Online Warfare: Definition, Drivers, and Solutions

Anjum Rahman and Mohan Dutta

As witnessed globally with the rise of hate[i], authoritarian populism[ii], racism, fascist movements, and calls to the genocide of minorities, online infrastructures exist at the core of campaigns threatening democracies and are directly linked to the violence experienced by communities at the intersectional margins. In this white paper, we outline the nature of online warfare, the mechanics underlying it, and its underlying drivers. Building on our analysis of the existing infrastructure that drives online warfare and the existing frameworks for responding to online warfare, we propose a community-led culture-centered approach to responding to online warfare, arguing that the empowerment of communities at the "margins of the margins,"[iii] combined with the development of infrastructures for critical literacy, are critical to addressing the local-national-global threat posed by online warfare. We argue that a culture-centered framework for data justice that empowers historically marginalised communities to participate in platforms, organise to challenge hate, and drive policies for regulating hate on platforms is critical to promoting and **sustaining sustainable development goal (SDG) 16: peace, justice, and strong institutions**. We foreground a culture-centered framework for digital data literacy that is rooted in community voice and ownership of storytelling processes.

Defining online warfare

This paper defines online warfare as organising online infrastructures for spreading disinformation and hate, threatening social cohesion, threatening democratic processes, attacking justice-based mobilisations emergent from marginalised communities, and calling for the deployment of violence targeting communities at the "margins of the margins."

We differentiate online warfare from cyber warfare, as the latter is focused on disrupting online infrastructure, hacking systems, breaching privacy or taking down websites through techniques such as distributed denial of service attacks. Online warfare is focused on content-based attacks.

The target communities chosen for attack are those with little perceived social power, political power, voice, and socio-economic power. Like schoolyard bullies, online hate networks pick targets who they believe are least likely to be able to fight back effectively while offering political and economic opportunity. This is a moving target: as one group is no longer deemed effective for their purposes, hate groups move to another target group to secure political and economic power.

The purposes for online warfare are many and often overlapping:

- **Silencing**: Individuals, organisations, and communities at the "margins of the margins" are targeted for the purpose of erasing their voices and perspectives from the online space. At its heart, this is a direct attack on the free speech rights of the target, through the use of communicative tactics outlined below.
- **Disrupting democracy**: State actors using these techniques and infrastructure seek to disrupt democratic processes in other countries, to impact the global balance of

power or to ensure the election of governments that will be sympathetic to the human rights abuses happening within their own country or perpetrated on other countries. This is an extension of “cold war” tactics moved to online spaces. This activity is particularly harmful when it is opaque to most users, who cannot, therefore, make informed decisions.

- **Income generation**: material focused on attacking vulnerable minority groups has been shown to generate large amounts of income, through the monetisation of accounts, the sale of merchandise and books, lecture tours, and more. The greater the online reach, the more advertising income is earned. Anti-vaccination campaigns are known to increase sales of alternative remedies, some of which can be actively harmful. Malicious actors earn millions of dollars with minimal effort[iv].
- **Political clout and power**: political actors, whether they be political parties, elected representatives, or aspiring candidates, use online warfare tactics to garner political support. By creating a perceived enemy that threatens the way of life, culture, and economic and social welfare of the majority groups, political actors are able to unify the majority population and build political support. Successful tactics have led to such actors gaining increased representation and sometimes being in power, thereby being able to use the state machinery to invest in online warfare.
- **Free advertising and reach**: malicious actors understand the impact of inflammatory speech. It is based on an advertising tactic that is tried and tested. The aim of this tactic is to rely on outrage to spread the visibility of the content and the actor(s) much further. Each share and/or outraged comment, each media article or blogpost provides free advertising and unpaid reach, regardless of the fact that such commentary may be critical and incredibly negative. The ensuing refutation increases reach and awareness further, thereby providing millions of dollars of free publicity.

Communicative Strategies of Online Warfare

Online warfare draws upon a range of interconnected communication strategies that often draw upon each other, build on each other, and magnify the effect of the targeted attack. Some of the communicative tactics of online warfare are as follows:

- **Disinformation**: Disinformation is information that is known to be incorrect, untrue, and/or misleading by the creator of content, and spread with the purpose to deceive the recipient of the information. In other words, disinformation is misinformation that is produced and disseminated strategically. Critical to the disinformation infrastructure online is the underlying objective. Producers of disinformation construct inaccurate information with the specific objective of misguiding the target audience, seeking to influence knowledge, attitude, and/or behavior.
- **Communicative inversion**: Communicative inversion is the deployment of symbols, narratives, and frames to turn materiality on its head. Misinformation is often constructed through communicative inversions. In the context of ongoing processes of marginalisation, majority communities are turned into victims of minority oppression, as reflected in hate narratives such as “The great replacement theory” and “The Muslim invasion” which are mobilised in calls for violence directed at minorities. Communicative inversions manufacture minority communities as threats to the majority culture. Consider here the mobilisation of the “racist” trope by white supremacists in settler colonies to target anticolonial critiques of whiteness, building campaigns targeting anti-racist activists, community members, and academics

speaking out against hate. Consider similarly the deployment of the term "Hinduphobia" by Hindutva groups as a hate construct to silence the critiques of Hindutva, its political project, and its targeting of Muslim individuals, households, and communities.

- **Framing**: Frames offer ways of seeing people, communities, and issues. Online warfare deploys framing to construct marginalised communities as threats to the status quo. Essential to the framing strategy is the deployment of the narrative of "the other," continually producing the other to generate emotions. Emotions such as anxiety, fear, and anger toward "the other" are produced through frames that narrate the story of cultural take-over or cultural loss. Consider for instance the construction of migrant activists as polluting threats to cultural purity, tied to anti-immigrant policies and threats to deport migrants. Consider the ongoing othering of Muslims that forms the core of the Islamophobia industry, built on the othering of Muslims as terrorists.
- **Dehumanisation**: Dehumanisation refers to the communicative process through which a target person or target group is deprived of human qualities. Dehumanisation is often directed toward minorities at the "margins of the margins," and is a critical step toward the organising of hate discourse toward violent action and is a critical feature of genocidal hate[v].
- **Equivocation**: Equivocation refers to the use of ambiguous language to conceal the truth. In the context of online hate, the strategy of equivocation signals and promotes hate while obfuscating the use of hate on the surface. The strategy of equivocation is often a critical resource for hate-based organisations that present the language of family, dialogue, and harmony, while promoting hate toward specific groups. Consider for instance the targeting of transgender communities by far-right Christian organisations that deploy the language of family, replete with images of happy families while simultaneously generating messages of hate[vi]. The family, and particularly children, are projected as at risk from transgender communities, forming the basis for mobilising violence.
- **Humour and plausible deniability**: Humour directed toward communities at the "margins of the margins" is often a strategy for online warfare[vii]. Hate is constructed as humour, coding calls to violence in the language of humour. Critical to the deployment of humour is the ability for hate messages to fall below the threshold of detectable hate speech on online platforms. Moreover, the deployment of hate as humour often offers plausible deniability, thus protecting the producer and disseminator of hate. Critical here is the role of media personalities, news anchors, and journalists in producing hate through humour, often mobilising the dehumanisation of marginalised communities.
- **Impersonation**: Fake websites are set up to impersonate individuals and organisations, with the process of impersonation working to silence critical voices. Some of these fake sites mimic media sites and have been effective in reputable media organisations reporting disinformation as news. In other instances, fake online sites, apps, and accounts impersonate activists, academics, community leaders from the margins, or working in solidarity with the margins. Consider here Hindutva-aligned hate producers that created an app and placed Muslim women activists, academics, and leaders on the app for sale by impersonating them.

