

GDI

Monetising Hate: How hate speech content generates revenue on YouTube

October 2026

Content Warning

This brief covers sensitive issues and might feature offensive language. GDI does not condone positions and language in the narratives, nor do they reflect the beliefs of GDI, its employees or partners. This document reflects the language and ideas prevalent in online media analysed by GDI. The content may be upsetting to some readers.



Table of Contents

Executive summary.....	1
What needs to happen.....	2
Recommendations.....	2
1. Introduction.....	5
2. Hate speech can pay, for creators and for YouTube.....	5
Engagement-driven algorithms reward harmful content.....	5
YouTube reaches 2.83 billion users and earned \$40 billion from ads in 2025.....	6
3. Monetisation estimates and methodology.....	7
Scope and method.....	7
What creators and YouTube earn.....	8
Geographic distribution of content creators.....	9
Modelled revenue scenarios coming from EU audiences.....	9
4. Who profits from hate speech and what they publish.....	11
Narrative analysis.....	13
5. Case studies: Monetised content that may breach YouTube's rules.....	14
Case study #1: US-based channel earning an estimated \$1.2M a year.....	15
Case study #2: UK podcast channel earning an estimated \$23k a year.....	16
6. Enforcement questions.....	17
YouTube's own rules.....	17
The legal framework in the EU.....	18
Grounds for investigation under the DSA.....	19
7. Recommendations.....	20
8. Conclusion.....	22
References.....	23
Annex 1: Methodology.....	26

Executive summary

Hate speech can be profitable for creators and for YouTube through its share of ad revenue.

A single channel publishing content GDI classified as hate speech can generate an estimated \$2.2 million a year in ad revenue: about \$1.2 million for the creator and \$1 million for YouTube, which keeps 45 cents of every advertising dollar. **Anti-immigrant hate earns the most, as videos on this topic garner the most views.** Most channels do not stop at one target: 14 of the 20 most-viewed channels (out of the 100 selected by GDI for this investigation) attack several groups, typically migrants, Muslims, and Black people together. Where hate pays, more content creators may be attracted into producing it unless action is taken.

YouTube is paying for content that may breach its own rules. YouTube's rules set a high threshold, so much of the hateful content we found does not meet YouTube's definition of hate speech^[1], and as such is permitted and monetised on the platform. But some content goes further. Our two case studies identify content that may breach YouTube's own rules, and may be illegal in some countries, yet was still online and earning money at the time of data collection:

- **A US-based commentary channel.** In one video, the host appears to encourage an arson attack on migrant camps being set on fire. The channel also describes migrants in dehumanising terms, calls for their forced removal, and promotes a clip portraying Islam and Muslim men in derogatory terms. The channel generates about **\$2.2 million a year in ad revenue, of which the creator keeps about \$1.2 million and YouTube about \$1 million.**
- **A UK-based podcast channel.** A guest on the channel portrays transgender people as potential perpetrators of mass shootings. On the same channel, the LGBTQIA+ movement is likened to terrorism and the Quran is portrayed as promoting terrorism. The channel makes about \$23k a year, and YouTube about \$19k.

GDI identified, selected and tracked 100 English-language YouTube channels in the United States and United Kingdom that publish content GDI classified as hate speech. In 2026, the 100 channels are on track to make an estimated **\$13.3 million in ad revenue alone, split \$7.3 million for the channel owner with YouTube taking an estimated \$6 million** from the same videos. Advertisers have limited control over where their adverts appear on YouTube and may be unaware they are funding content that targets marginalised groups, erodes social cohesion, or may be illegal or at odds with their brand values.

Some of this content may breach YouTube's Community Guidelines and could be illegal under German law. In September 2026 we accessed it from Germany, where it was still online and still showing ads. Under the European Union's Digital Services Act, content that is illegal in one member state counts as illegal content, which YouTube is required to act on when notified wherever the creator is based. YouTube's rules also let it penalise creators

who repeatedly target protected groups. At the time of data collection, neither channel appears to have been demonetised.

What needs to happen

In GDI's view on the basis of these findings:

- **The European Commission and the UK Office of Communications (Ofcom)** should use their powers under the EU Digital Services Act and UK Online Safety Act, respectively, to investigate whether YouTube is meeting its obligations in relation to the monetisation of hate speech, starting with the channels in this report.
- **YouTube** should ensure that it has effective procedures to identify content that is unlawful and/or breaks its rules, and to ensure that it does not pay creators for that content, or profit from it itself.

Recommendations

The **European Commission** should use the **Digital Services Act** to demand **YouTube** includes the following in **Article 34 and 35 risk assessment and mitigation reports**:

- **Require proof that YouTube enforces its repeat-offender rules.** YouTube's own rules and three-strike system^[2] let it penalise creators who repeatedly target protected groups.
- **Make hate monetisation part of risk assessments and audits.** YouTube should show that its Partner Program does not pay creators for content that meets its own definition of hate speech. It should also show how it detects such content and removes its monetisation.
- **Make sure hate speech is not a viable career.** Offenders who repeatedly breach YouTube's standards and guidelines should lose every route to earning, including through new channels.
- **Give advertisers placement data.** YouTube should show brands where their ads appear, at channel and video level.
- **Increase efforts to identify language gaps in enforcement, and close any that are found.** YouTube's risk assessment should show it enforces its rules consistently across all EU languages.

Ofcom should use the **Online Safety Act** to:

- **Check that YouTube's risk assessment covers monetisation.** Under sections 9 and 10 (which require platforms to assess the risk of illegal content on their service and to take proportionate steps to tackle it), YouTube should assess and if necessary reduce the risk that its Partner Program pays out on illegal hate content.
- **Hold YouTube to its own rules.** As a Category 1 service (the largest, highest-reach platforms that face extra duties under the Online Safety Act), YouTube must apply its repeat-offender penalties consistently.
- **Ask about these channels now.** Ofcom should use its information powers to ask whether the UK channels in this report have been reviewed against YouTube's rules.

- **Use transparency reports to make monetisation visible.** The first reports, expected in 2027, should show how many UK channels are demonetised or terminated for hate speech.

YouTube should:

- **Stop paying repeat offenders.** When its repeat-offender clause applies, monetisation should go. Where violations continue, YouTube should use the full penalties its rules allow, including channel termination.
- **Not profit from illegal content.** Content found illegal under national law should be removed or blocked where it is illegal, and demonetised everywhere. Neither YouTube nor the creator should keep the revenue it earned. YouTube should refund it to the advertisers whose ads ran on it.
- **Enforce penalties consistently, and publish how.** It should report how often penalties are applied, and give clear examples of content that does and does not break its hate speech rules.
- **Show advertisers where their ads run.** Channel- and video-level placement reports should be standard, with the option to exclude hateful channels.

Full recommendations are set out in section 7.

What does GDI mean by “hate speech”?

There is no single agreed definition of “hate speech”. In this report, the term can be used in three different senses:

- **GDI’s definition.** GDI uses the definition of Hate Speech and Bigotry set out in Table A1 (Annex 1), which is anchored in internationally recognised human rights standards, including the Council of Europe’s definition. GDI applied this definition to classify videos, both through automated labelling across the 100 channels and through analysts’ manual review of 200 videos.
- **YouTube’s definition.** YouTube’s Community Guidelines cover similar ground, but YouTube takes a more permissive view of where the line falls. Much of the content GDI classifies as hate speech is permitted on YouTube and can be monetised. Where content does appear to fall within YouTube’s own definition, in GDI’s view YouTube does not always enforce its rules rigorously or consistently.
- **Illegal hate speech.** What is illegal is set by national law and varies between countries, for example incitement to hatred under §130 of the German Criminal Code (see Section 5). Content can be lawful and still breach YouTube’s rules, or the other way round.

Unless stated otherwise, references in this report to “hate speech” are to hate speech as defined by GDI. The terms “Hate Speech and Bigotry”, “hateful content” and “hate-related content” have the same meaning. Where we refer to YouTube’s definition, we say so (for example, “YouTube’s hate speech policy”); where we refer to the law, we say “illegal hate speech”. Content that meets GDI’s definition does not necessarily breach YouTube’s rules or the law.

Revenue estimations

Every figure mentioned in this report in relation to revenues is to be considered an estimate and may not reflect actual values. Only the parties involved in the transaction, namely YouTube and channel creators, are privy to that information.

GDI produced these estimates using its monetisation methodology focused exclusively on YouTube advertising revenue. They exclude memberships, merchandise, sponsorships, and other monetisation options; hate speech’s true profitability for creators will be significantly higher. References in this report to possible breaches of YouTube’s Community Guidelines or of national law reflect GDI’s view that there are grounds for investigation. They are not findings asserting that any breach has occurred.

1. Introduction

Why estimating monetisation matters. Hate speech creates an unsafe environment for marginalised groups, exposing them to psychological harm and, in some cases, physical violence. It erodes social cohesion and can drive targeted groups out of public life altogether. Women, for example, increasingly cite online abuse as a reason for leaving, or never entering, public office. While hate speech earns money for creators and platforms, there is a financial incentive for it to continue. The monetisation programmes that sustain adversarial content run under opaque, shifting rules, and debates over enforcement go ahead with little sense of how much money is involved. Without credible, independent evidence of how much revenue harmful content generates for creators and for platforms, regulators and policymakers cannot design effective interventions.

