ISD | Powering solutions to extremism, hate and disinformation
CASM technology

Bi-monthly **Report #2,**
October-November 2024

US Online Domestic Violent Extremism Monitor

Bi-Monthly Data Snapshot of Cross-Platform Social Media Activity

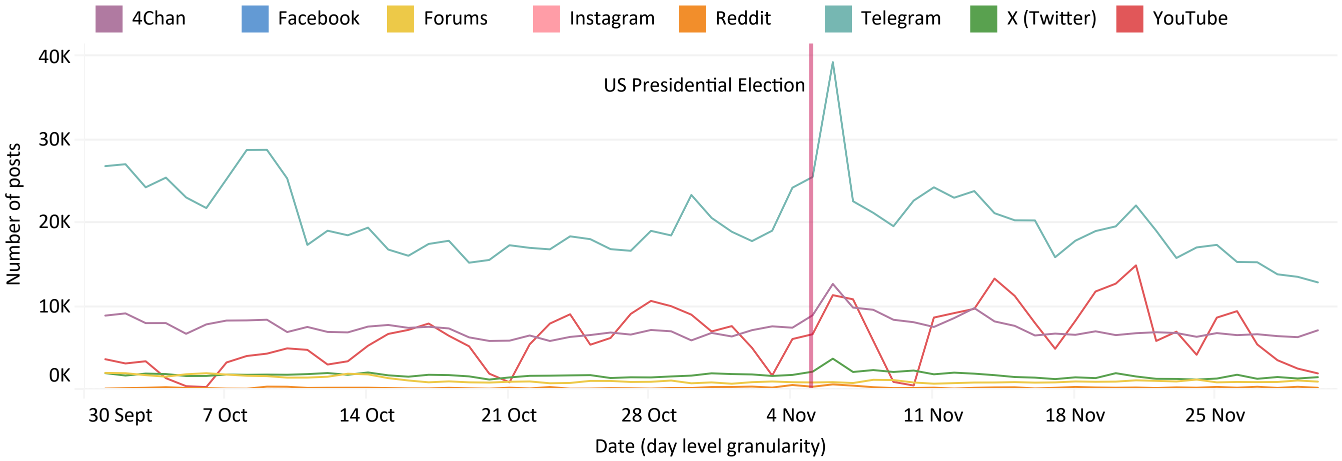
Executive Summary

This bi-monthly data snapshot overviews key trends in US domestic violent extremist (DVE) activity online, providing an up-to-date picture of the evolving digital threat landscape. This report is divided into the following sections: (1) key platforms; (2) threat actors; (3) semantic mapping of key topics; (4) targeted hate; (5) domestic and international events animating activity; and (6) methodology. Analysis is based on a dataset of over 1,000 US-linked accounts and channels across numerous platforms and ideologies that were manually vetted by experts as engaging in clear violent extremist behavior. All insights are anonymized and presented in aggregate, with personally identifiable information removed at the point of data collection. By identifying key themes and platforms, researchers aim to provide practitioners with data that can help inform prevention efforts. Understanding the relationship between patterns in online DVE activity and real-world developments can help practitioners allocate resources more efficiently. Analysis of the evolving narratives espoused by DVEs might allow practitioners to better understand potential entry points for helping individuals disengage from violence. Additionally, cross-platform data can inform practitioners' understanding of the different risks across online spaces, and where prevention efforts might be focused.

Findings

During October and November 2024, online DVE actors produced over 2.2 million posts, with activity peaking during the days around the US presidential election. A summary of key findings is provided below:

Platform Activity:
- DVE activity peaked on Telegram, 4chan, and YouTube on Election Day, which were the three most active platforms during this period. Across all platforms, Telegram had the most DVE activity, comprising 54% of all posts, while 4chan's /pol/ board came in second, comprising 20% of all activity.
- DVEs on YouTube generated substantial comment activity, with over 381,000 comments, followed by Instagram (76,327), Reddit (20,854), and Facebook (7,551).



Volume-over-time graph showing the total number of posts per platform

Threat Categories:
- Among threat actors, accounts classified as 'Other Domestic Violent Extremism' generated the most posts during this period, while activity by Racially or Ethnically Motivated Violent Extremists (REMVEs) declined by 35% compared to the last report, likely stemming from account takedowns or abandonment. Anti-Government or Anti-Authority Violent Extremists (AGAAVEs) continued to generate the most engagement.
- 'Other' DVEs were primarily motivated by a wide range of conspiracy theories, while Single-Issue Violent Extremists were most motivated by the Israel/Palestine conflict.

Key Narratives:
- Topic modelling analysis based on trained classifiers showed that DVE accounts were most frequently animated by discussions around targeted hate (13.8% of messages), government and politics (2.1%), conspiracy theories (1.0%), and conflict and geopolitics (1.0%).
- DVEs were particularly animated by the US presidential election, as well as by Hurricanes Helene and Milton and the expanding war in the Middle East.

Targeted Hate:
- Antisemitism continued to be the most prominent form of targeted hate within our dataset, with most of this hate being circulated by 'Other' DVE accounts.
- Anti-LGBTQ+ hate was the second-most prominent form of targeted hate, with a plurality of relevant activity occurring on 4chan. Mobilizing narratives included accusations associating this community with pedophilia and calls for the removal of trans individuals from the military.

Domestic and International Influences:
- Domestic DVEs were particularly animated by the widespread destruction caused by Hurricanes Helene and Milton, which drove conspiracy theories attacking federal response efforts. It is therefore crucial that in the wake of a natural disaster, practitioners prepare for an increase in conspiratorial narratives, schisms between communities, and potential disruptions to government aid.
- Internationally, US DVEs as well as US-based accounts showing support for Foreign Terrorist Organizations (FTOs) reacted strongly to the expanding war between Israel and Iranian-backed regional groups, with antisemitism and anti-Muslim rhetoric featuring prominently in such discourse.

Approach

Data was collected through platforms' public Application Programming Interfaces (APIs) and varies considerably by platform. Data analysis (explained in full in the methodological annex) incorporates the use of Large Language Models (LLMs) to identify key trends in violent extremist discourse, including prominent narratives, the targets of violent extremist activity, and the nature and extent of hate speech targeting minority communities. This research was supported by the U.S. Department of Homeland Security, Science and Technology Directorate, under Grant Award Number 23STFRG00021. Any opinions or conclusions contained herein are those of the authors and do not necessarily reflect those of the Department of Homeland Security, Science and Technology Directorate.

Platforms

Overview

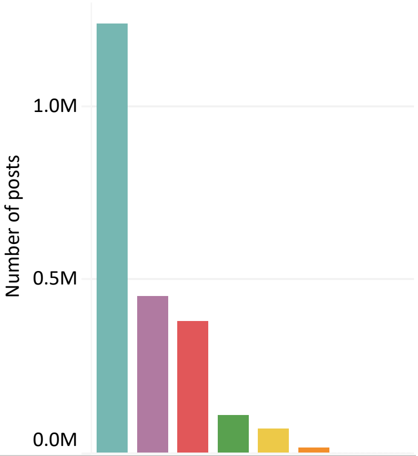
As with our last report, analysis showed Telegram and 4chan as the most active platforms in this monitoring period's dataset. Telegram remained the most active platform, featuring over 1.2 million posts by DVE accounts. Its ease-of-use and relative lack of content moderation continue to appeal to domestic extremists, despite them expressing concerns that the arrest of CEO Pavel Durov would jeopardize their operational security. In total, 54% of all online DVE activity occurred on Telegram, with total posts decreasing by 3.4% from the previous monitoring period.