- **Death and rape threats**: The use of death and rape threats has often proven effective in silencing the voices of the targets of such campaigns, promoting self-censoring and strategies such as locking accounts because of the targeted hate, thereby limiting their freedom of speech and ability to reach a wider audience. Explicit threats giving dates and times, or bomb threats for public speaking engagements have driven victims out of their homes, resulted in the cancellation of public speaking engagements, and deterred targets of the attacks from standing from public office[viii].
- **Targeting of livelihoods**: Online warfare targets the livelihoods of community members, activists, advocates, researchers, and academics that identify with intersectional marginalised positions or advocate for the rights of marginalised voices. The targets of such attacks are often activists, researchers, and academics who speak up and speak out against the hate. Organisers of online warfare post the email addresses and names of employers, encouraging others to write to the employer to raise complaints. Consider here swarms organised around complaint letters written to institutions demanding that academics be fired. The appearance of volume is meant to offer the semblance of widespread resentment, which then is intended to further amplify the pressure on the institution.
- **Mimicking research and accountability**: One of the core strategies for driving up online hate is the mimicking of research, drawing upon the narrative of accountability, to target critical voices from the margins. Hate influencers (more on this in a later section) play key roles in mobilising online trolls under the narrative of mobilising individuals to carry out public research, carrying out surveillance on targets under the guise of digging up information. Consider the ways in which far-right groups in Aotearoa have mobilised the performance of research to target academics, researchers, and activists. The framework of “digging up dirt” is critical in the organising of swarms and directing these swarms to destabilise institutional processes. Consider here the ways in which institutional complaint mechanisms are manipulated by online swarms to attack researchers and academics raising their critiques of power, and interrogating the mobilisation of hate. Fake websites that are then set up draw upon disinformation generated by online trolls to target critical voices from the “margins of the margins.”
- **Doxing**: publishing of physical addresses for home and work, along with email and phone contact information. This may be accompanied by photographs of the home and of the targets, possibly with family members, are provided with incitement to harass or commit explicit violence.
- **Volume**: Online warfare is characterised by the exponential volume of hate content. The harm that occurs from the hate is magnified manifold with the sheer volume and scale of comments, where each individual comment doesn’t breach any standards, but together, the volume of negative comments becomes overwhelming for the targeted individual or community. Volume creates psychological harm by overwhelming the target. The target and supporters do not have the capacity to respond to each individual commenter, and the comments may contain disinformation or low-level abuse.
- **Intertextuality**: Diverse actors deploying online disinformation and hate are often connected with each other, drawing upon and building on the hate, forming a networked infrastructure directed at the “margins of the margins.” The strength of

the hate discourse is multiplied manifold through the flows between diverse registers of hate. Note here for instance the ways in which Hindutva, far-right Zionist and white supremacist groups align in targeting individual activists, academics, and communities[ix]. Critical here is the Islamophobia that underpins the hate that is mobilised by these groups, engaging in dialogue with each other and building a networked infrastructure of hate that shares discursive resources, frames, and strategies.

- **Hate as free speech**: Various forms of hate discourses draw upon the articulation of free speech to build and sustain the infrastructures for producing and disseminating hate[x]. Particularly salient here is the role of white supremacist groups in settler colonies to draw upon free speech frameworks to legitimise the production and circulation of hate targeting Indigenous communities, migrants, gender diverse communities, and other communities at intersectional margins. The framing of free speech as an infrastructure for legitimising hate reflects the fundamental hypocrisy of whiteness as the organising ideology in settler colonies, simultaneously attacking violently and through many of the mechanisms outlined above Indigenous and minority voices exposing the hypocrisy of free speech. These attacks are often carried out through astroturfs and influencers, often drawing on narratives such as accountability to taxpayers while laying claims to supporting free speech as a democratic value.

Modes of Communication

Online warfare is carried out through multiple modes of communication, often working jointly or in dialogue with each other to amplify the volume of disinformation and hate.

- **Text**: Disinformation and hate are often conveyed through textual messages. These textual messages take diverse forms from short Twitter messages and Twitter threads to longer Facebook posts, to longer messages constructing narratives and shared through platforms such as Telegram and WhatsApp. Texts are
- **Memes, visuals, and videos**: Memes are critical communicative resources in the mobilisation of hate, recruiting members into hate groups, forming the basis of identities and connections, and building hate communities[xi]. Similarly, visuals powerfully draw audience members into the hate message, arousing emotions, and inviting them to act. Islamophobic propaganda films such as “The Kashmir Files” and “The Kerala Story” are examples of content that mobilizes hate towards minority communities (in this instance, hate toward Muslims in India, and toward Indian Muslim diaspora communities across the globe). Online gender-based violence is experienced with photoshopped photographs or altered videos superimposing the face of targets onto pornographic content, with the intent of humiliation and sexual harassment.
- **Music**: Music organises hate by generating and responding to affective registers. An example of the use of music is “Hindutva pop” music containing violent lyrics set to catchy rhythms. This kind of content is readily available on YouTube[xii] as well as available for download on apps. The Christchurch terrorist had a Serbian nationalist song knows as “Remove Kebab” playing in the background at the beginning of his livestream video of the mass murder[xiii].
- **Speeches**: Historically, speeches have been instrumental in the mobilisation of mobs. Violence targeting minorities has often been catalysed through speeches. Online

platforms exponentially multiply the reach of speeches, reaching a large audience and inviting the audience to participate in the violence. Recorded hate speeches played on cassette tapes have been instrumental in the mobilisation of genocidal violence.