Why is it hard. The data policymakers need is not public. While YouTube publishes its revenue split with creators, its public transparency reports focus heavily on content removal (deleted videos/channels) rather than publishing separate, specific data sets detailing how many videos were allowed to stay online but had their monetisation stripped away under their hate speech policy.

How we overcame it. GDI built a methodology using a panel of large language models that estimates how engagement on these channels turns into revenue for both the content producer and the platform, and which narratives that revenue funds. In this report, we estimate how much advertising revenue English-language YouTube channels based in the United States, the United Kingdom, Canada, and the European Union earn from hate speech and bigotry. The figures come from a single data collection run in September 2026, covering 4 January to 27 August 2026. Our unit of analysis is the channel and, within each channel, the individual video. The analysis works at three levels, from broadest to most specific:

1. a revenue estimate across all 100 channels in our sample and 19,297 videos;
2. a manual review of 200 videos from the 20 most-viewed channels (out of the 100 in our sample): 10 in the US and 10 in the UK;
3. two of those channels, examined in full as case studies.

Full methodology, including defined terms used in this report, is set out in Annex 1.

2. Hate speech can pay, for creators and for YouTube

Engagement-driven algorithms reward harmful content

Hate speech is omnipresent on social media platforms. This is well documented and has led to convictions across jurisdictions.^[3] Many reasons may incentivise actors to spread hate speech online, ranging from ideological to financial. Social media algorithms often prioritise content that maximises user engagement and time spent on platforms. However, engagement-based models for profit can leave room for creators and platforms to benefit financially from disseminating harmful content. This can create a self-perpetuating cycle in

which hate speech and bigotry content is leveraged by creators, and platforms profit from content that may, in some instances, violate their own moderation policies and national or supranational legislation.

YouTube reaches 2.83 billion users and earned \$40 billion from ads in 2025

In this report, English-language content has been reviewed as it remains one of the most used languages online. New automated dubbing and subtitling tools mean consumption of English-language content in non-English-speaking countries will likely increase. This first study looks at YouTube for three reasons: first, it has a huge user base with an overall global audience of 2.83 billion users (March 2026).^[4] Second, those users spend a lot of time on the platform: in the US, YouTube accounts for 13.8% of all time spent watching TV, more than any other media company (June 2026).^[5] Third, that makes it a major player in the advertising market with a global ad revenue of \$40.36 billion in 2025 alone.^[6]

How YouTube pays creators

Creators earn money on YouTube through its *YouTube Partner Program* (YPP). Joining is not automatic: channels must reach minimum thresholds for subscribers and for views or watch hours, then apply and be approved.

The main source of income is advertising. Creators earn from ads shown before, during, or alongside their videos, and from a share of YouTube Premium subscription fees when Premium members watch their content.

For long-form videos, YouTube keeps 45 cents of every dollar advertisers pay, and creators receive 55 cents.

For Shorts, the split is reversed: YouTube keeps 55 cents of every dollar allocated to creators, and the creator receives 45 cents.

Creators can also earn through brand sponsorships arranged via YouTube, merchandise and affiliate sales, paid channel memberships, and one-off donations from viewers. The estimates in this report cover *YouTube Ads & Premium* revenue only, so they leave out these other sources of income.

In August 2026, YouTube announced changes to the programme from February 2027, including stricter eligibility criteria and a new revenue split for Shorts. The changes are reported to favour established channels over new creators.

3. Monetisation estimates and methodology

Scope and method

This report presents estimates of the advertising revenue generated by 100 English-language YouTube channels based in the US, Canada, the UK, and EU member states. Channels were selected by matching hate-related keywords against video titles and channel descriptions, and video transcripts were subsequently classified via natural language processing to identify adversarial narratives. All figures are drawn from a single collection run in September 2026 and cover 4 January to 27 August 2026 inclusive [Figure 1]. GDI identified the channels through its own research and trusted third-party sources, and reviewed each one against its inclusion criteria using automated and manual methods. External experts from universities and civil society organisations reviewed the methodology, which is set out in Annex 1.

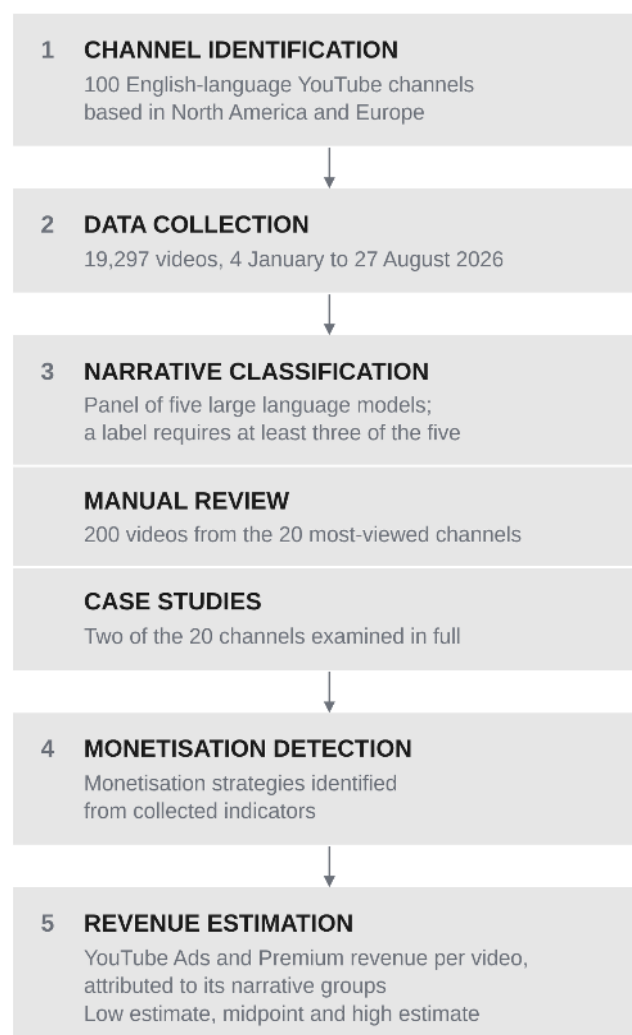


Figure 1. *Methodology overview.* 19,297 videos from 100 English-language YouTube channels (Jan to Aug 2026) labelled by a panel of five language models, checked against a manual review of 200 videos, and used to estimate ad revenue per video as a low, midpoint, and high range.

What creators and YouTube earn

An estimated \$5.5 million to \$46.1 million in advertising revenue, with a midpoint of **\$25.8 million**, was generated through YouTube Ads and Premium by the channels in our sample from January to August 2026.

Several factors influence the cost of running advertising on YouTube, and therefore the revenue generated by both the platform and creators, including audience location, content category, ad format, video length, time of year, and more. To account for such variability, our estimates are expressed using ranged values based on *conservative* and *aggressive* sets of variables. For transparency, we show the minimum, maximum, and midpoint of the resulting range; the figures quoted in this report use the midpoint unless stated otherwise. Revenue is estimated as:

$$\text{Revenue} = \text{Impressions} \times \text{CPM} \div 1000$$

Impressions are calculated by multiplying the number of views by monetisable ratios that span from 35% (conservative) to 85% (aggressive). Cost per mille (CPM) was derived from open-source industry reports and marketing research, and sanity-checked against self-reported numbers. The potential values range from \$1 to \$10 (conservative) and from \$5 to \$32 (aggressive). From this value, the revenue of creators and YouTube is derived according to the platform's own monetisation policies (45% retained by YouTube, 55% distributed to creators).

Videos labelled for Hate Speech and Bigotry content account for 63% (12k videos) of the sample, and generated an estimated \$1.8 million to \$15.3 million of the total advertising revenue, with a midpoint value of \$8.6 million [Figure 2].

An estimated \$2.5 million to \$20.7 million, with a midpoint of **\$11.6 million**, **was retained by YouTube**. An estimated \$3 million to \$25.4 million, with a midpoint of **\$14.2 million**, **was distributed to the channels**.

■ HATE SPEECH □ OTHER CONTENT

TOTAL AD REVENUE



Figure 2. Estimated advertising revenue of the 100 channels, January to August 2026, by content category and by recipient (US\$ million). Bars show the midpoint estimate, which is the mean of the low and high estimates.

Geographic distribution of content creators

Consistent with marketing trends observed over the past few years, the US market still holds a prominent position in YouTube’s global audience share, ranking second only to India^[7] in absolute number of users and first in ad-related revenues.^[8] This asymmetry is reflected in our sample as well. Despite a fairly balanced distribution of channels between the US (51%), the UK (36%), EU member states (9%), and Canada (4%), the US alone accounts for 59% of the total number of videos in our sample and is responsible for 81% of the estimated revenues. Creators based in the UK account for 32% of the total number of videos, while only producing 13% of the revenues. Creators based in EU member states produce 4% of the total number of videos and account for 3% of the estimated revenues.