Meanwhile, 4chan's /pol/ board remains a prominent platform for anonymous and unmoderated discussion among online extremists; given its size, we collected only from US-based users using keywords related to hate and extremism or associated with groups and communities that are regularly harassed on the board. Due to the anonymous and amorphous nature of accounts on the platform, individual 4chan users were not categorized into specific threat categories, as with other platforms.

Telegram, 4chan, and YouTube saw spikes in engagement by DVE accounts around the election compared to other platforms, while X saw only a modest increase and other platforms remained steady.

Our analysis also included over 70,000 posts from five forums that feature prominent levels of discussion related to domestic extremism (hereafter referred to collectively as "forums," which we have opted not to name to avoid undue amplification). These forums accounted for 3% of the messages in our dataset. Activity on these platforms peaked on October 6-7 and featured heavy use of antisemitic language, driven largely by the anniversary of the October 7 attacks.

Of note, there was a 33% decline in activity on X compared to the previous monitoring period, which was in large part due to platform moderation, deactivation, and abandonment of accounts. YouTube and Reddit activity declined from the last report following keyword filters being applied to comments on DVE-related YouTube videos and subreddits, thereby excluding posts not featuring keywords related to hate and extremism.



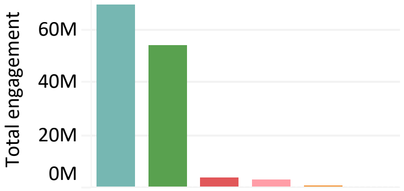
Telegram	1,234,176
4Chan	453,246
YouTube	381,517
X (Twitter)	108,081
Forums	70,999
Reddit	17,214
Instagram	697
Facebook	605

Bar graph & corresponding tabular view showing total posts per platform

Platforms

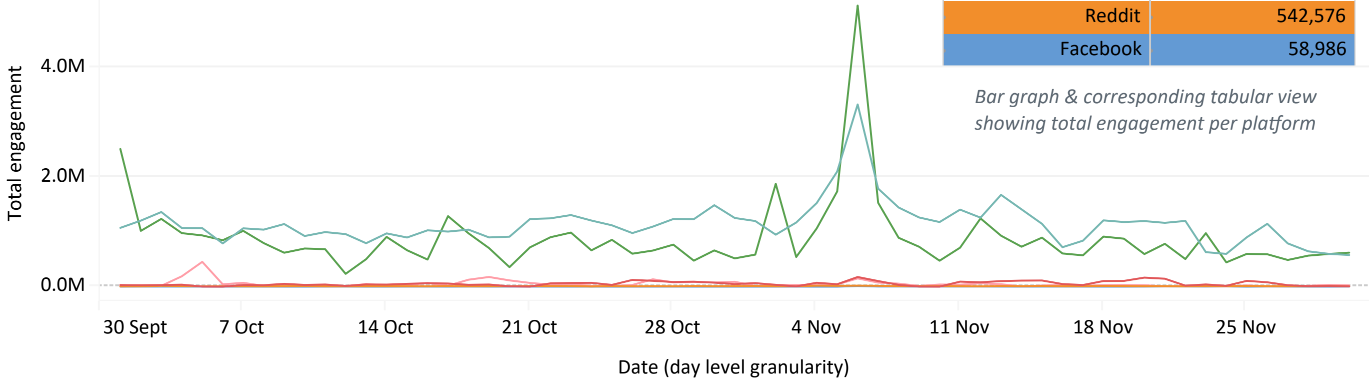
Engagement

The types and availability of engagement data (e.g. comments, likes, shares) vary dramatically by platform, but still provide valuable insight into narratives gaining traction among DVEs online. Across platforms, X and Telegram saw considerable engagement during this review period. DVE accounts on X totaled more than 13.3 million followers (down by nearly 400,000 followers from the last report), and their posts received nearly 54 million likes (up by nearly 30%). Meanwhile, on Telegram, DVE content received over 24.2 million likes and nine billion views, which constitutes an increase in views by over 31% since the last report. YouTube generated substantial engagement in the form of comments (over 381,000), followed by Instagram (over 76,000) and Facebook (over 7,500). The five forums analyzed in this report did not provide engagement metrics.



Telegram	69,547,295
X (Twitter)	53,800,176
YouTube	3,295,276
Instagram	3,025,039
Reddit	542,576
Facebook	58,986

Bar graph & corresponding tabular view showing total engagement per platform

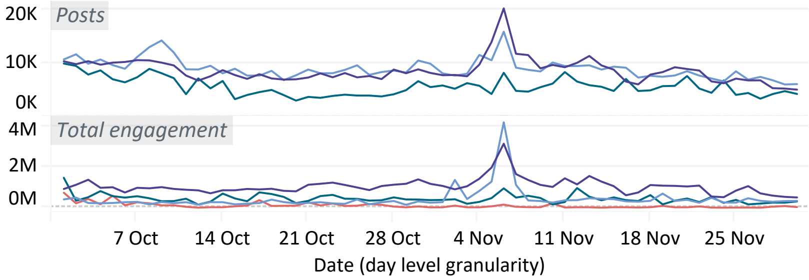


Volume-over-time graph showing total engagement per platform

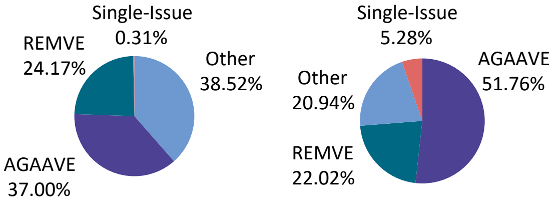
Threat categories

Overview

Drawing on terminology outlined by DHS and other US government agencies, ISD manually classified DVE accounts and channels into four threat categories: (1) Racially or Ethnically Motivated Violent Extremism; (2) Anti-Government or Anti-Authority Violent Extremism; (3) Single-Issue Violent Extremism (sub-categorized into Animal Rights, Environmental, Abortion, and Israel/Palestine-related Violent Extremism); and (4) Other Domestic Violent Extremism (including violent conspiracy movements).



Volume-over-time graphs showing the number of posts (top) and total engagement (bottom), by threat category



Circle charts showing the % of posts (left), and total engagement (right) by threat category

Because of how data was collected, the snapshots in this section exclude analysis from 4chan's /pol/ board, where anonymous posting and more ephemeral account identity make actor-based analysis more challenging. It also excludes YouTube and Reddit comment data, as accounts were categorized on a channel and subreddit-level respectively, rather than by individual commenters. Categories are ordered below from most to least active as revealed by analyzing the dataset:

Other Domestic Violent Extremism

545,172 *posts*
32 *active accounts*
26.48M *total engagement*

This category encompasses threats involving the potentially unlawful use or threat of force or violence in furtherance of ideological agendas which are not otherwise defined under or primarily motivated by one of the other domestic threat categories. In October and November, actors within this category were the most active in terms of messages produced, and their activity increased by over 20% compared to the last reporting period. However, this category had the second lowest number of active accounts, partly due to the presence of several highly active forums and group chats that centered around violent conspiratorial discussion. Within this category, 56% of posts were made by violent conspiracy theorists, while only 2% were made by accounts driven by misogynistic or anti-LGBTQ+ views. The lack of a clear animus and the prevalence of conspiracies in this cohort highlights the need for practitioners to craft targeted approaches based on critical, integrative thinking to allow for multiple perspectives and information from competing sources of information.