- **Mobs**: Mobs are crowds of people that are gathered usually within short timeframes, usually disorderly, and with the intent to cause violence. The mobilisation of mobs toward violence is a core strategy of violence.

Infrastructures for dissemination

- **Swarm structure**: Online misinformation and hate draws on the swarm structure of platforms[xiv]. The swarm structure multiplies the hate at an accelerated pace, growing it manifold and directed at a target. The target experiences the hate as overwhelming and exponentially growing volume of content.
- **New hate platforms**: Designers of hate campaigns continually innovate and develop new platforms through which they target communities, groups, and individuals at the margins. Consider for example the "Bulli Bai" app that was created in India by Hindutva-aligned hate groups to target Muslim women leaders, activists, artists, and other public figures.
- **Astroturfing**: Astroturfing is "organised activity that is intended to create a false impression of a widespread, spontaneously arising, grassroots movement in support of or in opposition to something (such as a political policy) but that is in reality initiated and controlled by a concealed group or organisation (such as a corporation)."[xv] Astroturfing in online spaces creates the infrastructure for mobilising around disinformation and hate, working actively to mislead the audience to recruit into hateful agendas. Salient here are the various organisations that are set up to mobilise around disinformation, funded by powerful political-economic infrastructures that directly gain from the circulation of disinformation.
- **Troll farms**: Institutional use of trolls[xvi], often paid, to disrupt discourse with the purpose of achieving political goals[xvii].
- **Automated bots**: The use of fake accounts that post auto-generated content. There is no human activity or intervention happening from these accounts, and other users are often not aware that the account is a bot. Therefore, any attempt to "debate" such accounts will not result in a change of opinion or an acknowledgment that the posted content is untrue or harmful.
- **Sock puppets**: These are accounts with fictitious identities that hide the real user. A single user may set up multiple sock puppet accounts to create the impression that there is much more widespread support for a particular point of view than actually exists. Far-right hate groups often draw on sock puppet accounts to drive up the disinformation and hate.
- **Anonymous accounts**: These accounts are not fake or fictitious but hide the identity of the real user. While such accounts are often necessary for legitimate reasons to protect the identity of, for example, activists or abuse victims, they are also used by malicious actors to avoid consequences for hateful content.

Targets of online warfare

- **Gender and gender-diverse communities**: Online disinformation and hate are disproportionately directed toward women and gender-diverse communities. The targeting of transgender communities with hate forms a critical resource in the far-right communicative infrastructure[xviii]. Women and girls are increasingly subject to a range of communicative tactics connected with hate, as highlighted by a recent UN report[xix]. Local women politicians, activists, and researchers in Aotearoa New Zealand have been similarly targeted with misogynist hate[xx].

- **Sexuality**: Far-right hate groups are increasingly targeting LGBTQI+ communities, portraying them as threats to children[xxi]. Consider here the collaborative intersections between various hate groups (more on this in the section on intersectional vectors).
- **Racial minority communities**: The far-right has historically organised through the targeting of minority communities, turning minorities into threats. The mobilisation of hate is based upon the vilification of minorities.
- **Religious minority communities**: Many different ideologies have targeted religious minorities to incite violence hate. Justification for the free spread of dehumanising language and hate is based on the notion that religious belief is a choice rather than an intrinsic characteristic.
- **Indigenous communities**: In Aotearoa New Zealand, Māori communities have borne the brunt of active online hate campaigns. Hateful content is focused on the use of te reo Māori, co-governance models, and culture[xxii]. Consider here the role of astroturfs, far right hate groups, and political parties organising attacks on Te Tiriti.
- **Intersectional vectors of harm**: The targeting of hate is often multiplied manifold at the intersectional margins. The concept “margins of the margins” draws attention to the intersectional targets of hate. Individuals and communities with multiple protected identities are likely to receive higher levels of abuse[xxiii].
- **Activists:** Hate groups that are deployed by populist and far-right authoritarian forces often seek to silence dissenting voices. The goal of repressive political actors working alongside hate groups is to produce a chilling climate, which forms a key ingredient in establishing fascist politics.
- **Academics and researchers:** Hate groups often directly target academic voices, with the goal of silencing academic voices. Hate campaigns are often directed at public facing academic voices. For instance, academics participating on platforms such as X challenging far-right, genocidal and authoritarian modes of power and control often become the targets of organised attacks. Hate campaigns often directly target disinformation researchers, seeking to silence them.
- **Politicians:** Politicians speaking in solidarity with marginalised communities, voicing the unmet needs of welfare recipients, advocating for social welfare policies often become the targets of hate campaigns. These campaigns are based on surveillance, alongside disinformation planted on online and offline spaces, and often directing the attacks at the broader familial and social networks of targeted politicians.
- **Education and welfare policies:** Neoliberal capitalist organisations including foundations and think tanks often fund and produce hate directed at welfare recipients, community members experiencing poverty, welfare policies, and education policies. Disinformation and hate work toward delegitimising public programmes, which in turn is a critical step toward dismantling welfare and privatizing education. Note here the linkages among hate content produced by far right think tanks, white supremacist groups, the extractive and tobacco industries, and right-wing politicians.

Types of platforms

- **Websites**: Websites of organisations and individuals can be used to host harmful content. Major platforms using automated content moderation have been exploited through content being held on websites, with only the URL being shared on these platforms. While this tactic is now being addressed by platforms, there are few effective mechanisms to take down websites that platform hate. Some content is hosted by websites such as WordPress or Substack, with minimal content guidelines and accountability[xxiv], with enforcement being patchy, and drawing on settler colonial constructions of free speech, drawing often from the US context. It is worth noting here that a large proportion of platform capital is headquartered in the US.
- **Gaming**: Extremists have used gaming sites in multiple ways, including the production of bespoke video games to promote their ideology; modification of existing popular games, and use of in-game chat functions[xxv].
- **Retail**: This is a subset of websites that sell merchandise that propagates extremist ideologies and symbols[xxvi]. Major and well-known sites like Amazon have had issues with selling merchandise that is used to fund extremist groups[xxvii], and payment services like PayPal and others have been exploited as well. Also critical to note here is the active relationship between some of the leaders in technology-based retail and far-right groups.
- **Social media**: These are platforms that allow users to share content and network with each other, creating a community around the sharing of information and support. They have been exploited in various ways to disseminate disinformation and hate, with social media platforms such as Facebook emerging as the dominant actors in the dissemination of disinformation and hate. Platforms such as Telegram and Discord, as well as private Facebook pages, are spaces for extremists to organise.
- **Messaging apps**: Messaging apps are powerful in disseminating disinformation and hate by drawing on the power of interpersonal relationships, leveraging intimacy to build trust. Apps such as WhatsApp, Signal, Messenger and Viber have end-to-end encryption, which is crucial for activism and privacy rights, but also hide the significant activity conducted in the online warfare space.

