February 2024
version 1

DISINFORMATION ON YOUTUBE: RESEARCH AND CONTENT MODERATION POLICIES

EU DisinfoLab

TABLE OF CONTENTS

Author: **Raquel Miguel Serrano**, Senior Researcher at EU DisinfoLab

Reviewer: **Guillaume Kuster**, CEO and Co-Founder of CheckFirst

The opinions expressed are those of the authors and do not necessarily reflect the position of EU DisinfoLab. This factsheet does not represent an endorsement by EU DisinfoLab of any organisation.

INTRODUCTION

This document shows how YouTube operates and can be exploited for disinformative purposes, but also provides guidance on conducting research on the platform. It is intended to serve as a valuable resource for the community dedicated to countering disinformation on YouTube.

The document is organised as follows: (I) It begins with a concise explanation of how YouTube works. (II) Subsequently, it introduces ways and tools for investigating on the platform. (III) It covers the process of reporting content, along with an overview of policy aspects related to enforcement. (IV) Lastly, it offers a repository of some examples of recent case studies on disinformation on YouTube.

As one of the largest and best-known online video sharing platforms, YouTube simplifies the process of creating, uploading, and watching videos online: the slogan "broadcast yourself" speaks for itself. Established by three former PayPal employees in February 2005, the platform was later acquired by Google in October 2006 and currently operates as one of its subsidiaries. The platform has fundamentally reshaped the way people engage with video content, including information, and has allowed new forms of monetisation, even creating the figure of professional content creators, dubbed "youtubers".

Expanded all over the world, India is the country with the largest YouTube audience by far, with approximately 462 million users engaging with the platform (as of October 2023). The United States follows, with around 239 million YouTube viewers, while Brazil comes third (144 million users). Within Europe, the United Kingdom emerges as the leader with around 56.2 million internet users. Among the European Union (EU) countries, Germany ranks highest.

Far surpassing the required 45 million users, YouTube was designated by the European Commission a very large online platform (VLOP) on 25 April 2023, requiring it to comply with the provisions of European Unions' Digital Services Act (DSA) from August 2023.

The platform faced with an outcry of criticism in 2022: over 80 fact-checking organisations collectively signed a letter addressed to YouTube's Chief Executive, Susan Wojcicki, describing the platform as a "major conduit" for falsehoods. The letter specifically called out YouTube for its role in spreading COVID-related misinformation, electoral narratives, and conspiracy theories. In response to these concerns, YouTube implemented various measures to combat misinformation. The reporting requirements by the DSA will be a good method to assessed if these measures taken had a discernible impact.

YOUTUBE: HOW DOES IT WORK?

YouTube can be accessed via its app or web interface. And there are two ways to engage with YouTube: as a content creator or as part of the audience.

1. **Content creators.** Content creators can easily create a channel in order to upload videos and monetise their content through advertisements or paid subscriptions, in case they reach certain level of audience engagement. Apart from the uploaded videos (long and short ones), they can create lists in their channels and offer information about their own communities. They can also refer to other channels or accounts to be followed on other platforms.
2. **Users/audience.** On the other side, users can watch videos on the platform for free, although there is a premium version available for ad-free viewing, and some specific content may require payment. Users can access content that is sorted by categories (all, music, psychology, live, mix, and videogames, among others) and by different sections: videos, short videos – with algorithmically recommended thumbnails to videos (if the access is via the web) and shorts (if access is via the mobile app) – the subscriptions section (with the content to which the user has subscribed) and the historical with the list of videos already watched by the user. The browsing history is managed through the Google account, which allows users to create playlists, archive videos they like, and receive recommendations from the platform's algorithm based on their preferences.

   Users can interact with content by liking or disliking it – indicated with thumbs up or thumbs down – leaving comments, sharing content, and subscribing to channels or accounts to stay updated on new content. Having an account is required to access these features. On their side, creators can provide feedback directly in the comments they receive.

