

Hate & Extremism Insights Aotearoa (HEIA)

Dr. Chris Wilson



Whakataki – Introduction

Hate and Extremism Insights Aotearoa (HEIA) conducts data-led research to measure and analyse harmful online rhetoric. HEIA is based at the University of Auckland and led by Dr Chris Wilson. This paper takes a snapshot in time – from 1 January through 30 June 2022 – to deliver insights into online extremist behaviour in Aotearoa New Zealand.

HEIA's research draws on data from a range of social media sites, including Telegram, Gab, 8Kun (formerly known as 8chan), Reddit, and 4Chan. Given the very large volume of contemporary online extremist interaction, HEIA uses various methods to identify genuine negative emotion and toxicity from general online rhetoric. The outcome is a more contextualised understanding of the online violent extremist landscape in Aotearoa.

By identifying a subset of posts which indicate higher levels of emotion, HEIA can focus on a body of data which is not only more manageable in size but comprises the most dangerous trends and topics among local violent extremists. In turn, this allows more targeted and effective interventions for preventing and countering online harm.

The Royal Commission into the terrorist attack on two Christchurch masjid outlined the pressing need for a better understanding of the local violent extremist landscape. HEIA is delivering a local capability in Aotearoa to inform this important mahi now and into the future.

This research was conducted by HEIA with the support of the Digital Safety Group in Te Tari Taiwhenua, Department of Internal Affairs.

Tāera – Methodological Note

This report uses two natural language classifiers to determine the emotion behind posts. These classifiers produce a 'negativity' score and a 'toxicity' score. The negativity score captures the intensity of negative emotions such as anger, fear, loneliness, guilt and jealousy conveyed by the language used in the post. The toxicity score captures how likely the post as a whole is to be regarded as rude, disrespectful or unreasonable to the extent that an ordinary reader would want to leave the conversation.

Both scores provide insight into the users' emotional investment in their posts, thereby identifying those which indicate higher levels of emotion, commitment to extremist ideas and potential for radicalisation and other harms. The emotions captured by the negativity score are those often experienced by individuals undergoing a process of extremist radicalisation. The toxicity score indicates a higher risk of online harm generated by the post.

Tirohanga – Report Insights

HEIA's study of online extremist rhetoric from 1 January to 30 June 2022 identified several key insights:

- We identify posts associated with three forms of extremist ideas and / or ideology: Far Right-related, Conspiracy Theory-related, and Misogyny-related;
- We identify a small subset of posts which can be classified as highly negative or toxic. We believe this smaller number of posts shows 1) a higher level of engagement with extremist ideas, 2) a higher level of emotion on the part of the posters involved and 3) indicate a greater risk of violent action;
- The three extremist ideologies peak and trough at different times, suggesting these ideological groups are at least partially distinct and respond to a range of different domestic and transnational ideas and events;
- Conspiracy Theory-related posts were the most numerous (53,246), followed by Far Right-related (15,039) and Misogyny-related posts (5,486);
- However, Far Right posters demonstrated the most negative emotion (45% of posts), followed by Conspiracy theory-related posts (36.39%) and Misogyny-related posts (32.68%);
- Misogyny-related posts were the most toxic (17%);
- There were fewer posters on fringe platforms (such as 4chan), but they are more active and negative than posters on more mainstream platforms (Telegram and Reddit);
- Overall, there has been a large drop in extremist rhetoric since mid-February 2022, driven largely by a steep decline in Conspiracy Theory-related discourse.

Wero - The Challenge

Aotearoa, like many countries, is awash in an online sea of extremist rhetoric. The anonymity and lightning speed of contemporary online interaction means that the threshold for expressing outrage, radicalism and hate is low. Thousands of individuals now regularly express extremist ideas online while acting in very different ways offline. Only a tiny minority of those who express hateful and extremist rhetoric online put their words into action.

Government agencies now face the extremely difficult task of identifying which of these numerous posts, topics, trends and interactions pose the greatest risk of violence and other harms. Agencies rarely possess the resources, time or social license to collate and analyse online data of this scale. What is required is a sophisticated method of identifying a subset of this data, a subset that is not only more manageable but comprises the most concerning and dangerous of online rhetoric.

