

Anti-trans hate in YouTube comments: Testing larger-scale social media content analysis in 1 language (English)

BRIEFING NOTE

Summary

The prevalence and frequency and forms of anti-trans discrimination and hate speech online are both important and complex to measure. This briefing summarises the steps taken and results of a programmatic test on a sample of 1,628 comments in English from 11 YouTube videos published between July 2023-2025, including by trans creators, anti-trans campaigners, and mainstream media outlets. An extensive list of 168 key words and phrases (informed by a survey of trans content creators from across Europe, conducted in 2025 by TGEU and IJSC) was used to test programmatic identification of hateful and discriminatory speech – alongside a mixed-methods approach including manual coding for comparison. This mixed-methods approach flagged approximately double the number of comments – and about 10% of the comments in total. A majority of the hateful and discriminatory comments detected by both methods were under videos published by mainstream media outlets. In these cases, there seems to be multiple accountability angles: both the platform (YouTube) as well as the publishers (the media outlets) have responsibilities and should be questioned as to why they have allowed such speech to remain online, including under the Digital Services Act (DSA) which came into force in the European Union in early 2024. After a brief introduction and background section, this document details the methods, sources and tools used, and the results obtained, before offering some conclusions about next steps.

Introduction and background

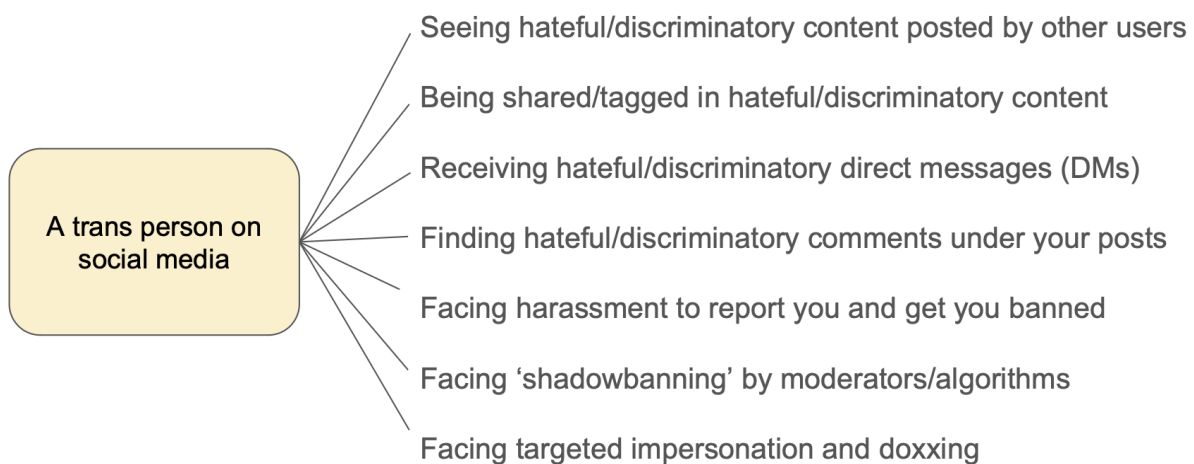
Trans people face hate speech and discrimination online at alarming levels. This was confirmed by the 2025 survey of trans content creators conducted by TGEU and IJSC as well as trans panellists at an online event co-hosted by the organisations in July 2025.¹ That survey as well as the event surfaced key insights including:

¹ [Event: Trans Creators Redefine Online Power, Joy, and Resistance.](#)

- **Misgendering, deadnaming, and denial of existence** were the most reported forms of hate speech.
- **Death threats, rape threats, and targeted reporting** were disturbingly common across platforms including Instagram, TikTok, and Twitter/X.
- **Hate speech wasn't just limited to "trans content"** – even mundane videos (e.g., walking on the street) drew attacks.
- **Platforms fail to respond appropriately to reported abuse**, with some creators even penalised for defending themselves.

Experiences and insights shared by survey respondents and event participants also reflected how hate and discriminatory speech can take a variety of forms (including many different terms and expressions) and can appear in a range of ways (including in comment threads and direct messages as well as in new content that's posted). What was in common was how trans people have been left to deal with hostile climates on their own. In some cases, they have left social media platforms altogether.