Anti-Government or Anti-Authority Violent Extremism (AGAAVE)

523,564 *posts*
109 *active accounts*
65.44M *total engagement*

AGAAVE encompasses the potentially unlawful threat of violence in furtherance of ideological agendas derived from anti-government or anti-authority sentiment, including opposition to perceived economic, social, or racial hierarchies, or perceived government overreach, negligence, or legitimacy. During this period, AGAAVE accounts were the second-most active among the threat categories based on messages produced (up 2.4% from the last report) and number of active accounts. Despite this, content produced by these accounts constituted nearly 85% of the views across all threat categories, showing considerable engagement. The overwhelming majority of AGAAVE actors continue to be motivated by grievances against their perceived political enemies, including US political parties. Notably, Proud Boys accounts produced nearly 19% more content compared to the last reporting period. These posts reflect the lead up to, and immediate aftermath of, the presidential election and practitioners should maintain an awareness of AGAAVE narratives, though it is feasible that animosity towards authority will reduce post-inauguration.

Racially or Ethnically Motivated Violent Extremism (REMVE)

342,066 *posts*
245 *active accounts*
27.84M *total engagement*

REMVE encompasses the potentially unlawful use or threat of force or violence in furtherance of ideological agendas derived from bias, often related to race or ethnicity, held by the actor against others or a given population group. During the monitoring period, REMVE was the third-most active threat category in terms of messages produced but had the highest number of active accounts. Compared to the last reporting period, posts by REMVE actors declined by nearly 35%, which appears to have been caused by the takedown or abandonment of certain prolific accounts. Relatedly, activity by neo-Nazi accelerationist accounts declined by nearly 72% from the last report. Of note, almost all REMVE accounts active in this period were motivated by white supremacy. Given the real-world harm implications of support or advocacy of REMVE violence, practitioners should consider prioritizing individuals interacting with these ideologies and focusing on interventions that focus on a community-based approach.

Single-Issue Violent Extremism

4,334 *posts*
31 *active accounts*
6.68M *total engagement*

This threat type is divided into the following sub-categories: Animal Rights-Related Violent Extremism; Environment-Related Violent Extremism; Abortion-Related Violent Extremism; and Israel/Palestine-Related Violent Extremism, a new category ISD defines as domestic violent extremist actors who are singularly motivated by the ongoing Israel-Hamas conflict. In October and November, actors motivated by Single-Issue Extremism were the least active in terms of both messages produced and number of active accounts. The most active accounts in this category were motivated by Israel/Palestine, followed by those motivated by the environment and abortion. Notably, despite the Israel/Palestine conflict driving considerable discussion across DVE actors, posts from Single-Issue accounts motivated by this issue declined by over 42% during this reporting period, likely stemming from account takedown or abandonment. Practitioners should consider exposing individuals to the deep, historical roots of antisemitism in an effort to separate the politics and conflicts of the Middle East from the faiths of those involved.

Semantic mapping

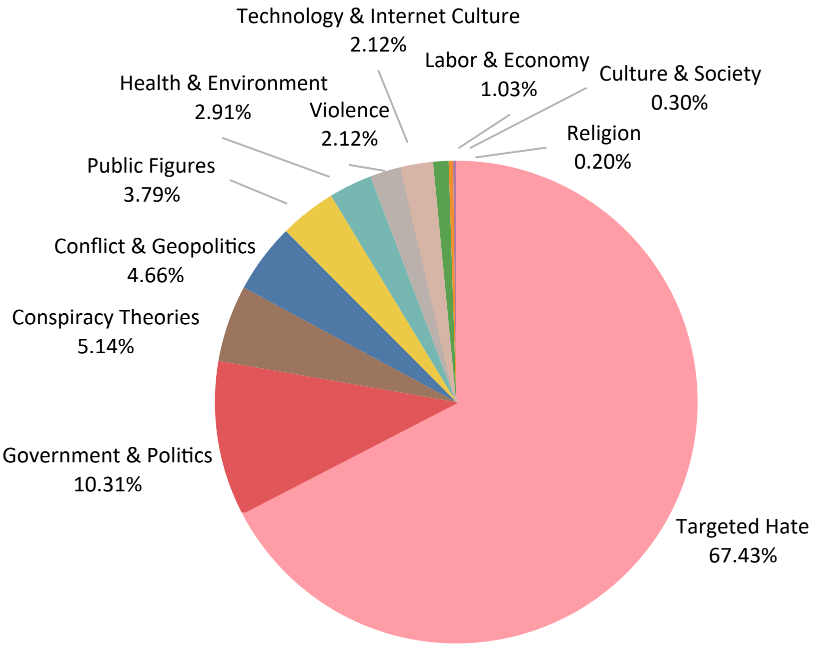
Summary

Bespoke Large Language Models (LLMs) were used to group messages into semantically distinct clusters to streamline the analysis of key narratives, with researchers now using LLMs to help characterize clusters of DVE messages to aid in identifying these themes. Analysis once again identified 11 overarching themes, encompassing 69 sub-themes, up from 44 in August and September. Within the Conflict & Geopolitics theme, for example, the LLM identified nine sub-themes that generated significant conversation, including discussions around Hamas, Israel, and Hezbollah.

As with the last briefing, many messages posted by DVEs were grouped into themes not directly related to hate, extremism, or violence, but were still deemed relevant for understanding the key narratives driving DVE activity. Further, semantic clusters were not treated as discrete categories—for example, anti-Muslim posts falling under the Targeted Hate theme may be imbued with racist rhetoric or feature discourse around the war in Gaza.

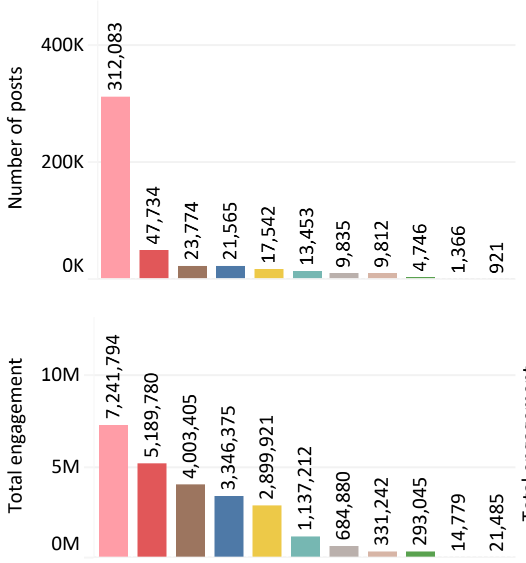
Of the 11 master themes, discussions relating to Targeted Hate were the most common theme identified within the data, followed by conversations around Government & Politics and Conflict & Geopolitics. In terms of engagement, content within the Targeted Hate theme received the most likes, with Government & Politics-related content earning the most comments and Conspiracy Theory-related activity earning the most views.

Analysis of the most prominent sub-themes showed that, as in our previous monitoring report, messages exhibiting antisemitism and anti-LGBTQ+ hate (both examples of it and discussions about it) were the most frequent in the dataset. Beyond Targeted Hate-related topics, conversations around Elections and Voting predominated in the days and weeks preceding and following the US elections. Notably, DVE discourse related to Elections and Voting received more than double the views that the next-closest sub-theme generated, which related to a prominent political figure.