INVESTIGATIONS ON YOUTUBE

YouTube indeed differs from other platforms in many aspects, primarily focusing on video content. The visual nature of videos can introduce additional challenges when researching disinformation, especially when dealing with videos in languages that the researcher doesn't understand. However, YouTube offers various elements beyond videos that can be explored for research. For those new to the platform, we highlight some essential characteristics to be considered when researching on YouTube:

- **Complex Search Challenges:** Videos pose an additional challenge when researching disinformation. Watching them can be more complex and time-consuming. That's why researching on YouTube often requires a creative approach. Beyond the video content itself, external elements like titles, channels, and comments can contain critical information for analysis. In addition, YouTube provides automatic translation of captions, which can be a valuable input when researching content in a non-spoken language.
- **Personalised Search Results:** YouTube's search results are influenced by a user's search history and prior video interactions. This personalisation means that search results for the same query can differ between users. Researchers must be aware of this factor while conducting their investigations.
- **Distinctive Video IDs:** Each video on YouTube has a unique ID, which serves both to protect creators' rights and as a valuable resource for research purposes. It enables precise identification and tracking of individual videos.
- **Recommendation Algorithms:** YouTube's recommendation system is a crucial aspect of the platform. Research into how this system functions and influences user behaviour and content dissemination is valuable for understanding the spread of disinformation.

Thus, there are multiple aspects that can be investigated on YouTube concerning disinformation. These include the content of the videos themselves, as well as factors like Coordinated Inauthentic Behavior (CIB) or the algorithmic recommendation schemes that present structural risks in the dissemination of disinformation.

Here are some guidelines for conducting searches and research on the platform.

1. SEARCH FUNCTION: SEARCH BOX AND ADVANCED SEARCH ON YOUTUBE

YouTube offers a search function, enabling users to find content based on keywords. This search box can help to search for video titles or channel names. According to YouTube, the platform considers several factors when returning search results, including content relevance, user interactions, and content quality. However, search results can be influenced by a user's search and playback history, provided they have this feature enabled. Therefore, search results may vary from one user to another.

For more precise searches, YouTube provides an advanced search feature. This feature allows users to search for videos, channels, or playlists and offers the option to refine search settings with filters like upload date, duration, features, or relevance. It is also possible to use Google operators to finetune the search. Valuable tips and tricks for optimising a YouTube search experience can be found in the following articles, that can help researchers effectively navigate and utilise the search features.

- How to Perform Advanced Search on YouTube
- YouTube Search Tricks
- How to Search YouTube Like a Pro with Google Advanced Operators

2. INVESTIGATING CHANNELS OR NETWORKS

In order to investigate a channel, YouTube can provide relevant information. In the information section, we can find inputs such us the creation date, views, or number of subscriptions, but also links to other accounts or channels that some creators promote in their bios or video captions, or alternative ways to contact them, such email addresses. These alternative accounts or channels can be very useful to track content that may have been removed from the platform but still exists on other platforms. The past experiences suggest that some creators have used such alternative accounts to guide their followers on where to find their content in case of moderation on YouTube.

Other visible details of the channels include the creator's lists or information about their community, which allows the creators to refer to like-minded individuals and expand their communities. From a research perspective, this information is invaluable for studying relationships within YouTuber communities.

This informative article by First Draft explores methods for uncovering and understanding networks on YouTube and fringe platforms.

3. INVESTIGATING ADS

Another element that can be a source of disinformation are the ads displayed on YouTube. These can be searched in two different libraries:

- YouTube's Ad Library. This library allows mainly commercial ads to be searched by various categories (apparel; beauty & personal care; car & trucks; food & drinks, health & fitness; home & outdoor; media & technology; pet accessories; shoes & accessories) and by brand.
- Google Ads Transparency Center. Google allows the search of ads served on Google platforms, including Google Search and YouTube. Researchers can obtain information about the ads run by one specific advertiser in a certain period of time (repository starting 31 of May 2018). Additional transparency information is available for political ads for certain regions. The information refers to advertisers who are verified through Google's advertiser verification program or the election ads verification program, and the ones who have not yet undergone or completed one of those processes.

  The search interface allows to differentiate between "all topics" and "political ads" and to search by advertiser, filtering the countries in which the advertisement is shown and the format of the advertisement (video in the case of YouTube). The interface also informs whether the advertiser has verified its identity or not.

  Google provides information such as the number of ads, the period when the campaign was active or the number of times the ad was shown in a specific country.

