Center for Countering Digital Hate

FUELLING HATE

One year after the UK summer riots, X still lets calls for violence spread unchecked

The Center for Countering Digital Hate works to stop the spread of online hate and disinformation through innovative research, public campaigns and policy advocacy.

Our mission is to protect human rights and civil liberties online.

Social media platforms have changed the way we communicate, build and maintain relationships, set social standards, and negotiate and assert our society's values. In the process, they have become safe spaces for the spread of hate, conspiracy theories and disinformation.

Social media companies erode basic human rights and civil liberties by enabling the spread of online hate and disinformation.

At CCDH, we have developed a deep understanding of the online harm landscape, showing how easily hate actors and disinformation spreaders exploit the digital platforms and search engines that promote and profit from their content.

We are fighting for better online spaces that promote truth, democracy, and are safe for all. Our goal is to increase the economic and reputational costs for the platforms that facilitate the spread of hate and disinformation.

If you appreciate this report, you can donate to CCDH at 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 nonprofit company limited by guarantee.

Contents

1 Report Summary 4
2 Introduction from Imran Ahmed 5
3 Key Findings 7
4 Violent Posts 8
5 Key Accounts 11
6 Case Study: Liverpool victory parade car crash 13
Appendix: Example violent replies to posts 15
Appendix: Example Liverpool crash posts 19
Appendix: Methodology 21
References 25

1 Report Summary

On 29 July 2024, three young girls - Alice Aguiar, Elsie Dot Stancombe, and Bebe King – were murdered by Axel Rudakubana, a teenager who stabbed them and injured ten others in a horrific attack in Southport. In the aftermath, high-profile hate figures on social media platforms like Elon Musk's X falsely linked the attack to Muslims and migrants, leading to widespread disorder and violent riots across the UK. Research from the Center for Countering Digital Hate (CCDH) found that these hateful influencers garnered millions of views per day on X.[1] The X platform utterly failed to moderate this explosion of dangerous, violent content.[2] CCDH's findings led the UK's online safety regulator Ofcom to conclude that there was a "clear connection" between posts on social media and the eruption of the 2024 riots.[3]

One year on from the 2024 riots, new CCDH research finds that little has changed

This study revisits six accounts on the X platform that CCDH identified as having posted influential hate and falsehoods surrounding the Southport attack. It finds that the same forms of violent and murderous rhetoric that precipitated and enflamed the 2024 riots are still widely circulating, with scant moderation by the platform.

Thousands of calls for violence are in reply to just six hateful influencers

The research shows how posts from six hate influencers – Tommy Robinson, Paul Golding, Ashlea Simon, Andrew Tate, Laurence Fox and Calvin Robinson – are the initial "parent" posts in an intensely violent series of replies from platform users. Reply posts encourage extreme acts of violence against Muslims and migrants: to shoot, to maim, and to kill.

X amplifies and monetises hateful influencers despite continued breaches of platform rules

This study makes clear the role of the social media platform X: harbouring these hate influencers, amplifying and monetising their dangerous content, and failing to enforce their own prohibitions against violent incitement as stated in conduct rules.[4]

Social media platforms like X have learnt nothing from the 2024 summer riots

The report contrasts the 2024 summer riots with the 2025 Liverpool parade car crash. Merseyside Police chose to share the ethnicity and nationality of the suspect in the Liverpool crash in an attempt to prevent false claims about either from gaining traction on social media and leading to offline violence, as happened following the Southport stabbing.[5] Unlike Merseyside Police, the X platform treated the 2025 Liverpool crash no differently than they treated the 2024 Southport stabbings – amplifying false incendiary information and failing to fact-check (or "Community Note") the vast majority of false claims about the Liverpool victory parade car crash.

2 Introduction from Imran Ahmed

One year ago, during a Taylor Swift–themed dance class attended by 26 young people, a 17-year-old British citizen walked into the studio and murdered three children: seven-year-old Elsie Dot Stancombe, six-year-old Bebe King, and nine-year-old Alice Aguiar. Ten others were injured in the attack.

In a moment of utter shock and mourning across the UK, a few high-profile hate-filled figures sought to swiftly hijack the tragedy to spread false claims online, by falsely identifying the perpetrator as a Muslim asylum seeker. These lies spread rapidly and widely on social media platforms, in particular X, fuelling violent anti-migrant and anti-Muslim riots across the UK.

In August 2024, CCDH reported that while affected communities and local authorities struggled to contain the violence both online and offline, social media platforms stood by and, in some cases, financially profited from the terror of these events.[6] X, in particular, emerged as a crucial vector of false information and hate. Rather than taking steps to mitigate harmful and illegal content, Elon Musk, owner of X, personally amplified conspiracy theories, warning of an impending "civil war" in Britain to his hundreds of millions of followers.[7]

Our researchers also found that hate preachers who had been banned from Twitter before being reinstated by Musk were receiving millions of views per day.[8] The platform was monetising their hate, running ads alongside content inciting violence, including posts encouraging mobs to "permanently remove Islam from Great Britain."[9]

One year on, we return to ask: what has changed? Have any of the accounts that spread lies and fuelled real-world violence faced consequences?

What we found is deeply disturbing.

Of the six prominent accounts identified by CCDH in the wake of the riots (Tommy Robinson, Paul Golding, Ashlea Simon, Laurence Fox, Andrew Tate, and Calvin Robinson), we found that each remains active on X. Today, their posts are still routinely followed by streams of replies containing threats, incitements to violence, and open expressions of hatred toward Muslims and migrants. These appear to remain largely unchecked, despite clearly violating X's own rules which prohibit such harmful content.

This new research makes one thing clear: one year on from the Southport riots, X remains the crucial hub for hate-filled lies and incitement of violence targeting migrants and Muslims. Anti-migrant and anti-Muslim influencers such as Tommy Robinson continue to play a central role in triggering waves of violent rhetoric. Their posts serve as rallying points for other hateful extremists, all too often leading to real-world harm.

This cannot be allowed to continue. Social media platforms must be held accountable and compelled to enforce their own rules with consistency and integrity. If they fail to do so, governments have a responsibility to step in by introducing and enforcing strong regulation that confronts the profit-driven amplification of violent and hateful content. Without intervention, the cycle of hate and violence will not just persist, it will deepen. The harm to targeted communities will grow and metastasize, with devastating consequences for British society and for all our communities.

Imran Ahmed
CEO and Founder, Center for Countering Digital Hate

3 Key Findings

CCDH identified 4,379 X posts promoting violence against migrants or Muslims

- Researchers revisited six hateful influencers who posted falsehoods and hate during the summer disorder of 2024.[10]
- This study finds that their posts on X are often followed with streams of replies promoting violence toward Muslims or migrants, including calls to kill.
- Using an AI model, researchers identified 4,379 violent replies targeting Muslims or migrants posted in response to 322 posts from the six accounts in the past year.[11]

X claims to prohibit posts promoting violence

- The 4,379 violent replies identified contained incitement to violence, threats of violence, and expressions of desire for violence.[12] Examples include:
    - "We need to shoot them at the point of entry, as they unload from the dinghies."[13]
    - "Dont deport them, let them come in and i will shot them all"[14]
    - "I would like to go postal on the Muslims! They are evil as hell!"[15]
- This type of violent speech is all prohibited by X with the platform's rules explicitly banning content that "threatens, incites, glorifies, or expresses desire for violence or harm."[16]

X is enabling hate by allowing key hateful accounts to benefit from X Premium

- All six accounts previously linked to the summer 2024 disorder are verified with a 'blue tick', which means they benefit from boosted reply visibility and advanced content tools.[17]
- Three of the accounts are allowed to offer paid subscriptions, enabling them to profit from their content.[18]

X is still failing to respond to hateful and misleading claims in emergency situations

- Despite law enforcement's updated strategy of transparent communication after a car struck spectators at Liverpool's victory parade, false claims thrived on X.
- False and misleading claims about the event reached 29 million views, often exceeding the reach of credible sources like Sky News and Merseyside Police.[19]
- Only 2 of 55 false or misleading posts displayed Community Notes.[20]

4 Violent Posts

CCDH identified over 4,000 posts promoting violence toward Muslims or migrants

This study revisits six accounts that CCDH previously identified as having posted influential hate and falsehoods on the social media platform X in the run-up to the disorder in summer 2024: Tommy Robinson, Paul Golding, Ashlea Simon, Andrew Tate, Laurence Fox and Calvin Robinson.[21]

It finds that in the year since the disorder occurred, their posts on X are often followed with streams of replies containing threats, calls for violence, and expressions of desire for violence directed towards Muslims and migrants.

