

Center for
Countering
**DIGITAL
HATE**

FAILURE TO PROTECT

**Social Media Platforms are Failing
to Act on Anti-Muslim Hate**



Contents

1 Introduction.....	4
2 Executive Summary	6
3 Our methodology.....	7
4 We recorded and reported 530 posts containing anti-Muslim hatred seen at least 25 million times.....	9
5 Platforms fail to act on 89% of anti-Muslim hatred reported to them	10
6 Platforms fail to act on 89% of posts promoting the “Great Replacement” conspiracy.....	14
7 Platforms are still failing to remove content glorifying the Christchurch terrorist	17
8 Instagram, TikTok and Twitter host hashtags used for anti-Muslim hatred	18
9 Facebook hosts pages and groups dedicated to anti-Muslim hatred.....	20
10 Recommendations	23
Recommendations for Lawmakers.....	24
Calls to Action	26
Appendix: Platform standards on hate.....	27

Center for
Countering
DIGITAL
HATE

The Center for Countering Digital Hate is a US-headquartered international non-profit NGO that disrupts the architecture of online hate and misinformation.

Digital technology has changed forever the way we communicate, build relationships, share knowledge, set social standards, and negotiate and assert our societies' values.

Digital spaces have been colonized and their unique dynamics exploited by malignant actors that instrumentalize hate and misinformation. These movements are opportunistic, agile, and confident in exerting influence and persuading people.

Over time these malignant actors, advocating diverse causes - from hatred of women to racial and religious intolerance to science-denial - have formed a digital Counter-Enlightenment. The disinformation they spread to bolster their causes has socialized the offline world for the worse.

The Center's work combines both analysis and active disruption of these networks. CCDH's solutions seek to increase the economic, political, and social costs of all parts of the infrastructure - the actors, systems, and culture - that support and profit from hate and misinformation.

If you appreciate this report, you can donate to CCDH at www.counterhate.com/donate

In the United States, Center for Countering Digital Hate Inc is a 501(c)(3) charity.

In the United Kingdom, Center for Countering Digital Hate Ltd is a non-profit company limited by guarantee.

1 Introduction

In a joint statement in 2019, Meta, Twitter, and Google committed to uphold the Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online. They stated that they would be “resolute in our commitment to ensure we are doing all we can to fight the hatred and extremism that lead to terrorist violence”.¹

Once again, their press releases prove to be nothing more than empty promises.

This report exposes that social media companies, including Facebook, Instagram, TikTok, Twitter, and YouTube, **failed to act on 89% of posts containing anti-Muslim hatred and Islamophobic content reported to them.** CCDH researchers, using platforms’ own reporting tools, reported 530 posts which contain disturbing, bigoted, and dehumanizing content that target Muslim people through racist caricatures, conspiracies, and false claims. **These posts were viewed at least 25 million times.** Many of the abusive content was easily identifiable, and yet there was still inaction. Instagram, TikTok and Twitter allow users to use hashtags such as #deathtoislam, #islamiscancer and #raghead. **Content spread using the hashtags received at least 1.3 million impressions.**

What is the impact of inaction? When social media companies fail to act on hateful and violent content, they know there is a significant threat of offline harm. Anti-Muslim hate seeks to dehumanize and marginalize communities of people who have historically been the subject of violent threats, attacks, discrimination, and hostility. Enabling this content to be promoted and shared on platforms without effective interventions and consequences, further endangers these communities by driving social divisions, normalizing the abusive behavior, and encouraging offline attacks and abuse. Worse still, platforms profit from this hate, gleefully monetizing content, interactions, and the resulting attention and eyeballs. For them, hate is good business.

Contrary to their press releases and pledges, **Facebook and Instagram failed to act on 89% of content promoting the “Great Replacement” conspiracy theory,** which inspired and was used by the terrorists who committed massacres at the Christchurch mosque attack in 2019 and the Tree of Life synagogue shooting in 2018. This is directly within scope of the Christchurch Call commitments, which the companies in this study committed to progress.

The conspiracist and racist content identified in this report spreads and perpetuates hatred of Muslims and their faith. It has a chilling effect on these communities, and prevents Muslim people from exercising their freedom of religion and speech online. Like Big Tech’s failures to act on antisemitism, anti-Black racism, misogynistic abuse, and misinformation, companies’ continued failure to act on anti-Muslim hate creates an ecosystem that restricts freedom of expression and pushes marginalized people off their platforms, all while allowing white supremacist, extremist, and hateful content to thrive and provide their shareholders with record profits.

Legislators, regulators, and civil society no longer believe social media companies when they promise to act on extremism and hate. Systemic and unchecked failures, like those identified in this report, must be addressed and technology companies must be held to account. Meta has been sued for their failure to address anti-Muslim attacks on their platforms by victims of the Rohingya genocide, and yet Facebook failed to act on 94% of posts in this sample. The status quo is insufficient to incentivize technology companies

from taking their responsibilities towards Muslim communities and other groups seriously.

Our experience as an organization suggests that three things are missing from existing powers globally:

- 1) The power to compel transparency around algorithms (which select which content is amplified and which is not); enforcement of community standards (which rules are applied and how and when); and economics (where, when, by whom, and using which data, advertising, which makes up the bulk of revenues for social media platforms, is placed).
- 2) The power to hold accountable social media platforms at an individual, community and national level for the impact of content they monetize
- 3) The power to hold accountable social media executives for their conduct as administrators of platforms that hold enormous power over discourse not just in terms of content moderation, but also: the amplification of content, institutional and user experience design of the systems through which discourse occurs, and equity in user experience for marginalized communities.

Speaking personally for a moment, if I may. My mom is Muslim. She is a good, hard-working, kind, and loving woman. She deserves better from those who have the power to protect her from the amplified hate of conspiracy theorists and devious, capable merchants of hate, and yet fail to do their bit. I cannot sleep when I see injustice. It makes me want to act. I cannot, for the life of me, fathom why the billionaires who own these platforms sleep at night when they know they could do so, so much more.