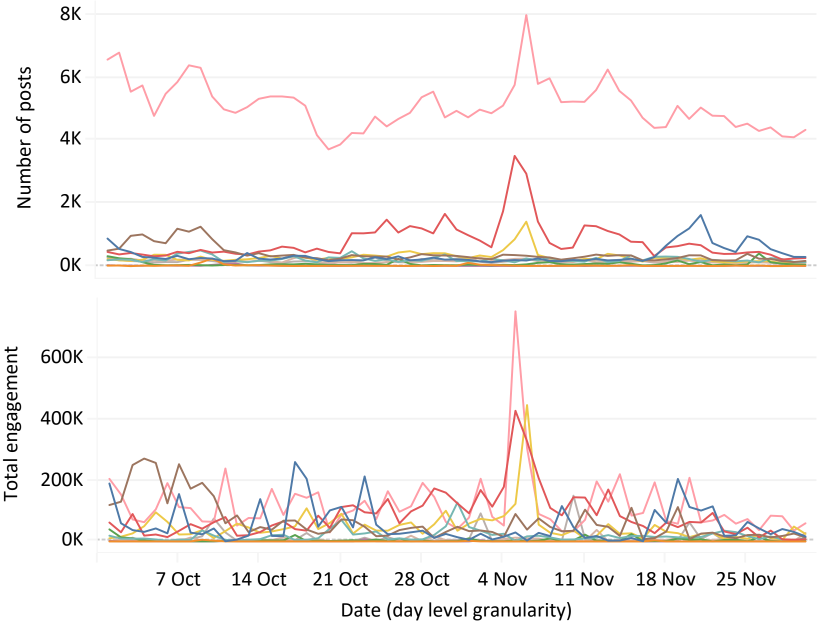


Circle chart showing the % of posts by theme

- Targeted Hate
- Government & Politics
- Conspiracy Theories
- Conflict & Geopolitics
- Public Figures
- Health & Environment
- Violence
- Technology & Internet Culture
- Labor & Economy
- Culture & Society
- Religion



Bar charts showing the number of posts (top) and total engagement (bottom) by theme

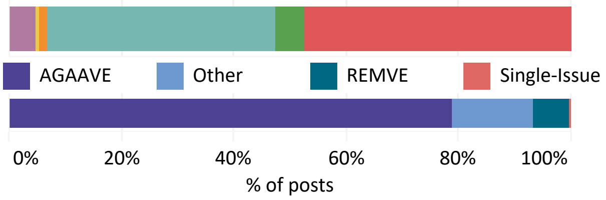


Volume-over-time graphs showing the number of posts (top) and total engagement (bottom) by theme, for the top 5 themes (by post count)

Semantic mapping - key themes

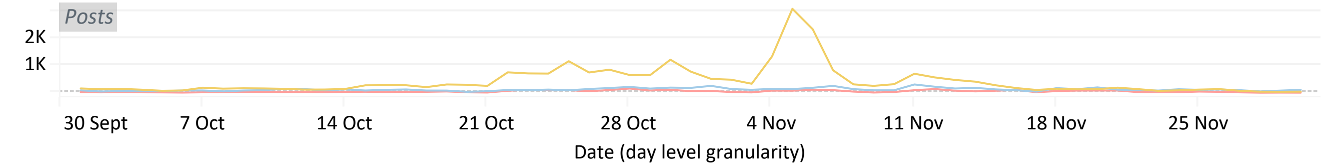
Government & Politics

With the US presidential election occurring in early November, messages related to elections and voting accounted for over 55% of all content within the government and politics master theme. Notably, AGAAVE accounts were the main drivers of election-related conversation, accounting for nearly 62% of the posts within this sub-theme. Collectively, messages from these actors generated nearly 320 million views. High-engagement posts included accusations of voter fraud occurring in swing states on Election Day. Notably, DVE conversations related to government and politics predominantly occurred within YouTube comments (50% of such discussion) and on Telegram (42%).



Rank	Sub-theme	Posts	Accounts
1	Elections and voting	26,095	12,120
2	Political Ideology	7,809	3,443
3	US Political Parties	3,520	2,555

Table showing the top 3 sub-themes by post count for this theme (above); & bar charts showing the % proportion of posts of this theme per threat category (left - bottom) and per platform (left - top)

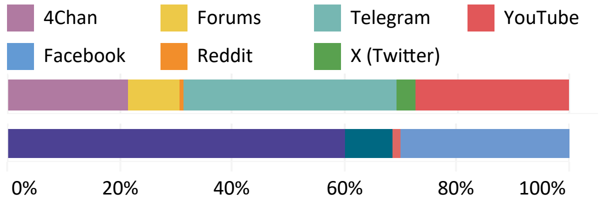


Volume-over-time graph showing the number of posts by sub-theme, for the top 3 sub-themes (by post count) for this theme

Semantic mapping - key themes

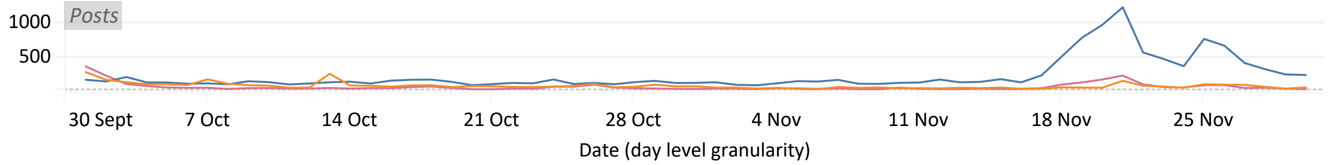
Conflict & Geopolitics

Online DVE activity related to conflict and geopolitics revolved primarily around the conflict between Russia and Ukraine, which accounted for 62% of the discussion within this theme, and the Israel-Palestine conflict and its spillover throughout the Middle East, which comprised 25% of such discussion. DVE posts related to Russia and Ukraine, largely criticizing Ukraine and US support for it, spiked in mid-November after the Biden Administration lifted restrictions on Ukraine's use of Western-made, long-range missiles to target Russia. Prominent narratives related to the Israel-Palestine conflict and its broader regional impacts included criticism of Israel's conduct in the war, which often utilized antisemitic rhetoric condemning Israelis and Jews writ large. Among platforms, Telegram featured the most discussion related to conflict and geopolitics (38%), followed by YouTube (27%) and 4chan (22%).



Rank	Sub-theme	Posts	Accounts
1	Russia/Ukraine	13,283	7,015
2	Israel/Palestine	3,804	1,099
3	War	2,869	1,401

Table showing the top 3 sub-themes by post count for this theme (above); & bar charts showing the % proportion of posts of this theme per threat category (left - bottom) and per platform (left - top)

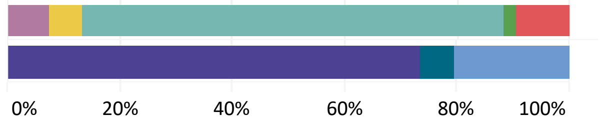


Volume-over-time graph showing the number of posts by sub-theme, for the top 3 sub-themes (by post count) for this theme

Semantic mapping - key themes

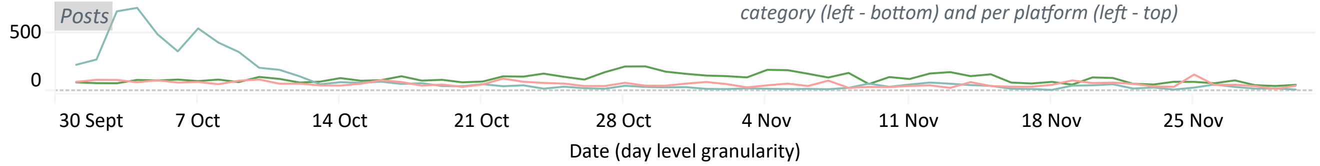
Conspiracy Theories

Conspiratorial discussion among DVEs was wide-ranging, encompassing topics ranging from climate-related conspiracies to antisemitic beliefs alleging Jewish control over the US government. However, QAnon-related discussions as well as conspiracy theories related to US federal agencies emerged as prominent sub-themes, together comprising over half of conspiracy-related posts. Specific narratives that gained considerable traction included accusations that FEMA prioritized the safety of illegal immigrants rather than US citizens when conducting disaster relief efforts in the aftermaths of Hurricanes Helene and Milton, as well as claims that Satanists are in control of important sites and facilities around the world. Of note, conspiracy-related discussions among DVEs largely occurred on Telegram, which accounted for nearly 75% of these discussions across platforms.