In total, CCDH identified 4,379 reply posts promoting violence towards Muslims or migrants that appear in response to 322 X posts from the six accounts. These reply posts amassed at least 383,102 views in total and encourage extreme acts of violence, including calls to shoot, to maim, and to kill.[22]

The violent posts target Muslims and migrants in differing ways. Some promote violence against broad groups, such as calls for violence against all Muslims or against certain groups of migrants such as asylum seekers. Others promote violence toward specific individuals, usually in response to an individual named in the parent post.

Researchers used AI to identify violent posts in three main categories:

1. **Incitement to violence**, such as: "We need to shoot them at the point of entry, as they unload from the dinghies."[23]
2. **Threats of violence**, such as: "Dont deport them, let them come in and i will shot them all"[24]
3. **Expressions of a desire for violence**, such as: "I would like to go postal on the Muslims! They are evil as hell!"[25]

All three of these types of violent post are explicitly banned by X, with the platform's rules stating that it does not allow content that "threatens, incites, glorifies, or expresses desire for violence or harm."[26] X further claims that they may remove or reduce the visibility of such content to, "ensure the safety of our users and prevent the normalization or glorification of violent actions."[27]

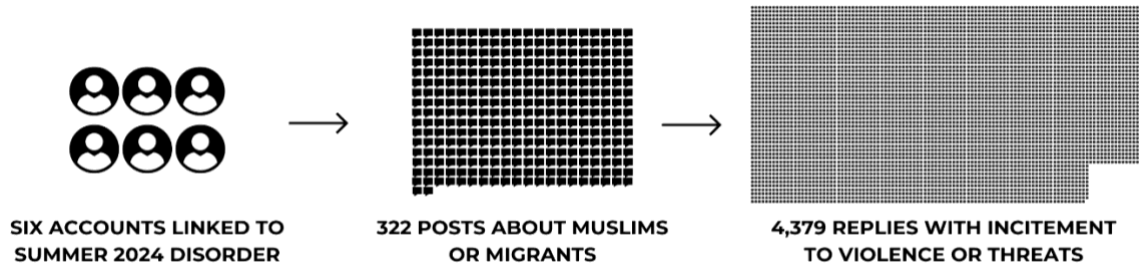
The findings of this study show that, one year on from the violent unrest of 2024, X remains a hub for violent speech towards migrants and Muslims. It quantifies the keystone role that anti-Muslim and anti-migrant influencers like Tommy Robinson play in triggering a series of violent calls from other users in replies to their posts.

How researchers identified 4,379 violent replies

Violent reply posts were studied using AI to analyse the replies to posts mentioning Muslims or migrants from six accounts that were previously identified by CCDH as posting falsehoods or hate during the summer disorder of 2024.[28]

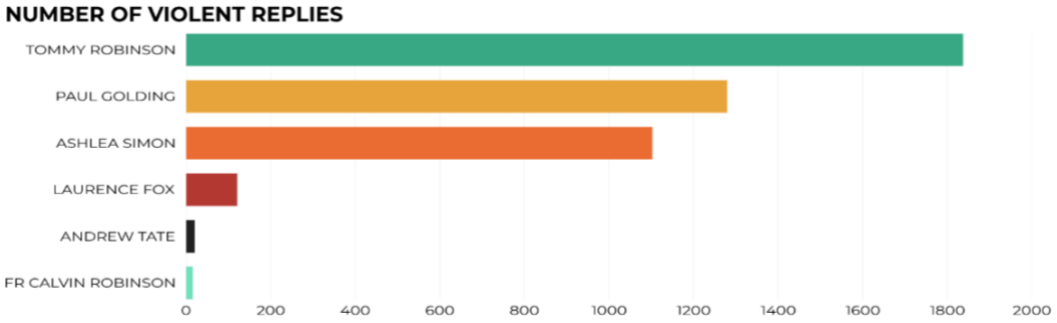
Researchers started by identifying posts from the six influential accounts that had over 250,000 views and contained keywords about Muslims or migrants, identifying a total of 322 such posts. They then analysed replies to these "parent posts", meaning the original post to which others respond, identifying a total of 4,379 violent replies.

The parent posts with violent replies were overwhelmingly anti-migrant or anti-Muslim in sentiment, often calling for mass deportations, containing negative depictions of Islam, or including examples of criminal behaviour purportedly carried out by either Muslims or migrants.



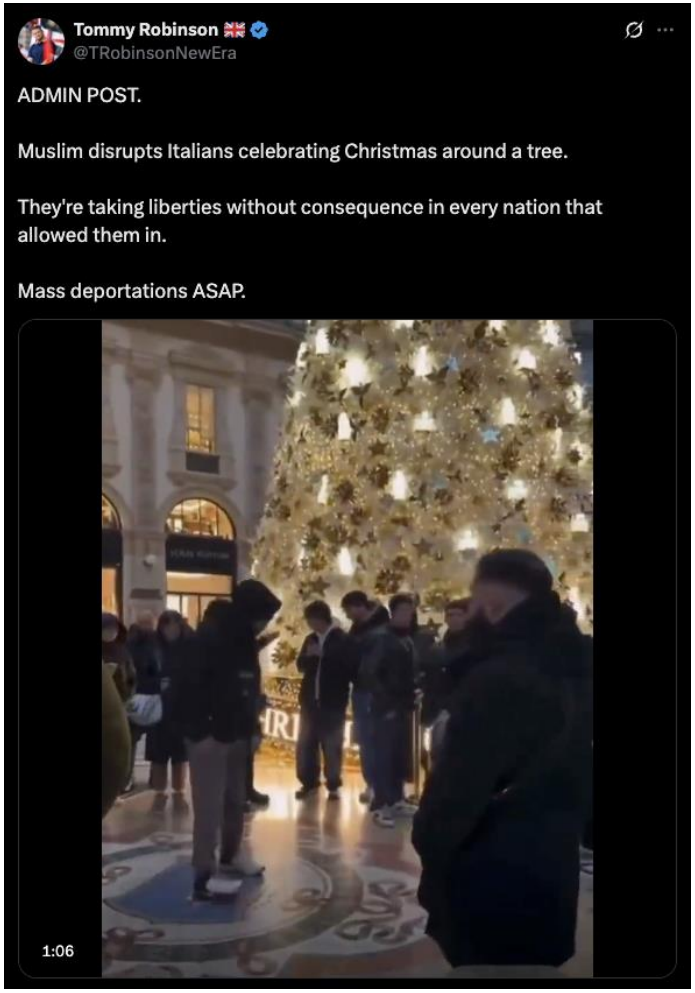
The AI system used for the analysis was designed to detect the promotion of violence towards Muslims or migrants in the context of post replies. The system worked with a high level of accuracy: our testing found that replies it flagged as promoting violence were identified correctly 95% of the time. Further details can be found in Appendix: Methodology at the end of this report.

Overall, most replies promoting violence were found under parent posts by Tommy Robinson with 1,838 violent replies, followed by Paul Golding with 1,280 violent replies, Ashlea Simon with 1,103 violent replies, Laurence Fox with 121 violent replies, Andrew Tate with 21 violent replies, and Calvin Robinson with 16 violent replies.



The next section shows an example of how a parent post by Tommy Robinson prompted violent replies. Examples for each of the other accounts can be found in an appendix.

Tommy Robinson: Example post with violent replies[29]



Post replies

Kill them all!	Come To My town and do this you will be hanging from the tree
Can we just have a army around boarders at this point to snipe and shoot people trying to invade.	It be a shame if a truck ran them all down the next time they do a group prayer.
Islam needs to be forbidden. Apply death penalty, no mercy	The other way round..... I KIL YOU, followed by beheading...
Drop a nuclear bomb on Mecca during the haj would be an appropriate response. https://t.co/hqsP3rnF4p	just shoot them on site. do not tolerate, its enough
Shoot them. Let's get this war started.	We need to put them all down like the animals they are
Should have cut his throat, only way for these animals to learn.	Shoot him on the spot!
The only good muslim is an unalived one.	Destroy the mother fuckers and mass deport them!
Need to send them home or Hitler them	Sh o o t them in the fucking face.
Fuk Islam this is war against the west. Eliminate them!	Round them up put em in ovens and let's get rid of them
execute these animals	Next time throw dead pigs in their mosques .
Why deportation? Shoot them. Killing such persons is the best example for others in line.	I can't wait until we start shooting. 😇
Fuck the Muslims, when can we go to war again and actually wipe these people of the planet. We can do without you.	They deserve death.
where are the facist italians from back in the day this shit should not happens we definitely need another hittler to get all these world right	Shoot them all instead. Fuck deportations
Imagine a sniper on the roof, bang.. bullet to the head. How satisfying	This world really doesnt need these people, mass genocide ASAP
Blow their heads as a sample. They will learn their lessons.	These people need to be eliminated
DEPORT THE SONS OF SATAN, DEPORT THEM ALL ASAP WITH KICKS AND POPS.. WITH NO LIE ISLAMOAPES DIES.	Their behavior is no longer acceptable. If these Muslim monkey dogs want war let's give them hell
That would be a great time to burn his ass alive the way he would've done to them in his shithole country of origin.	It is not deportations needed, it is just plain old simple death.