Imran Ahmed
CEO, Center for Countering Digital Hate

2 Executive Summary

- CCDH researchers identified over 500 (530) social media posts containing anti-Muslim hatred. Collectively, these posts were viewed at least 25.5m times. Our researchers reported these posts to the platforms using their own functionality for reporting hate speech.
- Posts were identified in February and March 2022 from Facebook, Instagram, TikTok, Twitter and YouTube.
- 89% of posts containing anti-Muslim hatred were not acted upon by social media companies. Twitter failed to act on 97% of posts, while none of the 23 videos reported to YouTube were acted upon. Content not acted upon includes:
 - False claims that Muslims are inherently violent
 - Conspiracies about a Muslim plan to “Islamize” Western countries
 - Depictions of Muslims as deceptive and untrustworthy
 - Racist caricatures depicting Muslims as inhuman
 - Sectarian Hindu nationalist hate narratives against Muslims
- Platforms fail to act on 89% of posts promoting the Great Replacement conspiracy theory which claims Muslims are working to ‘replace’ white people in the West. This conspiracy featured prominently in the ideology behind the Christchurch Mosque attacks in New Zealand and the Tree of Life synagogue attacks in Pennsylvania.
- Instagram, TikTok and Twitter allow users to use hashtags such as #deathtoislam, #islamiscancer and #raghead. Content spread using the hashtags received at least 1.3 million impressions.
- Facebook is hosting pages and groups dedicated to promoting anti-Muslim hatred with a total of 361,922 followers or members in the US, UK, Canada and Australia.
- Researchers identified 20 posts glorifying the terrorist behind the Christchurch massacre or featuring footage of the attack, with platforms failing to act in 70% of cases.
- Earlier reports by CCDH have shown platforms have similarly failed to act on antisemitism, anti-black racism, misogynist abuse and dangerous vaccine misinformation.

3 Our methodology

CCDH researchers assessed each platform’s enforcement effectiveness by (1) finding posts containing clear anti-Muslim hate, (2) recording and reporting them from ordinary user accounts and then (3) auditing what enforcement actions the platforms took.



1 Find Hate

In total 530 posts that breach platform standards anti-Muslim hatred were collected from Facebook, Instagram, TikTok, Twitter and YouTube over a period of three weeks, starting on 15th February and finishing on 9th March. Posts were collected from three different sources:

1. Searches for terms associated with anti-Muslim hatred, for example “raghead”
2. Hashtags associated with anti-Muslim hatred, for example “#deathtoislam”
3. Accounts and groups promoting anti-Muslim hate identified by our wider research

Defining “anti-Muslim hatred”

This report examines how platforms perform in acting on users reports of posts containing anti-Muslim hatred. We follow the Australian Muslim Advocacy Network (AMAN) in defining anti-Muslim hatred as follows:

“Anti-Muslim hatred” refers to the deliberate project of inciting hatred against Muslims, frequently conducted through dehumanization and conspiracy theory. It includes:

- Expressions of disgust towards Muslims
- Iteration of extreme right narratives about Muslims
- Expressions of wanting to expunge Muslims
- Expressions of wanting to kill or see Muslims dead
- As well as fantasies of violence against entire Muslim populations

In this context Islamophobia is a wider concept that covers discrimination and systemic bias as well as expressions of hatred.

2 Record & Report

For each post that contravened community standards, researchers collected the following information in our database:

- Screenshot of the post
- URL link to the post
- Date the content was posted
- Date the content was reported
- Platform on which it was posted

- URL and follower count of the account or group hosting the post
- Types of anti-Muslim hatred featured in the post
- Number of likes, comments, shares, and views
- Any further important observations of interest

These posts were reported to platforms using their own reporting tools as they were collected, using accounts established on each platform for the purposes of this research.

This set of 530 posts was then audited by another researcher who verified that the posts breached platform standards.

3 Audit Action

Once this had been completed, researchers checked every post and recorded any action taken by platforms, including whether the post had been removed or labelled, and whether the account or group that hosted the content had been removed.

4 We recorded and reported 530 posts containing anti-Muslim hatred seen at least 25 million times

In conducting our research, we recorded 530 posts from Facebook, Instagram, TikTok, Twitter and YouTube containing anti-Muslim hatred. The table below shows how many posts were recorded on each platform.

We also produced a minimum estimate for the number of times posts in the sample had been seen by users, using view numbers where platforms provide it, or interaction numbers where they don't. In total, posts in our sample had been seen at least 25.5 million times.

Examples of anti-Muslim hatred that were widely distributed

This YouTube video, originally posted in 2009, has amassed over 16 million views. It claims that the US, Canada and countries across Europe will become majority Muslim in the near future, and presents this as a deliberate plot on behalf of Muslims. YouTube failed to act on the video after it was reported to them.



Facebook failed to act on this video which has been viewed over 34,000 times and claims to explain "how Islam can destroy Europe". The animation depicts Muslims as terrorists.



5 Platforms fail to act on 89% of anti-Muslim hatred reported to them

We found that 88.7% of posts containing anti-Muslim hatred were not acted on by social media companies after our researchers reported them as breaching platform standards.

Of the 11.3% of reports that were acted upon:

- 4.9% resulted in the relevant post being removed
- 6.4% resulted in the posting account being removed
- None resulted in posts receiving warning labels

Platform	Reported	Post Removed	Account Removed	Acted On
Facebook	125	7 (5.6%)	0 (0.0%)	7 (5.6%)
Instagram	227	12 (5.3%)	20 (8.8%)	32 (14.1%)
TikTok	50	12 (24.0%)	6 (12.0%)	18 (36.0%)
Twitter	105	3 (2.9%)	0 (0.0%)	3 (2.9%)
YouTube	23	0 (0.0%)	0 (0.0%)	0 (0.0%)
Total	530	34 (6.4%)	26 (4.9%)	60 (11.3%)

While our sample of YouTube videos was small, the platform failed to act on any of the 23 videos reported to them. Twitter performed worst of the remaining four platforms, failing to act on 97% of anti-Muslim hatred reported to them.