Whakahoki - HEIA's Response

HEIA is a response to this challenge. Our response is built on five strategic pou (pillars):

- We are a local capability with Aotearoa-specific and global south expertise;
- We focus on Aotearoa-based online posters across a wide range of platforms;
- We don't stop at identifying the scale of the problem; we engage in sophisticated, theoretically and empirically informed analysis that can better assess potential online and offline harms to minorities in Aotearoa;
- We filter out the noise to focus on identifying the emotions, drivers and risk factors behind posts to identify a smaller but more dangerous subset of online extremism;
- We present our data in an easy-to-understand style.

Tātaritanga – Analysis

HEIA collected a total of 530,991 posts from 5 platforms over the first half of 2022. From these data, we identified posts discussing topics and keywords related to three forms of (potentially violent) extremist ideology: the Far Right, Conspiracy Theory and Misogyny-related posts. Using a range of methods, we identified which posts indicate concerning levels of negative emotion and toxicity.

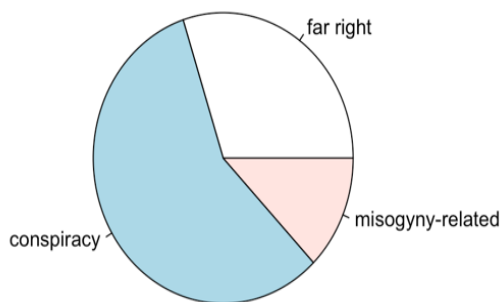
Three Extremist Ideologies

The HEIA dataset captured a high proportion of:

1. Conspiracy Theory-related ideas as well as;
2. Far Right rhetoric; and
3. Misogynistic narratives.

Figure 1

Relative Number of Users by Fringe View

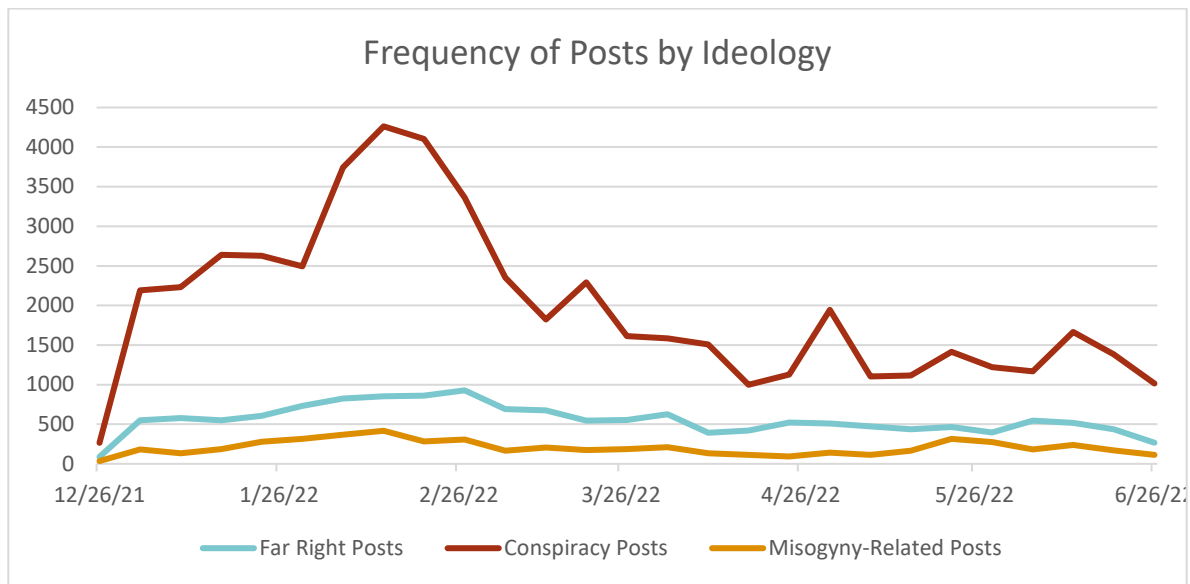


Users discussing Conspiracy Theory-related content were by far the most numerous over the first six months of 2022, followed by the Far Right and Misogyny-related posts.

That Conspiracy Theory-related ideas are over-represented in the data, together with other research, suggests that movements such as QAnon¹ remain a feature of an increasingly mixed local ideological extremist landscape.

¹ Refer ISD (2021) 'Understanding the New Zealand Online Extremist Ecosystem.'