4. INVESTIGATING FROM OUTSIDE: SEARCHES ON THE INTERNET AND CROSS-PLATFORM SEARCHES

Searches on the web. In some instances, searching within the YouTube platform can be challenging, especially when content has been deleted. However, there is a potential alternative to search for such content using Google's cache. It can be done by searching through the Google dork "site:youtube.com/ + keyword" and clicking the three points on the right that lead to the cache option. To optimise your search results, consider that:

- It's important to familiarise with the structure of YouTube URLs, which typically follows this format: https://www.youtube.com/@ChannelName/. This understanding will help you to get more precise searches.
- Each video has a unique ID. You can find it when sharing a video through a link, such as https://www.youtube.com/watch?v=--Oh8kd508M. Although this ID system was initially created to prevent unauthorised copying and protect copyright, it also creates a digital footprint that can be valuable for research purposes.

These strategies can be instrumental in helping researchers locate specific content on YouTube, even when it's challenging to access within the platform itself.

Custom search engines (CSEs) can also be helpful in seeking a more in-depth analysis beyond what YouTube's built-in search can offer. These CSEs are essentially Google web searches equipped with specific search operators like AND and OR. A comprehensive list of these operators is available here.

Cross-platform searches. Certainly, cross-platform searches are essential to recognise that content published on YouTube can often be found on other platforms. This becomes particularly valuable when content has been removed from YouTube. To learn more about this and gain deeper insights into tracking YouTube content, misinformation, and related topics, the following articles can be useful:

- Tracking YouTube Videos (First Draft News)
- Misinformation in YouTube Recommended Videos (First Draft News)
- Misleading YouTube Videos on Fringe Platforms (First Draft News)

These resources provide valuable information on conducting cross-platform searches and understanding the broader landscape of YouTube content dissemination and related challenges.

5. OTHER TOOLS TO RESEARCH ON YOUTUBE

Expanding your research on YouTube beyond videos, it's also possible to search for channels, investigate comments or conduct whole Open Source Intelligence (OSINT) investigations. Here are some valuable resources and tools to aid in this task.

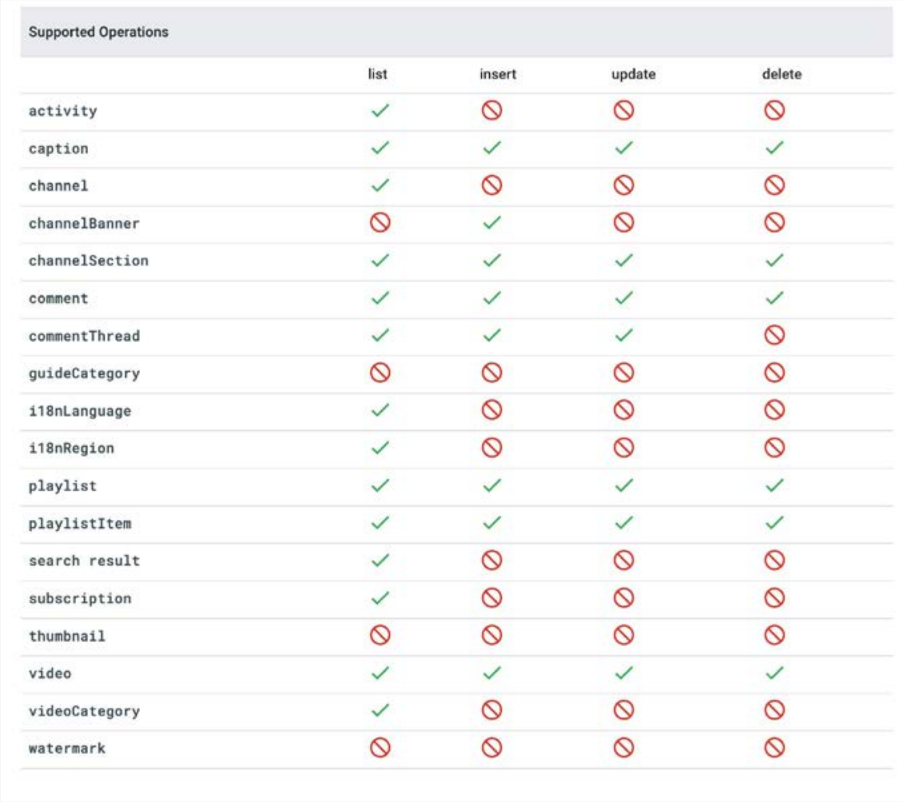
- To download videos: Youtube-dl is one of the most powerful command line tools enabling users to download videos and metadata in automated way.
- To search for YouTube channels by geolocation: YouTube Geolocation Finder.