5 Key Accounts

Six key accounts studied still enjoy privileges on X despite role in 2024 disorder

The six accounts at the core of this study continue to subscribe to X Premium, awarding them the privileges of 'blue tick' verification status, and three offer paid subscriptions, despite playing a role in fanning the flames of the summer 2024 disorder.

CCDH previously identified the following six accounts for spreading false claims about the Southport attacker's identity.[30] Online rumours suggesting that the attacker was Muslim or a migrant reportedly fuelled subsequent violence, including attacks on mosques and asylum seeker accommodations.[31] Each of the six influencers posted at least once on X making false claims about the attacker or inciting hatred toward Muslims or migrants during the unrest:

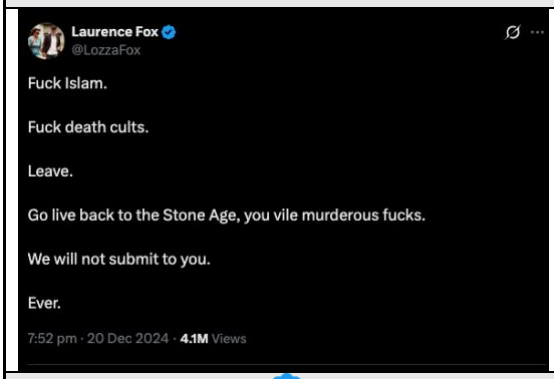
- **Tommy Robinson**, a far-right anti-Islam activist, who posted a video about the Southport attack with the caption: "There's more evidence to suggest islam is a mental health issue rather than a religion of peace."[32]
- **Paul Golding**, co-leader of Britain First who posted "the evidence is stacking up that the Southport attack was carried out by a migrant".[33]
- **Ashlea Simon**, co-leader of Britain First who said "We all know who's responsible for the Southport attack. White Christian teens don't stab children, Islamists do."[34]
- **Andrew Tate**, a misogynistic influencer whose post stating "ILLEGAL MIGRANT STABS 6 LITTLE GIRLS. WAKE UP." received 15.2 million views.[35]
- **Laurence Fox**, a former GB News presenter who reacted to news of the attack by posting "we need to permanently remove Islam from Great Britain."[36]
- **Calvin Robinson,** another former GB News presenter who posted during the disorder that "Islam is evil and needs removing from the UK."[37]

All the accounts retain verified 'blue tick' status which shows they pay X to have access to X Premium benefits like boosted reply visibility and advanced content tools. X has also allowed three of these users – Laurence Fox, Calvin Robinson, and Andrew Tate – the ability to offer subscriptions, enabling them to profit from their content.[38]

This is despite X's guidelines which state that premium accounts can lose their 'blue tick' for going against X's rules, including posting hateful content, such as using fearful stereotypes or dehumanizing language.[39] Not only did these accounts play a role in inspiring the summer 2024 disorder, but they have also posted hateful content that violates X's guidelines in the year since. Examples of violating posts are in the following section.

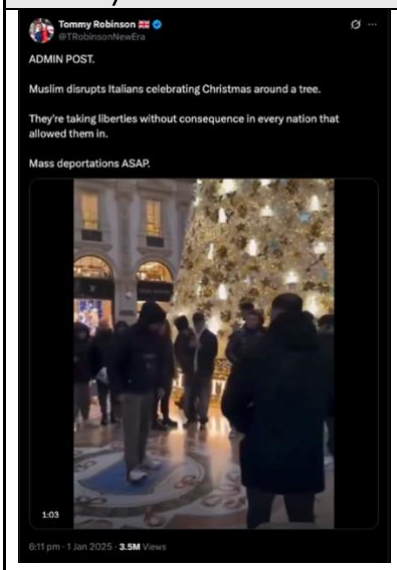
Examples: Posts from key accounts that breach X's Community Guidelines