Effects of online hate

Online hate produces a range of effects, from leading to real-life violence to impacting the psychological well-being of the targeted individuals, groups, and/or communities. Below are some examples of widespread offline violence that is directly tied to online discourse.

- **Violence:** Online hate results in various forms of violence, with disinformation often being mobilized to catalyse mobs that carry out the violence[xxviii]. Digital platforms have emerged as key features in the mobilisation and organisation of mobs around specific events. Consider here the role of rumours spread over online platforms in escalating violence, resulting in the targeting of individuals, groups and/or communities. The violence is disproportionately directed toward marginalised communities.

  - **Andrew Tate and pick-up artist culture**: There is growing evidence that young men and boys are actively harassing teachers and students, influenced by viral influencers such as Andrew Tate[xxix]. The virality of this content is by design. “Evidence obtained by the Observer shows that followers of Tate are being told to flood social media with videos of him, choosing the most controversial clips in order to achieve maximum views and engagement.”[xxx]
  - **Hindutva**: In India and in the Indian diaspora, far-right Hindutva discourses have mobilised around the production of the Muslim other, generating the fear of the Muslim to mobilise violence[xxxi]. Extremist Hindutva discourses online dehumanise Muslims, pathologize Muslims, articulate rape threats directed disproportionately at Muslim women, and call for violence directed at Muslims. In India, Hindutva discourses mobilised online have led to the organising of mobs, often under the umbrella of organisations such as the Vishwa Hindu Parishad (VHP).
  - **Myanmar**: Meta played a key role in the organising of violence targeting Rohingya Muslim communities in Myanmar. An Amnesty International report states that “Meta knew or should have known that Facebook’s algorithmic systems were supercharging the spread of harmful anti-Rohingya content in Myanmar, but the company still failed to act.... Meta repeatedly failed to heed the warnings, and also consistently failed to enforce its own policies on hate speech.”[xxxii]
  - **Kenya case**: This case shows the alleged direct impact of posting online resulting in violence. “Court filings say that in October 2021, militants followed Meareg home from work, shot him in the back and leg, and left him to bleed. Meareg, a well-respected chemistry professor according to the lawsuit, was targeted after Facebook posts spread his name, photo, and false allegations that he was associated with a deadly rebel group because of his ethnicity as a Tigrayan, the country's minority demographic. In an affidavit, Abrham says he asked Facebook multiple times to remove posts about his father — both before and after his death.”[xxxiii]
  - **Other examples**: in Germany, there was a correlation between anti-refugee Facebook posts and attacks on refugees[xxxiv]; the Charleston Church shooter engaged in an online self-learning process; the Pittsburgh Synagogue shooter espoused conspiracy theories on the “great replacement” of white people on Gab; this same trope was used as justification by the Christchurch terrorist who credited YouTube for shaping his views; and in Sri Lanka the Tamil Muslim minority was targeted based on rumours spread online[xxxv].

- **Mental health**: Online warfare directly impacts the health and well-being of targeted individuals, groups, and communities[xxxvi]. Even if there is no such translation, online harm can result in psychological trauma responses, radicalisation, and desensitisation[xxxvii].

- **Voter suppression**: Online warfare increasingly plays a key role in targeting communities at the “margins of the margins” with disempowering messages, directed at keeping them away from participating in democratic processes. In electoral processes, far-right groups have deployed hate toward voter suppression, mobilising to build sense of disempowerment in marginalised communities and in discouraging community members from participating in voting. In the US elections of 2016 for instance, the Russian Internet Research Agency purchased ads targeting African American audiences and urging them to "boycott the elections."[xxxviii]

Drivers of online warfare

Underlying the mobilization of online warfare is a political-economic infrastructure. The colonial-capitalist model that seeks to maximise engagement at the expense of safety, the right to life and freedom from discrimination.

- **Machine learning and algorithms**: recommender algorithms increase reach by recommending similar content. Liking, sharing and commenting on posts increase reach and visibility. The design of some platforms ensure that posts designed to garner outrage will get higher reach than those which receive a "like" response[xxxix]. Content moderation algorithms can reduce reach or remove content. Another example is in the Amnesty International report on Myanmar, which shows how algorithms were promoting content instead reducing reach[xl]. Algorithmic transparency.
- **Generative artificial intelligence**: there have already been widespread discussion of copyright breaches[xli] and existential threats to humanity[xlii]. There is also the risk that generative AI will exacerbate the spread of disinformation[xliii] and can also similarly be used for radicalisation and the spread of hate ideologies[xliv]. Generative AI includes the production of fake videos, which has the potential to call into question evidence or create incidents that never occurred for the purpose of inciting violence. There have been calls for transparency, so people are aware that the content is generated by AI, as well as a whole body of work on ethics in AI[xlv].
- **XR (Extended Reality)**: "XR is an emerging umbrella term for all the immersive technologies. The ones we already have today—augmented reality (AR), virtual reality (VR), and mixed reality (MR) plus those that are still to be created. All immersive technologies extend the reality we experience by either blending the virtual and "real" worlds or by creating a fully immersive experience."[xlvi] As with multiplayer role-playing video games, there is the possibility of avatars or online personas being beaten, killed or subjected to sexual violence, resulting in psychological trauma. Targeting of marginalised groups for this type of violence as well as using it to desensitise and to spread violent ideologies is a risk.
- **Rewarding outrage and hate**: Outrage and hate are economically profitable, for the producers and disseminators of hate, as well as for the platforms that circulate the hate. Hate content that generates reactions of outrage is much more likely to be circulated compared to other forms of content. Discourses that organise hate are viral by nature, drawing on the capacity of hate to mobilise attention, generate reactions and shares, catalyse the formation of mobs, and mobilise the mobs to act.
- **Monetising hate**: many platforms allow users to earn income based on reach, such as YouTube. Other platforms provide users with large following with sponsorships for promoting products or particular types of content. With outrageous and borderline content gaining greater reach both through algorithms and the infrastructures for dissemination, those promoting hate are earning significant amounts of income[xlvii]. For example, 25 YouTube accounts were able to earn an estimated USD3.5million for a hate campaign targeted at Meghan Markle[xlviii].