Rank	Sub-theme	Posts	Accounts
1	QAnon	6,385	291
2	Federal Agencies	6,259	797
3	Child Sexual Abuse	3,550	742

Table showing the top 3 sub-themes by post count for this theme (above); & bar charts showing the % proportion of posts of this theme per threat category (left - bottom) and per platform (left - top)



Volume-over-time graph showing the number of posts by sub-theme, for the top 3 sub-themes (by post count) for this theme

Targeted hate

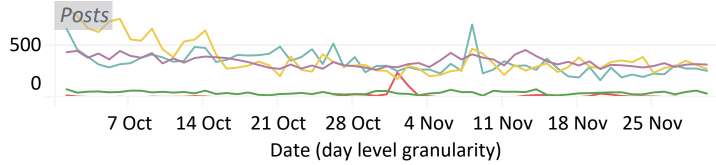
Antisemitism

Antisemitism again represented the most prominent form of targeted hate in our dataset, with such language appearing in 2.6% of all messages. Notably, 40% of all antisemitic discourse in the dataset occurred on DVE-centric forums, while 4chan (filtered by a set of keywords likely to surface extremism and targeted hate) and Telegram messages made up 36% and 18% of antisemitic activity respectively. 78% of all antisemitic posts came from 'Other' DVEs driven by violent conspiracy theories, while 17% came from REMVE actors. Posts with the highest engagement included conspiratorial messages alleging Jewish control of media, financial, and government institutions as well as Israeli influence over US politics and foreign policy.

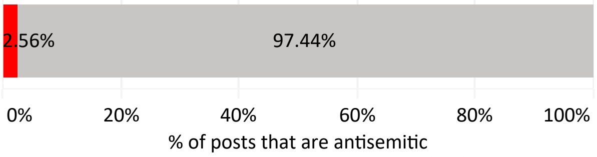
57,959
posts

20,768
active accounts

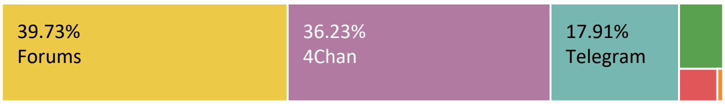
2,201,413
total engagement



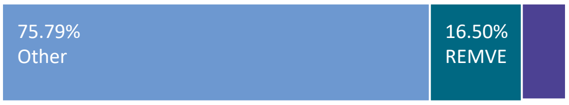
Volume-over-time graph showing the number of antisemitic posts by platform



Bar chart showing the % of antisemitic (red) posts



Treemap showing number of antisemitic posts per platform



Treemap showing number of antisemitic posts per threat category

🟪 4Chan	🟧 Reddit	🟩 X (Twitter)
🟨 Forums	🟩 Telegram	🟥 YouTube

🟪 AGAAVE 🟦 Other 🟦 REMVE 🟥 Single-Issue

Targeted hate

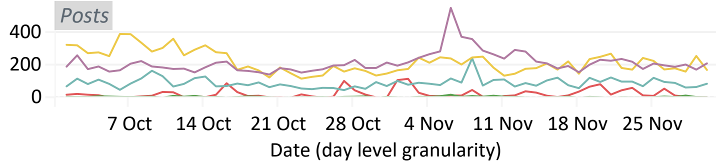
Anti-LGBTQ+ Hate

Anti-LGBTQ+ hate represented the second-largest form of targeted hate in our dataset, appearing in 1.6% of messages. Hateful rhetoric targeted exclusively at trans individuals comprised roughly a third of anti-LGBTQ+ content. Such content was most prominent on 4chan and DVE-centric forums and generated more engagement than anti-Muslim posts, but significantly less than antisemitic material. Prominent narratives featured in the dataset included vague calls for violence against LGBTQ+ people, conspiracy theories associating this community with pedophilia, and calls for the removal of trans individuals from the military. 'Other' DVE accounts were the most active among threat actors in posting anti-LGBTQ+ content, a sign of the online DVE landscape becoming increasingly defined by unclear ideological affiliations.

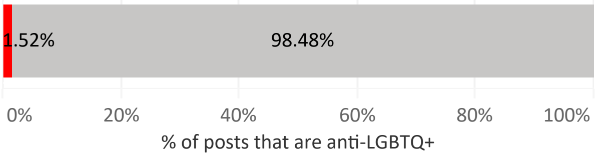
34,423
posts

14,480
active accounts

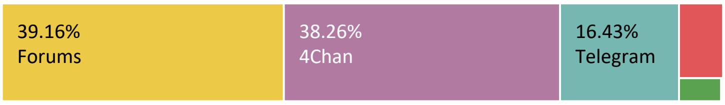
1,140,762
total engagement



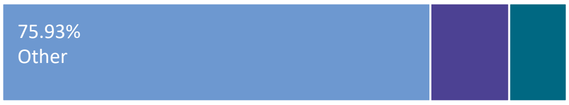
Volume-over-time graph showing the number of anti-LGBTQ+ posts by platform



Bar chart showing the % of anti-LGBTQ+ (red) posts



Treemap showing number of anti-LGBTQ+ posts per platform



Treemap showing number of anti-LGBTQ+ posts per threat category

Targeted hate

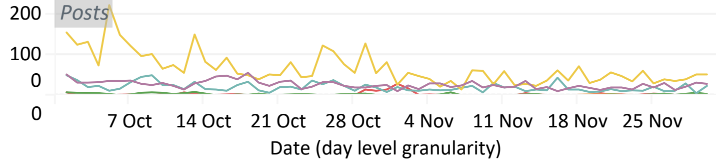
Anti-Muslim Hate

Anti-Muslim language appeared in 0.3% of the messages in our dataset, with 55% of messages coming from DVE-centric forums. 22% of relevant activity occurred on 4chan, while 17% was observed on Telegram. Unlike the previous monitoring period, 'Other' violent extremists posted the vast majority of anti-Muslim hate, constituting 89% of such activity. REMVE accounts were a distant second, accounting for 9% of all anti-Muslim messages. Prominent anti-Muslim narratives included calls for the deportation of Muslims, the use of racist and Islamophobic language to describe Islamic culture and practices, and references to Muslims as "invaders." Overall, anti-Muslim posts received significantly less engagement than those that were antisemitic or anti-LGBTQ+ in nature.

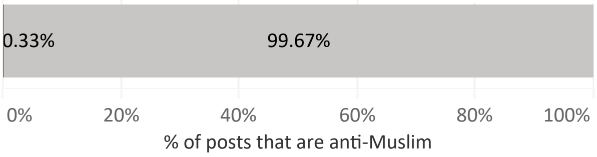
7,390
posts

1,873
active accounts

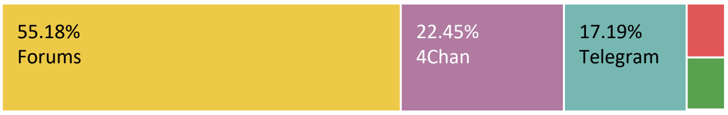
29,219
total engagement



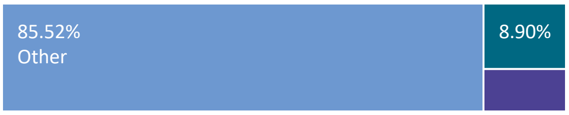
Volume-over-time graph showing the number of anti-Muslim posts by platform