Instagram performed significantly better than Facebook, despite both platforms being owned by Meta. Instagram failed to act on 86% of anti-Muslim hatred, while Facebook failed to act on 94%. Across both platforms, Meta had a failed to act on 89% of this content. TikTok performed best but still failed to act on 64% of anti-Muslim hatred.

Platforms failed to act on multiple types of anti-Muslim hatred

To investigate platforms' actions against different types of anti-Muslim hatred, we tagged posts in our sample according to their content. Posts were tagged with each type of anti-Muslim hate they displayed: for example, a post might brand Muslims as inherently violent at the same time as promoting conspiracy theories about them.

Type	Inherently Violent	Muslims as terrorists	Anti-Muslim migrants	Muslims as a disease	Racist caricature
Actioned	15 (9.8%)	12 (11.2%)	8 (10.0%)	3 (7.5%)	3 (15.0%)
Not Actioned	138 (90.2%)	95 (88.8%)	72 (90.0%)	37 (92.5%)	17 (85.0%)
Total	107	107	80	40	20

Examples of content platforms failed to act on

Many posts in our sample featured racist caricatures of Muslims. Platforms failed to act on 85% of such posts, including the Instagram post opposite which depicts Muslims as pigs and calls for them to be ejected from Europe.



Content claiming that Muslims are inherently violent was common in our sample. Platforms failed to act on 90% of this content, even where it featured violent videos or racist caricatures like the Instagram post featured opposite.



Nearly one in ten posts in our sample likened Islam to a disease, with platforms failing to act on 93% of such content. Instagram failed to act on the post opposite which likens Islam to cancer and implies that Muslims should be confronted with violence.



Platforms failed to act on 89% of posts that claimed ordinary Muslims are terrorists or that terrorism justified attacks on Muslims. Instagram failed to act on this post which used the 9/11 attacks to justify a call to “eradicate Islam”.



Many posts in our sample focused their attacks on Muslim migrants. Instagram failed to act on this post which portrays Muslim migration as “an invasion” along with the hashtags #Eurabia and #Islamification.



Some English-language posts in our sample had been shared by Hindu nationalist accounts. Instagram failed to act on this post from a Hindu nationalist account which claims Hindu women marrying into Muslim families will be beaten and carried that #islamiscancer hashtag.

Pic.1- Moslim girl after marrying in other religion

Pic.2- Hindu girl after marrying in Islam



10:22 pm · 06 Jan 21 · Twitter for Android

Platforms failed to act on 94% of posts that claim Muslims are deliberately concealing extreme or dangerous views. Instagram failed to act on this post which misrepresents the theological concept of “taqiyya” as the “Islamic doctrine of deception”, suggesting that Muslims routinely lie about their true beliefs. Far right and alt-right figures are known to have promoted similar misrepresentations of taqiyya.²



Both Facebook and Instagram failed to act on posts featuring this meme suggesting that Muslims “pretend” to be secular until they become a majority, at which point they look to “evict or kill non-Muslims”.



6 Platforms fail to act on 89% of posts promoting the “Great Replacement” conspiracy

Analysis of nearly 100 posts featuring elements of the “Great Replacement” conspiracy theory shows that platforms failed to act on 89% of them.

The “Great Replacement” conspiracy claims that non-white immigrants are ‘replacing’ white people and culture in western countries. Posts in our sample identified Muslim migrants in particular as a threat to white majorities, in many cases insinuating that this is part of a deliberate agenda or part of Islamic doctrine.

Platform	Facebook	Instagram	TikTok	Twitter	YouTube	Total
Actioned	1 (2.5%)	8 (20.0%)	2 (40.0%)	0 (0.0%)	0 (0.0%)	11 (11.1%)
No Action	39 (97.5%)	32 (80.0%)	3 (60.0%)	6 (100.0%)	8 (100.0%)	88 (88.9%)

While Instagram performed better than Facebook, the average performance of both platforms owned by Meta shows a failure to act on 89% of content promoting elements of the “Great Replacement” conspiracy. This is particularly disappointing given Meta’s pledge to improve its performance in removing hateful content following the Christchurch attack in 2019, where the terrorist justified their actions by referencing the conspiracy.³

YouTube failed to remove any of the eight videos we reported which promoted the “Great Replacement”. The platform’s failure to act over many years means that these videos have amassed nearly 19 million views and serve as reference points for the hateful conspiracy.

What is the “Great Replacement” conspiracy theory?

“The Great Replacement” conspiracy theory was first popularized in Europe by the French writer Renaud Camus. Camus wrote that “replacist elites” seek to replace white Europeans with non-white immigrants, particularly by Muslims who he claimed had a higher birth rate than white people. He claimed that immigration was being used as a deliberate policy to induce “genocide by substitution”. His conspiracy theory was widely adopted by White Supremacists around the world. In some versions, it is clear that the “replacist elites” are Jews, deliberately promoting non-white and Muslim immigration to bring about the downfall of white western culture.

Several extremists have cited the “Great Replacement” in their justification for mass killings. The white supremacist behind the 2018 Tree of Life synagogue shooting that killed 11 people in the US had posted to social media blaming Jews for bringing non-white immigrants into the US. The Christchurch terrorist who killed 51 and injured 50 people at Mosques in New Zealand in 2019 had published an online manifesto titled “The Great Replacement”.⁴

Examples of “Great Replacement” content not acted on

Instagram failed to act on this post which claims that Muslim families have “outbred” white European families, and will eventually look to remove white families from society.



Twitter failed to act on this post which claims that Muslim migration is part of a plot to change the demographics and culture of western countries.