Laurence Fox [Subscribe] 40

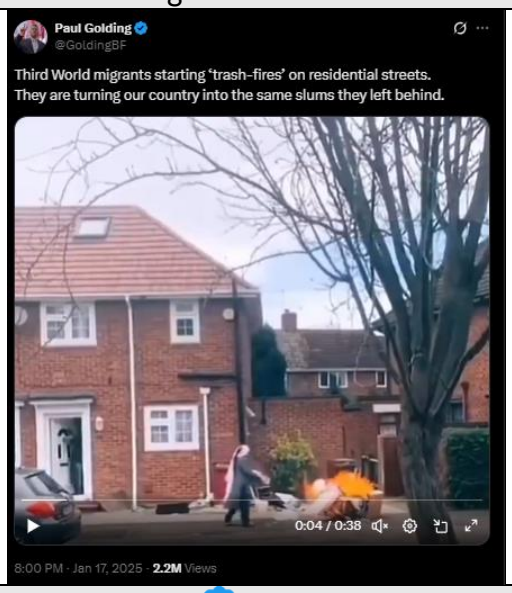


Calvin Robinson [Subscribe] 41

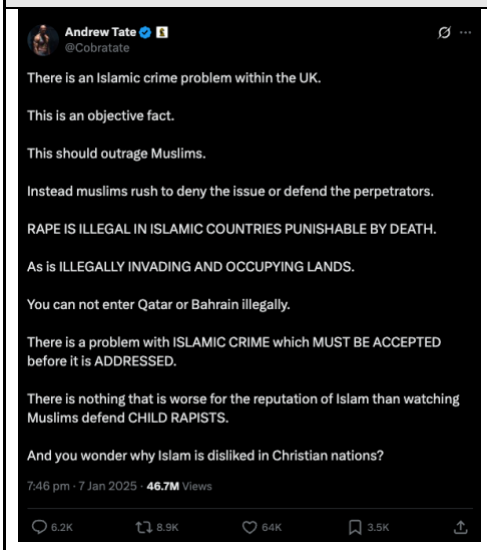
Tommy Robinson 42



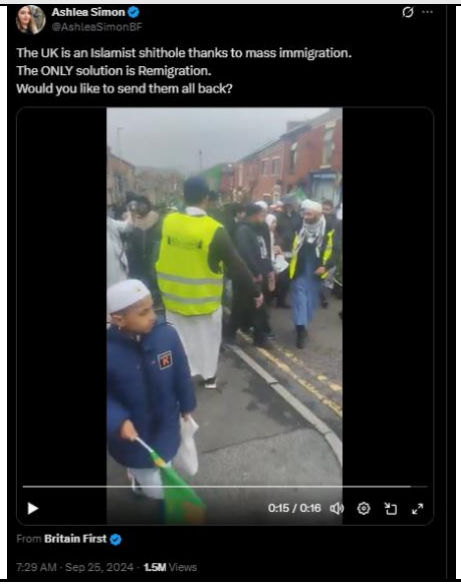
Paul Golding 43



Andrew Tate [Subscribe] 44



Ashlea Simon 45



6 Case Study: Liverpool victory parade car crash

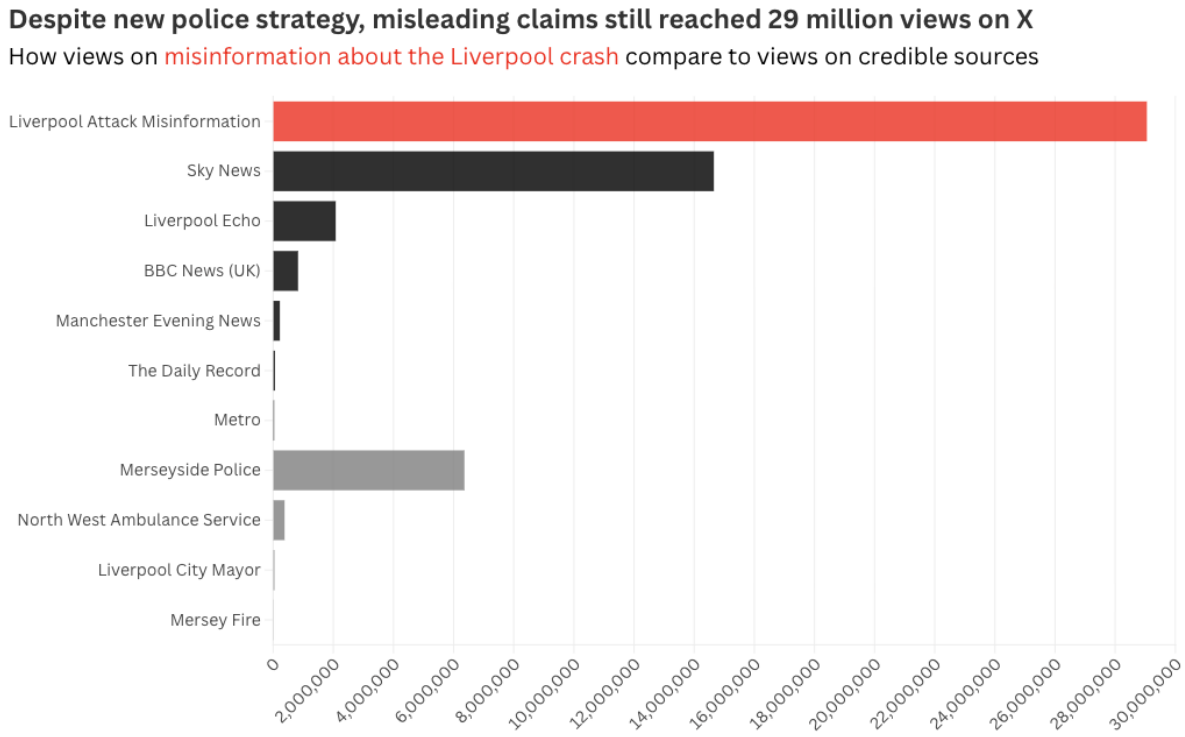
Recent incident shows that X is still allowing divisive claims to spread unchecked

On 26th May 2025, a driver struck spectators during Liverpool's Premier League victory parade.[46] The incident left 50 people injured, with several sent to hospital.[47]

Recognising the potential for further harm from the spread of conspiracies, speculation and false information, Merseyside Police swiftly shared details of the suspect's ethnicity and nationality, stating that the suspect was a "53-year-old white British man from the Liverpool area".[48]

This reflects an adapted approach from the police in light of the summer 2024 disorder, with official advice stating that police forces "must provide a true narrative online to reach people who may be searching for information" to help prevent future disorder.[49]

The response was also praised by Liverpool's mayor, who emphasised that it was important for Merseyside Police to act quickly to quell speculation and misinformation, saying: "The speculation was that it was part of something much bigger… the police acted very, very quickly, really swiftly, to put paid to some of that speculation."[50]



Despite Merseyside Police's decision to act quickly, false and misleading claims about the Liverpool crash still reached 29 million views on X, often exceeding the reach of credible news sources like Sky News and Merseyside Police.

The research shows that even when the police provided timely information, this did not prevent X becoming a space where divisive false and misleading claims were able to spread unchecked to reach large numbers of views.[51] While credible information was available, many were still being exposed to false or misleading claims.

In total, researchers identified 55 posts from the two days after the crash that contained false and misleading claims about the perpetrator. These 55 posts were collectively viewed 29 million times.[52] Credible posts examined for this case study received 24.7 million views, including posts from news sources with 17.9 million views and posts from official sources and emergency services with 6.8 million views.[53]

Misleading claims about the incident fell broadly into the following themes:

1. That the police were trying to cover up the truth about the perpetrator
2. That that the person in the vehicle was not actually white nor 53 years old
3. A conspiracy that a man involved in a separate incident was the one arrested

Not only did the misleading claims reach high numbers of views, but just two of the 55 false and misleading posts displayed Community Notes, showing that X's crowdsourced fact-checking system is not working consistently.

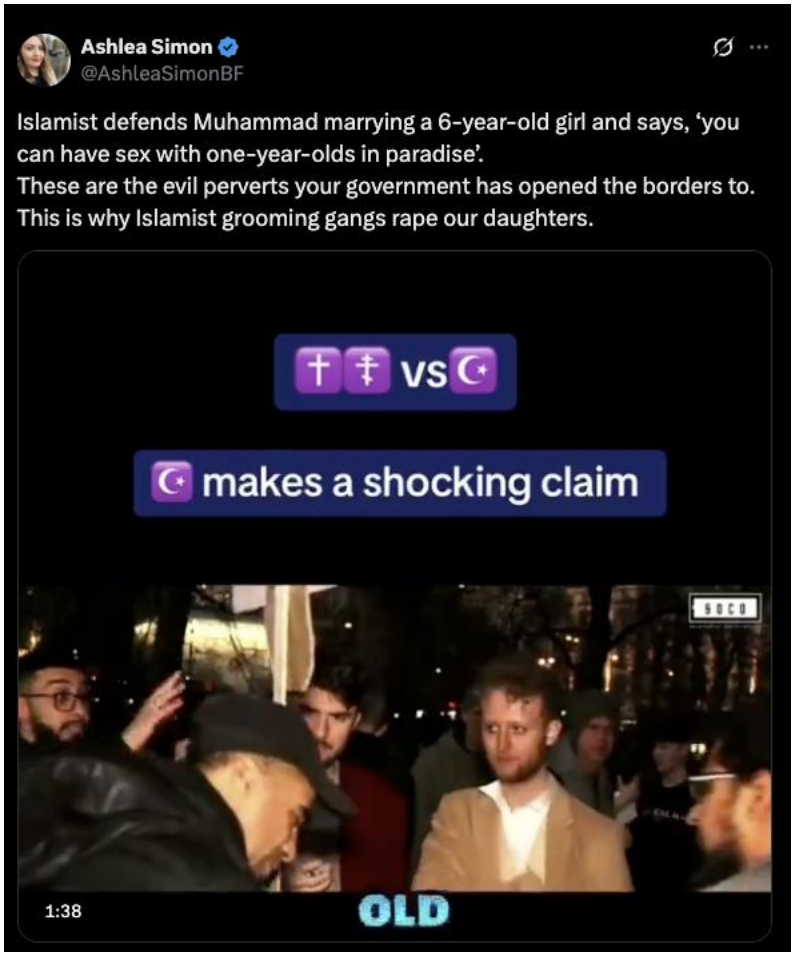
Comparing virality of misinformation to credible sources

CCDH researchers compared how false and misleading claims about the Liverpool attack performed in the two days after the attack when compared to content from more reliable sources like news and emergency response accounts.

News sites in the comparison were chosen by looking at the UK news sites with the highest web traffic that also achieved high scores from NewsGuard, a rating system that judges news reliability.[54] Estimates for views on their relevant X posts were calculated by aggregating the views on posts containing keywords about the attack.

Appendix: Example violent replies to posts

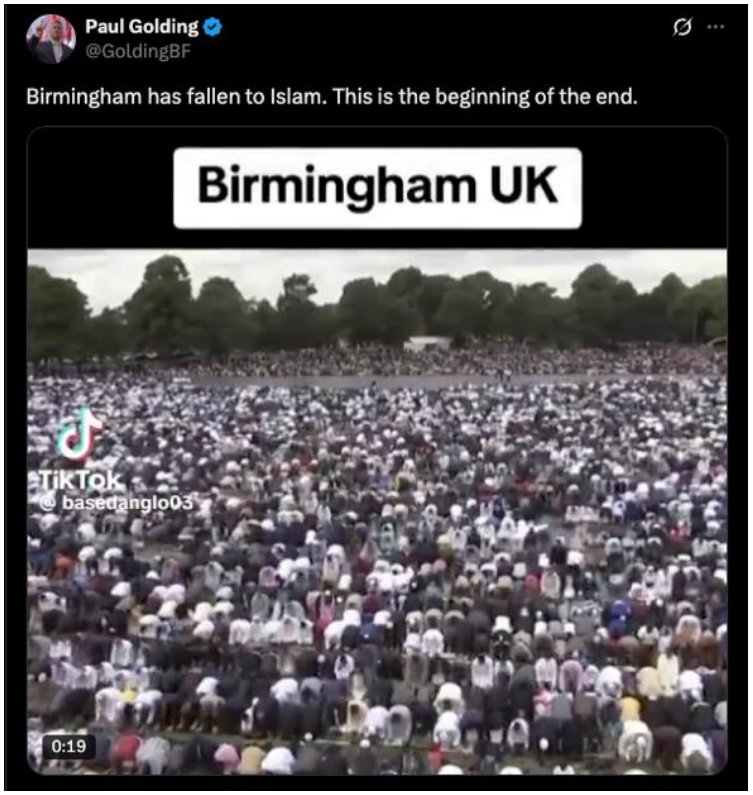
Ashlea Simon: Example post with violent replies[55]



