

An inclusive
Aotearoa New
Zealand where
everyone belongs.



Inclusive
Aotearoa
Collective
Tāhono

Regulating online harm

Anjum Rahman (she/her)

www.inclusiveaotearoa.nz

Disinformation and Russia's war of aggression against Ukraine

Threats and governance responses

03 November 2022

“the campaign consisted of spreading pro-Russian content; impersonating media such as Le Monde, Le Figaro and Le Parisien, as well as government websites including France's ministry of European and foreign affairs; creating websites on francophone news with polarizing angles; and coordinating fake accounts to spread the content created.”

“Der Spiegel warned about natural gas shortages. In the United Kingdom, the Guardian, another media outlet, raised doubts about Russian war crimes in Ukraine. In Italy, Ansa, a leading news agency, criticized Kyiv's storage of much-needed grain...”

<https://www.oecd.org/ukraine-hub/policy-responses/disinformation-and-russia-s-war-of-aggression-against-ukraine-37186bde/>

<https://www.politico.eu/article/russia-influence-ukraine-fake-news/>

<https://www.politico.eu/article/france-accuses-russia-of-wide-ranging-disinformation-campaign/>

Russian trolls attack Syrian humanitarians to cover up war crimes

Network of Syria conspiracy theorists identified - study

Campaign disseminating disinformation sent thousands of tweets, often targeting the White Helmets, research finds

[A new report](#) reveals how bots and trolls linked to Russia have reached an estimated 56 million people with tweets attacking Syria's search and rescue organisation, the Syria Civil Defence – also known as the The White Helmets – during ten key moments of 2016 and 2017.

<https://diary.thesyriacampaign.org/killing-the-truth/>

<https://www.theguardian.com/world/2022/jun/19/russia-backed-network-of-syria-conspiracy-theorists-identified>

India Undergoing Troll Farm Arms Race: European Experts Caution Against Information Manipulation

10 Nov'22 | By Laxitha Mundhra

...social media users in India are unguarded against the manipulation of digital information by troll armies, creating political and ideological propaganda.

...a troll farm is a group of internet trolls that are either paid by certain groups or formed on their own accord to interfere with political opinions and decision-making.

Besides India, governments of countries such as China, Brazil, Russia and the USA, among others, have been accused of building troll farms. Most of these bots operate on Meta platforms Facebook, Instagram and Twitter with barrages of hashtags and propaganda.

<https://inc42.com/buzz/india-undergoing-troll-farm-arms-race-european-experts-caution-against-information-manipulation/>

Study Shows How Russian, Chinese Disinformation About COVID-19 Evolved During The Pandemic

China and Russia have pushed disinformation and propaganda about the origins of COVID-19, unproven cures for the disease, and the efficacy of vaccines aimed at winning over foreign audiences and sowing distrust toward Western governments since the emergence of the deadly virus in the Chinese city of Wuhan two years ago, a new study shows.

<https://www.rferl.org/a/russia-china-covid-disinformation-campaigns/31590996.html>

Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows

"The 'Disinformation Dozen' produce 65% of the shares of anti-vaccine misinformation on social media platforms," said Imran Ahmed, chief executive officer of the Center for Countering Digital Hate, which identified the accounts.

These figures are well-known to both researchers and the social networks. They include anti-vaccine activists, alternative health entrepreneurs and physicians. Some of them run multiple accounts across the different platforms.

<https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes>

Who Is Behind QAnon? Linguistic Detectives Find Fingerprints

Using machine learning, separate teams of computer scientists identified the same two men as likely authors of messages that fueled the viral movement.

<https://www.nytimes.com/2022/02/19/technology/qanon-messages-authors.html>

5. Harmful behaviours, right-wing extremism and radicalisation

...the extreme right-wing has exploited the power of the internet through an array of online platforms and spaces, which it uses to connect with like-minded people and ultimately to recruit new members...¹⁰⁷

Indeed, lone actor right-wing extremist terrorists ... are never quite alone as they are often part of virtual communities on the internet.¹⁰⁸

As extremist content is increasingly removed from the major platforms, right-wing extremists are gathering more and more on unmoderated fringe platforms.