Such asymmetry can be attributed to several factors, including:

- a more sizable and established audience in the US, by virtue of home advantage;
- a higher number of native English speakers in the North American and UK markets compared to the European Union;
- higher cost per mille (CPM) and revenue per mille (RPM) rates in the US compared to the UK and EU markets, leading to greater payouts.

Modelled revenue scenarios coming from EU audiences

YouTube does not disclose public information related to audience demographics, and its quarterly/annual financial disclosures report numbers by macroeconomic region (e.g. EMEA), making it virtually impossible to provide estimates on how much non-EU content is consumed by an EU audience. Nonetheless, channels in the data sample were checked for

ad-specific indicators on sampled videos, confirming that EU users were presented with advertisements at the time the data was acquired (see Annex 1).

EU users frequently consume US-produced content because of high English literacy rates and cultural overlap. Streaming services based in the US reach up to **75% of the EU audience** on a daily basis,^[9] with YouTube outperforming the top national streaming service in several countries,^[10] and US-specific keywords often feature among YouTube trending topics in both the UK and EU member states.

With approximately 460 million monthly active users,^[11] EU member states are particularly of interest to US creators. While CPM and RPM values (and therefore potential revenues) perform slightly worse on average compared to Anglophone markets,^[12] its user base is 78% larger than the US's 259 million.^[13] With the advent of AI-driven features such as auto-dubbing and auto-transcription, language barriers are increasingly losing relevance, representing an additional financial incentive for both adversarial and non-adversarial actors to intercept such an audience.

YouTube does not disclose where a channel's viewers are located, so the EU share of this revenue cannot be measured directly. To illustrate the scale, Figure 3 applies a range of assumptions to display the potential revenue generated by hateful content (i.e. content that in GDI's view should be regarded as hate speech, even if YouTube does not regard it as such) served to an EU audience, using arbitrary segments set at 5%, 10%, and 15%. Even if applied conservatively, if 10% of these channels' traffic came from the EU, that would equate to an estimated \$2.6 million in advertising revenue, of which **\$0.9 million is associated with hate speech** from content accessible to users based in the EU and potentially illegal under member states' laws.

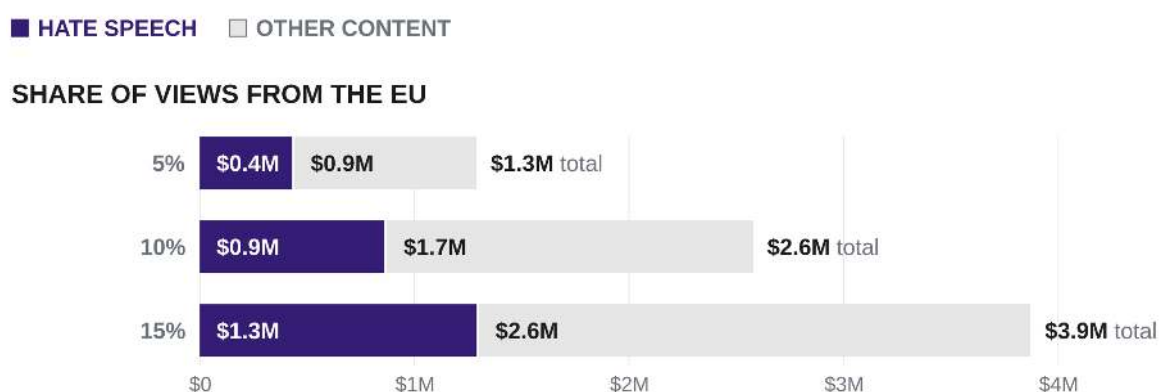


Figure 3. Modelled advertising revenue from EU audiences of the 100 channels, January to August 2026, under three scenarios for the share of views from the EU: 5% (low), 10% (mid), and 15% (high), applied to the midpoint estimate of total revenue, US\$25.8 million.

YouTube's lack of transparency on ad-related demographics poses a serious challenge for institutional agencies and watchdog organisations involved in monitoring, reporting, or enforcing existing laws and regulations. Without granular, country-level data, regulators are unable to reliably audit platform algorithms, assess market dominance, or verify compliance with legal frameworks such as the Audiovisual Media Services Directive (AVMSD), Digital Services Act (DSA), and Online Safety Act (OSA). As a result, regulators and researchers are increasingly forced to rely on third-party estimates, digital trace data, and independent research techniques to map the true extent of US influence and platform penetration across EU digital spaces.

4. Who profits from hate speech and what they publish

The 20 most-viewed channels (out of the 100 selected by GDI for this investigation), 10 based in the US and 10 based in the UK, are mostly online shows commenting on current affairs. Revenue generated content classified by GDI as hate-related is estimated at a midpoint of \$4.2 million, of which \$2.3 million distributed to creators and \$1.9 million retained by YouTube [Table 1].

The highest monthly earner from hate-related content is estimated to generate \$54k out of the channel's total monthly revenue of \$101k (53%) by spreading content featuring Islamophobic, anti-immigration, and anti-Black narratives. Similarly, the second-highest earner is estimated to generate \$38k out of a total monthly revenue of \$103k (37%) with content featuring anti-LGBTQIA+ discourse and white supremacy rhetoric. By contrast, the third-highest displays a significantly lower volume of hate-related revenue: 34k out of a monthly total of \$472k (7%), with most of its revenue coming from content not classified by GDI as hate-related.

	Subscribers	Views (2026)	Creator monthly revenue estimates (total / hate)	YouTube monthly revenue estimates (total / hate)	Description
1.	5M - 10M	250M - 500M	\$101k / \$54k	\$82k / \$44k	Political commentator and online show host based in the US. Nine of the top ten videos with the most interactions featured Islamophobic and anti-immigration rhetoric, as well as anti-Black discourse.
2.	1M - 5M	100M - 250M	\$103k / \$38k	\$84k / \$31k	Online show host based in the US. Five of the top ten videos with the most interactions contained hate speech, including anti-LGBTQIA+ rhetoric and white supremacist narratives.
3.	1M - 5M	500M - 750M	\$472k / \$34k	\$386k / \$28k	Online show host based in the US focusing on finance. The top ten videos for this creator did not contain adversarial narratives.
4.	5M - 10M	100M - 250M	\$64k / \$28k	\$52k / \$23k	Political commentator and online show host based in the US. Three of the top ten videos included hate speech against the LGBTQIA+ community.
5.	1M - 5M	100M - 250M	\$44k / \$20k	\$36k / \$17k	Online show host based in the US. One of the top ten videos was labelled as spreading anti-LGBTQIA+ and anti-immigration narratives.
6.	1M - 5M	100M - 250M	\$40k / \$20k	\$33k / \$17k	Online show based in the US. Four out of the top ten videos contained hateful and bigoted speech. Videos disseminated anti-Black, anti-LGBTQIA+, and anti-feminist/misogynistic rhetoric.
7.	1M - 5M	50M - 100M	\$24k / \$15k	\$20k / \$12k	Creator based in the US. Seven out of the top ten videos featured hate speech, including Islamophobic and anti-immigrant rhetoric.
8.	<1M	100M - 250M	\$40k / \$15k	\$33k / \$12k	Online show based in the UK. Five of the top ten videos featured hateful speech, including anti-immigrant and Islamophobic discourse. Examples include criminalising immigrants, portraying the UK as overtaken by the Muslim community, and targeting a politician with Islamophobic rhetoric.
9.	5M - 10M	250M - 500M	\$66k / \$14k	\$54k / \$12k	Online show host based in the US. Among the top ten videos, one was labelled as Hate Speech and Bigotry. It promoted the Great Replacement conspiracy theory.
10.	1M - 5M	100M - 250M	\$38k / \$13k	\$31k / \$11k	Online show based in the UK. Four of the top ten videos were labelled as containing hate speech, including anti-immigration rhetoric.
11.	1M - 5M	100M - 250M	\$17k / \$11k	\$14k / \$9k	Content creator based in the UK. All top ten videos featured highly adversarial white supremacist and anti-immigration rhetoric, portraying people of colour as uncivilised.
12.	<1M	50M - 100M	\$20k / \$9k	\$16k / \$8k	Content creator based in the UK. None of the top ten videos were labelled as containing hate speech, although one video contained adversarial rhetoric targeting the Israeli population.
13.	1M - 5M	10M - 50M	\$6.3k / \$3k	\$5k / \$3k	Online show. Seven of the top ten videos of this channel spread anti-immigrant, Islamophobic, and white supremacist discourse. In multiple instances, immigrants were described as a threat to society.
14.	1M - 5M	10M - 50M	\$5.3k / \$3k	\$4k / \$2k	Online show host based in the US. Among the top ten videos, five were labelled as spreading hateful speech and bigotry. Various videos featured multiple forms of hateful speech, including anti-immigrant and anti-feminist/misogynistic discourse.
15.	1M - 5M	10M - 50M	\$5.1k / \$2k	\$4k / \$2k	Online show based in the US. Seven out of the top ten videos were classified as spreading hateful and bigoted speech. This includes disseminating anti-immigrant, Islamophobic, anti-Black, anti-Asian, anti-LGBTQIA+, anti-feminist, and misogynistic rhetoric.
16.	1M - 5M	50M - 100M	\$5k / \$2k	\$4k / \$1k	Content creator based in the UK. Although the top ten videos were mostly deemed non-adversarial, the channel platformed creators using anti-LGBTQIA+, anti-Black, Islamophobic, and misogynistic rhetoric.
17.	<1M	10M - 50M	\$13k / \$1k	\$11k / \$1k	Online show host based in the UK. None of the top ten videos contained hateful speech, instead featuring global conspiracy theories and geopolitical disinformation narratives.
18.	5M - 10M	100M - 250M	\$4.7k / <\$1k	\$4k / <\$1k	Content creator. One of the top ten videos featured anti-feminist/misogynistic discourse, including gender essentialist rhetoric often promoted by various factions of the manosphere.
19.	1M - 5M	50M - 100M	\$2.5k / <\$1k	\$2k / <\$1k	Content creator based in the UK. Four out of the top ten videos were labelled as containing hate speech, including anti-LGBTQIA+ and anti-feminism.
20.	<1M	10M - 50M	\$1.8k / <\$1k	\$1.5k / <\$1k	Online show based in the UK. Five of the top ten videos featured hateful speech, including anti-immigrant, Islamophobic, anti-feminist and misogynistic, anti-LGBTQIA+ and antisemitic rhetoric.