Post replies

Fucking kill them all that God sort them out Cut all their throats filthy bastards I hope you rot in hell! Children?? Wtf is wrong with you
People like this should be wiped from this earth i swear to god UTILIZE THE BOBBETT EFFECT, ONLY RECOURSE, CASTRATE ALL MALE MUSLIMS They deserves to be exterminated We need to bring back concentration camps and dump these sickos in oil drums and flush them in the sea.
Kill him.he is a danger to all children ISLAM IS EVIL☠ DEATH TO ISLAM☠ Wish y'all that's supporting Hamas and Palestine PS let y'all daughters being from a year old till adult,be raped by all this blood 🩸 suckers Kill them all We need to take care of those mslim burry them with pig 😈🐖🐖🐖
Kill them all Castration or death. Deport the rest We will slaughter every sick religious mfr eventually! Sick fucken race 🤬🤬🤬 they need to be exterminated.
Islam in one sentence U can have sex with one year old in paradise U can not fix this mentality We have to obliterate it WE NEED CAPTIAL PUNISHMENT FOR THIS IN THE UK, if I made the law's he would have his tongue cut out and mouth stitched up just for speaking like that or even thinking in such manner disgusting and vile, what about us being offended by this in our country @ukhomeoffice @GBNEWS GET ARMED SHOT THEM Bomb the whole lot!
Please destroy these muslim demons.
They need to confront these terrorists with swords and guns and not with words just as they do Another dirty Muhammad nonce, I would skin this fucker alive if he came near my children. If in heaven I saw you having sex with a one year old girl..I personally executed you and send you to hell of fire and feed you to my wonderful dogs, Cerberus🤣🤣🤣 stay there eternally suffer

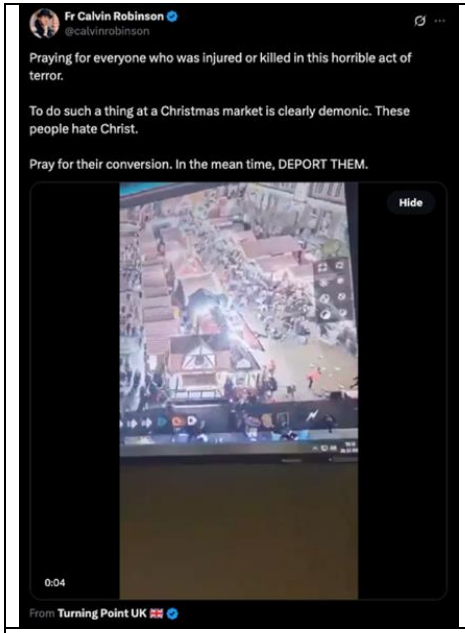
Paul Golding: Example post with violent replies[56]



Post replies

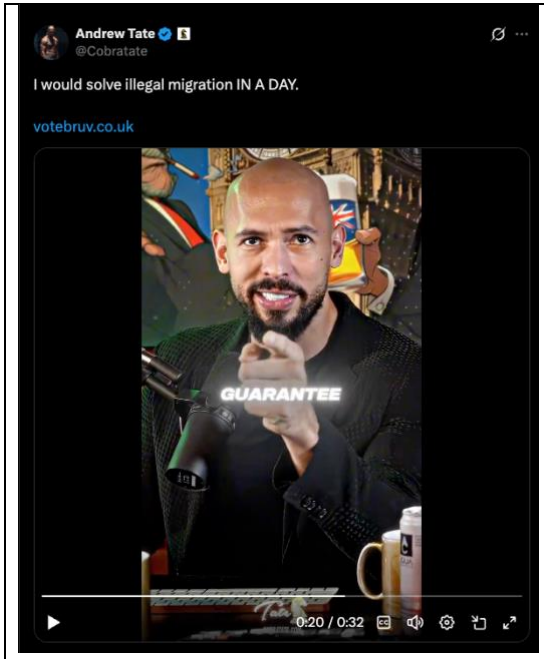
Blow them up	Good time and place to drop a bomb.
If they don't want to use their mosques , burn them fucking down. A football field of peados that looks like to me	Surely you Britons have heard about Napalm? You must know how this kind of a gathering gives the best results in the use of Napalm? Simple solution, in'nit?
A few passes with Apaches and A10's would solve that pile of sub human shit in a few seconds. https://t.co/At2Zu32Ecm	It's cheaper to genocide them if you don't have to go to the middle east to do it. I wonder if they trust the food and medicine the locals give them?
Drop just one jdam or Moab and the problem will be sorted	True there is no way back without bloodshed. It's either that, forced conversion, acceptance of a Dihmmi status or death.
What a travesty it would be if a low yield nuke landed right in the center of them	Large quantities of napalm would solve this problem easily.☺
Bomb where they worship	drop a bomb right in the middle of them...
Small tactical nuke would be worth using	Napalm from air will stop that in a hurry.
Deport or die by the Machete	Looks like a great spot for some bomb testing
Perfect time to bomb	Get violent or die
Good place for a cluster bomb	We realy need next hiroshima and nagasaki in UK
No. All you need is 100 men, armed with bats, armed with gasoline, organized, with nothing to lose. Solve the problem in months.	A couple of machine guns could solve this specific problem right there. When the time come, these animals will have no mercy and will fire them over native english. Is what they do for 1.400 years. But you guys seem to be more afraid of being called "racist". Have a good ride.
So, we need to blow up all of the mosques since the Muslims are NOT using them? We need a plan to stop the Islamification of the United Kingdom. Ideas please!!	too many tough mofos in birmingham. just gotta organize like the muslims. take em out ONE AT THE TIME calmly, quietly and asurely.
Drop a bomb..... problem solved	With those demographics, can #remigration work? Or would it be more effective to just nuke the place?

Calvin Robinson: Example post with violent replies[57]



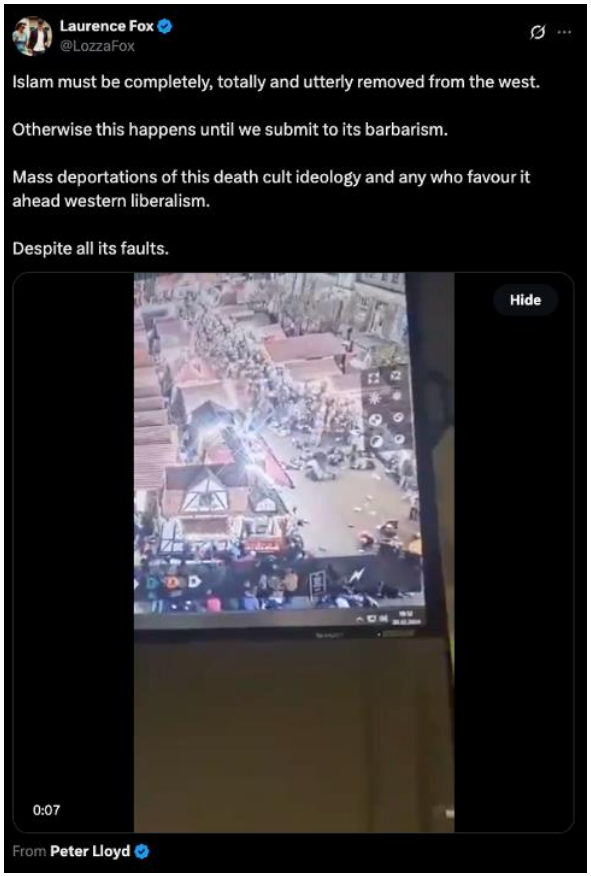
Father, if these attacks continue then the narrative of deporting them will be abandoned and we will move on to killing them, it's them or us.
They shouldn't be deported but the human filth should be exterminated.
You can pray for them, but I'd much prefer them to be hanged.
Deport? NO! EXECUTE
No. Kill them now. Execute!
Execute them not deport them..
I pray for their death not their conversion, sorry
Utter cunt, fucker needs a bullet in his head.
A bullet in the head should do the trick

Andrew Tate: Example post with violent replies[58]



Hell yes top g brother! Fuck those immigrants! We stick to our own policy and rape them!
ugly but efficient truth in history, you want fast deportation, tell unwanted illegal immigrants they either leave or get killed, it has worked in all countries who applied it, and theu usualy leave. No need for fancy plans or much money spent.
id shoot them all personally
If you come within ten miles of our shore, we will sink the boat.
All you have to do is open fire on the first boat the Navy comes across, put that on the it all stops. Nothing says do not come like automatic machine gun fire.
You and $EITLER have the same idea for illegals. Straight to the gas chambers.
If the boats can't confirm themselves send them back or shoot to kill. Problem solved