So true! and nobody mentions how he's brainwashing and dividing people more with Islam! Central-America will soon have the same problems with Islam as here in Europe [#StopsIslam](#) [#IslamIsCancer](#)

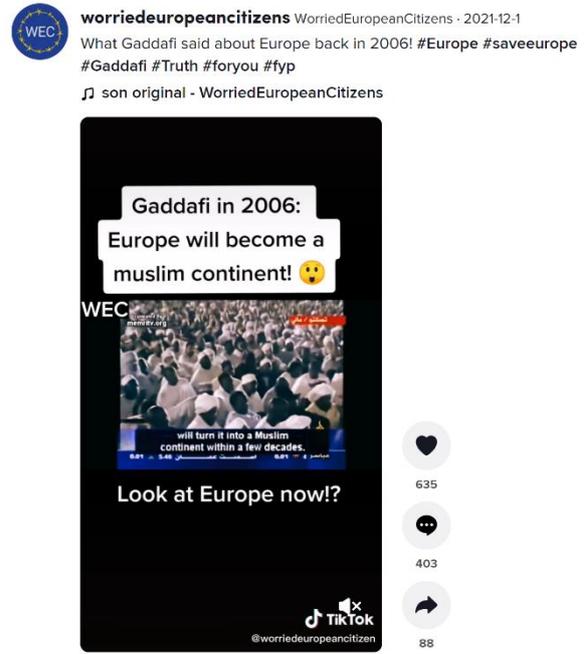


6:03 PM · Jan 23, 2022 · Twitter Web App

Facebook failed to act on this post which claims that Muslims are seeking a demographic majority in western countries, at which point they will seek to “institute sharia law”.



TikTok failed to act on this video which uses a speech by Libyan dictator Muammar Gaddafi to claim that Muslims are to turn Europe into a “Muslim continent”.



7 Platforms are still failing to remove content glorifying the Christchurch terrorist

Posts collected during the course of our research revealed that content glorifying the Christchurch terrorist, including footage of their attack, is still being posted to social media platforms and in many cases is not being acted upon.

In total, researchers identified 20 posts featuring the Christchurch terrorist, of which just 6 were acted upon. Facebook, Instagram and Twitter, who have all committed to promptly removing terrorist and extremist content as part of the Christchurch Call, failed to remove any the content we identified.⁵

Platform	Facebook	Instagram	TikTok	Twitter	YouTube	Total
Actioned	0 (0.0%)	0 (0.0%)	6	0 (0.0%)	0 (N/A)	6
No Action	2 (100.0%)	2 (100.0%)	3 (33.3%)	7 (100.0%)	0 (N/A)	14

The Christchurch Call

On 15 March 2019, a single gunman entered two Mosques in Christchurch, New Zealand during Friday Prayer and perpetrated a terrorist attack that killed 51 people and injured 50. Many more suffered from mental health impact including post-traumatic stress disorder (PTSD), including the families of victims and first responders.⁶ This attack has become known as the “Christchurch shooting” or the Christchurch Mosques terrorist attack.

This offline attack had strong links to online hate and extremism. The terrorist live-streamed the first shooting on Facebook, and published an online manifesto justifying the attack on Twitter and to the anonymous message board 8chan.⁷ Examination of the shooter’s manifesto and online presence revealed a commitment to a white supremacist and Islamophobic ideology, as well as admiration for other far-right terrorists.⁸

Two months after the attack on 15 May 2019, New Zealand’s Prime Minister Jacinda Ardern with French President Emmanuel Macron, led an international summit aiming to eliminate terrorist and violent extremist content online known as “The Christchurch Call”.⁹ To date, the Call has been joined by 55 countries and 10 service providers including Google, Meta, Twitter and YouTube.¹⁰ The US joined as a member of the Call in 2021.

Service providers supporting the Christchurch Call commit to “prioritising moderation of terrorist and violent extremist content” and to “closing accounts where appropriate”.¹¹

8 Instagram, TikTok and Twitter host hashtags used for anti-Muslim hatred

Hashtags allow users on Instagram, TikTok and Twitter to tie their posts to topics, making it easy for users to find other posts on the same topic. Platforms use hashtags to direct users to more content, keep them on platform and serve them more ads that generate revenue.

There are 285 posts analyzed by this report that feature hashtags used to share anti-Muslim hatred. Together they have generated at least 1.3 impressions on Instagram, TikTok and Twitter.

Instagram hashtags used for anti-Muslim hatred host 131,000 posts

This report analyses Instagram posts featuring hashtags such as #deathtoislam, #islamiscancer and #stopislamization. Instagram's own analytics show that these hashtags have been used in 131,365 posts across the platform.¹²

Instagram failed to act on this post which suggests that Muslims have a deliberate plan to “infiltrate” western societies and “establish Sharia Law”. It was shared using the hashtags #saveindia, #fuckislam, #stopislan and #stopislamization.



Instagram failed to act on this French language post promoting the “Great Replacement” conspiracy theory. It was shared using the hashtags #grandremplacement, #remplacement, #islamisation, #stopislamisation and #stopislam.



TikTok hashtags used for anti-Muslim hatred have had 1.2 million views.

Some of the TikTok posts featured in this report use the hashtags #raghead and #saveeurope. Content using these hashtags has been viewed 1.2 million times according to TikTok’s own analytics.¹³

This TikTok post promotes claims that Muslims will ‘replace’ white European demographics. It was shared using the hashtags #saveeurope and #nationalism.



Twitter allows anti-Muslim hashtags such as #deathtoislam and #islamiscancer

Some of the tweets in our sample were shared using anti-Muslim hashtags, such as #deathtoislam, #fuckmuslims and #islamiscancer. Other posts featured racist hashtags such as #raghead or #sandn****r.¹⁴

Twitter failed to act on this tweet containing a number of images claiming that Muslims are inherently violent and intolerant. The tweet carries the hashtags #RejectIslam and #IslamIsCancer.



9 Facebook hosts pages and groups dedicated to anti-Muslim hatred

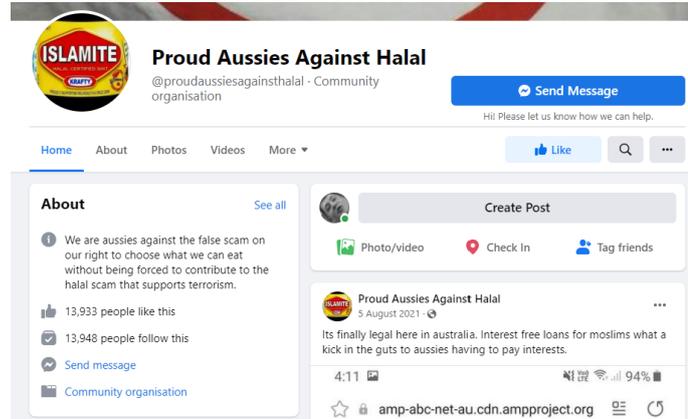
Many posts collected from Facebook for this report were from pages or groups dedicated to spreading anti-Muslim hatred. Collectively they have 361,922 followers or members.