Table 1. The 20 most-viewed English-language channels in the sample, by estimated monthly revenue from hateful content, January to August 2026.
Estimated monthly averages over 8 months.

Narrative analysis

GDI found that 79 out of 200 manually reviewed videos in our sample disseminated content GDI would regard as hate or bigoted speech, produced by 17 of the 20 channels reviewed (85%) [Table 2]. While most channels featured hate speech in at least one of their top ten videos of 2026, only a few featured it in the majority of their videos.

Three channels did not feature hate speech in their ten most-engaged videos, despite hateful content accounting for 14%, 33%, and 56% of all the videos published between January and August 2026, respectively. This suggests that adversarial content may not always feature prominently in a content creator's average output, increasing the difficulty to identify hate speech.

Type of hate speech	Count
Overall Hate Speech and Bigotry	79
Anti-immigrant	35
Islamophobia	17
Anti-LGBTQIA+	15
White supremacy	11
Anti-feminist/misogyny	9
Anti-Black	7
Anti-Asian	1
Antisemitism	1

Table 2. Videos classified by GDI as hate speech videos by type, among the 200 most-engaged videos from the 20 most-viewed channels in the GDI sample, January to August 2026. 79 videos contained hate speech; a video can carry more than one type.

Within this sample, anti-immigrant and Islamophobic rhetoric regularly appeared together. Various videos framed immigrants and Muslim communities as a threat to Western societies in the US and across Europe, rooted in the Great Replacement conspiracy theory, which claims that white European populations are being purposefully replaced by immigrant populations, directly threatening their national and cultural identity.^[14] More specifically, multiple channels attributed the “erasure” of white European cultural identity and the diminishing prosperity of a country to Black, brown, and/or immigrant communities. Several videos stated there was a supposed “anti-white bias” and unfair treatment of white people by police, frequently citing the case of Henry Nowak, a British student who was fatally stabbed by Vickrum Digwa in December 2025.^[15] Bodycam footage showed Nowak’s arrest despite requiring urgent medical attention, causing outrage among the British public and

leading to violent protests.^[16] This case has been used to propel narratives around two-tier policing in the UK, and accusations that Nowak's death was the result of an "invasion" of immigrants. The content claimed that diversity training in the police had led to institutional bias against white people.

In multiple instances, videos used derogatory language portraying Black, brown, immigrant and Muslim communities as sexual predators and criminals, in order to dehumanise them. Various videos disseminated femonationalist^[17] rhetoric, which overwhelmingly portrays Black, brown, Muslim and immigrant men as inherently violent towards women, and responsible for gender-based violence. This echoes findings from GDI's previous work in other geographies (e.g. Germany).^[18]

Similarly, channels spreading anti-Black narratives reinforced racist stereotypes, portraying Black communities as orchestrating social decay. Anti-Black narratives were almost exclusive to channels based in the US. This rhetoric was most commonly attributed to channels spreading several forms of hate speech, most notably anti-immigrant and Islamophobic narratives. Anti-Black narratives primarily centred on anti-Black tropes, including claims that Black people are overwhelmingly responsible for crime, which is commonly used to dehumanise and portray Black communities as a threat to public safety.

Multiple channels also disseminated anti-LGBTQIA+ discourse, most of it targeting transgender people (12 of 15 videos). This content frequently drew on "gender ideology" rhetoric,^[19] which presents trans identities as an ideological movement, and portrays transgender people as a threat to women's and children's safety, including through claims that they "groom" children. These narratives echo GDI's findings in Germany^[20] and Canada,^[21] illustrating the transnational circulation of hate narratives.

5. Case studies: Monetised content that may breach YouTube's rules

GDI cross-referenced videos in the sample of 200 videos with YouTube's Community Guidelines to identify potential violations. Most of the hate speech and bigotry we found falls below the threshold set by YouTube's rules: it is permitted on the platform, and it is monetised. However, we found it remains challenging to assess potential violations as YouTube's Community Guidelines remain broad. Nonetheless, GDI identified multiple instances of potential Community Guidelines breaches. Despite potentially breaching YouTube's own guidelines, all these videos were still online and showing ads at the time of data collection in September 2026.

From these channels, GDI selected two for closer examination, one based in the US and one based in the UK. GDI is not naming channels and has paraphrased content and omitted identifying details for security reasons and to protect anonymity. The examples below illustrate why GDI considers there are grounds to investigate whether the content breaches YouTube's rules. Some of the following content is disturbing. Full details of GDI findings have been shared with regulators.

Table 3 brings together, for each featured channel, what it earns, what YouTube earns from it, specific content that may breach YouTube’s own rules (with the rule quoted), and its status at GDI’s September 2026 data collection. Together they suggest that YouTube may be paying, and profiting from, creators whose content may breach its own rules: and that this may be a pattern among high-reach, monetised channels rather than a one-off.

Channel	What the content says	YouTube rule possibly breached (quoted)	Status (Sept 2026)	Metrics	
Case study 1: A US based commentary channel; GDI analysts labelled 9 of 10 videos as featuring hate speech (as defined by GDI).	Describes migrants in dehumanising terms, calls for their forced removal, and promotes a clip portraying Islam and Muslim men in derogatory terms	“Praise or glorification of violence against individuals or groups based on their protected group status.” “Use of racial, religious, or other slurs and stereotypes that incite or promote hatred based on protected group status.”	Online Showing ads Accessible in the EU (accessed from Germany)	Creator earnings (monthly)	\$101k
				YouTube earnings (monthly)	\$82k
				Subscribers	5M - 10M
				Views	250M - 500M
Case study 2: A UK based podcast channel; GDI analysts labelled 5 of 10 videos as featuring hate speech (as defined by GDI).	Host and guest portrays transgender people as potential perpetrators of mass shootings. On the same channel, the LGBTQIA+ movement is likened to terrorism and the Quran is portrayed as promoting terrorism	“Conspiratorial claims that individuals or groups are evil, corrupt, or malicious based on their protected group status.” “Praise or glorification of violence against individuals or groups based on their protected group status.”	Online Showing ads Accessible in the EU (accessed from Germany)	Creator earnings (monthly)	\$1.8k plus audiences driven to other platforms and paywalled content
				YouTube earnings (monthly)	\$1.5k
				Subscribers	<1M
				Views	10M - 50M

Table 3. The two case-study channels: content that may breach YouTube’s Community Guidelines, the rule concerned, channel status at data collection in September 2026, and estimated monthly earnings for the creator and YouTube, January to August 2026. Earnings are monthly averages over 8 months.

Case study #1: US-based channel earning an estimated \$1.2M a year

This case study examines the YouTube channel of a US-based political commentator (Table 3). The creator is also active on other social media platforms.

Of the 10 videos analysed from this channel, all were deemed adversarial. Nine pertained to the category Hate Speech and Bigotry, promoting narratives focusing on Islamophobia, anti-immigration and anti-Black sentiment, which sometimes intersect. One such video expresses support for a far-right protest in the UK. In the video, the host portrays the UK as overrun by migrants characterised as criminals and sexual offenders and references the Great Replacement conspiracy theory

Another video discusses unrest in Europe following a large-scale migrant arrival in 2026. In it, the host characterises migrants as invaders who pose a sexual threat to children, encouraged by humanitarian organisations. Responding to footage of locals setting fire to migrant camps, the host makes a remark that could reasonably be read as encouraging real-world violence against migrant populations. The video's AI-generated cover image references historical military conflict between Christians and Muslims, and its title calls for migrants to be forcibly removed.