Laurence Fox: Example post with violent replies[59]



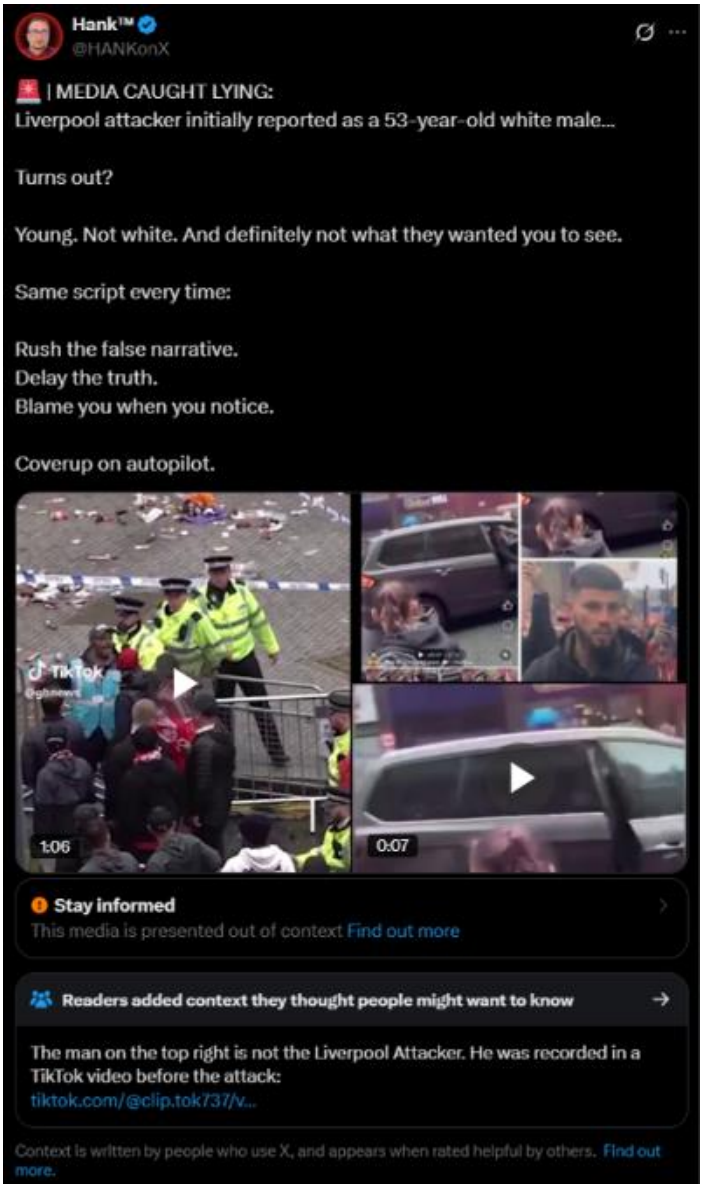
Post replies

Wipe them off the face of the fucking earth!!!	If it turns out the person's an islamist, he should be executed in the streets, and it should be broadcast to the country from which he came from.
Ah, the religion of peace strikes again...eliminate them all.	We MUST eradicate ISLAM!! #Startdeportations
Mass Deportations won't work. You don't understand that yet, do you? There is but one solution. The Serbian Solution.	Mass deportation, confinement in detention camps and castration are the only option to contain them. And if found involved in criminal activity, then minimum 20+ years imprisonment to chuslim criminal. Do it for just one year and thank me after that.
Find him. Incinerate him in the public square AFTER,...after he watches his entire family emolated in public. Then,...punish him.	Exterminate them all
Nuke the Middle East. Been saying it for years. There's too many bad eggs to care about the rest of them.	Always filthy terrorists should be executed without delay.
Bomb them straight to HELL 😡🔥	Exterminate them all
Hope the killed the driver too. Tie him and smash with a car.	Actually Islam needs to be removed from earth completely. It will never end till they're all gone.
A full culling, the western world should wipe them out	We should drive them into the sea along with the bed wetting lefties who defend them! Fuck Islam!!

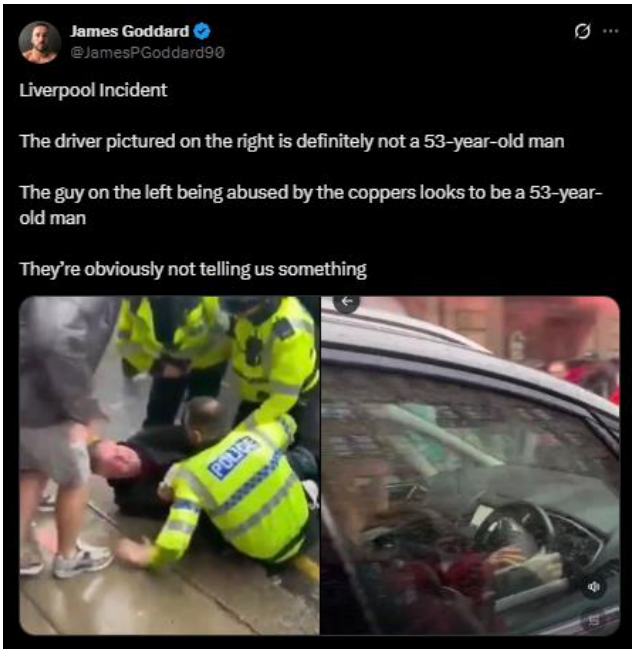
Appendix: Example Liverpool crash posts

This appendix contains examples of false or misleading claims about the identity of the suspect behind the Liverpool parade car crash on 26 May 2025. False or misleading posts about the crash studied by this report amassed 29 million views in total.

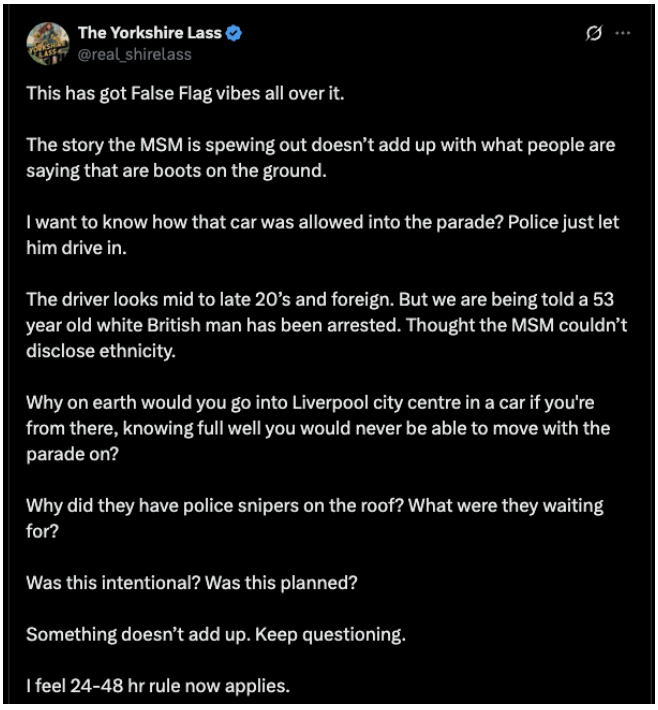
- This post with 3.1 million views promotes a false claim that the driver in the Liverpool car crash incident was not a 53-year-old white male but a young person who was not white, claiming that it was a "coverup" and the media has been "caught lying".[60] This post did not have a community note until 5 days after the incident and it already had 1.6m views less than 24 hours after it was posted[61].



- This post, with 3.8 million views, does not have a community note and claims that the 53-year-old man arrested was involved in another incident and was not the driver of the car. Saying that there is something “they’re obviously not telling us”.[62]



- This post, which has almost 1 million views and no community note, says that the driver of the car looks “mid to late 20’s and foreign’. They ask where this was “intentional” or “planned”.[63]



Appendix: Methodology

This appendix outlines the methodology used to develop the findings of this report, including the terms used to identify posts for study, how violent threats were identified using AI, and how false or misleading posts about the Liverpool crash were identified.

Studying parent posts from six accounts about Muslims and migrants

To analyse a set of relevant posts, researchers used a licensed third-party tool to study posts from six accounts filtered down by keywords related to migrants and Muslims.

The study was designed to capture activity on X in the time since the summer 2024 disorder. As such, posts in the study were from 9 August 2024 through to 28 May 2025, where the 9 August was chosen as a start date as the disorder had stopped by this stage and there had already been 600 arrests.[64] Researchers focussed only on those posts that reached over 250,000 views.