Pages and groups were identified as mainly or wholly dedicated to anti-Muslim hatred based on their names, descriptions and content. Locations were determined by transparency or profile data about their administrators, or by the countries typically mentioned in posts if that data was not available. None have been removed to date, despite having their content reported to Facebook.

Name	Type	Location	Followers
A Totalitarian Agenda Exposed	Page	US	11,114
A few THINGS about Islam	Page	US	720
Anti Islam - Australia	Page	Australia	7,724
Boycott Halal Certification in Australia	Page	Australia	86,397
CANADA'S DEFENCE LEAGUE	Page	Canada	3,821
Cancer of Civilization Exposed	Page	India	984
Christian Defence League-United States	Page	US	1,744
Fight Against Liberalism, Socialism & Islam	Private Group	South Africa	4,943
Idaho ACT for America	Page	US	1,774
Infidel Task Force	Page	US	13,029
Infidels Against the Islamification of the West	Page	US	2,719
ISLAM means Terrorism	Page	UK	1,335
It isn't Racist to question the Islamification of your country.	Page	Australia	524
Knights of St. Michael the Archangel	Page	US	10,392
Latrobe Valley Patriots	Page	Australia	3,368
No Compulsion	Page	US	13,088
Patriots Defence League Australia- Townsville chapter	Page	Australia	1,236
Proud Aussies Against Halal	Page	Australia	13,948
Stand Up For Australia- Melbourne	Page	Australia	64,362
Stop Islamization of America	Page	US	78,916
Stop The Islamization Of America	Public Group	US	95
The Strange Death Of Europe	Page	UK	9,930
Wake up Australia - My unpopular opinion.	Page	Australia	29,759
Total			361,922

Examples of anti-Muslim hatred from Facebook pages and groups

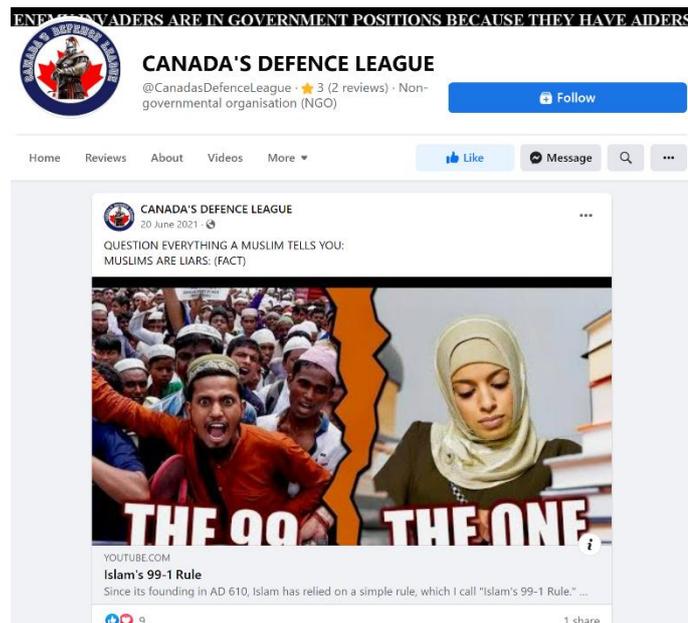
Some groups used false claims about Halal food in order to spread anti-Muslim hatred. The description of this Australian page with nearly 14,000 followers falsely claims that Halal certification “supports terrorism”.



Facebook has failed to act on this Private Group which makes its purpose of spreading anti-Muslim hatred clear in its banner image which portrays Islam as a gun pointed at the head of a woman representing “western civilization”.



Facebook failed to act on this post from a Canadian Page dedicated to anti-Muslim hatred. It states that “Muslims are liars” and links to a YouTube video with over 310,000 views that claims Muslims silence people who question their faith “through pressure, ridicule, intimidation, abuse or violence.”¹⁵



Facebook failed to act on this post from a Page dedicated to spreading anti-Muslim hatred called “Cancer of Civilization Exposed”. The graphic finishes with a call to “destroy Islam, save humanity”. Transparency information for the Page shows that it used to be called “Islam Exposed” and that the Page’s manager is based in India.

Cancer of Civilization Exposed
7 October 2019 · ⚙

**There are almost
SIXTY Islamic nations
in the world.**

**Yet Muslims have a problem
with there being ONE Jewish
state, and ONE Hindu state.**

**Then they come to the West,
preach radical hatred, and
try to change our laws and culture.**

POINTING THIS OUT THEN MAKES US RACIST!

Destroy Islam, save HUMANITY

42 4 comments 86 shares

Page transparency ✕

Page information for Cancer of Civilization Exposed ⓘ

 **Cancer of Civilization Exposed**
Interest

Page history ⓘ

-  Changed name to Cancer of Civilization Exposed
4 July 2020
-  Page created - Islam Exposed
26 April 2018

People who manage this Page ⓘ

-  Primary country/region location for people who manage this Page includes:
India (1)

10 Recommendations

Promises were made to the Muslim community - and everyone who is subject to terrorist and violent extremist content - when the Christchurch Call was signed by Big Tech. Promises were made in the policies and terms and conditions of use of Big Tech about racism and hate speech. Platforms have repeatedly stated that they would not allow hate and extremist actors to exploit their platforms and share their divisive and harmful propaganda. Platforms have promised that extremist content would be acted upon. Big Tech have repeatedly pushed PR that their platforms - marketed for mainstream audiences - would be cleaned up; that safety comes first. Through this study, we have found a massive gap between these promises and the reality.