In another video, the host shares a clip from another creator who describes Muslim men in derogatory terms and portrays Islam as evil. The host then suggests that Islam cannot co-exist with Western values, including freedom and modern-day feminism. This is not an isolated instance of the host repurposing Islamophobic content from other creators, which supports femonationalist arguments with statements advocating the exclusion of Muslim communities.

Given the above examples, in GDI's view these videos likely violate YouTube's Community Guidelines, particularly with reference to the following sections:

- *"Dehumanisation of individuals or groups by calling them subhuman, comparing them to animals, insects, pests, disease, or any other non-human entity based on their protected group status".^[22]*
- *"Praise or glorification of violence against individuals or groups based on their protected group status. Use of racial, religious, or other slurs and stereotypes that incite or promote hatred based on protected group status. This can take the form of speech, text, or imagery promoting these stereotypes or treating them as factual".^[23]*

According to YouTube's Community Guidelines, penalties or content removal may occur if a creator *"repeatedly targets, insults, and abuses a group based on protected group status across multiple uploads, [or] creates content that harms the YouTube ecosystem by persistently inciting hostility against a group with protected group status for personal financial gain."* In GDI's assessment, there are grounds to investigate whether this channel has breached YouTube's Community Guidelines in multiple instances; at the time of data collection, it remained monetised. Despite what appears to be repeated targeting of groups based on immigration status, race, and religion, this channel earned an estimated \$101k a month on average between January and August 2026 (about \$1.2 million a year), while YouTube earned an estimated \$82k a month (about \$1 million a year) from it.

Case study #2: UK podcast channel earning an estimated \$23k a year

The second case study focuses on the YouTube channel of a UK-based podcast. The podcast is also available on other platforms, including behind a paywall. Out of the top ten videos, five were labelled as featuring Hate Speech and Bigotry.

The channel's most-engaged video in our data sample also raised potential concerns. To protect anonymity, GDI has not described it here; details have been shared with YouTube and regulators

In GDI's view, this video likely violates YouTube's Community Guidelines, including under the following:

- *"content that denies or minimizes someone's role as a victim of a well-documented, major violent event";*^[24]
- *"denial or minimization of a well-documented, major violent event or the victimhood of such an event";*^[25]

Additionally, in this account's top ten videos, there were other instances that potentially violate YouTube's Hate Speech Community Guidelines. In one extract, the host and a guest portray transgender people as potential perpetrators of mass shootings, liken the LGBTQIA+ movement to terrorism, and compare LGBTQIA+ people to sexual predators, a narrative often leveraged to vilify the LGBTQIA+ community.

In another extract, the host and a guest spread anti-immigrant and Islamophobic discourse. The guest portrays the Quran as promoting terrorism", while both leverage femonationalist rhetoric to overwhelmingly blame gender-based violence, including sexual assault, on immigrants and the Muslim community. In this extract, the guest also makes remarks that could be read as endorsing violence. GDI assesses that these statements are likely to violate YouTube's Community Guidelines in multiple instances, including the following:

- *"Conspiratorial claims that individuals or groups are evil, corrupt, or malicious based on their protected group status."*
- *"Use of racial, religious, or other slurs and stereotypes that incite or promote hatred based on protected group status."*
- *"Praise or glorification of violence against individuals or groups based on their protected group status."*

In its Community Guidelines regarding monetisation, YouTube mentions the platform may "remove content or issue other penalties when a creator [...] repeatedly targets, insults, and abuses a group based on protected group status across multiple uploads." While this channel has repeatedly targeted various groups based on their protected characteristics, it earned an estimated \$1,8k a month on average between January and August 2026, while YouTube earned an estimated \$1,5k a month from the channel.

6. Enforcement questions

YouTube's own rules

What YouTube's rules say. YouTube's hate speech policy clearly states that "hate speech is not allowed on YouTube".^[26] It bans content that encourages or glorifies violence against protected groups, dehumanises them, or makes conspiratorial claims that they are evil, corrupt, or malicious. It also states that, "in some rare cases", YouTube may remove content or issue other penalties when a creator repeatedly targets, insults, and abuses a group based on protected group status across multiple uploads. Our findings suggest that these

rules may not be consistently enforced among high-reach monetised channels: we found no evidence of action having been taken against the content we reviewed. The wording is part of the problem: penalties for repeated abuse are discretionary, reserved for “rare cases”, and terms like “repeatedly” and monetisation thresholds are undefined. This makes consistent enforcement harder and makes it difficult for advertisers and regulators to hold YouTube to its own standards. Every video cited among the case studies in this report was still online, showing ads, and accessible in the EU when GDI collected data in September 2026. GDI accessed both case-study channels from the EU. This may bring them within the scope of the Digital Services Act, regardless of where the creators are based. Article 14(4) of the DSA requires providers to apply and enforce the restrictions in their terms and conditions in a diligent, objective and proportionate manner, with due regard to the fundamental rights of users. Articles 34 and 35 require very large platforms such as YouTube to assess and mitigate systemic risks of illegal content and harms to fundamental rights.

The legal framework in the EU

The EU approach to illegal hate speech. The DSA does not define illegal hate speech. “Illegal content” is any information that breaches EU law or the law of a member state that complies with EU law (Article 3(h)), so national criminal law decides what counts. At a minimum, every member state must criminalise public incitement to violence or hatred on grounds of race, colour, religion, descent, or national or ethnic origin (Framework Decision 2008/913/JHA), and many go further, for example by covering sexual orientation and gender identity. In 2021 the Commission proposed adding hate speech to the list of “EU crimes” so that common minimum rules could cover more grounds, but this requires unanimity in the Council and the consent of the European Parliament, and it has not been adopted.^[27] Because YouTube offers its service in the EU, these obligations apply to content accessible there, whether or not EU users can be shown to have viewed it. YouTube would only need to restrict access within the EU, not remove the content worldwide.

The DSA does not ban monetising content that is hateful but lawful. However, YouTube must enforce its own terms diligently, objectively and proportionately (Article 14(4)), process notices of illegal content and explain moderation decisions, including demonetisation (Articles 16–17), and, as a very large online platform (VLOP), assess and mitigate systemic risks, including those arising from its advertising systems (Articles 34–35). The Commission, which has exclusive powers to enforce VLOP-specific obligations, would need to establish any breach, but it can require YouTube to hand over data, including on viewer location and revenue (Article 67).

What counts as illegal hate speech in Germany. The two case-study channels were accessed from Germany. The central offence is incitement to hatred (Volksverhetzung) under §130 of the German Criminal Code. It is committed by anyone who, in a way capable of disturbing the public peace, incites hatred against a national, racial, religious or ethnic group, or against a section of the population or an individual because they belong to one;

calls for violent or arbitrary measures against them; or attacks their human dignity by insulting, maliciously maligning or defaming them. German courts have treated groups such as refugees and migrants as “sections of the population”, and the term can also cover other groups such as LGBTQIA+ people. Disseminating such content, including making it publicly available online, is a separate offence under §130(2). Holocaust denial and the approval or glorification of Nazi rule in a way that violates the dignity of its victims are also crimes. Other offences can apply alongside §130, including public incitement to commit crimes (§111), publicly approving certain serious crimes (§140) and insulting religious beliefs in a way that disturbs the public peace (§166). The threshold is high in order to protect free speech: criticism of immigration or religion is protected, but incitement, calls for violence, and attacks on human dignity are not. Since May 2024, the Bundesnetzagentur has been Germany’s Digital Services Coordinator under the DSA.

Possible illegality in Germany. The threshold is deliberately high, and whether a statement meets it depends on its context and on how a court would assess it. Whether any of the content in this report is illegal is a matter for the German authorities to assess. Both channels were accessed from Germany, so German law is the most relevant test, and several passages may meet the elements of §130. Regarding the US-based channel’s apparent encouragement of an arson attack on migrant camps and its calls for migrants’ forced removal, we would encourage the authorities in Germany to consider whether this content amounts to incitement to hatred, and a call for violent measures, against a section of the population defined by origin. The clip the channel endorses, portraying Islam and Muslim men in derogatory terms, may attack the human dignity of a religious group. The UK-based podcast’s episodes, portraying transgender people as potential mass shooters and the Quran as promoting terrorism, may maliciously malign sections of the population. Statements approving or calling for violence may also engage §111 and §140 of the German Criminal Code. Whether creators based outside Germany could be prosecuted does not change YouTube’s position: under the DSA, content illegal in one member state is illegal content, which YouTube must act on within Germany when notified.

Grounds for investigation under the DSA

A wider pattern. Seventeen of the 20 most-viewed channels in the sample published hate speech within 79 of the 200 videos our analysts manually reviewed. Examples include claims that people of colour are uncivilised, Islamophobic attacks on a senior UK politician, and promotion of the Great Replacement conspiracy theory, among others. Across the full sample of 100 channels, our classification models detected hate-related discourse in 63% of videos published between January and August 2026. While both manual and automated approaches present limitations, our findings suggest that potential compliance gaps may be systemic rather than exceptional, which in GDI’s view warrants investigation.