Examples of ways hateful/discriminatory speech is experienced on social media



Approaches to tracking transphobic hate speech

There is a growing body of academic research into hate speech online – using a range of methods including computational analyses, lexicon-based filtering, machine learning and natural language models – including some work specifically focused on transphobic speech.² Reviewing 10 studies published since 2017, they have notable commonalities including the selection of focuses (platforms, time periods, keywords). Additionally, even when studies deployed (advanced) computational methods, they still included elements of manual coding/annotations before and/or after programmatic work i.e. a person had to carefully review the inputs and outputs. They were also all upfront about limitations with these methods and did not claim to be comprehensive or conclusive – providing indications instead. The most media-relevant and/or impact-oriented of these studies a) focused on platform accountability for hate speech and/or b) honed in on key moments such as relevant legislative debates or events.

One of the key limitations in conducting such research of hate speech online is access to data because of platform decisions – ex. API restrictions placed by Twitter/X, which is otherwise unique among social media platforms for not allowing users to delete comments/replies under their posts. However, other platforms allow for creators/publishers to delete comments under material they have posted. This means that while trans content creators may see hateful/discriminatory speech under their posts (and be affected by it) they may be able to delete it. It also means that when other users leave such comments online, they have chosen to do so (and the platforms have not required them to do otherwise). From a research perspective, we cannot measure the frequency and prevalence of speech that we cannot see (comments that have been deleted). However, from an impact-oriented perspective, comments suggest two accountability angles: the platforms as well as the publishers of the content under which comments appear, if they have chosen to keep hateful/discriminatory speech up.

This is why we chose to focus our programmatic test on YouTube comments – because they were possible to programmatically extract and because of these dual targets for accountability. The methods used for this test are summarised step-by-step in the section below.

² See separate Literature Review document for an indication of the similar studies that have been conducted to monitor hate speech on the basis of SOGIESC..

Methods

1. Sample Selection

Eleven English-language YouTube videos published within the last three years were selected to give us a sample from a range of different European content creators/publishers including trans individuals; anti-trans campaigners; and mainstream media. They each had titles clearly referencing their coverage of trans rights or related issues in Ireland, Malta, Sweden, Brussels and Berlin as well as across Europe and in the UK.³

Screenshots of the 11 YouTube videos in our sample

Irish Trans and Intersex Pride march.
 COPYPOC
 568 subscribers
 1.1K views 1 month ago

Visiting Malta with Windy - Transgender Travel Vlog
 Windy in Paris
 88K subscribers
 15K views 1 year ago #transgender #transgendervlog #windyinparis

Things I didn't Expect HRT to Change | mtf transgender
 Mathilda Hogberg
 180K subscribers
 24K views 3 weeks ago

The YOUNG LGBs fighting trans ideology in Ireland | Not All Gays Interview
 mrmanno
 34.3K subscribers
 9.3K views 3 months ago #LGB

How Brussels funds the Trans lobby | Huge exposé | Deep Dives
 MCC Brussels
 30K subscribers
 2.4K views 3 months ago #GenderIdeology #MCCBrussels #EUScandal

Taoiseach Micheál Martin asked about the Irish Government's support for trans people
 The Journal
 38.6K subscribers
 2.6K views 5 months ago

³ The UK-related video was published on YouTube by DW (German).

Sponsored by it.shenyun.com

"Do we feel safe?" Why Trans & Intersex Pride is more important than ever | Ireland AM

Virgin Media Television
53.4K subscribers

Subscribe 87

4.4K views 2 years ago #Ireland #IrelandAM #VMTV

Sponsored by amazon.it

How Berlin became a hub for trans people | DW History and Culture

DW History and Culture
303K subscribers

Subscribe 493

21K views 3 years ago #lgbtqia #transgender #cscd

Sponsored by amazon.it

UK Supreme Court rules definition of a 'woman' is based on biological sex at birth | DW News

DW News
5.99M subscribers

Subscribe 1.1K

DW is a German public broadcast service. [Wikipedia](#)

48K views 4 months ago #transgender #women #uk

Sponsored by reddit.com

Young and trans: How does access to gender-affirming care for minors differ across the EU?

euronews
2.48M subscribers

Join Subscribe

32

Euronews is funded in whole or in part by the European Union. [Wikipedia](#)

1.9K views 2 years ago #World

What a new pope could mean for trans Catholics | Focus on Europe

DW News
5.99M subscribers

Subscribe 102

DW is a German public broadcast service. [Wikipedia](#)

5.3K views 4 months ago #PopeFrancis #Trans #CatholicChurch

2. Sample Analysis (View Counts and Publication Dates)

Python was used in Google Colab to analyse these videos and extract the number of views each had received as well as their publication dates.⁴ The exact publication dates of videos on YouTube are not readily visible from their pages (which instead indicate how many months or years ago they went up). The script searched the HTML and extracted JSON data to retrieve these dates.

