



# ONLINE HATE AND OFFLINE HARM

This research was funded by InternetNZ, and prepared by Jess Berentson-Shaw & Marianne Elliott of The Workshop.

*“The web is under threat — but the web we want is not out of reach. It’s up to us to overcome these threats and ensure the web remains an open platform that is truly a force of good for everyone.”*

- *The Case for the Web*

# CONTENTS

<b>BACKGROUND</b>	<b>04</b>
<b>OVERVIEW</b>	<b>05</b>
<b>1. HATEFUL AND RACIST DISCOURSE AND HATEFUL CRIMES OR TERRORISM</b>	<b>06</b>
<b>2. FREEDOM OF PEOPLE TO ACCESS GOOD INFORMATION</b>	<b>08</b>
<b>3. THE SPREAD OF RACIST IDEAS</b>	<b>09</b>
<b>4. IMPACTS ON PEOPLE</b>	<b>11</b>
4.1 DIRECT BENEFITS FROM THE INTERNET	11
4.2 DIRECT HARM FROM THE INTERNET	11
4.3 INDIRECT HARM FROM RACIST DISCOURSE ON POLICY TO ADDRESS THE PROBLEM	12
<b>5. TE TIRITI O WAITANGI</b>	<b>13</b>
<b>CONCLUSIONS</b>	<b>14</b>

# BACKGROUND

The shock and horror felt by many New Zealanders at the Christchurch mosque massacres has reignited a debate in our country about the role of the internet in fostering and spreading hatred generally, and extremist violence specifically. Much of the public commentary on this issue is currently framed as a discussion about content moderation, with little reference to or understanding of the larger challenges facing the open web.

This presents an important opportunity for New Zealand to contribute to a broader, more well-informed and evidence-based conversation about the relationship between online hatred and real-life harm.

In its report, 'The Case for the Web' the World Wide Web Foundation, founded by Sir Tim Berners-Lee, outline the major threats currently facing the open internet. Those threats include:

- a massively disproportionate power over online spaces in the hands of a few, huge companies.
- online decisions with serious real-life consequences are increasingly being made by algorithms and machines, replicating biases and reinforcing inequalities found offline.
- the web is not “for everyone” and most of the unconnected are marginalised populations in low- and middle-income countries, and most are women.
- a decline in the ability of anyone, anywhere to create and share content, ideas and information freely, driven by exposure to online harm including through misuse of data, malinformation and online hatred and abuse.

This paper focuses largely on the fourth of these main threats, the role that online harm plays both in undermining the open nature of the web and in creating offline harm. However, inevitably, a consideration of any one of these threats overlaps with the others, since the ways in which online hate has been spread is linked to both the role of algorithms in online decisions and the dominant role a handful of companies, including Google and Facebook, increasingly play in most people's experience of the web.

# OVERVIEW

This paper looks at the evidence as to the role online hate plays in spreading racist thinking and actions offline, including hate and terror crimes. It explores the evidence for whether harmful action online has increased both harmful ideas and actual harm against minority ethnic and racial groups.

Specifically we cover

- Hateful discourse on the internet and acts of violence,
- Freedom of all people to access and use quality information (including that related to issues of race and ethnicity),
- Spread of racist ideas and ways of thinking across populations and,
- Impacts on people from minority ethnic and religious groups in everyday life.

We start with a discussion on direct links between hateful discourse and group membership on the internet and acts of extremist violence. However, we emphasize that this is a very high burden of proof: most people do not commit acts of extreme violence making it harder to measure and research. The research evidence shows a complex systems effect at play in the relationship between online hate and offline harm, which requires we look beyond direct surface links between hate on the internet and acts of extreme violence when exploring the role and responsibilities of the open internet.

# 1. HATEFUL AND RACIST DISCOURSE AND HATEFUL CRIMES OR TERROSIISM

Empirical literature shows extremism has multiple causes like most human behaviour. The question is: does the internet play a role? The answer is certainly yes according to those who research these connections<sup>1</sup>. The mainstream social discourse, of which a large component now happens online, plays a fundamental part in acts of hate, violence and terrorism.<sup>2</sup> Mainstream discourse on the internet acts to network hate, not disinfect it via sunlight (as we discuss later).

The question is not if the internet plays a role in violent extremism, but how large a role it plays specifically in the contemporary radicalisation process<sup>3</sup>. Research from 2017 showed “while self-radicalisation is possible through the medium of the internet, physical contact played a significant role in the case of those interviewed. This evidence, while based on a small number of cases, also shows that while the internet facilitates the radicalisation of individuals it is not the sole driver of the process.”

Chan et al studied the link between 14,000 internet sites containing hate related content and official racial hate crimes and found high correlations, most evident in areas with higher racism and segregation. They also found a link between online hate sites and those racially motivated crimes by lone wolf perpetrators<sup>4</sup>. The Southern Poverty Law Centre noted in 2014 that active users of the website Stormfront.org were allegedly responsible or the murders of nearly 100 people in the previous 5 years.<sup>5</sup>

Raphael Cohen-Almagori has made extensive case studies of acts of race hate and violence and concludes “there is sufficient evidence to suggest that speech [on the internet] can and does inspire crime” and “Bigots, inspired by what they have read online, went on to inflict violence on their targets.” He includes analysis of the connection

---

1 Raphael Cohen-Almagori. “Taking North American White Supremacist Groups Seriously: The Scope and the Challenge of Hate Speech on the Internet”. International Journal of Crime, Justice, and Social Democracy (2018).