Figure 2

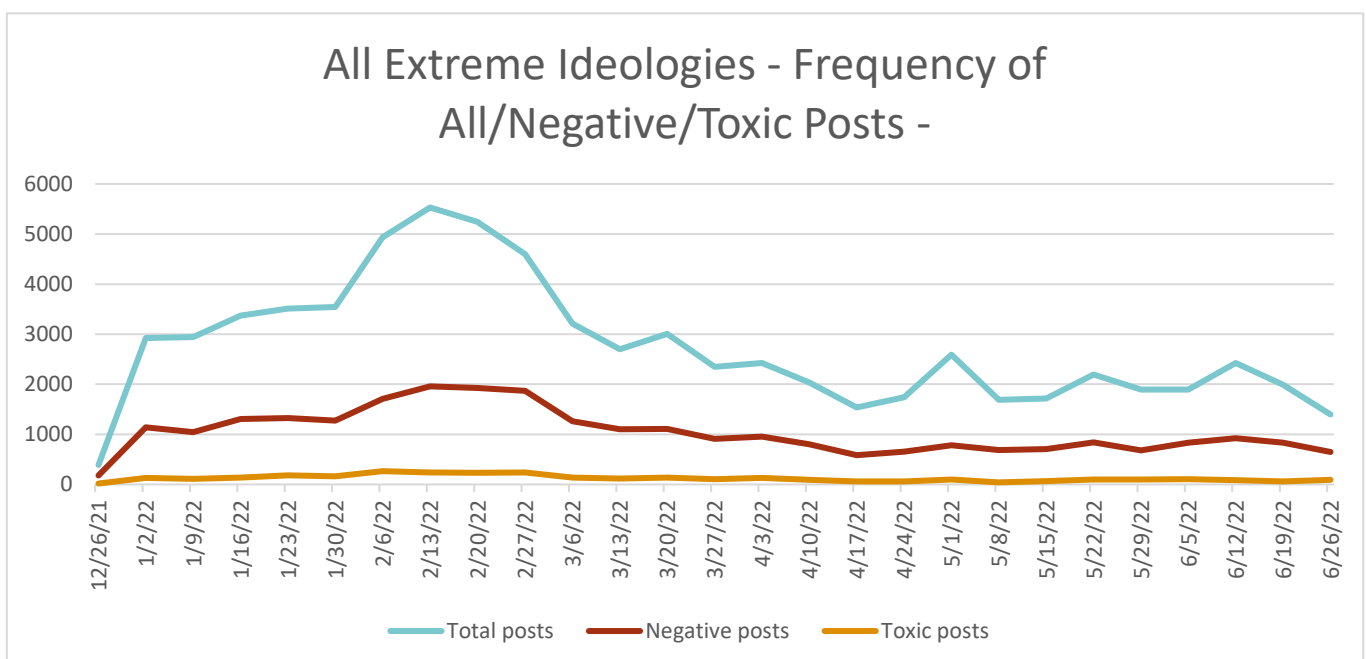


While Figure 2 shows the total number of posts associated with the three extremist ideologies, this only tells us part of the story.

It is important that we identify which of these posts indicate a strong commitment to the ideology and higher than normal levels of negative emotion. We manually set a threshold for concerning levels of negative emotion and toxicity through a qualitative examination of a random selection of posts.

In the following figures, we present our analysis of the proportion of those posts which are strongly negative and toxic. Each shows the number of posts related to the three ideologies at weekly intervals from 26 December 2021 to 26 June 2022. The first graph combines all extremist ideologies, and the following three present each ideology individually. Each figure also adds lines demonstrating how many posts meet our thresholds of concerning levels of negative emotion and toxicity.