Bar chart showing the % of anti-Muslim (red) posts



Treemap showing number of anti-Muslim posts per platform



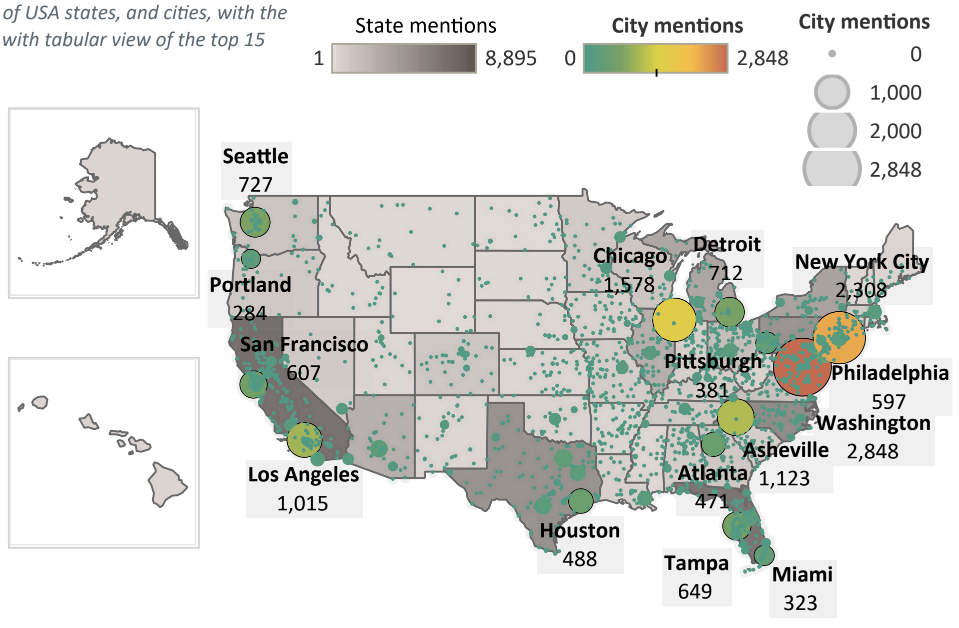
Treemap showing number of anti-Muslim posts per threat category

Domestic lens

Geographic overview

Choropleth showing the number of mentions of USA states, and cities, with the top 15 most-mentioned cities labelled; along with tabular view of the top 15 most-mentioned states

Rank	US state	Mentions
1	California	8,895
2	Florida	8,461
3	North Carolina	6,478
4	Texas	5,994
5	Pennsylvania	5,808
6	New York	4,513
7	District of Columbia	4,339
8	Michigan	4,163
9	Illinois	3,384
10	Arizona	3,369
11	Ohio	2,189
12	Virginia	1,867
13	Washington	1,842
14	Colorado	1,619
15	Wisconsin	1,589

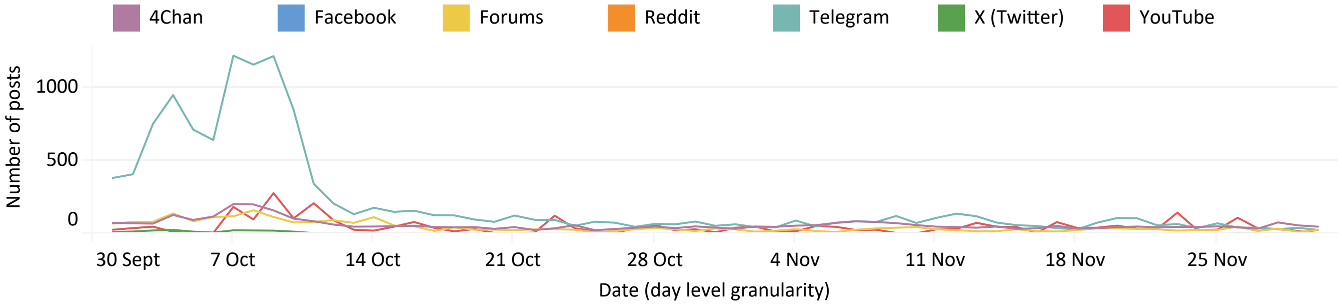


Domestic lens

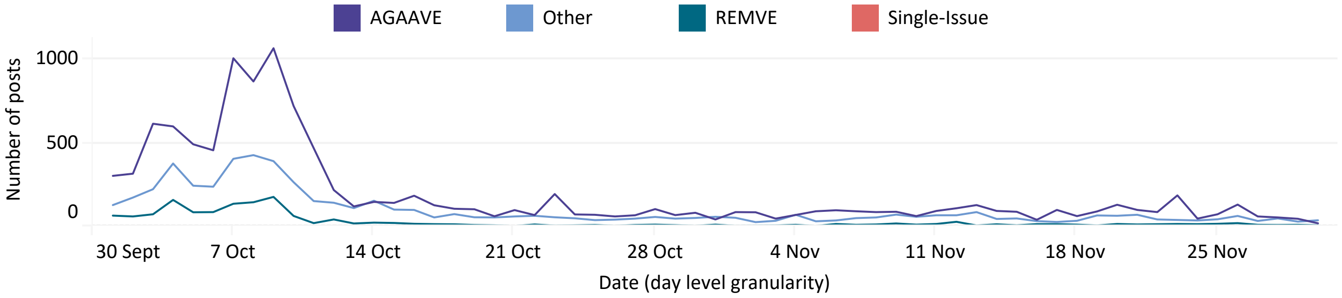
Spotlight - DVE Responses to Hurricanes Helene and Milton

The widespread destruction caused by Hurricanes Helene and Milton and the ensuing disaster response efforts carried out by the federal government triggered a sharp increase in online discussion among DVEs in early October. These conversations were led by AGAAVE accounts and were particularly prominent on Telegram. Collectively, DVEs produced over 20,000 messages related to the hurricanes, accruing more than 100 million views.

Relevant posts surged in early October, eventually petering out and stabilizing from October 12 onward. An analysis of these messages revealed that DVEs capitalized on the destruction to spread conspiracy theories, disinformation, and extremist rhetoric intended to fuel hostility and violence against the federal government. A REMVE account on X claimed that the US government provided ten times the amount of money to every Israeli citizen this year as it did to Hurricane Helene victims, concluding with the suggestion that the US government hates its own citizens. This post accumulated over 30,000 likes and 700,000 views. Additionally, an AGAAVE account on Telegram claimed that FEMA's evacuation orders in Florida were part of a broader plan to lock people out of the state, which would involve force if necessary. This post garnered nearly 340,000 views. Meanwhile, multiple accounts attempted to showcase disaster response efforts conducted by extremist groups, stating that such efforts were superior to those of the federal government. This activity shows that practitioners must be prepared for the potential for natural disasters to exacerbate conspiratorial thinking, stir conflict between certain constituencies, and possibly even impede the delivery of critical aid.