All of the platforms whose content was assessed in this study state that the “report” button is their first line of defence against harmful content. All of them make commitments similar to Facebook (in its user Help Center), which promises “when something gets reported to Facebook, we review it and remove anything that goes against [our] Community Standards.”

But our research has proven time and again that this simply isn’t true.

In our earlier reports, we have drawn attention to consistent failures by Big Tech to act on a range of harmful hate and misinformation, for example:

- 87.5% of Covid and vaccine misinformation¹⁶
- 84% of content featuring anti-Jewish hate¹⁷
- 94% of users sending racist abuse to sportspeople¹⁸
- 90% of misogynist abuse sent to high-profile women over DM¹⁹
- Users who repeatedly send hateful abuse²⁰

This report shows that platforms are failing to act on 89% of anti-Muslim hatred. It is not a one-off circumstance. It is evidence of wider system failure and inaction. It is also evidence that the companies are not appropriately incentivised to invest in safety and put in place the appropriate levers for action, policies and processes by themselves. Self-regulation has failed.

Big Tech needs to be both:

- Proactive in addressing the online harm and misinformation before it becomes viral and communities are exposed to it; and
- Responsive to reports from users who have been exposed to harmful content and misinformation, and are raising the alarm.

But, currently when users pull that alarm, nothing happens. That’s for two reasons.

1. Platforms profit from getting users to view ads alongside content, even harmful content, and are reluctant to do anything that disturbs the flow of ad revenue.
2. Platforms refuse to invest in the moderation staff needed to respond to user reports appropriately and on time.

It suits platforms to pretend that this task is impossible to solve or that they are doing something because they have a pretty policy. Despite the press releases, neither of these things are true. Where governments have recognised that self-regulation has failed and

put in place basic transparency and safety standards, platforms have been forced to invest in better moderation.²¹

Recommendations for Lawmakers

These recommendations are specific to our report, but they also include core features and principles that are generally applicable, and technical recommendations that should be built into the platform's systems and processes to ensure the safety of other groups in society, including children.

The report shows that core features of a regulatory framework and enforcement are both missing and needed for platforms, including:

- **Transparency:** We know from our research, and others, that there are limits on the types of information that are currently publicly available, even with knowledge and tools. This information is held by the Big Tech companies who can track and trace what is happening online and the impact in changes to algorithms, processes, or safety features. This information asymmetry has devastating impacts for individuals, communities, and society, particularly where the same vested interests who hold most of the information are making all the decisions and profiting from them.

In terms of this report, the findings support a need to ensure that key data, patterns, and trends about abuse online [and engagement with that content] can be easily accessed, identified, analyzed, and addressed. Not just the data that Big Tech decides to share publicly or in voluntary working groups, but the data that exposes the problems (and possible solutions).

A regulator should have the power to require this information to be given and enable it to be shared with experts, including independent researchers, academics, and civil society organizations. This could be through anonymized data. The UK's Online Safety Bill could, if passed by Parliament, and if the new regulator of digital spaces, OFCOM, does its job as intended, be a watershed moment for social media companies. Regulators will only be able to carry out their new responsibilities with investment in independent research and civil society watch-dogs that can work with them to identify harmful content and trends.

Only then will we really be able to understand the full extent of the harm being experienced and continue to identify effective intervention points and tools. Relying on studies released by the company or organizations funded by Big Tech is like asking Big Oil to produce a feasibility study on climate change.

- **Responsibility:** a clear, proactive duty of care that is placed on platforms to ensure that their services and products are safe for all users, including children, before those users are exposed to harm. This includes any substantial changes made to those products, services, and processes. The weight should not fall on individuals to address online harm, particularly where they don't have the privilege of access to the underlying causes of that harm or the resources to design interventions that would change the operating environment. There should be safety by design. Social media should be safe before it is used, in the same way as we demand from those producing food, cars or pharmaceuticals. Clearly, unregulated Big Tech is not motivated to do this by itself - as evidenced in this report. Removing any general and unjustified exceptions to negligence law and rebalancing the

investment that is spent on engagement with safety features will lead to a safer environment for all users, including children.

- **Behavior that is illegal offline should be illegal online:** A lot of the content that we access and analyze in our reports, including this one, involves behavior that is illegal offline. This report shows that threats of violence and extremist content is allowed to flourish unfettered on platforms - despite the promises that have been made by Big Tech
- **Controls on harmful content:** There is a large amount of content shared on platforms, including content assessed in this report, which may fall short of a criminal speech standard but nonetheless is harmful because of its nature, intensity, or repetition or because some users may be more sensitive to that content, such as children or because it is targeted at people with particular characteristics or vulnerabilities. Much of this content has already been recognized as harmful content by the platforms themselves, which is why they specifically address it in their terms and conditions and press releases.
- **Complaints systems:** Clear, easy to access and responsive complaint systems - with oversight accountability of the platform by an independent government regulator. Our report shows significant failings in all of the platforms' current reporting pathways for complaints.
- **Accountability:** An effective and resourced regulator, and process for appeal or prosecution through tribunals or a court, will help to ensure that core responsibilities are being met.
- **Financial incentives:** Big Tech profits from both hate and the controversy surrounding that hate, which drives attention and traffic to their platforms and makes them money from ad revenue. There are different models to change the financial incentives:
 - **Negligence / Breach of statutory duty:** When a platform fails to act to remove harmful content (e.g. after warnings from the regulator), and when it fails to deploy its vast resources to avoid harms generated on their platform, courts could be free to decide if someone harmed by their inaction, for example, the families of the victims of the Christchurch and Tree of Life massacres, deserves restitution and compensation. This creates an economic incentive for action. In the US context, litigation (and the risk of litigation) is a strong motivator for shifting corporate behavior.
 - **Financial Penalties:** In Germany, under the "NetzDG law", platforms which fail to swiftly remove harmful content face significant financial penalties. Opponents warned that the NetzDG law would lead to overly sensitive censorship and infringements on free speech. So far, it has not. But it certainly has protected millions of people from vicious racism. Once there is a financial incentive to comply with local laws, the social media companies are suddenly able to act: it is no coincidence that Facebook was quick to establish a vast hub of moderators in Berlin.
 - **Corporate responsibility:** Consequences and offences also need to apply to companies and senior officials within the company - corporate liability for failure to discharge the responsibilities will help to ensure that the

investment is made to improve the functions and processes and to put safety first. A similar model of incentives operates in health and safety law in many jurisdictions, and it is one of the underlying premises behind corporate manslaughter.