Grounds to investigate. YouTube keeps a share of the ad revenue on every monetised video, so it may profit from creators whose content may breach its own rules. This content was still online and showing ads in September 2026, weeks or months after publication, and accessible in Germany, where some of it may be illegal. This may point to gaps in

YouTube's enforcement and demonetisation systems, and in GDI's view gives grounds to investigate compliance with Articles 14(4), 34 and 35 of the DSA.

Questions about YouTube's assurances to the Commission. In its 2025 risk assessment, YouTube describes a three-strike system for repeat violators of its Community Guidelines.^[28] Yet its independent DSA audit found gaps in enforcement. Two trusted-flagger notices alleging illegal hate speech were not prioritised.^[29] Twenty-four of its 33 in-scope terms were not available in all official EU languages,^[30] and the auditor did not test YouTube's fixes to trusted-flagger handling.^[31] Google also acknowledges that its automated systems may perform less well in some languages.^[32] If there are enforcement gaps in English, they may be wider in other languages used across YouTube's average of 462 million signed-in EU accounts.^[33]

7. Recommendations

The European Commission should:

- **Enforce the repeat-offender clause.** Both case studies cite YouTube's clause allowing penalties for creators who repeatedly target a protected group. Under Article 67, the Commission should ask how often YouTube has applied it to EU-accessible channels, how many channels it has demonetised or terminated under this clause, and why the channels in this report were not. GDI can share these channels confidentially.
- **Assess and audit the monetisation of hate content.** Require YouTube's next risk assessment under Article 34 to examine how the Partner Program does not pay creators for content that meets its own definition of hate speech, and require mitigation under Article 35, such as dedicated monetisation review for repeat offenders. The Article 37 audit should test whether policy-violating content is still monetised and verify that trusted-flagger reports are being effectively prioritised.
- **Make ad placements visible.** Under Article 35, require YouTube to give advertisers channel- and video-level placement reports, so brands can see where their ads ran.
- **Require data by language.** Enforce reporting by EU language under Article 42(2), covering removals, demonetisation and moderation accuracy, and confirm all terms are published in all 24 official languages. The Commission should insist that YouTube's assessment breaks down removals, demonetisation and moderation accuracy by EU language. Where gaps appear, Article 35 should require fixes, such as more language-specific moderation resources. The Commission should also confirm that YouTube publishes its terms in all 24 official languages, as Article 14(6) requires

Ofcom should use the Online Safety Act to:

- **Check whether YouTube's risk assessment covers its business model.** The Act requires YouTube's illegal content risk assessment to consider how its business

model affects risk (s. 9). Ofcom should assess that the assessment covers how the YouTube Partner Program pays out on priority illegal hate content, and require YouTube to reduce that risk under section 10.

- **Hold YouTube to its own rules.** As a Category 1 service, YouTube must apply its terms on suspending and banning users consistently (s. 72). Ofcom should consider whether a repeat-offender clause reserved for “rare cases” supports consistent enforcement. Ofcom’s Terms of Service Guidance, expected in early 2027, should make this clear.
- **Ask about these channels now.** Ofcom should use its information powers to ask YouTube whether the UK channels in this report have been reviewed against its policies.
- **Use the first transparency reports to make monetisation visible.** Category 1 services must publish their first transparency reports in 2027, on information Ofcom chooses (s. 77). Ofcom should require data on how many UK channels are demonetised or blocked in the UK for hate speech, and how often the repeat-offender clause is used.

YouTube should:

- **Act on the channels in this report.** Review the channels identified here against its hate speech policy, and remove or demonetise content that breaches it. Where creators repeatedly target protected groups, apply the penalties its own repeat-offender clause allows.
- **Close the gap between monetisation and enforcement.** Stop paying creators, and taking its own share, on content that breaches its Community Guidelines. Review monetised channels proactively rather than relying on user reports, prioritising channels with large audiences and repeat violations.
- **Publish how it enforces the repeat-offender clause.** Report how often the clause is used, how many channels have lost monetisation or been terminated under it, and the criteria applied, so advertisers, researchers and regulators can hold YouTube to its own rules.
- **Give advertisers full visibility.** Provide channel- and video-level placement reports as standard, and let advertisers exclude channels that publish hate speech, so brands are not unknowingly funding it.
- **Fix trusted-flagger handling and test the fix.** Make sure notices of illegal hate speech by trusted flaggers are prioritised without undue delay, and have the improved process independently tested in the next DSA audit.
- **Enforce consistently across languages.** Publish moderation and demonetisation data by EU language, and make sure all terms are available in all 24 official languages.

8. Conclusion

Countering the dissemination of hate speech on YouTube should not solely be a reactive process solved with user reporting and post hoc moderation; a systemic approach targeting the financial incentive of spreading hate speech is necessary. This report documents that hate speech generates a revenue stream, shared between creators who produce it and the platform that distributes it. Across the 100 channels in our sample, content labelled as hate speech is on track to earn creators an estimated \$7.3 million in 2026, and YouTube an estimated \$6 million. These figures cover advertising alone. The real incentive is larger.

The money is not spread thinly. A small number of channels earn more than \$1 million a year, and the most lucrative videos target migrants, Muslims and Black communities, often together. The case studies show that some of this content likely breaches YouTube's own rules, may be illegal under EU member state law, and was still online and earning money when we collected our data.

Two problems sit behind these findings. The first is the rules themselves. Most of the hate speech we found does not breach the threshold YouTube's Community Guidelines set, and even trained analysts often could not tell where that threshold lies. The second is enforcement. Where content does appear to cross the line, YouTube has the tools to act, including penalties for repeat offenders; nonetheless, YouTube has failed to enforce them. While both issues persist, producing content featuring hate speech remains profitable for both creators and YouTube. These incentives do more than reward the creators already producing hate speech. As it can be highly lucrative for some, it draws new people, often young men, into making harmful content for money. As long as hate pays, it will keep attracting new people to produce it. The findings likely understate the problem, as this report only covers ad revenue; complementary work covering other types of monetisation would be necessary to have a fuller picture of the revenue generated by spreading hate speech on YouTube.

Additionally, this report focuses on English-language content, where YouTube's moderation should be at its strongest. Google itself acknowledges that its automated systems perform less well in other languages. If there are enforcement gaps in English, it is reasonable to ask whether they are wider in the other languages spoken by YouTube's 462 million average number of signed-in EU accounts.

References

1. Google. "Hate speech policy - YouTube Help." (n.d.). <https://support.google.com/youtube/answer/2801939#zippy=%2Cother-types-of-content-that-violates-this-policy>.
2. Google. "Community Guidelines Strike Basics on YouTube - YouTube Help," n.d. <https://support.google.com/youtube/answer/2802032>.
3. Spray, Hannah. "Saskatoon Man Sentenced to House Arrest for Inciting Hatred Against Jewish People." *CBC*, October 2, 2026. <https://www.cbc.ca/news/canada/saskatoon/brandon-taylor-moore-antisemitism-inciting-hatred-sentencing-9.7361359>.
4. Shewale, Rohit. "YouTube Statistics 2026: Users, Demographics, Usage & More." *Resourcera* (blog), March 16, 2026. <https://resourcera.com/data/social/youtube-statistics/>.
5. Honeyshah. "FOX And NBCUniversal Score With FIFA World Cup 2026™ Success in Nielsen's June Gauge Reports." *Nielsen*, August 18, 2026. <https://www.nielsen.com/news-center/2026/fox-and-nbcuniversal-score-with-fifa-world-cup-2026-success-in-nielsens-june-gauge-reports/>.
6. Shewale, Rohit. "YouTube Revenue (2025): Latest Data and Growth Breakdown." *Resourcera* (blog), March 25, 2026. <https://resourcera.com/data/social/youtube-revenue/>.
7. Statista. "YouTube Ad Reach by Country 2026," April 27, 2026. <https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users>.
8. "Alphabet Announces Second Quarter 2026 Results." Press release, July 2026. s206.q4cdn.com/479360582/files/doc_financials/2026/q2/2026q2-alphabet-earnings-release.pdf.
9. The DPP. "US Giants Are Winning Streaming Wars in Europe — but Where Are Local European Services Fighting Back?," DPP <https://www.thedpp.com/news/european-tv-streaming-outlook-2026>.
10. *Ibid.*
11. SocialMediaTransparency.org. "YouTube Statistics & DSA Compliance in the EU" n.d. <https://socialmediatransparency.org/platform/youtube>.
12. Kyriienko, Anastasiia. "In Which Countries Are YouTube CPM Rates the Highest in 2026?" *MilX*, October 21, 2026. <https://milx.app/en/cases/in-what-countries-cpm-are-the-highest>.
13. Statista. "YouTube Ad Reach by Country 2026," April 27, 2026. <https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users>.
14. Institute for Strategic Dialogue. "The 'Great Replacement' - Institute for Strategic Dialogue," January 31, 2026. <https://www.isdglobal.org/isd-explainer/the-great-replacement/>.
15. Bodycam footage showed police responding to the incident, which had initially been reported to police as a racially motivated attack against Digwa. Unaware of Nowak's grave injuries, police placed Nowak under arrest before he lost consciousness and later passed away.
16. White, Marcus, and Curtis Lancaster. "Five Jailed Over Henry Nowak Police Protest in Southampton." *BBC News*, June 10, 2026. <https://www.bbc.co.uk/news/articles/cwye9v0dy3yo>.
17. Mainz, Sophie. "Inclusion to Exclude: How Femonationalism Impacts Policy Preferences." *Political Science Research and Methods*, January 26, 2026, 1–24. <https://doi.org/10.1017/psrm.2025.10068>.
18. Global Disinformation Index. "State of Disinformation in Germany - June 2025." *The Global Disinformation Index*, September 13, 2025. <https://www.disinformationindex.org/research/2025-09-13-state-of-disinformation-germany-june-2025/>.
19. GLAAD. "Gender Ideology: Definition, Meaning, and Origin in Anti-LGBTQ Hate | GLAAD."