Volume-over-time graph showing the number of posts by platform, filtered to just posts mentioning at least one hurricane-related keyword (i.e., FEMA; hurricane; hurricanes; Helene; and Milton; in any casing)



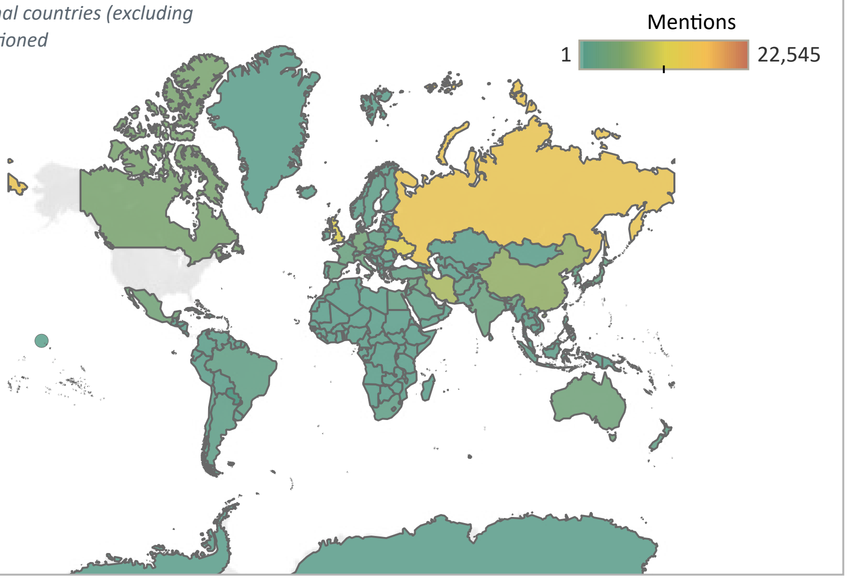
Volume-over-time graph showing the number of posts by threat category, filtered to just posts mentioning at least one hurricane-related keyword (i.e., FEMA; hurricane; hurricanes; Helene; and Milton; in any casing)

International influence

Geographic Overview

Choropleth showing the number of mentions of international countries (excluding USA), along with tabular view of the top 15 countries mentioned

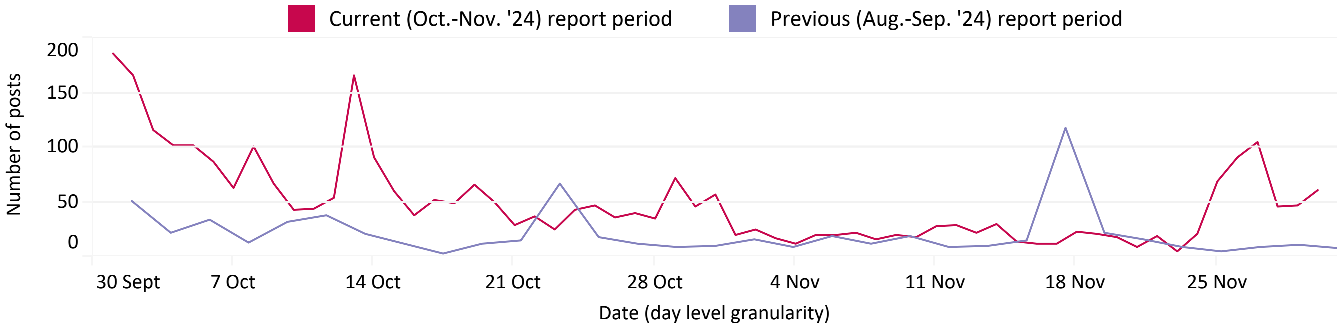
Rank	Country	Mentions
1	Israel	22,545
2	Russian Federation	14,845
3	Ukraine	12,416
4	United Kingdom	11,818
5	Iran	8,193
6	China	6,821
7	Canada	4,829
8	Germany	3,618
9	Mexico	3,608
10	Palestine	3,515
11	Australia	3,280
12	Italy	2,678
13	Lebanon	2,612
14	India	2,319
15	France	2,229



International influence

Spotlight - DVE Responses to Expanding Middle East Conflict

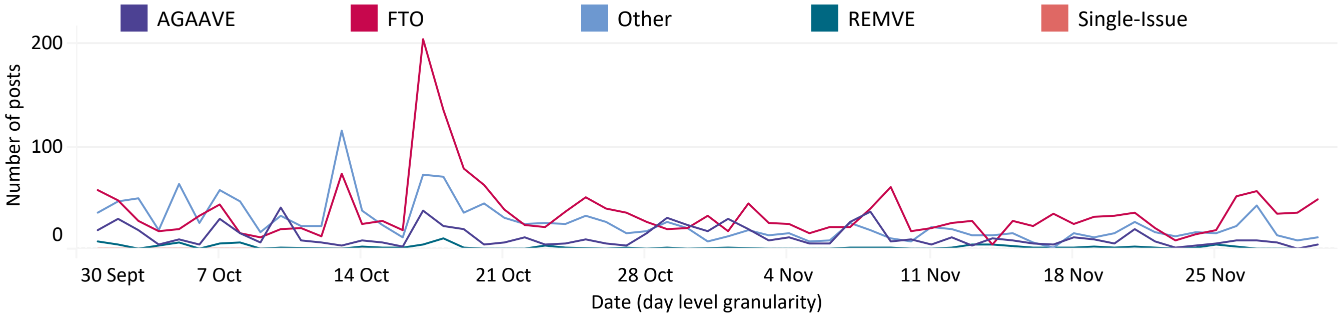
The expansion of the Israel-Hamas war into a broader regional conflict triggered strong reactions from US-based DVEs, with 'Other' DVE accounts representing the most active threat category in mentions of Israel (62%), Hamas leader Yahya Sinwar (52%), Hezbollah (66%), and Hamas (73%). In early October, DVE mentions of Israel spiked after it invaded Lebanon to push Hezbollah back from the Israel-Lebanon border, prompting retaliatory strikes from Iran. Notably, mentions of Israel increased nearly 27% from the previous monitoring period, while Hezbollah mentions increased by 34%.



Volume-over-time graph showing the number of posts, filtered to just posts mentioning "Hezbollah" (in any casing), comparing the current (Oct.-Nov. '24) report period to the previous (Aug.-Sep. '24) report period.

One Telegram post that received significant engagement celebrated an Israeli soldier's death and condemned the soldier's alleged killing of children, one of many posts celebrating any loss or failure by Israel as a win against a perceived pariah state. Another Telegram post proclaimed that Israelis "are not people," while another called for Israel's destruction. On X, DVE posts with high engagement called Israel "evil" and used racist and antisemitic language, with one post receiving tens of thousands of views alone.

We also analyzed US-centric accounts that did not fit into a DVE threat category but showed support for Foreign Terrorist Organizations (FTOs), most of which showed support for the broader Iranian resistance network, the "Axis of Resistance" (91%). Activity occurred exclusively on X, with most content expressing anger towards Israel and praising Hamas and Hezbollah for their resistance. Pro-FTO accounts were most active on Oct 1, with some highly engaged-with posts celebrating Yahya Sinwar as a leader and martyr. For example, one post received more than 160,000 views after posting "Long live Sinwar!".



Volume-over-time graph showing the number of posts by threat category, filtered to just posts mentioning "Hamas" (in any casing). The largest spike was driven by FTO-supporting DVE accounts on the day after Hamas leader Yahya Sinwar was killed.

Appendix

Methodology

Account Selection

ISD's and CASM's research drew on data from over 1,000 violent extremist accounts and forums. To be included in analysis, accounts or forums had to meet the following criteria:

1. Advocating for an extremist ideology or worldview.
2. Promoting terrorism or unlawful violence, are operated by a group or movement with a history of violence, or are supporting designated Foreign Terrorist Organizations (FTOs).
3. Are operated by individuals or groups based in the US or which produced content primarily focused on the US.

All accounts were rigorously reviewed based on these criteria by at least two expert analysts prior to inclusion. To compile this list of accounts, analysts:

1. Began by using existing lists of accounts maintained by ISD from previous research into domestic violent extremism.
2. Identified additional accounts through targeted keyword searches in posts and user biographies, aimed at capturing violent extremist issues and groups across a broad ideological spectrum.
3. Expanded this existing list of accounts using network analysis to identify additional relevant accounts. Specifically, we analyze interactions between the original accounts and others they engage with, such as those they share, reply to, or mention. By identifying the most frequently linked accounts, we generate a larger pool of candidate accounts. These accounts are then manually reviewed by analysts for relevance, ensuring that only pertinent accounts are added to the updated list.