- To search for comments there are several comment finder tools.
  - YouTube Comment Search Tool
  - Discovering Valuable Conversations: A Guide to YouTube Comment Search Tools
- Resources to conduct Open Source Intelligence (OSINT) on YouTube:
  - YouTube Search Tool by Aware Online
  - OSINT Essentials
  - Digital Methods Initiative YouTube Data Tools
  - YouTube Scraper – tools to scraping data

These resources and tools are valuable for conducting in-depth investigations, geolocation searches, and OSINT on YouTube, helping researchers extract critical insights and information beyond the video content itself.

6. DATA ACQUISITION: YOUTUBE FOR RESEARCHERS

YouTube offers a program for academic researchers that provides access to a scaled and expanded dataset of global video metadata from the entire public YouTube corpus through its Data API. Researchers must request access and describe how they will use the data and YouTube must approve their applications. To learn more about the available data and how to access it, you can explore the YouTube API reference.



Supported Operations

	list	insert	update	delete
activity	✓	⊘	⊘	⊘
caption	✓	✓	✓	✓
channel	✓	⊘	⊘	⊘
channelBanner	⊘	✓	⊘	⊘
channelSection	✓	✓	✓	✓
comment	✓	✓	✓	✓
commentThread	✓	✓	✓	⊘
guideCategory	⊘	⊘	⊘	⊘
i18nLanguage	✓	⊘	⊘	⊘
i18nRegion	✓	⊘	⊘	⊘
playlist	✓	✓	✓	✓
playlistItem	✓	✓	✓	✓
search result	✓	⊘	⊘	⊘
subscription	✓	⊘	⊘	⊘
thumbnail	⊘	⊘	⊘	⊘
video	✓	✓	✓	✓
videoCategory	✓	⊘	⊘	⊘
watermark	⊘	⊘	⊘	⊘

Figure 1. Supported Operations using the YouTube API.

HOW TO FLAG CONTENT ON YOUTUBE AND ITS ENFORCEMENT

1. RELEVANT POLICIES AND STRATEGIES FOR CONTENT MODERATION

Users who witness content that is blatantly illegal or violates any of YouTube's Community Guidelines can report it to the platform. Community guidelines are sorted in six main areas (spam & deceptive practices; sensitive content; violent or dangerous content; regulated goods; educational, documentary, scientific, and artistic (EDSA) content; and misinformation).

Focusing on misinformation, YouTube bans "certain types of misleading or deceptive content with serious risk of egregious harm", including content that "can cause real-world harm, certain types of technically manipulated content, or content interfering with democratic processes." Aside from these general considerations, YouTube relies on specific policies for medical misinformation and a elections misinformation, which shows the relevance given to those topics.

In addition, there are other policies that could play a role when fighting disinformation, such as the guidelines against fake engagement, impersonation, or spam, deceptive practices and scams, technics that are also used in disinformative campaigns. The guidelines against hate speech, harmful or dangerous content, or harassment can also be a reference when the disinformative content has these connotations.

In addition to YouTube's guidelines referring directly or indirectly to mis- and disinformation, the platform has published several documents outlining its strategies to tackle misinformation. The framework "Four Rs" was developed in 2019 as a general framework, consisting on removing violative content; reducing the spread of problematic content and raising and rewarding authoritative sources. In 2022, YouTube admitted new challenges – catching new misinformation before it goes viral; the cross-platform problem and expanding the efforts around the world – in another document and decided to broaden partnerships and fund Poynter's International Fact-Checking Network, as well as to introduce a long term vision for medical misinformation policies. These policies and strategies serve as guidelines for content moderation on YouTube that users can exploit to report content.