The full keyword search is in the box below:

migrant|muslim|muslims|illegal|islam|islamic|migrants|immigration|immigrants|islamist|foreign|remigration|deport|immigrant|deportations|islamists|invaders|refugees|illegals

Identifying violent threats in replies using AI

Using the same licensed third-party tool, researchers studied replies to all the parent posts about Muslims or migrants, resulting in an overall analysis of 700k replies. Researchers analysed these reply posts using an AI-assisted system that was built to identify posts promoting violence towards Muslims or migrants.

To identify violent posts accurately and efficiently, the AI system used a two-stage process. Firstly, a lightweight model, OpenAI's GPT-4o-mini, scanned all replies to flag those most likely to contain violent content. Subsequently, a more powerful model, GPT-4o, reviewed this high-risk subset in detail. This approach allowed researchers to focus on the most serious cases while keeping the process resource efficient. It also made it possible to refine the second stage specifically for the replies that were most likely to contain the promotion of violence toward migrants or Muslims, improving both precision and reliability.

In the second stage, GPT-4o is prompted to assign a 0 to 10 risk score to each high-risk reply. Researchers refined this prompt using a labelled sample of 400 replies,

with positive examples representing violent threats targeted at Muslims or migrants and negative examples representing replies that either did not contain violent threats or the targets of a violent threat were not Muslims or migrants. Based on this refinement and empirical testing, a classification threshold of 8 was chosen as it provided the best balance between capturing true violent threats and minimizing false positives.

Overall, the two-stage system results in an F1 score of 80%. An F1 score is used to measure a model's accuracy and is a weighted average of the following two metrics:

- **Precision:** the number of posts labelled as belonging to a category by the model that were actually correct.
- **Recall:** the number of posts for a category that were correctly identified by the model as belonging to that category.

The precision score was 95%, meaning when the model identified a post as a violent threat towards migrants or Muslims, it was correct 95% of the time. Recall was 68%, meaning out of every 100 violent threats, the model successfully found about 68.

These accuracy scores are based on a weighted analysis of 400 high-risk replies and 400 lower risk replies. The scores indicate that the model was extremely accurate when positively identifying a post as containing a violent threat, but did miss a number of posts that should have been identified as violent. This should make the numbers used throughout our report conservative estimates of the scale of violent threats.

In total, this methodology identified 4,379 replies that promote violence against Muslims or migrants that were posted in response to 322 'parent posts' from the six key accounts studied. View figures were available for just 517 of the 4,379 violent replies in our study. Those 517 posts amassed 383,102 views in total.

How researchers identified false and misleading posts about Liverpool car crash event

The Liverpool car crash incident occurred at 6pm on the 26 May 2025.[65] To find relevant posts, a third-party tool was used to study posts based on a keyword search query likely to identify false or misleading claims around the incident. Posts studied were from the hour before the incident through 28 May 2025.

The top 1,000 posts by views were reviewed and classed based on whether they contained false or misleading claims concerning the driver of the incident. Each post classed as false or misleading was then checked by an additional researcher to

confirm that it was correctly categorised. This led to a final set of 55 posts with total of 29m views that were classed as false or misleading.

The full keyword search is in the box below:

(liverpool) & (“car attack”|crowd|driver|car|parade|muslim|immigrant|migrant|illegal|brown|islam|foreign|Taliban|coverup)

How researchers compared the false and misleading posts to trusted sources

To understand how the spread of false or misleading claims from the Liverpool incident compared to reliable news sources, posts were studied using a third-party tool from top UK news sites and emergency responders. These posts were studied over a similar timeframe as the false and misleading Liverpool posts, from the time of the incident at 6pm on 26 May 2025 through 28 May 2025.

To analyse the views on trusted news sites, researchers created a list of 6 large, reliable news sites and estimated the number of views on their posts about the Liverpool car crash incident. To create this list, researchers selected the biggest news sites by traffic in the UK during May that had a NewsGuard “reliability” score of over 90. Data on the top UK news sites by traffic was based on statistics from Semrush.[66] NewsGuard's scores assess sources based on a set of apolitical criteria of journalistic practice, with a high score reflecting strong reliability.[67] To get an estimate of the number of views each publication received on posts about the Liverpool incident, researchers gathered all posts from these outlets between the same time period and filtered using the keyword “Liverpool” along with any of the following keywords: car, parade, crash, incident.

These keywords were derived by analysing the 200 most frequently occurring words across all studied posts and selecting the terms relevant to the incident. This allowed researchers to focus on content likely to be related to the event in Liverpool and calculate the total views per outlet for comparative analysis.

To analyse the views on posts by emergency responders, all posts on X were studied from four accounts which were likely to provide accurate, real-time information during the incident. These accounts were Merseyside Police, Merseyside Fire & Rescue Service, North West Ambulance Service and the Mayor of the Liverpool City Region, Steve Rotheram.

Researchers were able to confirm that the keyword approach performed with a high level of accuracy by testing a random sample of 300 newspaper posts. The random sample was made up of 150 posts that were identified as pertaining to the incident through keywords and 150 that were not to ensure that there was a representative sample of both outcomes.

Researchers then manually labelled whether they were related to the Liverpool incident or not. These labels were compared to posts identified using the keywords approach, resulting in an overall F1 Score of 96% with a precision of 99% and a recall of 93%.

References

[1] "X ran ads on five accounts pushing lies and hate during UK riots", The Center for Countering Digital Hate, 19 August 2024, https://counterhate.com/research/x-ran-ads-on-five-accounts-pushing-lies-and-hate-during-uk-riots/#about
"Musk's X helps Tommy Robinson rack up 434 million views during UK riots", The Center for Countering Digital Hate, 7 August 2024, https://counterhate.com/research/musks-x-helps-tommy-robinson-rack-up-434-million-views-during-uk-riots/
[2] "Social Media's Role in the UK Riots", Center for Countering Digital Hate, 27 Aug 2024, https://counterhate.com/research/social-media-role-in-the-uk-riots-osa-amendments/
[3] "Ofcom: Clear link between online posts and violent disorder", BBC News, 22 Oct 2024, https://www.bbc.co.uk/news/articles/c70w0ne4zexo
[4] "The X Rules: Safety, privacy, authenticity", X Help Centre, accessed 9 July 2025, https://help.x.com/en/rules-and-policies/x-rules
[5] "Update: Police incident, Water Street, Liverpool city centre - 26 May 2025", Merseyside Police, 26 May 2025, https://www.merseyside.police.uk/news/merseyside/news/2025/may-2025/update-police-incident-water-street-liverpool-city-centre/
[6] "Musk's X helps Tommy Robinson rack up 434 million views during UK riots", Center for Countering Digital Hate, 7 August 2024, https://counterhate.com/research/musks-x-helps-tommy-robinson-rack-up-434-million-views-during-uk-riots/
"X ran ads on five accounts pushing lies and hate during UK riots", Center for Countering Digital Hate, 19 August 2024, https://counterhate.com/research/x-ran-ads-on-five-accounts-pushing-lies-and-hate-during-uk-riots/
[7] "How Elon Musk Helped Fuel the U.K.'s Far-Right Riots", Rolling Stone, 31 Oct 2024, https://www.rollingstone.com/politics/politics-features/elon-musk-uk-far-right-riots-us-election-2024-1235148476/
[8] "X ran ads on five accounts pushing lies and hate during UK riots", Center for Countering Digital Hate, 19 August 2024, https://counterhate.com/research/x-ran-ads-on-five-accounts-pushing-lies-and-hate-during-uk-riots/
[9] ibid.
[10] "X ran ads on five accounts pushing lies and hate during UK riots", The Center for Countering Digital Hate, 19 August 2024, https://counterhate.com/research/x-ran-ads-on-five-accounts-pushing-lies-and-hate-during-uk-riots/#about
"Musk's X helps Tommy Robinson rack up 434 million views during UK riots", The Center for Countering Digital Hate, 7 August 2024, https://counterhate.com/research/musks-x-helps-tommy-robinson-rack-up-434-million-views-during-uk-riots/

[11] More details available in section 'Violent posts'
[12] More details available in section 'Violent posts'
[13] @noRegrekt, X, 27 September 2024, https://x.com/noRegrekt/status/1839462203868557517
[14] @PhuongLoc3, X, 25 August 2024, https://x.com/PhuongLoc3/status/1827642316653580340
[15] @mrollins1952, X, 1 January 2025, https://x.com/mrollins1952/status/1874550861101502905