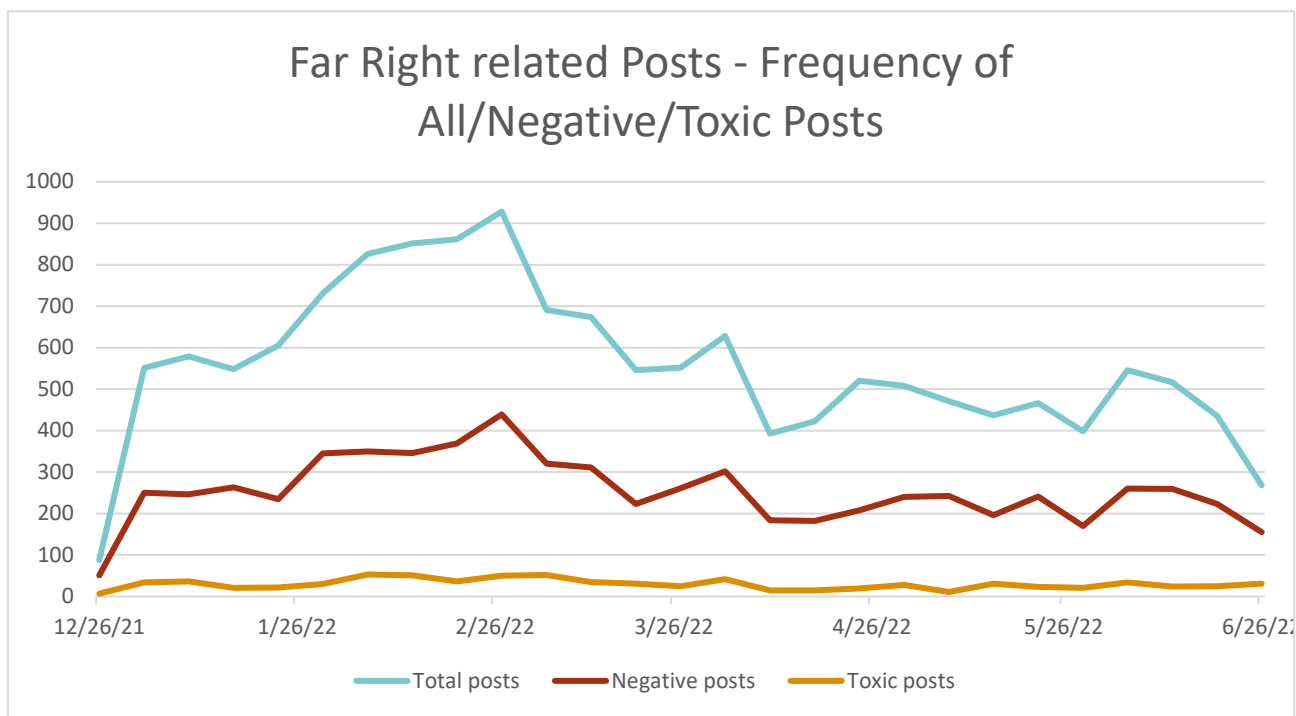
Figure 3



Over the six-month period, 73,711 posts discussed key topics related to either far right, conspiracy theory or misogynistic ideologies. As indicated, 28,040 (38%) posts demonstrate concerning levels of negative emotion, and 3,286 (4.4%) demonstrate concerning levels of toxicity.

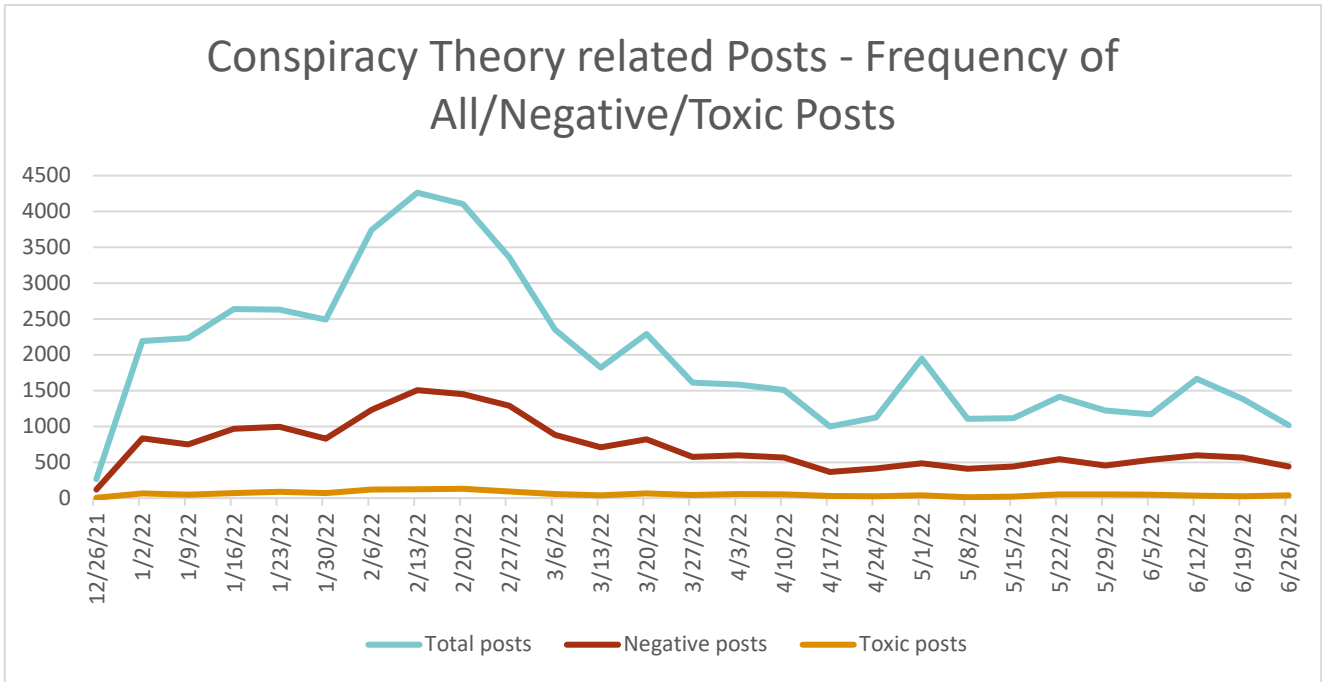
While these proportions are not inconsequential, identifying this subset allows for a more targeted analysis of trends, topics, convergence between ideologies, and risks of extremist violence. This analysis is continued for each ideology separately.