- GLAAD | GLAAD Rewrites the Script for LGBTQ Acceptance, June 17, 2026.
<https://glaad.org/gender-ideology-definition-meaning-anti-lgbt-online-hate/>.
20. Global Disinformation Index. "State of Disinformation in Germany - June 2025." *The Global Disinformation Index*, September 13, 2025.
<https://www.disinformationindex.org/research/2025-09-13-state-of-disinformation-germany-june-2025/>.
 21. Global Disinformation Index. "Hate Speech and Bigotry in Canada - August 2025." *The Global Disinformation Index*, September 15, 2025.
<https://www.disinformationindex.org/research/2025-09-15-hate-speech-bigotry-canada-august-2025/>.
 22. Google. "Hate speech policy - YouTube Help." (n.d.).
<https://support.google.com/youtube/answer/2801939#zippy=%2Cother-types-of-content-that-violates-this-policy>.
 23. *Ibid.*
 24. Google. "Harassment & cyberbullying policies - YouTube Help." (n.d.).
<https://support.google.com/youtube/answer/2802268?sjid=15008812601719613090-EU#zippy=%2Cother-types-of-content-that-violate-this-policy>.
 25. Google. "Hate speech policy - YouTube Help." (n.d.).
https://support.google.com/youtube/answer/2801939?hl=en-419&ref_topic=9282436#zippy=%2CEducational-documentary-scientific-and-artistic-content%2Cother-types-of-content-that-violates-this-policy.
 26. "Hate Speech Policy - YouTube Help," Google
<https://support.google.com/youtube/answer/2801939>.
 27. European Commission. "COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime." *EUR-Lex*, December 9, 2021.
<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021DC0777>.
 28. Google. "Report of Systemic Risk Assessments 2025," December 2025.
https://storage.googleapis.com/transparencyreport/report-downloads/dsa-risk-assessment_2025-8-28_2025-8-28_en_v1.pdf.
 29. Google Ireland Limited. "DSA Audit Implementation Report 2025," 2025.
https://storage.googleapis.com/transparencyreport/report-downloads/dsa-audit-g-implementation_2024-6-1_2025-5-31_en_v1.pdf.
 30. *Ibid.*
 31. *Ibid.*
 32. Google. "Report of Systemic Risk Assessments 2025," December 2025.
https://storage.googleapis.com/transparencyreport/report-downloads/dsa-risk-assessment_2025-8-28_2025-8-28_en_v1.pdf.
 33. Google. "Information about Monthly Active Recipients under the Digital Services Act (EU)," August 14, 2026.
https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-24_2026-1-1_2026-6-30_en_v1.pdf.

Annex 1: Methodology

Table A1 defines the terms used in this report.

Definitions

Adversarial narrative	A narrative that meets two or more of these criteria: it is false or misleading; it is financially or ideologically motivated; it is antagonistic towards people and at-risk groups protected under international human rights standards, towards democracy and the institutions needed to fulfil human rights obligations, or towards current scientific consensus; it aims to foster long-term social, political or economic conflict; it poses a risk of real-world harm.
Hate Speech and Bigotry	In this report, “hate speech and bigotry” (or “hate speech”) means any type of expression that incites, promotes, spreads or justifies violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status, such as “race”, colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity or sexual orientation. [This includes expressions of this kind whether stated outright or conveyed through framing, repetition or selective use of evidence.] This definition adopts the definition of hate speech used by the Council of Europe’s Committee of Ministers (Recommendation CM/Rec(2022)16 on combating hate speech), which is grounded in internationally recognised human rights standards.
Channel	A public YouTube account through which content is published.
Actor	An entity assessed or externally attributed as operating, controlling, or systematically coordinating one or more channels. It may be an organisation, a company or media structure, a networked group, or a natural person.
Engagement	Content metrics counting user interactions: views, likes, and comments.
Impression	The number of times an ad is displayed or served to a user.
CPM	<i>Cost Per Mille (thousand)</i> , the cost paid by advertisers to YouTube for every thousand impressions.
RPM	<i>Revenue Per Mille (thousand)</i> , the revenue a publisher earns for every thousand impressions. The publisher is the actor operating a channel enrolled in the YouTube Partner Programme.

Scope, period and levels of analysis

We estimate how much advertising revenue English-language YouTube channels earn from hate speech and bigotry content, looking at channels based in North America and Europe. The figures in this report come from a single collection run in September 2026 and cover 4 January to 27 August 2026. Our unit of analysis is the channel, and within each channel, the individual video. Figure A1 summarises each step.

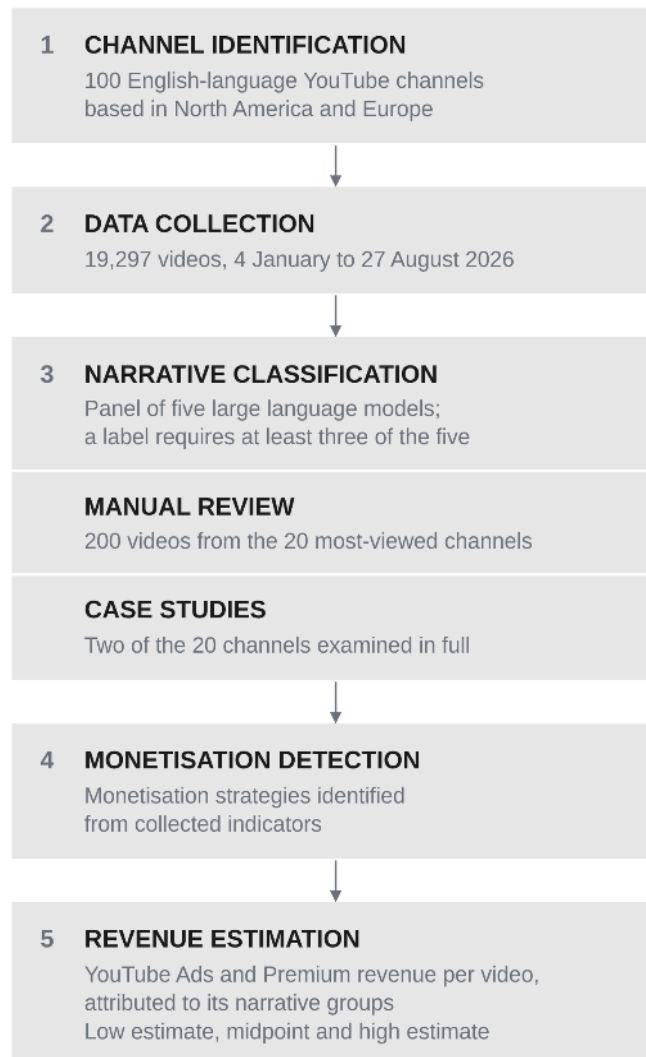


Figure A1. Methodology overview. 19,297 videos from 100 English-language YouTube channels (January to August 2026) were labelled by a panel of five large language models, checked against a manual review of 200 videos, and used to estimate ad revenue per video as a low, midpoint and high range.

This analysis has three levels, from broadest to most specific: (i) a revenue estimate across all 100 identified channels and 19,297 collected videos; (ii) a manual review of 200 videos from the 20 most-viewed channels, 10 from the United States and 10 from the United Kingdom; (iii) two of those channels examined in full as case studies. This annex follows that order: how we built the corpus and classified it, then how we selected and examined the most-viewed channels in depth, then how we built the two case studies, and finally how we detected and measured monetisation across the corpus.

Channel identification and data collection

A channel is added to the list only once analysts confirm that its content meets the adversarial narrative criteria we track (see Table A1). This holds whether the narrative is stated outright or carried through framing, repetition, and selective use of evidence. Hate speech and bigotry, the narrative group examined in this report, is one of the narrative groups GDI tracks.

We find candidates by monitoring public channels for adversarial keywords in their video metadata (titles and descriptions) and by adding those named in external reporting, then extend the list by following the links, references, and shared infrastructure that connect a known channel to others.