Other platforms that were not set up for hosting right-wing extremist content – including Reddit and 4chan and some online gaming platforms – have been “hijacked” by right-wing extremists.

<https://christchurchattack.royalcommission.nz/the-report/part-2-context/harmful-behaviours-right-wing-extremism-and-radicalisation/>

Extremist recruitment reaching young Australian gamers



The AFP has seen a concerning trend of members and associates of extremist groups targeting young people to expose them to dangerous content - including violent recreations of actual terrorist events - across online gaming platforms.

“Our concern is extremist groups are exploiting these platforms to target a very young group of Australia’s population, by creating content to share and encourage far-right/extremist ideologies and abhorrent violence against others.”

<https://www.afp.gov.au/news-media/media-releases/extremist-recruitment-reaching-young-australian-gamers>

Figure 5: Spectrum of harmful behaviours

Racism and micro-aggressions

- Biased or belittling jokes
- Stereotyping – Muslim students being asked to explain and account for the behaviour of Islamist extremists
- Micro-aggressions – someone not from the majority group being told they speak English well, which sends a message that they are a foreigner

Religiously and/or ethnically motivated harassment and abuse

- Physical abuse – women having their religious attire pulled off in public
- Verbal abuse – people being told in aggressive and threatening ways to “go back to where they came from”
- Online harassment – Facebook messages making derogatory remarks about a person’s religion

Hateful extremism

- Coordinated online or offline campaigns aimed to convince their audience to adopt hateful or discriminatory attitudes
- Protests that promote white identity
- Distributing pamphlets denigrating certain ethnic or religious communities

Violent extremism

- Terrorism – advocating for, planning, materially supporting or carrying out terrorist attacks
- Assault – serious physical violence committed by members of a white supremacist group at a protest
- Vandalism, property damage and arson – setting fire to a masjid

Online hate can have an impact in three interconnected ways:

- the harm done to its targets, either from personal harassment or from online spaces being experienced as hostile;
- the risk that those who encounter it may be radicalized by it, becoming more sympathetic and possibly even active; and
- the effect that it has on the values and culture of the online spaces in which it happens.

Not just online, but flows out to offline values, culture and behaviour too.

<https://mediasmarts.ca/online-hate/impact-online-hate>

MYANMAR: FACEBOOK'S SYSTEMS PROMOTED VIOLENCE AGAINST ROHINGYA; META OWES REPARATIONS

Rohingya sue Facebook for £150bn over Myanmar genocide

Victims in US and UK legal action accuse social media firm of failing to prevent incitement of violence

<https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/>

<https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence>

the social media network's algorithms amplified hate speech and the platform failed to take down inflammatory posts, according to legal action launched in the US and the UK.

Facebook admitted in 2018 that it had not done enough to prevent the incitement of violence and hate speech against the Rohingya, the Muslim minority in Myanmar. An independent report commissioned by the company found that "Facebook has become a means for those seeking to spread hate and cause harm, and posts have been linked to offline violence"

Meta faces \$1.6bn lawsuit over Facebook posts inciting violence in Tigray war

Legal action backed by Amnesty alleges hateful posts inflaming war in northern Ethiopia were allowed to flourish on platform

In February [an analysis](#) by the Bureau of Investigative Journalism (TBIJ) and the Observer found that Facebook was letting users post content inciting violence through hate and misinformation, despite being aware that it helped directly fuel tensions in Tigray, where thousands have died and millions been displaced since war broke out in late 2020.

The research found that one post from a local influencer, calling for people to “cleanse” the area of supporters of Tigrayan forces, stayed up for four months after it was reported to the company.

<https://www.theguardian.com/technology/2022/dec/14/meta-faces-lawsuit-over-facebook-posts-inciting-violence-in-tigray-war>

UN report highlights growing online violence against women and girls, related research

Studies show that online violence against women and girls often precedes offline violence against women and girls.

Ethnic women, indigenous women, lesbian, bisexual and transgender women, and women with disabilities are at greater risk for experiencing online violence.

Online violence causes significant harm and impacts on women's participation in public life.

<https://www.nzfvc.org.nz/news/un-report-highlights-growing-online-violence-against-women-and-girls-related-research>