiers in comparison to more logical implementations.

## **Undisclosed major advertising changes**

The 2nd project made by Sealow and Vk focus on a new advertising tier called “Limited inventory”. The ad tier was first discovered by seeing a clear difference in

CPM between videos containing swearing and no swearing in early 2018. This pattern was later confirmed via clear testing results in late spring of 2018. A direct link between DoubleClick revenue, Google's premium advertising brand and higher CPM videos was also established during this testing period.

The diagram illustrates three inventory types for advertising, each represented by a dollar sign icon and a progress bar. The 'Limited inventory' type is highlighted with a blue border and a checkmark in the top right corner.

Inventory Type	Icon	Progress Bar	Recommendation
Expanded inventory	Yellow circle with white '\$'	Blue bar (approx. 75%)	Not recommended
Standard inventory	Green circle with white '\$'	Blue bar (approx. 50%)	Recommended
Limited inventory	Green rectangle with white 'CPM'	Blue bar (approx. 25%)	Recommended (highlighted)

**Expanded inventory**

Maximize available inventory by showing ads on some sensitive content

Excludes extremely sensitive content, for example:

- Excessive profanity
- Graphic sexual content and nudity
- Graphic violence and serious injury

**Standard inventory**

Show ads on content that's appropriate for most brands

Same exclusions as Expanded, and also excludes, for example:

- Repeated strong profanity
- Strong sexual content and discussions of sex
- Violence either real or dramatized

**Limited inventory**

Exclude most types of sensitive content and limit your available inventory

Same exclusions as Expanded and Standard, and also excludes, for example:

- Moderate profanity
- Moderate sexually suggestive content

The system was later added to all advertisers in early September later that year. Despite this new system being publicly available to all advertisers for a year now, Youtube have never disclosed this change to the community nor give them any indication or way to request a review. The reason why you might already know about this change is because of our original testing last year and it's subsequent spread among youtubers.

So to make it clear, no official communication channel by Youtube have ever acknowledged this change. Neither Youtube's recently updated advertiser friendly guideline nor Creator insider's video about demonetization ever mention this. Despite the former video being endorsed by Youtube's CEO Susan Wojcicki. They are in fact giving outright misleading information that would severely hurt a channel's income by not being cleared for the "Limited ad inventory". Which is known to give more than double the ad revenue compared to the "Standard inventory".

## Inappropriate language

Content that contains frequent uses of strong profanity or vulgarity throughout the video may not be suitable for advertising. **Occasional use of profanity won't necessarily result in your video being unsuitable for advertising,** but context matters.

Examples (non-exhaustive)

Category	Limited or no ads 💰 💰
Inappropriate language	<ul style="list-style-type: none"> <li>Profanity in the title or thumbnail image</li> <li><b>Strong profanity</b> used repeatedly in a hateful or derogatory way</li> <li><b>Strong profanity</b> used during the beginning part of the video</li> <li><b>Strong profanity</b> used multiple times throughout the video even if bleeped or for comedy, documentary, news, or educational purposes</li> </ul>

As an example: Youtube's [help page](#) make it clear that "Occasional use of profanity won't necessarily result in your video being unsuitable for advertising, but context matters". They also directly state that "Strong profanity" is disallowed while "Light profanity" is allowed on their [content examples help page](#).

## Profanity

Video content	Expanded inventory	Standard inventory	Limited inventory
<b>Light profanity</b> used in a non-hateful, comedic or artistic way	Included	Included	Included
<b>Moderate profanity</b> used in a non-hateful, comedic, or artistic manner, or a music video with frequent profanity	Included	Included	<b>Excluded</b>
<b>Strong profanity</b> used throughout or at the very beginning of the video in comedy, documentary, news, or education	Included	<b>Excluded</b>	<b>Excluded</b>

Google ad:s [help page](#) however clearly differentiate between light, moderate and strong profanity. Where moderate profanity is excluded for the "limited inventory" ad tier alongside "moderate sexually suggestive behavior, or a music video containing sexual content". This ad tier difference of "moderate profanity" and "moderate sexually suggestive content" have never been disclosed to the community by Youtube.

By not disclosing this ad tier Youtube have left upcoming and well established channels to lose out on upwards of double the revenue for no good reason at all. A difference that can be crucial between being able to survive as an independent creator or having to fire people as larger company.

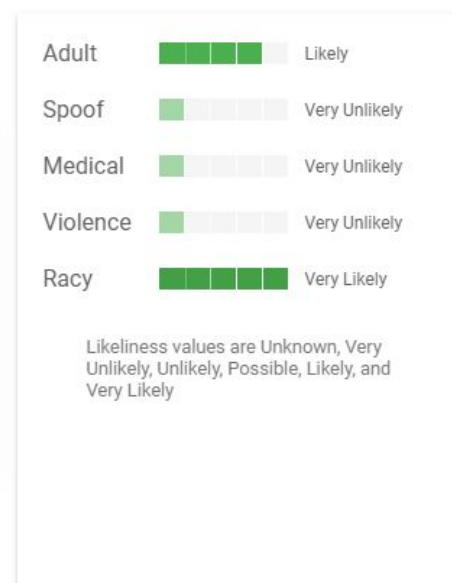


The former is actually well documented in the video [“Where is Mariah?”](#) by the multi channel owner Matthias. Who starting at 13:52 explain that he had to fire an employee of the channel REKT due to the channel operating at a loss. Not due to the viewership being low but specifically that the CPM was significantly lower than any of his other channels. In the video Matthias state that the channel’s CPM was “A tenth of what it should have been”. A clear victim of horrendous communication by Youtube that resulted in an employee having to be fired for no good reason at all.

There is absolutely no good reason for Youtube keeping this new ad tier a secret. Even if that means that we don’t have a manual reviews. Just knowing about this being a thing have been enough for hundreds of creators to massively increase the amount of money they make by actively trying to never swear in their videos.

## Getting into the “Limited inventory”

- Don’t swear in your videos. Even a single swear word other than “damn, hell, etc” will get you out of the limited inventory ad tier.
- “kill, shoot, died, etc” are all fine for videos in the gaming category.
- Changing the captions or self certifying your video will not affect your CPM.
- Vision AI “safe search” scores higher than or equal to 3 for “racy” or “adult” will also result in a removal from the limited inventory ad tier.



This video had less than half of it’s expected ad revenue due to the AI ratings. At the time of publication the video has 21 million views of mostly US viewership. I will refrain from making estimations of revenue lost but I will insist that keeping a system

with such a massive impact on revenue a secret is severely hurting the Youtube ecosystem. Especially as had he known, he wouldn't have used this topless version.