2. HOW TO REPORT CONTENT

2.1 Who can report content? Content creators, audiences, and vetted flaggers.

Content creators. A content creator has some moderation competencies when it comes to comments received on their videos, which they can delete, report, or hide from the public using their YouTube Studio account. When reporting the comments, they must specify in which sense the content is harmful or inappropriate and choose between the following options:

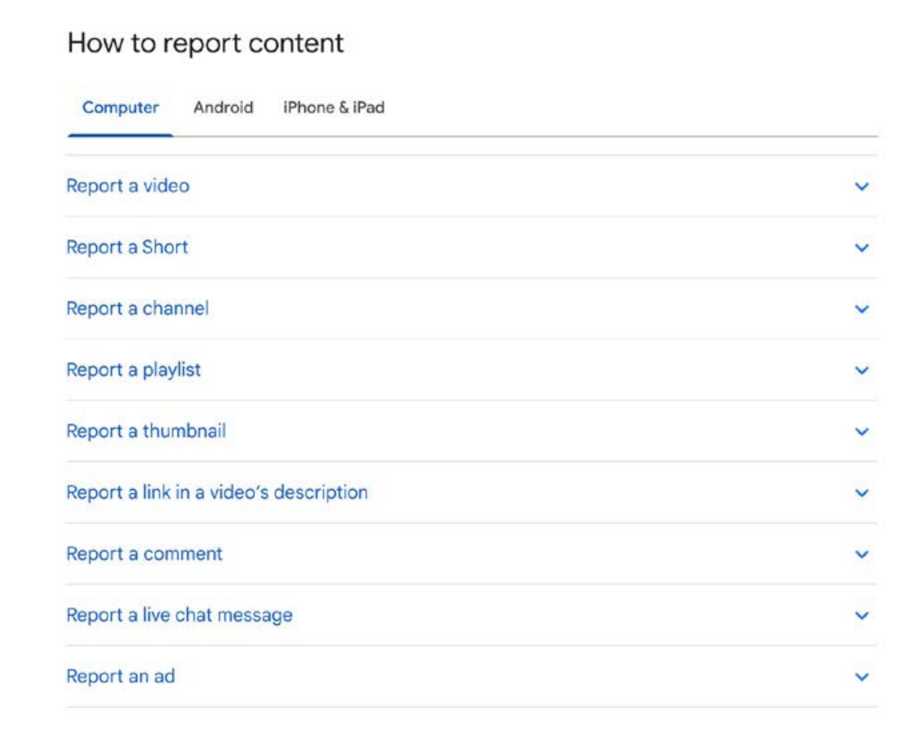
- Unwanted commercial content or spam
- Pornography or sexually explicit material
- Child abuse
- Hate speech or graphic violence
- Harassment or bullying

Users/audience. As a video viewer, content or channels can be reported according to different categories or formats following processes that can have little variations also according to the device used, as we will explain in the next sections.

Vetted flaggers. Apart from the platform's users and community reporting, YouTube relies on disinformation experts to detect disinformation content, such as fact-checkers (limited to some countries) and "priority flaggers". Priority flaggers are experts who participate in a homonymous YouTube program, specifically government agencies and non-governmental organisations (NGOs), that count on direct contact to the platform, prioritised reviews of the content flagged and more visibility into decisions on the content, as well as ongoing discussion and feedback.

2.2 Different types of content and devices

YouTube offers options to report different content on YouTube, specifically: a video, a short, a channel, a playlist, a thumbnail, a link in a video's description, a comment, a live chat message, and an ad. It also explains how to do it on four different devices: computer, Android, iPhone & iPad and from TV devices. The processes are detailed online in this link.



How to report content

Report a video
Report a Short
Report a channel
Report a playlist
Report a thumbnail
Report a link in a video's description
Report a comment
Report a live chat message
Report an ad

Figure 2: Types of content to report on YouTube on different devices.