⁴ [Code used for extracting video publication dates and video counts](#)

3. Comment Extraction

Python was also used to extract all of the comments under the sample videos and store them in a CSV file.⁵ The *youtube-comment-downloader* package was used for downloading Youtube comments without using the Youtube API. Preferences were set to download English-language comments.

4. Comment Coding: Programmatically with Keywords/Phases

Python was again used to programmatically search the extracted comments and identify instances of key words and phrases in five different categories (structural denial; dog whistles; slurs; threats; and neutral).⁶ The script used added columns to our CSV to flag rows containing these key words and phrases and their relevant categories. In total, there were 168 key words and phrases (including common or expected variations of spelling and formatting) fed into this stage. These were identified or informed by the previous survey of trans content creators and the online event held with some of the first responders.

5. Comment Coding: Manually for Comparison

Each of the 1,626 comments extracted was also manually reviewed by two researchers and coded first into simple categories: neutral; supportive; negative; hateful/discriminatory; unclear; and off-topic. The hateful/discriminatory category was then broken up into others to match the programmatic analysis for comparison (structural denial; dog whistles; slurs; threats; and other).

6. Comment Analysis

Spreadsheet pivot tables were produced to analyse the comment data to identify, for example: the total number and proportion of comments in our sample with hateful/discriminatory speech; the break-down of that total number by type of publisher; and key words and phrases that appear most often in such comments.

7. Comparison of Programmatic vs Manual Coding

We compared the results of both coding approaches side-by-side to evaluate what they picked up by category, video and type of publisher. We also checked whether the programmatic approach caught anything it should not have.

⁵ [Code used for extracting comments from all of the videos](#)

⁶ [Code used for identifying comments with key words and phrases](#)

Results

The 11 videos in our sample were published between July 2022-July 2025 and had a total of 136,000 views among them – and 1,626 comments (as of 12 September 2025).

<i>category</i>	<i>views</i>
mainstream media	84,564
trans	39,428
anti-trans	11,784
local media	1,114
Grand Total	136,890

Of these 1,626 comments, only 80 of them contained any of the 168 key words and phrases that were programmatically searched for. More than half of these 80 comments (46) were underneath 5 videos published by mainstream media organisations (DW, Euronews, The Journal-Ireland, and Virgin Media Television-Ireland). DW is a German public broadcaster and Euronews receives funding from the European Union.

<i>category</i>	<i>comments</i>
mainstream media	46
anti-trans	27
trans	4
local media	3
Grand Total	80

Structural denial and slurs were the top categories that came up in this programmatic analysis, with the key words and phrases that appeared most frequently including: “gender ideology”; “disgusting”; “mutilation”; “satanic”; “mentally ill” (and variations). Others that appeared in comments that were picked up programmatically include: “trans agenda”; “pervert” (and variations); “men in dresses”; “freak”; “demon”; “groomer”.

Our mixed-method comparison approach, involving manual review and coding of the comments in our sample, yielded additional findings and nuances. In particular:

- **It identified approximately double the number comments with hateful/discriminatory content** (163 comments and about 10% of the total sample of 1,626), compared to our programmatic approach (80, and 5%)
 - Large gaps like this exist across the different types of publishers, including mainstream media (102 comments vs 46).
 - Six of the eleven videos in our sample had hateful/discriminatory content in at least 10% of their comments, according to our manual review. Meanwhile, our programmatic analysis only identified one video like this (posted by an anti-gender group in Brussels, MCC).
 - MCC was the top video for the percentage of comments with hateful/discriminatory content, across both approaches. Our manual review flagged more than 25% of its comments (programmatic: 13.5%)
 - Apart from this video published by MCC, four of the other five videos which our manual review flagged 10%+ comments were all published by mainstream media outlets (including DW and Euronews)