- **Smaller platforms**: smaller platforms do not have the resource base or the same level of expertise to effectively control exploitation by extremist groups. Often such groups are using smaller platforms to organise campaigns for larger social media platforms. Smaller platforms also have less capacity to respond to a crisis event, where livestreaming or manifestos related to an offline live event are hosted on these sites. These platforms require support, and organisations like Tech Against Terrorism have a focus on providing this through upskilling and a range of tools.
- **Virtual Private Networks**: Virtual private networks (VPNs) create secure connections between a computing device and a computer network, or between two networks, using public networks such as the Internet. On one hand, VPNs create pathways for bypassing censorship in authoritarian regimes, and for resisting surveillance. On the other hand, they are often used as infrastructures in the dissemination of disinformation and hate.
- **Unmoderated platforms**: certain platforms have become sites for extremists to organise, share memes that dehumanise and desensitise, and host content that includes threats of violence and doxing. Such sites refuse to moderate content and are sometimes based in countries where there are no effective legal regulations, or where such regulations are not enforced. Websites like 4chan and Telegram deliberately exclude themselves from international mechanisms such as the Global Internet Forum to Counter Terrorism or other organisations providing support such as Tech Against Terrorism.
- **Media influencers**: Media influencers such as Alex Jones of Infowars are powerful drivers of disinformation and hate. Disinformation and hate form the primary drivers of revenue for these influencers. In Aotearoa, media influencers such as the far-right activist Kelvyn Alp use platforms such as Counterspin Media to circulate disinformation and hate. Note here the range of far-right influencers that play key roles in mediating the disinformation-based hate discourses from the far-right spaces into the mainstream. Similarly, in the Hindutva ecosystem in Aotearoa, media influencers such as Roy Kaunds, collaborating with Counterspin Media, are key actors in the dissemination of Islamophobia.
- **State actors**: In disinformation campaigns related to the Ukraine war, Russian disinformation tactics included the use of fake websites to mimic Western news websites[xlix]. Other tactics documented were the use of paid trolls, fake Facebook profiles that posed as journalists, payment of TikTok influencers, and amplifying authentic messages that supported the government's viewpoint[l]. Similar tactics were used in Syria to attack humanitarians and spread disinformation on the use of chemical weapons[li]. The Indian government is known to use troll farms, and other countries have been accused of doing the same, including China, Brazil and the United States[lii]. China and Russia also ran disinformation campaigns and propaganda related to Covid-19[liii].
- **Political actors**: Disinformation and hate-based campaigns form core infrastructures of authoritarian populist politics. Politicians on the far-right draw upon hate discourses and collaborate with far-right hate groups to target minorities at the "margins of the margins."
- **Removing access**: Internet Shutdowns, geo-blocking, takedown notices are techniques used by governments to silence dissent and prevent opposing views being shared with communities. This allows harmful and dehumanising content to

flourish without being challenged or refuted. India is well known for shutting down access in Kashmir, Manipur and other regions. Several countries limit access to certain platforms, such as China and countries in the Middle East.

Strategies for challenging online warfare

We propose a culture-centered approach to data justice, drawing on the concept that co-creating voice infrastructures at the "margins of the margins" forms the basis for challenging disinformation and hate. Based on the observation that online warfare directly mobilises violence targeting communities at the "margins of the margins," seeking to erase the articulations emergent from the "margins of the margins," co-creating voice infrastructures in partnership with the "margins of the margins" forms the basis for resisting online warfare. Culture-centered processes of co-creating voice infrastructures at the margins challenge the disinformation and hate that is disseminated online, and contribute toward building sustainable peace, strengthening institutions, and sustaining democracies.

Empowerment processes

Because online warfare is critically organised to silence the voices of communities at the "margins of the margins," empowering communities to understand and recognise disinformation and hate, respond to disinformation and hate, and develop community-led interventions is critical.

- **Critical digital literacy**: Critical digital literacy builds the capacities to critically evaluate and process information, attending to information quality, evaluating the underlying mechanisms that drive platform hate, and analysing the power and control that shape how information is produced and disseminated. Developing culture-centered processes of empowerment grounded in critical media literacy strengthen the capacities of communities to challenge the disinformation and hate.
- **Education and awareness**: Ensuring that there is widespread public awareness of the way online spaces are manipulated is critical to inoculating societies and preventing unsuspecting users from being taken in by online warfare tactics. Organisations in Aotearoa New Zealand such as The Disinformation Project, Antifascist Aotearoa, and Fight Against Conspiracy Theories (FACT) are providing valuable services in this space.
- **Fact-checkers**: Fact-checkers form critical infrastructures in countering disinformation-based campaigns. Fight Against Conspiracy Theories in Aotearoa New Zealand ran a fact-checking campaign during the local body elections in 2022. The group effectively exposed candidates belonging to various anti-vaccination groups or other organisations that promoted conspiracy theories[liv]. Organisations such as Alt News in India play key roles in challenging the disinformation-based hate structure of Hindutva.
- **Community mobilisation**: campaigns that bring together community organisations and individuals to counter and speak back to visible and organised hate campaigns have had some level of success. Tauiwi Tautoko[lv] is a programme focused on training non-Māori effective techniques to engage online and speak back to hate directed towards Māori. Inclusive Aotearoa Collective Tāhono is in partnership with other organisations (including Tohatoha and Amnesty International) to build a civil society group known as the Coalition for Better Digital Policy. The work of the Center for Culture-centered Approach to Research and Evaluation (CARE)

- **Increase diversity**: Empowerment processes are vital to fostering diversity and pluralism in decision-making and at all levels of government, regulatory bodies, platforms and community-led approaches. Centering the lived experience of those who have suffered the impacts of online warfare is critical, as well as the need to ensure diverse perspectives from different world views. Evidence suggests more diverse groups make better decisions[lvi], but there is a more important human rights imperative to ensure diverse communities can participate in making decisions that affect them.
- **Legal empowerment**: Participation in legal processes is critical to enabling communities at the margins to launch effective challenges to platform-based disinformation and hate. The complicity of Big Tech in the production and dissemination of disinformation and hate implies that the challenge to Big Tech ought to be mobilised through legal processes. It is critical to build a legal infrastructure that is connected to communities and that works alongside communities to challenge the hate.

Decolonising processes

Recognising that the proliferation of hate online is directly tied to the colonial-capitalist ideology that underlies the architecture of platforms shapes the organising of data justice around decolonising registers. In Aotearoa New Zealand, Te Tiriti-based mobilisation draws on the concept of data sovereignty to offer a framework for organising against online disinformation and hate.