2.3 Type of violations to be reported

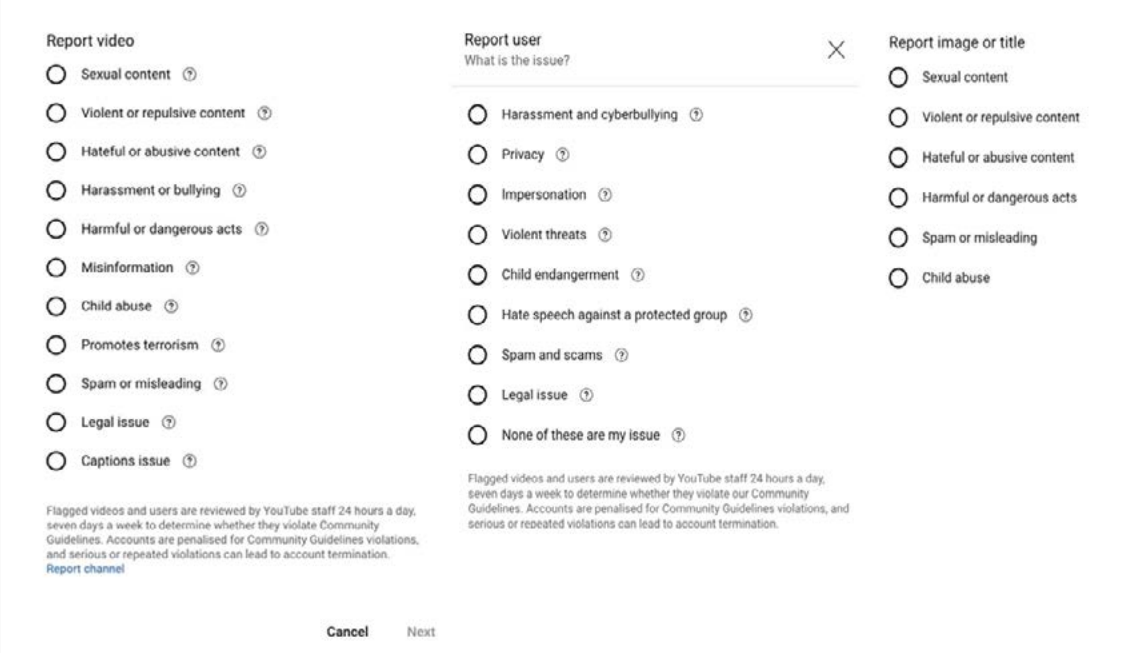
When reporting a piece of content, specifications about the policy violations must be made. YouTube makes a distinction between policy violations and legal violations and provides specific instructions to report each type of content.

- Policy violations refer to content that breaches YouTube's Community guidelines and privacy guidelines. When reporting content, users must specify, for instance, if the violation is related to sexual, violent content, hate speech or harassment, harmful content, misinformation, child abuse, terrorism, or spam, among others. For violations of privacy guidelines, a privacy complaint can be filled.

- Legal violations encompass various legal issues related to content, such as:
    - Copyright: content that infringes on your copyright or intellectual property rights.
    - Trademark: content that infringes on your trademark rights.
    - Counterfeit: content related to counterfeit goods or products.
    - Defamation: content that contains defamatory or false statements about individuals or entities.
    - Stored music policy: content that violates YouTube's policy on using copyrighted music in videos.
    - Other legal complaints: content that violates other legal aspects. In this link there is information about other legal issues to take into account.

The categories that can be referred to when denouncing content can slightly vary depending on the type of reported content. For instance, for reporting playlist there is no request to specify the reasons. Here is an overview of the options offered for the rest of content:

VIDEO	**SHORT**	**CHANNEL/USER**	**THUMBNAIL**	**COMMENT**	**AD**
Sexual content	Sexual content		Sexual content	Pornography or sexually explicit material	Sexual content
Violent or repulsive content	Violent or repulsive content	Violent threats		Violent or repulsive content	Violent or dangerous
Hateful or abusive content	Hateful or abusive content	Hate speech against a protected group	Hateful or abusive content	Hate speech or graphic violence	
Harassment or bullying	Harassment or bullying	Harassment or cyberbullying		Harassment or bullying	
Harmful or dangerous acts	Harmful or dangerous acts		Harmful or dangerous acts	Suicide or self injury	Violent or dangerous
Misinformation	Misinformation			Misinformation	Misleading or scam
Child abuse	Child abuse	Child endangerment	Child abuse	Child abuse	
Promotes terrorism	Promotes terrorism			Promotes terrorism	
Spam or misleading	Spam or misleading	Spam and scams	Spam or misleading	Unwanted commercial content or spam	
Legal issue	Legal issue	Legal issue		Legal issue	Restricted product or service
Captions issues	Captions issues				
		Privacy			
		Impersonation			
		None of these are my issue			Something else



Report video

○ Sexual content

○ Violent or repulsive content

○ Hateful or abusive content

○ Harassment or bullying

○ Harmful or dangerous acts

○ Misinformation

○ Child abuse

○ Promotes terrorism

○ Spam or misleading

○ Legal issue

○ Captions issue

Flagged videos and users are reviewed by YouTube staff 24 hours a day, seven days a week to determine whether they violate Community Guidelines. Accounts are penalised for Community Guidelines violations, and serious or repeated violations can lead to account termination.
Report channel

Cancel Next

Report user
What is the issue?