PII Removal

This work deployed technological approaches for removing personally identifiable information (PII) at the point of data collection, with several robust measures taken so sensitive data was properly anonymized while maintaining the integrity of the dataset. We focused on the removal of locations (to the zip code level and below), names, and other obvious PII like credit card information. This included both metadata and free-text fields.

- For free-text data, we employed Microsoft's Presidio anonymization tool, which is specifically designed to identify and remove PII from text. Presidio allowed us to automate the detection and removal of various PII elements, including personal names, locations, and other sensitive identifiers.
- At the same time, a curated list of over 1,000 public figures—primarily key political figures—was compiled and integrated into the process. These names were not redacted due to their analytical utility.
- Outbound URLs were shortened to preserve the information but not allow potential PII to remain in free text or metadata.

Our approach leaned towards over-removal of content to ensure compliance and protect privacy. While this occasionally resulted in the inadvertent removal of words that were not PII, the overall impact on the research was minimal.

Data Collection

- Data was collected from October 1st – November 30th, 2024:
- Through official API endpoints for Telegram, Reddit, YouTube and 4Chan.
- Through third-party tool BrandWatch for X, Facebook, Instagram, which employs the platforms' APIs.
- Through third-party tool BrandWatch for Forums.

Our collection and storage of data was compliant with GDPR, the US. Privacy Act, and all platforms' terms of service.

Datasets

Different subsets of data are used in this report:

1. The "wider dataset" refers to all messages collected from accounts within the report time window, and includes comments collected on Reddit posts and YouTube videos. Additionally, this dataset contains posts from forums and US tagged accounts on the 4Chan /pol/ board, messages from these sources are only included if they match hate target keywords.
2. The "hate analysis subset" includes all messages from the "wider dataset", except a 10% random sample of 4Chan is used for tractability.
3. The "sampled subset" includes a stratified sample of 322,147 messages from the "wider dataset" and is used to build the thematic classifier.

To generate the "sampled subset", data is randomly sampled from the "wider dataset" on a per-platform and message-type basis, with the aim to include a reasonable number messages from the "wider dataset" while making processing tractable.

The sampling process takes:

- 100,000 X/Twitter messages
- 100,000 Telegram messages
- 70,999 (all) Forum posts
- 45,262 4Chan messages (a random 10 percent of data matching hate target keywords)
- 578 (all) YouTube videos
- 1,790 YouTube comments (all data matching hate target keywords)
- 1,525 (all) Reddit posts
- 691 Reddit comments (all data matching hate target keywords)
- 697 (all) Instagram messages
- 605 Facebook (all) messages

Named Entity Recognition (NER)

This process extracts mentions of people, locations, and organizations from the text after the PII removal process to identify references to prominent figures and places above the ZIP code level.

- **Organizations**: We used a language model from SpaCy (en_core_web_lg) to automatically find all organization names in the text.
- **Persons**: We compared the text to a pre-approved list of people's names and added both the version of the name found in the text and the official name it matches.
- **Locations**: A combination of Microsoft's Presidio and CLIFF geoparser were used to identify and preserve location text that was considered sufficiently coarse granularity to not be considered PII, such as third-order administrative divisions and above and capitals of political entities. Granularity was determined using Geonames feature codes; one of ADM1, ADM1H, ADM2, ADM2H, ADM3, ADM3H, PCL, PCLD, PCLF, PCLH, PCLI, PCLS, PPLA, PPLA2, PPLA3, PPLC, PPLCH, or PPLG. For further categorising locations into countries, US states, and US cities, we applied the Mordecai3 geoparsing tool on this redacted text, which extracted: **Countries**: The country code and the original country name mentioned; **US States**: The state code and the original state name mentioned; **US Places**: The formal name of any U.S. location found and its original mention in the text.

Semantic Mapping

We followed a topic modelling approach in which semantically related messages are placed into topic clusters which are then manually assessed and broken into sub-themes (e.g. Christianity) and themes (e.g. Religion). Topic modelling represents each message numerically such that messages with similar (mathematically close) representations have similar semantics. Such numerical representations are referred to as embeddings.

Appendix

Methodology

CASM computed embeddings using “BAAI/bge-m3” due to its open accessibility, widespread adoption and ability to handle longer text length than many other models. We began by applying semantic mapping to a sample dataset of 322,147 messages. This ultimately yielded 294 clusters of messages with semantically similar text, corresponding to varying narratives, claims, themes, tactics, etc.

For assessing topic clusters and deriving the themes and subthemes, we randomly sampled 100 messages per topic cluster. Analysts then identified sub-themes by reading through messages in the sample, and removing clusters of noisy data (e.g. clusters with just one word of text) or those with completely irrelevant discussion (e.g. skateboarding).

Having isolated dozens of analytically useful sub-themes, analysts then discussed and came to agreement on how to break these sub-themes into 11 overarching themes. The most relevant themes – Government & Politics, Conflict & Geopolitics, and Conspiracy Theories – were analyzed further in a dashboard allowing analysts to layer on additional analytically useful data, such as Named Entity Recognition, engagement data, volume over time, hate classification and others.

Thematic Classification

A nearest neighbour classifier was used to assign topic clusters, subthemes, and themes to messages in the “wider dataset”. In this approach, messages are classified based on their proximity to messages in the “sampled subset”, analysed in semantic mapping. We use scikit-learn’s NearestNeighbors implementation with ‘cosine’ metric and 10 neighbours. We set a minimum cosine similarity threshold of 0.65 for neighbour retrieval. Labels are assigned using the majority label of neighbours.

Hate Classification

Hate classification was performed over the following three phases:

1) Keyword Filtering

The “hate analysis subset” was annotated via keywords / phrases. This subset of data includes all messages from the “wider dataset”, except a 10% random sample of 4Chan is used for tractability.

- For antisemitic content, 164 exact words/phrases and 176 partial matches were applied to this filtered dataset to identify discussions about Jews or Judaism, resulting in 97,862 flagged messages.
- For anti-Muslim content, 56 exact keywords and 181 partial matches were used to identify mentions of Muslims or Islam, yielding 35,009 flagged messages.
- For anti-LGBTQ+ content, 229 exact keywords were used to identify mentions of the wider community or specific sub-groups (e.g. transgender individuals), yielding 18,398 anti-trans and 57,749 anti-LGBQ+ flagged messages. While analysis grouped trans and the LGBQ+ communities into a wider LGBTQ+ community, two separate classifiers were used for each of these two sub-groups. This is due to wide linguistic differences between discussion of LGBQ+ and transgender individuals and associated issues.

2) Classifier Selection and Evaluation

Samples of messages from each of the keyword-filtered datasets (Antisemitic, anti-Muslim, anti-LGBQ+ and anti-trans) were manually labelled. The labelled samples were then used to evaluate a combination of LLM models, prompts (tailored to each target using keywords and examples), and hyperparameters to optimize the precision and recall for each target group. The precision and recall for hateful messages for each classifier, as evaluated on a evaluation set is as follows:

- Antisemitic: 0.78 precision, 0.91 recall
- Anti-Muslim: 0.86 precision, 0.84 recall
- Anti-LGBQ+: 0.74 precision, 0.88 recall
- Anti-trans: 0.88 precision, 0.97 recall

3) Classification

The optimal setup from the previous step is then used to label each entire dataset. The final messages from each set deemed hateful are as follows:

- Antisemitic: 68,511 (70% hateful)
- Anti-muslim: 7,846 (22% hateful)
- Anti-LGBQ+: 23,352 (40% hateful)
- Anti-trans: 12,707 (69% hateful)