- Land rights: Recognising that historically disinformation formed the architecture of colonial land grab informs the organising of resistance to disinformation and hate in land rights.
- Challenging incarceration: Incarceration is an essential resource in upholding the politics of hate. Through processes of incarceration, hegemonic power structures enact control over communities at the margins. A decolonising approach to challenging online hate therefore is deeply rooted in the organising work of dismantling the military-prison-industrial complex.
- Data sovereignty: Communities at the margins, Indigenous and local communities in the Global South are turned into exploitable sources of data in hate campaigns. Centering data sovereignty prioritizes questions of data ownership and control, noting that communities at the margins ought to own the data that is gathered from them. Centering community control resists the gathering of data for manipulation by far-right hate campaigns. Platforms such as Meta and X are held accountable to communities. Such processes of community control foreground decolonising values of love, connection and community, fundamentally undoing the colonising work of polarising hate content.
- Data sharing: Data sharing refers to the sharing of data by placing data in community values, community norms, and community conversations. Principles of sharing are guided by values of building connections, rooted in Indigenous and Global South registers of organising data. These connections sustain community-led efforts of social cohesion and peace building, defining data in values rooted in peace and community. The privatisation and extraction of data that shape disinformation-based hate campaigns is resisted through community-led processes of data sharing. The

extractive logics that fuel the profiteering models of Big Tech are resisted through principles of sharing.

- Decolonising free speech: The hegemonic registers of free speech are rooted in whiteness, embedded within colonial formations. That organisations built around free speech advocacy in Western democracies are often formed under frameworks mobilised to protect and promote white supremacy reflects the underlying colonial logic. Decolonising free speech therefore calls for building openings for Indigenous and Global South knowledge systems, locating concepts of freedom and speech in justice-based registers[lvii].

Building information cooperatives

- Building alternative news media models: The hegemonic model of news production and circulation continues to circulate and uphold state and capitalist power. The nature of news itself serves hegemonic actors, often platforming dominant voices. The ideology of hate is normalised and upheld in colonial and racial capitalist processes, including in hegemonic processes of news production and dissemination. It is therefore critical to build community-led culture-centered news infrastructures grounded in the voices of communities at the margins. Decolonising the disinformation and hate infrastructure therefore is tied to co-creating voice infrastructures owned by communities at the margins, enacting story sovereignty. Hegemonic news values are de-centered through community participation and mobilisation.
- Building community media: Community media are vital to community participation in democratic processes. Building community media turns the power of challenging disinformation and hate in the hands of communities. The capacity of communities to tell their own stories counters the disinformation and hate circulated through online platforms.

Accountability processes

Algorithms are the key drivers of disinformation and hate, and therefore, it is critical to establish processes of accountability that monitor and regulate algorithms.

- **Algorithmic audits**: auditing the impacts of algorithms is a useful tool for holding platforms to account. There are objections around commercial confidentiality and intellectual property, however it's possible to use the financial audit process as a model to develop an infrastructure for algorithmic audits[lviii].
- **National security strategies**: need to explicitly include tactics to counter disinformation, troll armies, astroturfing and other being used by states. State disruption requires a state response through national security strategies that are well resourced and effectively implemented.
- **Platform regulation**: Various jurisdictions have implemented legal regulatory regimes to reduce the impact on online harm. Various examples include the European Union's Digital Safety Act[lix], Australia's Online Safety Act[lx], and Ireland's Online Safety and Media Regulation Act[lxi]. The UK's Online Safety Bill is currently going through the legislative process[lxii], while regulation in Aotearoa New Zealand has not yet reached the phase of draft legislation[lxiii]. Many of the existing regimes are new, so it will take time to see how effective they are in reducing online warfare tactics.

- ***Product liability***: platforms have not generally been held legally liable through civil or criminal legislation for the harm they cause in society. Section 230 of the US Communications Act has generally placed liability on the creators of content rather than the platforms. This ignores the systemic issues and the impact of the volume of accounts attacking a single user. The EU's Digital Safety Act, on the other hand, has now included provisions for product liability.
- ***Resourcing***: many of the solutions to online warfare rest on community activism, as well as well as regulatory infrastructures which are expensive to effectively maintain. Often lack of enforcement from law enforcement is the direct result of the lack of resourcing to deal with the sheer volume of content that includes death and rape threats, and other harmful content. One way to provide resourcing is through the fair taxation of multi-national platforms. This would require international agreements and possibly covenants, due to the ability of multinational platforms to change locations of registration and head office to avoid tax. Currently there is plenty of evidence of tax avoidance[lxiv].
- ***Taxonomies and definitions***: the terms "terrorism" and "violent extremism" are contested, with no internationally agreed definitions. Similarly, clear and concise definitions of hate speech, legal and illegal content can be complex and context dependent. This is part of the complexity of this space, and a challenge to regulators, policy makers and content moderators[lxv].
- ***The Trump test***[lxvi]: when considering regulatory frameworks, it is important to consider how any content moderation or censorship laws would be used by a hostile government. In the US, President Trump was known for targeting minorities such as Muslims, Mexicans, disabled people and others in his political rhetoric. Having such a figure have power over the machinery of the state runs the risk of laws and regulations being applied most harshly against the very groups it is designed to protect. (Free speech versus over-censorship)

Collaborative processes

- ***International collaboration***: there are many instances of multistakeholder forums that comprise the public and private sector, along with civil society, to seek solutions to online warfare. This is a participatory approach, recognising that the public sector has regulatory power, the private sector control online platforms, and civil society groups hold both accountable for upholding civil, political and human rights. Examples of such forums are the Christchurch Call to Action, the Global Internet Forum to Counter Terrorism, OECD multistakeholder work on a Voluntary Transparency Reporting Framework and the Global Partnership on Artificial Intelligence. These processes are useful to present a range of views on various issues, and to inform public policy. Enforcement of decisions can be a problem, particularly when there is no effective method beyond diplomacy to do this.
- ***Positive interventions***: are used for the promotion of credible, positive alternatives or counter narratives, and other forms of digitally distributed user-facing messaging, with the goal of counteracting the possible interest in terrorist and violent extremist groups[lxvii]. These can use tactics such as flagging harmful content with content warnings, reminders to click on the link in a post prior to sharing, providing warnings around sharing misleading information or redirecting to scholarly articles when

particular search terms are entered. These techniques are designed and implemented by platforms.