○ Harassment and cyberbullying

○ Privacy

○ Impersonation

○ Violent threats

○ Child endangerment

○ Hate speech against a protected group

○ Spam and scams

○ Legal issue

○ None of these are my issue

Flagged videos and users are reviewed by YouTube staff 24 hours a day, seven days a week to determine whether they violate our Community Guidelines. Accounts are penalised for Community Guidelines violations, and serious or repeated violations can lead to account termination.

Report image or title

○ Sexual content

○ Violent or repulsive content

○ Hateful or abusive content

○ Harmful or dangerous acts

○ Spam or misleading

○ Child abuse

Figure 3: Different reporting options depending on the type of content to report.

2.4 Follow the report process and appeal the decision

When content is reported, it's not automatically taken down, but reviewed by YouTube. It is, however, possible to follow the process and check the status in your Report history.

In addition, the DSA obliges platforms to make available to users a mechanism for appealing their decisions in this respect. European citizens therefore have this appeal mechanism at their disposal filling this reporting appeal.

2.5 Moderation-related actions: remove, downranking, strike policy, demonetise

Depending on the violation of policies, YouTube content may be removed or downranked, receiving less visibility, while channels may be terminated. At the same time, YouTube has policies that regulate the demonetisation of content that violates its policies.

3. OTHER MECHANISMS ABIDE BY YOUTUBE

Apart from its own policies, YouTube is bound by others that can have a major impact on the fight against disinformation.

3.1 DSA

Far surpassing the required 45 million users, YouTube was designated by the European Commission a very large online platform (VLOP) on 25 April 2023, requiring it to comply with the provisions of European Unions' Digital Services Act (DSA) from August 2023.

Among others, YouTube is bound to additional transparency reporting requirements under Articles 15, 24, and 42 of the mentioned regulation. In response to it, Google recently published its first Transparency Report (a joint report for Google and YouTube) for the period from 28 of August 2023 to 10 September 2023. YouTube is also required to make a systemic risks assessment, including on disinformation, and propose appropriate mitigation measures.

3.2 Strengthened Code of Practice on Disinformation

YouTube is also signatory of the strengthened Code of Practice on Disinformation – future Code of Conduct under DSA. Although a voluntary tool, it would still incentivise the platform to increase its efforts to fight disinformation through some public scrutiny.

RELEVANT CASES ON HOW YOUTUBE IS USED IN DISINFORMATION CAMPAIGNS

This final section is by no means exhaustive but simply wishes to convey with a few examples the extent of YouTube has been used to deceive and mislead in the past.

- (2019) Plasticity.AI discovered a sophisticated YouTube political troll campaign with 8M+ views one year out from the 2020 U.S. presidential election.
- (2020) More than one in four of the most viewed COVID-19 videos on YouTube since March to May 2020, in spoken English, contained misleading or inaccurate information, reveals one study published by BMJ Global Health.
- (2021) This Mozilla Investigation shows how YouTube recommends videos that violate the platform's very own policies such as against misinformation, violent content, hate speech, and scams. The research also finds that people in non-English speaking countries are far more likely to encounter disturbing videos.
- (2022) This study analyses COVID-19 and vitamin D misinformation on YouTube, showing how infodemic has the potential to increase avoidable spread as well as engagement in risky health behaviours.
- (2022) Are YouTube algorithms addicted to state-controlled media? This report of the Crossover Project showed how on the topic of Russia, Belgian users of YouTube were more likely to be widely exposed to state-controlled media, despite the ban of Sputnik and RT France.
- (2022) This study investigates the connections between polarisation and misinformation and measures the flow of misinformation about COVID-19 in the comment sections of four popular YouTube channels.
- (2023) YouTube was one of the platforms used to broadcast deepfake videos of Slovak politicians ahead of the 2023 elections. According to the article, "YouTube didn't respond to reports of the posts within 48 hours from when they were reported via their respective reporting mechanisms".
- (2023) YouTube still ran ads on content that rejects mainstream climate science, even after Google said it wouldn't allow that, according to a research by Climate Action Against Disinformation (CAAD) coalition.