- **Consequences for bad actors:** Bad actors should not have free rein to abuse people online, but this is the status quo. Our work, as well as other research, has proved that deplatforming has been effective at reducing the online harm from bad actors. This could sit alongside a warning and suspension system. There are different ways that this could be actioned, including regulator direction powers or oversight powers with a positive obligation on platforms to take action.

Calls to Action

1 Legislators should ensure that the core legislative framework (above) is implemented, including changing financial incentives for companies

If legislators want to stop hate corroding our society, social media platforms which give a megaphone to racism, abuse, and hate. Ensure the regulatory framework is fit for purpose: transparency requirements, responsibility (duty of care), behavior that is illegal offline is illegal online, disincentivize distribution of harmful sub-criminal content, effective complaints systems, accountability through an independent regulator and the courts, rebalancing financial incentives, consequences / offences for bad actors, platforms, and senior officials within the company.

2 Platforms should hire, train and support moderators to remove hate

The clear examples of anti-Muslim hatred exposed in this report show that tech giants' current efforts to moderate their platforms are not fit for purpose.

Platforms need to re-invest some of their immense earnings to hire, train and support the moderators needed to effectively remove dangerous anti-Muslim hate. In the long-term, platforms must also be held to account on their speed in dealing appropriately with harmful content, as platforms are designed to put content, including harmful content, into users' feeds while it is still new.

3 Facebook should remove groups dedicated to anti-Muslim hate

In conducting our research for this report, we identified Facebook Groups dedicated to sharing anti-Muslim hate. There is no moral justification for allowing public and private groups dedicated to fomenting hatred of Muslims - these groups and their organisers should be removed from Facebook's platform.

4 Instagram, TikTok and Twitter must act on hashtags promoting anti-Muslim hate

Similarly, this report identifies hashtags that are viewed by millions and regularly used to share anti-Muslim hate. Platforms should be bearing down on anti-Muslim hatred, not facilitating its spread. These racist hashtags must be banned.

5 Close user accounts which publish racism on social media platforms

This should apply to anti-Muslim hate and any other form of racism. There should be no tolerance of racism in public spaces. There is no constitutional right to use a privately-owned platform to publish racist material. Platforms and the hateful accounts they host must stop profiting from the viral spread of racist content.

Appendix: Platform standards on hate

All five platforms studied by this report – Facebook, Instagram, Twitter, YouTube and TikTok – have rules that prohibit hate speech against individuals or groups based on protected characteristics including ethnicity, nationality, race, or religion.

According to these standards, all the platforms studied by this report should remove clear examples of anti-Muslim hatred when they are reported to them.

facebook

Facebook’s community standards prohibit hate speech, defined as “a direct attack against people on the basis of... race, ethnicity, national origin, disability, religious affiliation, case, sexual orientation, sex, gender identity and serious disease.” It further defines “attacks” as “violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation.”²²

Facebook’s standards also ban “organizations that proclaim a violent mission or are engaged in violence to have a presence on Facebook” including “organized hate”.²³ Facebook stated that this policy is intended to cover “white nationalism and white separatism” as well as “white supremacy”.²⁴



Like its parent company Facebook, Instagram’s community guidelines prohibit “content that contains credible threats or hate speech” adding that “it’s never OK to encourage violence or attack anyone based on their race, ethnicity, national origin, sex, gender, gender identity, sexual orientation, religious affiliation, disabilities or diseases.”²⁵

Facebook’s standards also ban “organizations that proclaim a violent mission or are engaged in violence to have a presence on Facebook” including “organized hate”.²⁶ Facebook stated that this policy is intended to cover “white nationalism and white separatism” as well as “white supremacy”.²⁷



Twitter’s hateful conduct policy states that users “may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease.” It adds that Twitter “[does] not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.”²⁸

Twitter restricts the use of the following hateful images in some tweets and unsolicited messages, but does not ban them outright:

- Symbols associated with hate groups, such as the swastika
- Images depicting others as less than human, for example as animals²⁹

YouTube

YouTube states that “hate speech is not allowed on YouTube” and prohibits “content promoting violence or hatred against individuals or groups” based on ethnicity, nationality, race, or religion, amongst other protected characteristics.³⁰

YouTube does not explicitly define any forms of anti-Muslim hatred, but does give the following examples which are applicable to some anti-Muslim content:

- Describing groups of people as “a disease”
- Claiming that people with particular attributes “threaten our existence, so we should drive them out at every chance we get.”
- Content containing hateful supremacist propaganda including the recruitment of new members or requests for financial support for their ideology.³¹

TikTok

TikTok’s community guidelines state “we do not permit content that contains hate speech or involves hateful behavior, and we remove it from our platform.” It also promises to “suspend or ban accounts that engage in hate speech violations, or which are associated with hate speech off the TikTok platform.” It lists race, ethnicity, national origin, and religion amongst its protected attributes.³²

TikTok does not explicitly define any forms of anti-Muslim hate speech, but does give the following examples which are applicable to some anti-Muslim hatred:

- Depicting groups of people “as animals... or other non-human entities”
- “Promoting or justifying exclusion, segregation, or discrimination against them”
- “Conspiracy theories used to justify hateful ideologies”
- “Content that praises, promotes, glorifies, or supports any hateful ideology”³³

Published 28 April 2022

© 2022 Center for Countering Digital Hate Inc

¹ “Facebook Joins Other Tech Companies to Support the Christchurch Call to Action”, Meta, 15 May 2019, <https://about.fb.com/news/2019/05/christchurch-call-to-action/>

² “Taqiyya: How An Obscure Islamic Concept Became An Obsession Of Anti-Muslim Activists”, BuzzFeed, 12 April 2018, <https://www.buzzfeednews.com/article/ishmaeldaro/taqiyya-explained>