- ***UN processes***: existing UN processes are available to hold nation states accountable for upholding human rights or other conventions. Examples of these are the Universal Periodic Review run by the United Nations Human Rights Council, or the reporting process for the UN Convention on the Elimination of all forms of Discrimination Against Women. For each of these processes, both civil society and governments report on their progress in upholding their commitments and are questioned by other nation states. Such processes could be used, for example, to hold states accountable for their commitments through the Christchurch Call to Action. This option has not yet been explored.
- ***Hate index*** (SDG 19): We propose community-led culture-centered processes for measuring online hate, monitoring online hate, and labelling the political and economic structures that drive hate.

Conclusion

In this white paper, we have outlined the nature and mechanics of online warfare, the drivers of online warfare, and potential strategies for challenging online warfare. Drawing on the CCA, we argue that empowering communities at the "margins of the margins" to lead advocacy is a critical to challenging online disinformation and hate. Moreover, we note that a decolonising register for countering online warfare locates disinformation and hate amidst the interpenetrating forces of colonialism, imperialism, and racial capitalism. To counter disinformation and hate at the structural level therefore calls for building and sustaining anticolonial practices of resistance that center values of love, connection and community.

[i] George, C. (2016). *Hate spin: The manufacture of religious offense and its threat to democracy*. MIT Press.

[ii] Waisbord, S. (2020). Mob censorship: Online harassment of US journalists in times of digital hate and populism. *Digital Journalism*, *8*(8), 1030-1046.

[iii] We define "margins of the margins" as intersectional positions that continue to experience the violence of erasure, attending to the inequities that exist within communities at the margins; Elers, C., Jayan, P., Elers, P., & Dutta, M. J. (2021). Negotiating health amidst COVID-19 lockdown in low-income communities in Aotearoa New Zealand. *Health Communication*, *36*(1), 109-115.

[iv] see for example https://www.rollingstone.com/politics/politics-news/alex-jones-infowars-store-165-million-1281059/; https://www.theguardian.com/media/2022/aug/20/andrew-tate-money-making-scheme-for-fans-of-extreme-misogynist-closes ; https://www.businessinsider.com/white-supremacists-made-massive-profits-through-bitcoin-new-analysis-suggests-2021-12 "In addition to cryptocurrency, Molyneux appears to make money through soliciting donations on his website, from fans of his purchasing membership to his virtual "Freedomain Community," and by selling books. Molyneux has 10 philosophy-themed books listed for sale on his website."

[v] De Ruiter, A. (2023). To be or not to be human: Resolving the paradox of dehumanisation. *European Journal of Political Theory*, *22*(1), 73-95.

[vi] Py, F. (2020). Bolsonaro's Brazilian Christofascism during the Easter period plagued by Covid-19. *International Journal of Latin American Religions*, *4*(2), 318-334.

[vii] Bliuc, A. M., Faulkner, N., Jakubowicz, A., & McGarty, C. (2018). Online networks of racial hate: A systematic review of 10 years of research on cyber-racism. *Computers in Human Behavior*, *87*, 75-86; Sakki, I., & Castrén, L. (2022). Dehumanization through humour and conspiracies in online hate towards Chinese people during the COVID-19 pandemic. *British Journal of Social Psychology*, *61*(4), 1418-1438; Schmid, U. K., Kümpel, A. S., & Rieger, D. (2022). How social media users perceive different forms of online hate speech: A qualitative multi-method study. *new media & society*, 14614448221091185.

[viii] see https://www.amnesty.org/en/latest/news/2018/03/online-violence-against-women-chapter-5-5/; https://blogs.lse.ac.uk/politicsandpolicy/ignoring-online-abuse-of-women-mps-has-dire-consequences/

[ix] Sen, S. (2015). Fascism without Fascists? A Comparative Look at Hindutva and Zionism. *South Asia: Journal of South Asian Studies*, *38*(4), 690-711; Pate, T. (2023). Hostile homelands: the new alliance between India and Israel. Pluto.

[x] Vats, A., & Dutta, M. J. (2020). Locating freedom of speech in an era of global white nationalism. *First Amendment Studies*, *54*(2), 156-180.

[xi] Kiela, D., Firooz, H., Mohan, A., Goswami, V., Singh, A., Ringshia, P., & Testuggine, D. (2020). The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in neural information processing systems*, *33*, 2611-2624; Lee, R. K. W., Cao, R., Fan, Z., Jiang, J., & Chong, W. H. (2021, October). Disentangling hate in online memes. In *Proceedings of the 29th ACM international conference on multimedia* (pp. 5138-5147).

[xii] https://www.cjr.org/tow_center/youtube-allows-hindutva-pop-videos-that-violate-its-hate-policies-auto-generates-more.php

[xiii] https://www.rferl.org/a/christchurch-attacks-yugoslavia-tarrant-inspiration-suspect-new-zealand/29823655.html

[xiv] Ganesh, B. (2018). The ungovernability of digital hate culture. *Journal of International Affairs*, *71*(2), 30-49.

[xv] https://www.merriam-webster.com/dictionary/astroturfing

[xvi] ”Trolling is when someone posts or comments online to ‘bait’ people, which means deliberately provoking an argument or emotional reaction. In some cases they say things they don’t even believe, just to cause drama. In other cases, they may not agree with the views of another person or group online, so they try to discredit, humiliate or punish them.” from https://www.esafety.gov.au/young-people/trolling

[xvii] https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election/

[xviii] Hattotuwa, S., Hannah, K., & Taylor, K. (2023). Transgressive transitions. The Disinformation Project. Retrieved from https://thedisinfoproject.org/wp-content/uploads/2023/05/Transgressive-Transitions.pdf

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

[xx] https://www.nzherald.co.nz/nz/intense-threats-deputy-pm-grant-robertson-flags-fears-for-mps-safety-on-election-campaign/75HTZMKVL6TV5RV3UPHGUII4QY/

[xxi] https://www.adl.org/resources/blog/what-grooming-truth-behind-dangerous-bigoted-lie-targeting-lgbtq-community; https://www.hrc.org/press-releases/new-report-anti-lgbtq-grooming-narrative-surged-more-than-400-on-social-media-following-floridas-dont-say-gay-or-trans-law-as-social-platforms-enabled-extremist-politicians-and-their-allies-to-peddle-inflamatory-discriminatory-rhetoric

[xxii] https://www.newshub.co.nz/home/new-zealand/2022/09/increase-in-online-racism-towards-m-ori-concerning-experts-say.html

[xxiii] https://digitalcommons.schulichlaw.dal.ca/cgi/viewcontent.cgi?article=1274&context=cjlt; https://www.engender.org.uk/news/blog/gendered-online-harassment--whats-law-got-to-do-with-it/

[xxiv] https://substack.com/content; https://wordpress.com/support/user-guidelines/

[xxv] https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/221005_research_launch_on_gaming_ve.pdf

[xxvi] https://nz.news.yahoo.com/report-reveals-major-e-commerce-sites-profit-from-selling-extremist-merch-140055603.html

[xxvii] https://www.splcenter.org/hatewatch/2022/08/23/white-nationalist-group-exploits-amazon-fund-their-cause

[xxviii] Ward, M. (2020). Walls and cows: social media, vigilante vantage, and political discourse. *Social Media+ Society*, *6*(2), 2056305120928513.