Figure 4



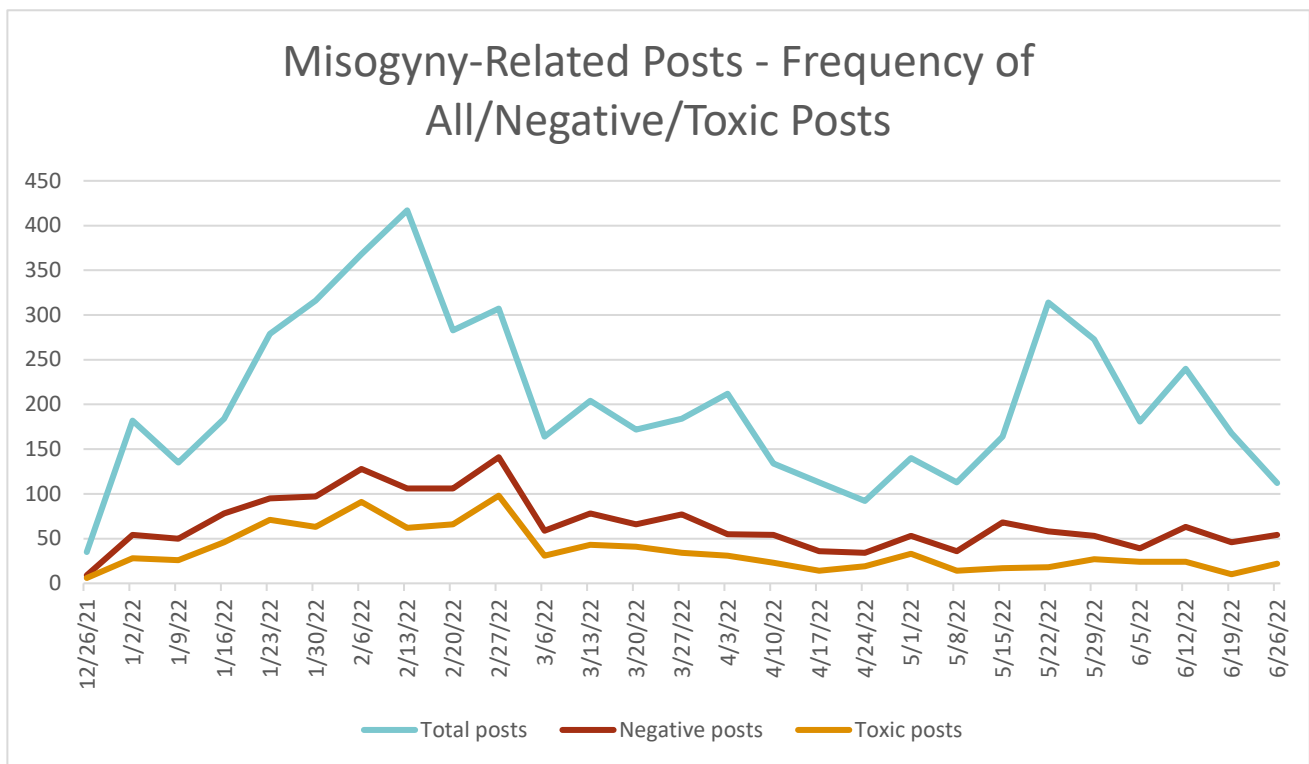
Of the 15,039 posts discussing far-right-related topics, 6,871 (45.68%) demonstrated concerning levels of negative emotion, and 802 (5.3%) demonstrated concerning levels of toxicity.