³ “By Working Together, We Can Win Against Hate”, Meta, 29 March 2019, <https://about.instagram.com/blog/announcements/sheryl-sandbergs-statement-on-new-zealand-shooting>

⁴ “The Great Replacement: An Explainer”, ADL, retrieved 13 March 2022,

<https://www.adl.org/resources/backgrounders/the-great-replacement-an-explainer>

“New Zealand Terrorist Manifesto Influenced by Far-Right Online Ecosystem, Hatewatch Finds”, SPLC, 15 March 2019, <https://www.splcenter.org/hatewatch/2019/03/15/new-zealand-terrorist-manifesto-influenced-far-right-online-ecosystem-hatewatch-finds>

⁵ Christchurch Call, retrieved 14 March 2022, <https://www.christchurchcall.com/supporters.html>

⁶ “Royal Commission of Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019” <https://christchurchattack.royalcommission.nz/publications/what-we-heard-from-the-whanau-of-the-51-shuhada-survivors-and-witnesses-of-the-christchurch-terrorist-attack/impact-of-the-terrorist-attack/>

⁷ “New Zealand mosque attacker's plan began and ended online”, Reuters, 15 March 2019, <https://www.reuters.com/article/us-newzealand-shootout-internet-idUSKCN1QW1MV>

“Shitposting, Inspirational Terrorism, and the Christchurch Mosque Massacre”, Bellingcat, 15 March 2019, 26 November 2020, New Zealand Department of Internal Affairs,

<https://www.bellingcat.com/news/rest-of-world/2019/03/15/shitposting-inspirational-terrorism-and-the-christchurch-mosque-massacre/>

⁸ “Shitposting, Inspirational Terrorism, and the Christchurch Mosque Massacre”, Bellingcat, 15 March 2019, <https://www.bellingcat.com/news/rest-of-world/2019/03/15/shitposting-inspirational-terrorism-and-the-christchurch-mosque-massacre/>

⁹ Christchurch Call, retrieved 14 March 2022, <https://www.christchurchcall.com/supporters.html>

¹⁰ *ibid.*

¹¹ The Christchurch Call, retrieved 14 March 2022, <https://www.christchurchcall.com/christchurch-call.pdf>

¹²

Hashtag	Posts
#antiislam	23,699
#banislamworldwide	4,442
#deathtoislam	1,306
#eurabia	771
#fuckislam	58,478
#greatreplacement	742
#islamiscancer	4,268
#islamisevil	10,529
#islamistheproblem	5,172
#notareligionofpeace	302
#ragheadsmustdie	10

#refugeesgohome	700
#religionofpeacemyass	885
#stopislam	17,177
#stopislamization	2,884
Total	131,365

Instagram, retrieved 14 March 2022

¹³

Hashtag	Views
#raghead	51,700
#saveeurope	1,100,000
Total	1,151,700

TikTok, retrieved 14 March 2022

¹⁴

Hashtag
#deathtoislam
#fuckislam
#fuckmuslims
#islamiscancer
#raghead
#sandnigger

Twitter, retrieved 14 March 2022

¹⁵ "Islam's 99-1 Rule", Acts17Apologetics, YouTube, 7 January 2020,

<https://www.youtube.com/watch?v=plhYzOKIrf0>

¹⁶ "Marketplace flagged over 800 social media posts with COVID-19 misinformation. Only a fraction were removed", CBC, 30 March 2021, <https://www.cbc.ca/news/marketplace/marketplace-social-media-posts-1.5968539>

¹⁷ "Failure to Protect", Center for Countering Digital Hate, 30 July 2021,

<https://www.counterhate.com/failuretoprotect>

¹⁸ "Instagram fails to take down more than 94% of racist abuse accounts targeting England players after Euros", iNews, 15 July 2021, <https://inews.co.uk/news/technology/instagram-racist-abuse-posts-england-players-after-euros-1102896>

¹⁹ "Hidden Hate", Center for Countering Digital Hate, 6 April 2022,

<https://www.counterhate.com/hiddenhate>

"I get abuse and threats online - why can't it be stopped?", BBC, 18 October 2021,

<https://www.bbc.co.uk/news/uk-58924168>

²⁰ "Twitter fails to remove 100 abusive misogynists", The Times, 13 January 2022,

<https://www.thetimes.co.uk/article/twitter-fails-to-remove-100-abusive-misogynists-z7nwg6d9t>

²¹ "Germany Acts to Tame Facebook, Learning From Its Own History of Hate", New York Times, 19 May 2018, <https://www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html>

²² Community Standards: Hate speech, Facebook, retrieved 13 July 2021,

https://www.facebook.com/communitystandards/hate_speech

²³ Community Standards, Facebook, retrieved 13 July 2021,

https://www.facebook.com/communitystandards/dangerous_individuals_organizations

²⁴ Standing Against Hate, Facebook, 27 March 2019,

<https://about.fb.com/news/2019/03/standing-against-hate/>

²⁵ Community Guidelines, Instagram, retrieved 13 July 2021,

<https://help.instagram.com/477434105621119>

²⁶ Community Standards, Facebook, retrieved 13 July 2021,

https://www.facebook.com/communitystandards/dangerous_individuals_organizations

²⁷ Standing Against Hate, Facebook, 27 March 2019,

<https://about.fb.com/news/2019/03/standing-against-hate/>

²⁸ Hateful conduct policy, Twitter, retrieved 13 July 2021, [https://help.twitter.com/en/rules-and-](https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy)

[policies/hateful-conduct-policy](https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy)

²⁹ *ibid.*

³⁰ Hate speech policy, YouTube, retrieved 13 July 2021,

https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436

³¹ *ibid.*

³² Community Guidelines, TikTok, retrieved 13 July 2021, [https://www.tiktok.com/community-](https://www.tiktok.com/community-guidelines?lang=en#38)
[guidelines?lang=en#38](https://www.tiktok.com/community-guidelines?lang=en#38)

³³ *ibid.*