[xxix] https://www.nzherald.co.nz/world/teachers-and-girls-call-out-andrew-tate-influence-as-rape-threat-revealed/QKJUCTSLOB4HIA4HLWJM6CB2MU/

[xxx] https://www.theguardian.com/technology/2022/aug/06/andrew-tate-violent-misogynistic-world-of-tiktok-new-star

[xxxi] Nizaruddin, F. (2021). Role of public WhatsApp groups within the Hindutva ecosystem of hate and narratives of “CoronaJihad”. *International Journal of Communication*, *15*, 18.

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

[xxxiii] https://www.npr.org/2022/12/17/1142873282/facebook-meta-lawsuit-ethiopia-kenya-abrham-amare

[xxxiv] Hanzelka, J., & Schmidt, I. (2017). Dynamics of Cyber Hate in Social Media: A Comparative Analysis of Anti-Muslim Movements in the Czech Republic and Germany. *International Journal of Cyber Criminology*, *11*(1).

[xxxv] https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons. See also https://www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/

[xxxvi] Keipi, T., Räsänen, P., Oksanen, A., Hawdon, J., & Näsi, M. (2018). Exposure to online hate material and subjective well-being: A comparative study of American and Finnish youth. *Online Information Review*, *42*(1), 2-15; Keipi, T., Näsi, M., Oksanen, A., & Räsänen, P. (2016). *Online hate and harmful content: Cross-national perspectives* (p. 154). Taylor & Francis.

[xxxvii] https://mediasmarts.ca/online-hate/impact-online-hate

[xxxviii] Overton, S. (2019). State Power to Regulate Social Media Companies to Prevent Voter Suppression. *UC Davis L. Rev.*, *53*, 1793.

[xxxix] https://thehill.com/changing-america/enrichment/arts-culture/578724-5-points-for-anger-1-for-a-like-how-facebooks/. Note that this information is only available due to exposure by a whistleblower. Similar information is not available for other platforms.

[xl] In 2014, Meta attempted to support a civil society led anti-hate initiative known as 'Panzagar' or 'flower speech' by publishing a sticker pack for Facebook users to post in response to content which advocated violence or discrimination. The stickers bore messages such as, 'Think before you share' and 'Don't be the cause of violence'. However, activists soon noticed that the stickers were having unintended consequences. Facebook's algorithms interpreted the use of these stickers as a sign that people were enjoying a post and began promoting them. Instead of diminishing the number of people who saw a post advocating hatred, the stickers actually made the posts more visible
See https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/

[xli] https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem

[xlii] https://abcnews.go.com/Technology/dispute-threat-extinction-posed-ai-looms-surging-industry/story?id=101495898

[xliii] https://www.theguardian.com/commentisfree/2023/mar/03/fake-news-chatgpt-truth-journalism-disinformation

[xliv] see also https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/

[xlv] see for example https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

[xlvi] https://www.forbes.com/sites/bernardmarr/2019/08/12/what-is-extended-reality-technology-a-simple-explanation-for-anyone/?sh=64ecf2872498

[xlvii] https://www.wcvb.com/article/exclusive-part-2-monetization-of-hate-online-soars/36409300#

[xlviii] https://www.newsweek.com/meghan-markle-trolls-hate-3-million-industry-youtube-1670272

[xlix] https://www.politico.eu/article/russia-influence-ukraine-fake-news/

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

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

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

[liii] https://www.rferl.org/a/russia-china-covid-disinformation-campaigns/31590996.html

[liv] https://factaotearoa.nz/local-body-elections-2022-what-fact-saw/

[lv] https://www.tauiwitautoko.com/

[lvi] https://hbr.org/2019/03/when-and-why-diversity-improves-your-boards-performance ; https://www.dimins.com/blog/2022/02/15/diversity-matters-in-decision-making/

[lvii] Dutta, M. J. (2023). Theorizing Southern Strategies of Anti-Racism: Culturally Centering Social Change. In *The Routledge Handbook of Ethnicity and Race in Communication* (pp. 301-314). Routledge; Vats, A., & Dutta, M. J. (2020). Locating freedom of speech in an era of global white nationalism. *First Amendment Studies*, *54*(2), 156-180.

[lviii] https://internetnz.nz/news-and-articles/algorithmic-audits-an-accountants-view/

[lix] https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348

[lx] https://www.esafety.gov.au/newsroom/whats-on/online-safety-act

[lxi] https://www.oireachtas.ie/en/bills/bill/2022/6/

[lxii] HYPERLINK "https://www.weforum.org/agenda/2023/06/united-kingdom-uk-online-safety-bill-internet-privacy-parliament/"https://www.weforum.org/agenda/2023/06/united-kingdom-uk-online-safety-bill-internet-privacy-parliament/

[lxiii] https://www.dia.govt.nz/safer-online-services-media-platforms-consultation

[lxiv] https://www.hindustantimes.com/technology/microsoft-avoids-paying-tax-in-many-countries-study-101665633479612.html; https://www.rnz.co.nz/programmes/the-detail/story/2018804542/taxing-the-big-tech-companies; HYPERLINK "https://www.salon.com/2021/06/01/amazon-facebook-and-other-tech-giants-paid-almost-100b-less-in-taxes-than-they-claimed-analysis_partner/"https://www.salon.com/2021/06/01/amazon-facebook-and-other-tech-giants-paid-almost-100b-less-in-taxes-than-they-claimed-analysis_partner/

[lxv] see this report from the Global Internet Forum to Counter Terrorism as an example of a taxonomy: https://gifct.org/wp-content/uploads/2022/12/HSDB-Taxonomy-FOR-PUBLICATION-Dec-2022-1.pdf

[lxvi] Demaske, C. (2022). Trump's hateful rhetoric and First Amendment failures: Re-envisioning incitement, true threats, and hate speech. *Communication and Democracy*, *56*(2), 91-116.

[lxvii] from https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAPI2-2021.pdf