- **It identified more instances of ‘dog whistles’ in particular** (61 vs 11) and structural denial (90 vs 39). Examples include:
 - “Grown men are determined to fight their way into the little girls’ changing room. We don’t have a society here”
 - “We need to WARN our children to watch out for those type of people”
 - “They are trying to trans out homosexuality in childhood”
 - “No TQ in LGB”

- **It also identified more instances of ‘threats’** (6) while our programmatic approach identified none (0). Examples include:
 - “There is a reason society in the generations gone by would remove the poison permanently from society in years gone by”
 - “Let all the Muslims come in they will trash the hell out of you guys”
 - “Now they need to find out what A CRIME is.”

- **It identified approximately double** the number of comments with hateful/discriminatory content overall and across several key categories of publishers, including mainstream media (102 vs 46).

Table: Comparison of comments flagged by approach and publisher category

Publisher Category	Manual Review Flagged	Python Flagged	Total Comments
Mainstream media	102	46	948
Trans	9	4	223
Anti-trans	41	27	401
Local media	11	3	54
Grand Total	163	80	1626

Table: Comparison of comments flagged by approach, by video

With percentages of hateful/discriminatory comments over 10% highlighted

Publisher Category	Video	Total Comments	Manual Review Flagged	Python Flagged	MR %	Python %
Mainstream media	DW Report UK	562	42	19	7.47%	3.38%
Mainstream media	Euronews Report	136	26	12	19.12%	8.82%
Anti-trans	Irish Young LGBs	327	22	17	6.73%	5.20%
Anti-trans	MCC Brussels	74	19	10	25.68%	13.51%
Mainstream media	Irish TV Trans Safety	88	17	6	19.32%	6.82%
Local media	Irish Trans Pride Parade	54	11	3	20.37%	5.56%
Mainstream media	DW Report Berlin	94	11	8	11.70%	8.51%
Trans	Swedish Trans Experience	119	9	4	7.56%	3.36%

Mainstream media	DW Report Pope	46	6	0	13.04%	0.00%
Mainstream media	Irish Government News	22	0	1	0	4.55%
Trans	Malta Travel	104	0	0	0	0.00%
	Totals	1626	163	80	10.02%	4.92%

Conclusion

This test reflects the high rate of hateful and discriminatory content in YouTube comment threads. The mixed-methods approach (combining programmatic extraction of comments with manual review and coding) flagged about 10% of the total 1,626 comments – with up to 20-25% for some specific videos. Where videos are published by anti-trans campaigners, the videos themselves may also be problematic (they were not studied as part of this research). When they are published by trans people, it is likely that other hateful/discriminatory comments have already been deleted by the creators (and thus our figures likely underestimate what they face). Finally, when they are published by mainstream media organisations on their social media pages this suggests multiple accountability angles for possible next steps: including via the platforms, and under the DSA, as well as via stronger media regulation that focus on the responsibility of legacy media that are platforming hate speech and permitting hate speech in comments to remain on their pages.

The Python scripts produced for this test can be reused for additional and expanded analysis. In this test, they were more effective in extracting comment data than coding it for hateful and discriminatory speech. Manual review and coding of these comments, by comparison, unearthed more cases of hateful and discriminatory speech, reflecting the limitations of programmatic methods on their own (even if it is supplied an extensive list of 168 key words and phrases to search for). Further tests could be done – including using sentiment analysis tools⁷ and pre-trained models to classify comments into categories – although they will likely require customisation for the specific case of anti-trans speech and, again, will likely require or benefit from an element of manual review and coding.

This document has been produced with the financial assistance of the Council of Europe. The views expressed herein can in no way be taken to reflect the official opinion of the Council of Europe.

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

⁷ For instance: [YouTube Sentiment Analyzer by vengotimuktha \(Github\)](#) classifies comments as Positive, Neutral, or Negative, and [YouTube Comment Analyzer by michaelswisa \(Github\)](#) analyses sentiment of comments and tone, but may need customisation to detect specific discriminatory language. [YouTube CommentGuard by HasanCan6241](#) classifies comments into categories like Hate Speech, Sexism, and Racism using a pre-trained model, and could be usefully tested on the example of anti-trans speech.