Figure 5



Of the 53,246 posts discussing conspiracy theory-related topics, 19,376 (36.39%) demonstrated concerning levels of negative emotion, and 1,502 (2.82%) demonstrated concerning levels of toxicity.

Figure 6



Of 5,486 posts discussing misogynistic extremism-related topics, 1,793 (32.68%) demonstrated concerning levels of negative emotion, and 962 (17.90%) demonstrated concerning levels of toxicity.

General findings

By far the highest number of total posts related to any ideology relate to conspiracy theories. Conspiracy Theory-related extremism also has the largest number of negative and toxic posts. However, as a proportion of total posts, Far Right-related posts demonstrate the highest levels of negative emotion. Misogynistic extremism-related posts have by far the highest proportion of toxic posts.

While the frequency of posts for each ideology follows a similar pattern over the six-month period, peaking in February 2022 and exhibiting a steady decline since then, differences can be discerned between the ideologies in shorter-term peaks and troughs.

Aside from a spike in all three ideologies in February (driven by the protests in Wellington and elsewhere), we have identified several differences. For example,

- Far Right-related posts spiked between 10 and 24 April and 29 May – 5 June
- Conspiracy-Theory posts spiked between 13 and 20 March 24 April – 1 May
- Misogyny-related posts rose between 8 and 22 May.

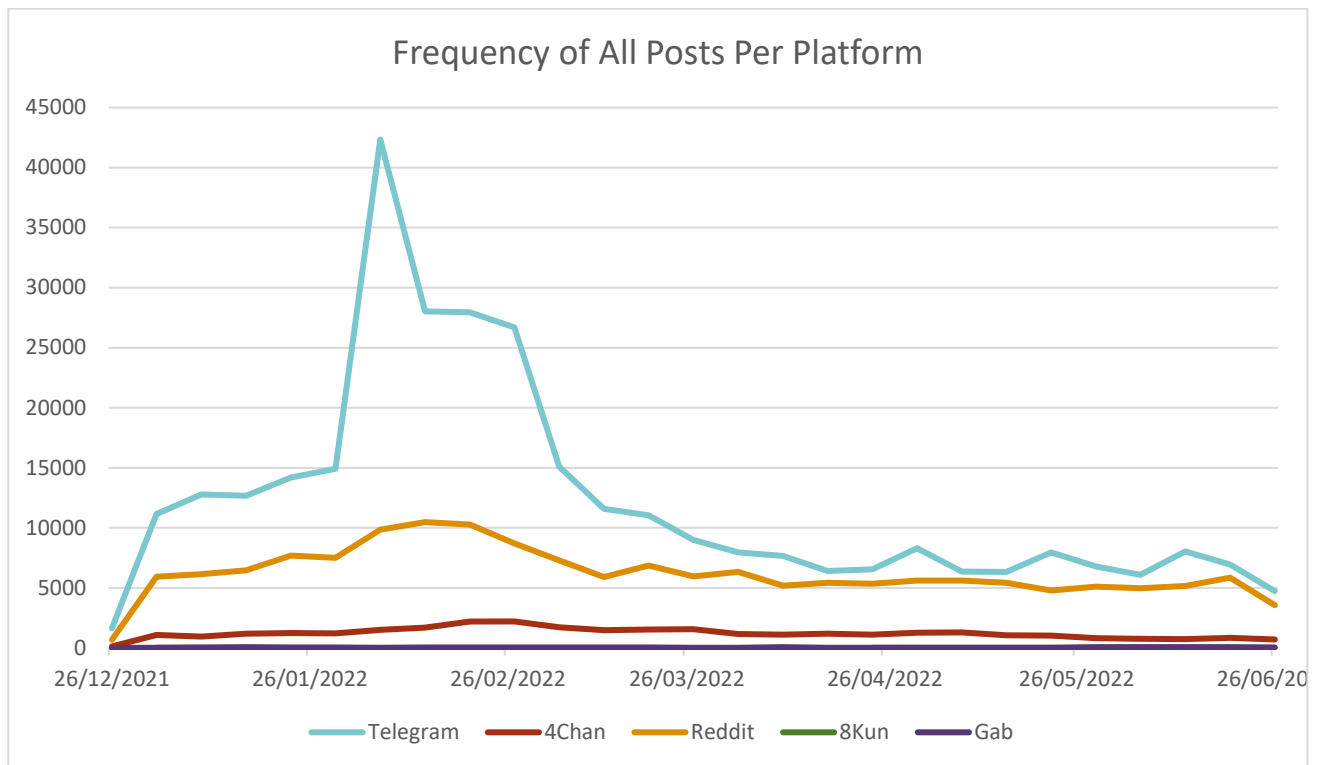
We hypothesise that this variation indicates that while the three ideologies have some overlaps, they are largely discrete, with subscribers responding to different online and offline events.