[16] "Violent Content", X Help Center, February 2025, https://help.x.com/en/rules-and-policies/violent-content
[17] More detail available in section 'Key Accounts'
[18] More detail available in section 'Key Accounts'
[19] More detail on this finding can be found in the section 'Case Study: Liverpool victory parade car crash'
[20] More detail on this finding can be found in the section 'Case Study: Liverpool victory parade car crash'
[21] "X ran ads on five accounts pushing lies and hate during UK riots", The Center for Countering Digital Hate, 19 August 2024, https://counterhate.com/research/x-ran-ads-on-five-accounts-pushing-lies-and-hate-during-uk-riots/#about
"Musk's X helps Tommy Robinson rack up 434 million views during UK riots", The Center for Countering Digital Hate, 7 August 2024, https://counterhate.com/research/musks-x-helps-tommy-robinson-rack-up-434-million-views-during-uk-riots/
[22] View figures were available for just 517 of the 4,379 violent replies in our study. Those 517 posts amassed 383,102 views in total. The number of views for the full set of 4,379 is unknown given that views data was not accessible for many posts, but would be higher.
[23] @noRegrekt, X, 27 September 2024, https://x.com/noRegrekt/status/1839462203868557517
[24] @PhuongLoc3, X, 25 August 2024, https://x.com/PhuongLoc3/status/1827642316653580340
[25] @mrollins1952, X, 1 January 2025, https://x.com/mrollins1952/status/1874550861101502905
[26] "Violent Content", X Help Center, February 2025, https://help.x.com/en/rules-and-policies/violent-content
[27] "Violent Content", X Help Center, February 2025, https://help.x.com/en/rules-and-policies/violent-content
[28] "How the far right hijacked Southport protests", Sky News, 6 August 2024, https://news.sky.com/story/how-the-far-right-hijacked-southport-protests-13191492
[29] Tommy Robinson GB, X, 1 January 2025, https://x.com/TRobinsonNewEra/status/1874518878757511591
[30] "X ran ads on five accounts pushing lies and hate during UK riots", The Center for Countering Digital Hate, 19 August 2024, https://counterhate.com/research/x-ran-ads-on-five-accounts-pushing-lies-and-hate-during-uk-riots/#about
"Musk's X helps Tommy Robinson rack up 434 million views during UK riots", The Center for Countering Digital Hate, 7 August 2024, https://counterhate.com/research/musks-x-helps-tommy-robinson-rack-up-434-million-views-during-uk-riots/
[31] "The real story of the news website accused of fueling riots", BBC, 8 August 2024, https://www.bbc.co.uk/news/articles/c5y38gjp4ygo
"UK's worst riots in years were incited by online disinformation about asylum seekers", NPR, 10 August 2024, https://www.npr.org/2024/08/10/nx-s1-5066896/uks-worst-riots-in-years-were-incited-by-online-disinformation-about-asylum-seekers
"From rumours to riots: How online misinformation fuelled violence in the aftermath of the Southport attack", Institute for Strategic Dialogue, 31 July 2024, https://www.isdglobal.org/digital_dispatches/from-rumours-to-riots-how-online-misinformation-fuelled-violence-in-the-aftermath-of-the-southport-attack/

"Social media fuelled Southport misinformation, UK home secretary says", Financial Times, 5 August 2024, https://www.ft.com/content/6b886570-1b55-4647-8b6a-c4ad4c5b6925

[32] Tommy Robinson, X, 30 July 2024, https://x.com/TRobinsonNewEra/status/1818328857587413359

[33] Paul Golding, X, 29 July 2024, https://x.com/GoldingBF/status/1817944026122715319?lang=en-GB

[34] Ashlea Simon, X, 29 July 2024, https://x.com/AshleaSimonBF/status/1817985447999717593

[35] Andrew Tate, X, 29 July 2024, https://twitter.com/any/status/1818011145548218552

[36] Laurence Fox, X, 29 July 2024, https://x.com/LozzaFox/status/1818038598069002467

[37] Calvin Robinson, X, 2 August 2024, https://x.com/calvinrobinson/status/1819403003742638178

[38] "About Subscriptions creators", X, accessed 30 April 2025, https://help.x.com/en/using-x/subscriptions-creator

[39] "X Verification requirements", X, accessed 4 April 2025, https://help.x.com/en/managing-your-account/about-x-verified-accounts
"The X Rules", X, accessed 4 April 2025, https://help.x.com/en/rules-and-policies/x-rules

[40] Laurence Fox, X, 20 December 2024, https://x.com/LozzaFox/status/1870195584902086735

[41] Fr Calvin Robinson, X, 3 November 2024, https://x.com/calvinrobinson/status/1853026047987531951

[42] Tommy Robinson, X, 1 January 2025, https://x.com/TRobinsonNewEra/status/1874518878757511591

[43] Paul Golding, X, 17 January 2025, https://x.com/GoldingBF/status/1880344484954710506

[44] Andrew Tate, X, 7 January 2025, https://x.com/Cobratate/status/1876717170069221553

[45] Ashlea Simon, X, 25 September 2025, https://x.com/AshleaSimonBF/status/1838828057030795479

[46] "Nearly 50 injured after car ploughs into crowd at Liverpool FC victory parade", The Guardian, 26 May 2024, https://www.theguardian.com/uk-news/2025/may/26/man-arrested-car-crowd-liverpool-fc-football-fans

[47]"Nearly 50 injured after car ploughs into crowd at Liverpool FC victory parade", The Guardian, 26 May 2025, https://www.theguardian.com/uk-news/2025/may/26/man-arrested-car-crowd-liverpool-fc-football-fans

[48] "Why police released details about Liverpool crash suspect so quickly", BBC, 27 May 2025, https://www.bbc.co.uk/news/articles/cvgv4ddpyddo

[49] "Police told to do better to fill information void like one that fuelled Southport riots", Press Gazette, 9 May 2025, https://pressgazette.co.uk/news/police-told-to-do-better-to-fill-information-void-like-one-that-fuelled-southport-riots/

[50] Facebook, Liverpool City Region Combined Authority, 27 May 2025, https://www.facebook.com/watch/?v=1415798062661074

[51] "Did social media fan the flames of riot in Southport", BBC, 31 July 2024, https://www.bbc.co.uk/news/articles/cd1e8d7llg9o

[52] False or misleading posts in the case study received 29,149,098 views in total. This compares to 24,661,539 views for credible posts, including posts from news sources

that received 17,871,997 views and posts from officials or emergency services that received 6,789,542 views.

[53] Credible posts in the case study received 24,661,539 views in total, including posts from news sources that received 17,871,997 views and posts from officials or emergency services that received 6,789,542 views. This compares to 29,149,098 views for false or misleading posts in the case study.

[54] More details on how the news sites were selected can be found in the Methodology. "News Reliability Ratings", NewsGuard, accessed 17 June 2025, https://www.newsguardtech.com/solutions/news-reliability-ratings/
Top websites in the United Kingdom (Newspapers Industry), Semrush, accessed 17 June 2025, https://www.semrush.com/trending-websites/gb/newspapers

[55] Ashlea Simon, X, 4 September 2024, https://twitter.com/AshleaSimonBF/status/1831270984370045304/video/1

[56] Paul Golding, X, 30 November 2024, https://x.com/GoldingBF/status/1862881793747100102

[57] Calvin Robinson, X, 20 December 2024, https://x.com/calvinrobinson/status/1870198404401479903

[58] Andrew Tate, X, 23 January 2025, https://x.com/Cobratate/status/1882447329074307379

[59] Laurence Fox, X, 20 December 2024, https://x.com/LozzaFox/status/1870195169611644944

[60] Hank, X, 26 May 2025, https://twitter.com/HANKonX/statuses/1927096037258244518

[61] Hank, archive.today, 27 May 2025, https://archive.ph/7hkp9

[62] James Goddard, X, 26 May 2025, https://twitter.com/JamesPGoddard90/statuses/1927097883565736403

[63] The Yorkshire Lass, X, 26 May 2025, https://x.com/real_shirelass/status/1927096023484182991

[64] "Why are there riots in the UK?", BBC, 7th August 2024, https://www.bbc.co.uk/news/articles/ckg55we5n3xo

[65] "Liverpool FC parade: How the crash unfolded", Sky News, 27 May 2025, https://news.sky.com/story/liverpool-fc-parade-how-the-crash-unfolded-13375262

[66] Top websites in the United Kingdom (Newspapers Industry), Semrush, accessed 17 June 2025, https://www.semrush.com/trending-websites/gb/newspapers

[67] "News Reliability Ratings", NewsGuard, accessed 17 June 2025, https://www.newsguardtech.com/solutions/news-reliability-ratings/

Published July 2025
© Center for Countering Digital Hate Inc