This is a purposive, non-probability sample; we do not claim to have identified every channel that earns revenue from hate speech and bigotry content, nor that the sample is typical or representative. What YouTube audiences see is set by algorithmic ranking, so exposure and revenue concentrate in a small number of high-performing channels and videos. We follow that concentration: among channels that meet the adversarial narrative criteria, discovery prioritises those that draw the most engagement and revenue.

From the channel list, we build the dataset. Collection covers public data of the identified channels using third-party services. For each video collected, we keep its engagement metrics (views, likes, and comments), its text (title, description, and transcript where available), and other signals showing how the channel is additionally monetised (e.g. an external shop link or a donation link).

Narrative classification

Each video's title and transcript are classified by a panel of five generative pre-trained large language models from different developers: Gemini 3.5 Flash-Lite (Google), Llama 3.3 70B Instruct (Meta), Ministral 8B (Mistral AI), Nemotron Nano 3 30B (NVIDIA) and gpt-oss-120b (OpenAI). Each model is given the label definitions but no prior training or worked examples for this task (a zero-shot approach). Classification is multi-label: each model assesses every video independently against the definitions in this report, deciding separately for each narrative group whether the video carries it, so a video can carry more than one. A video is labelled as Hate Speech and Bigotry when at least three of the five models identify it as such. When applicable, the panel assigns the subnarrative; for Hate Speech and Bigotry these are anti-feminism and misogyny, anti-immigrant, anti-LGBTQIA+, anti-Black, Islamophobia, antisemitism, anti-Asian, or white supremacy. For every video and every narrative group, we record how many of the five models identified that narrative group: the more models agree, the more confident the panel's label. Each video's estimated revenue is then attributed to the narrative groups the panel labelled it with. We tested the panel against our manual labels for the 196 of the 200 reviewed videos (see below) that have a transcript: the panel labelled 65 of our 78 Hate Speech and Bigotry videos among them as such (83% recall), and 65 of its 88 Hate Speech and Bigotry labels matched ours (74% precision), giving an F1 score of 0.78.

Manual review

Video views are aggregated by channel and channels ranked by their cumulative views. Because advertising revenue scales with views, views serve as a proxy for earnings, so this ranking approximates the most lucrative channels. The 20 most-viewed qualifying channels in the sample, 10 from the US and 10 from the UK, are retained for closer review. From each, we take its ten videos with the highest engagement, 200 in all, and label every one by hand against the definitions above, giving each video its narrative group and subnarrative.

Case studies

Two of these 20 channels are examined in full as case studies. Their videos are read closely for claims and imagery and checked against YouTube's Community Guidelines, and their presence on other platforms (X, Instagram, Spotify, Patreon) and any merchandise, podcast, or subscription revenue is traced from public sources.

Monetisation detection

Channels can earn from several monetisation strategies on YouTube, including ad revenue sharing, merchandise, paid subscriptions, and donations through its YouTube Partner Programme (YPP). We detect which strategies a channel employs by using multiple indicators, e.g. the presence of a donation button, the rendering of pre-video advertising, and performance metrics meeting the eligibility criteria, among others. The presence of one or more indicators can predict whether a channel is currently enrolled in the YPP; however, the specific details of the agreement between YouTube and the creator, as well as its monetisation status, are known only to the parties involved.

Revenue estimation

In this report, we focus on estimating revenues generated by YouTube ad revenue-sharing programme, known as *YouTube Ads and Premium*. Revenue was estimated using the formula:

$$Revenue = Impressions \times CPM \div 1000$$

Where:

- *Impressions* is the number of qualified views, calculated using each video's reported views and multiplied by a monetisable ratio;
- *CPM* (cost per thousand) is the cost paid by advertisers to display ads on YouTube.

The qualified views ratio and CPM values are not disclosed publicly by YouTube, and are visible only to parties involved in the transaction. The values used in our estimation model were derived from external, reputable sources (see below). Moreover, both the qualified views ratio and CPM values are subject to heavy fluctuations, as they depend on a number of factors, including: audience location, content category, ad format, video length, time of year, use of ad-blocking technology, watch time, and more.

To account for these limitations, each variable in our estimation model is divided between a *conservative* and *aggressive* value, and every estimate is expressed as a min-mid-max ranged value. When an estimation is not explicitly expressed with the min-mid-max range in this report, it is assumed to always reference the midpoint value.

Our qualified views ratio and CPM variables were collated from industry reports, marketing research, and marketing statistics published by vetted sources, including: MilX (YouTube MCN), AIR Media-Tech (YouTube MCN), Statista (Market Research), Blscience (Media Intelligence), SproutSocial (Marketing Platform), JBX Digital Lab (Marketing Agency), and Gupta Media (Marketing Agency).

Additionally, revenue estimates were subject to sanity checks against self-reported numbers by creators who publicly disclosed their earnings on YouTube and Reddit threads. These tests did not inform our estimation model or variables; their purpose was to attest whether the model produced realistic results within a tolerable margin of error.

Methodology review

We thank the external reviewers who examined the full methodology underlying this report: Kalina Bontcheva (Professor of Text Analytics, School of Computer Science at the University of Sheffield and Chair of Working Group 1 for the EU Code of Practice on Transparency of AI-Generated Content under the EU AI Act), Jeff Allen (Co-founder and Chief Research Officer, Integrity Institute), Dr Leticia Bode (Research Director, Knight Georgetown Institute), Mark Scott (Senior Resident Fellow at the Digital Forensic Research Lab's Democracy + Tech Initiative, Atlantic Council), Dr Oliver Marsh (Demos Fellow and UCL Hon. Research Associate).

None of the reviewers was involved in the research or writing of this report. Their review covered channel identification, narrative classification, and revenue estimation, and was limited to the methodology.

Background: why harmful content earns more

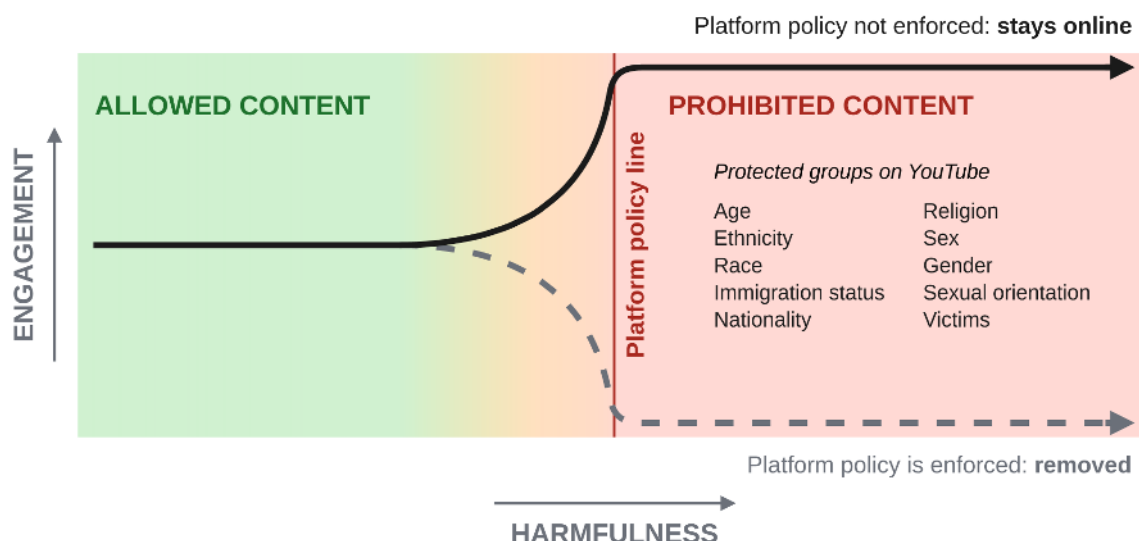


Figure A2. Chart inspired by research on *Negativity Bias* (Rozin and Royzman, 2001) and adapted from Meta's *Natural engagement pattern* (Zuckerberg, 2018). On the horizontal axis, content becomes more harmful from left to right, moving from allowed content towards the red policy line, past which YouTube's hate speech policy prohibits it. On the vertical axis, the grey path shows engagement when content is demoted as it approaches the line and removed once it crosses it, and the black path shows engagement when content stays online.

Figure A2 illustrates [the natural engagement pattern](#): content that approaches or crosses a platform's policy line tends to attract higher levels of engagement than content further from that line. A ranking system designed to maximise engagement therefore recommends content closer to the line more often, generating more views and, in turn, more advertising revenue. For the content examined in this report, YouTube's policy line is the [hate speech](#) policy of its Community Guidelines, which prohibits content that promotes violence or hatred against groups based on protected group status, such as race, religion or immigration status.

Figure A2 combines the natural engagement pattern, shown in black, with a second chart from the same publication, shown in grey, which illustrates the proposed effect of platform enforcement. Under this model, content approaching the red policy line is demoted, reducing its visibility, while content that crosses the line is removed. Content that approaches the line without crossing it therefore remains online but is made less visible. YouTube claims that it has [recommended such content less often](#) since 2019.

The black path shows the relationship when this enforcement is absent or fails: the ranking system continues to recommend content near or beyond the policy line, so it keeps receiving views and generating advertising revenue.