We further hypothesise that while all three ideologies respond to a mix of international and domestic events, the domestic context played a greater role in driving Conspiracy Theory-related extremism than it did with the two other forms of extremism. This is illustrated by the large spike surrounding the anti-mandate and associated protests in Wellington in February and March and the steep decline in Conspiracy Theory-related posts since then. In contrast, the number and emotion of far right and misogyny-related posts remained more consistent over the study period.

Analysis of Posting by Platform

The following section briefly presents patterns of posting behaviour by the platforms studied. Graph 5 illustrates the number of posts by platform over the six-month period.

Figure 7



As seen in Figure 7, while Telegram was the most used platform, this usage declined sharply from the start of March (the end of the Wellington protest). The other platforms have remained far more consistent over the study period. The following two figures show the number of posts categorised as demonstrating concerning levels of negative emotion or toxicity on each platform. Unsurprisingly, more fringe platforms such as 4chan and Gab have a higher proportion of negative and toxic posts.

Figure 8: Total number of negative posts by platform

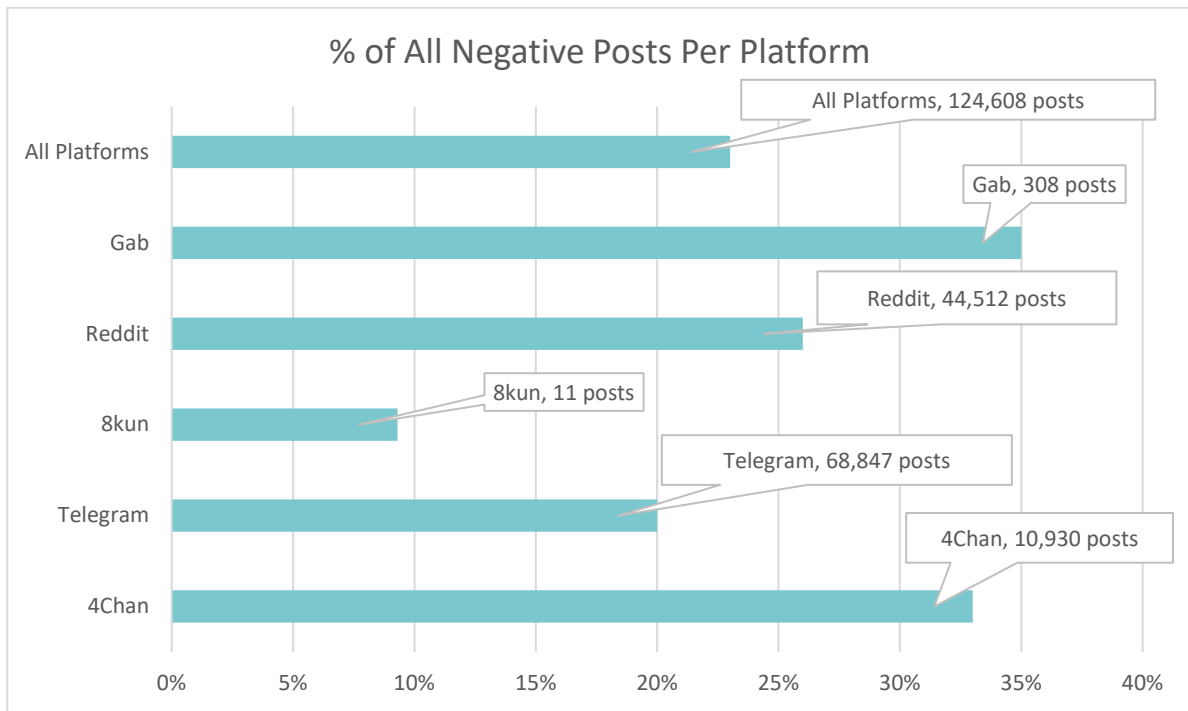
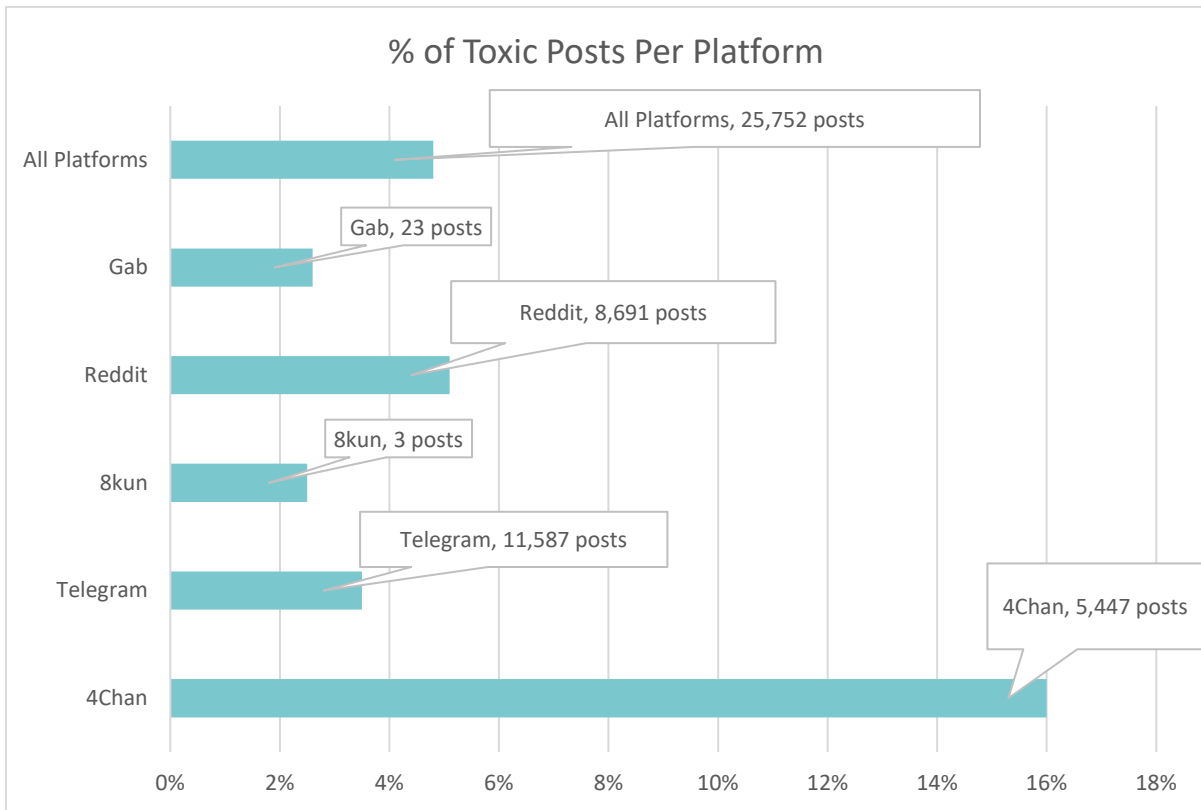


Figure 9: Total Number of toxic posts by platform



Data Privacy and Ethics

HEIA takes all practical steps to ensure the privacy of research subjects is maintained. All data are anonymised, with personally identifying information removed. We limit the metadata collected from platforms to the minimum level necessary to conduct our research. We do not identify or track the activities of individuals or groups. We restrict data collection in line with the policies of the platforms and only access publicly available data. We do not use fake accounts to access closed groups/pages.

This research was approved by the University of Auckland Human Participants Ethics Committee on 9 August 2021 for three years. Reference Number UAHPEC22980.

Contact

Dr Chris Wilson – chris.wilson@auckland.ac.nz