2 See the safehousealliance toolkit <https://www.safehousealliance.org/wp-content/uploads/2012/10/Tools-for-Liberation-Packet-SPAN.pdf>, based on 2005 research, by Paula Rothenberg. <http://fabletext.info/paula-s-rothenberg-race-class-and-gender-in-the-united-states-an-integrated-study-eighth-edition-fable-text.pdf>

3 Conway, Maura. “Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research.” Studies in Conflict & Terrorism 40.1 (2017): 77-98.

4 Chan J, Ghose A and Seamans R (2016) The Internet and racial hate crime: Offline spillovers from online access. MIS Quarterly 40(2): 381-403.

5 Beirich H (2013) Hate across the waters: The role of American extremists in fostering an international white consciousness. In Wodak R, Mral B and KhosraviNik B (eds) Right Wing Populism in Europe. Politics and Discourse: 89-104 London: Bloomsbury.

between hate groups and biochemical warfare in particular.<sup>6</sup>

One study looked at a sample of fifty-one Canada extremists to examine the role of social media both in the process of radicalization, and how extremists use social media after they become radicalized.<sup>7</sup> The results confirm that social media played a role either during or after the radicalization process of the majority of the sample and converts are more vulnerable to online radicalization than non-converts.

Nikita Malik, Director of the Centre on Radicalisation and Terrorism (CRT) at the Henry Jackson Society, has reported on recent (unpublished) research into which social media platforms were being utilized to spread hatred.<sup>8</sup> She found that people who were convicted for spreading material and planning hate attacks in the Far Right tended to use only two platforms to do this – Twitter and Facebook. By comparison, her previous research found that Islamist related content was spread on a variety of platforms including the Darknet and encrypted messaging services.<sup>9</sup>

In general, what the observational research shows is that case for the link between hate speech on the Internet and hate crimes has been well made, however more research is needed to understand the details.

We can also look to intervention research for information that observational studies may not give. Evidence shows that intervening in the spread of racist and hateful ideas and ways of thinking at multiple levels (including in public discourse which includes the internet) works to prevent terrorist acts. Figure 2 below outlines the macro, meso and micro level interventions that have evidence of impact on countering extremism (translated from German).<sup>10</sup>

---

6 Cohen-Almagor, Raphael. "Balancing Freedom of Expression and Social Responsibility on the Internet." *Philosophia* 45.3 (2017): 973-985.

7 Mehmet F. Bastug, Aziz Douai, Davut Akca (2017), Exploring the "Demand Side" of Online Radicalization: Evidence from the Canadian Context. *Studies in Conflict & Terrorism*.

8 Malik, N. Unpublished research reported in Forbes (2018), The Fight Against Terrorism Online: Here's The Verdict. Retrieved from <https://www.forbes.com/sites/nikitamalik/2018/09/20/the-fight-against-terrorism-online-heres-the-verdict>

9 Malik, N. (2018) Terror in the Dark: How Terrorists use Encryption, the Darknet and Cryptocurrencies Retrieved from <http://henryjacksonsociety.org/wp-content/uploads/2018/04/Terror-in-the-Dark.pdf>

10 Daniel Köhler & German Institute on Radicalization and De-Radicalization Studies (GIRDS). Structural Quality Standards for work to intervene with and counter violent Extremism. A handbook for practitioners, state coordination units and civil society programme implementers in Germany. <http://girds.org/publications>

Figure 2:

## THE COUNTER-TERRORISM NETWORK

	MACROSOCIAL	MESOSOCIAL	MIKROSOZIAL
PREVENTION (GENERAL, SPECIFIC)	e.g. education, research, youth work, social work	e.g. local crime prevention, federal state action plans (LAPs) etc.	e.g. workshops in schools with former extremists
REPRESSION	Legislative, executive, nationwide security architecture	e.g. banning orders, neighbourhood officers	e.g. banning orders, neighbourhood officers
INTERVENTION	Counter-narratives	e.g. family support	e.g. individual exit programmes

The following sections on the role of the internet in blocking access to good information, mainstreaming racist and white supremacist ideas from hate groups, and influencing public discourse and support for policies that address bigotry, racism and their impacts, highlight the importance of the role of the internet on hateful discourse and actions.

We emphasize the need to examine closely arguments that require a absolute 'proof' for the internet being a sole cause of an act of terrorism in order to take action. It does not reflect best scientific practice in an area as complex as this. And it may reflect the effects of the laundering and spread of harmful ideas and ideology into the mainstream public discourse, which plays a key role in undermining support for policies and practices that could limit bigotry and racism, and their harmful impacts, as we discuss later.

## 2. FREEDOM OF PEOPLE TO ACCESS GOOD INFORMATION

The evidence is clear that the internet has democratised the availability of information. However, the benefits derived from greater availability of information is countered by an infrastructure that make it increasingly difficult for people to locate and recognise good quality information.

Specifically, research shows that both the technical effect of the attention economy (advertising revenue linked to attention) and social psychological processes work to direct people using search engines and social media towards disinformation, misinformation, and material that confirms their existing incorrect or conspiratorial beliefs.<sup>1</sup> Echo chamber effects on Twitter, for example, work to restrict people's exposure to ideologically challenging discourse in particular.<sup>2</sup>

False information, particularly that of a polarized and political nature, evokes strong emotions, and hold people's attention. Taking advantage of basic human social and psychological processes, leading internet platforms have developed their core technology to capitalize on this behaviour, as spreading such information drives advertising revenue.<sup>3</sup> Cambridge Analytica shows that targeted political advertising is a highly profitable business.<sup>4</sup> Researchers found that false political news and viral types, notably terrorism, diffuses faster, deeper and more broadly than the truth on the internet.<sup>5</sup>

Even when people are made aware of misinformation and false information, their individual ability to identify and counter this is limited and often ineffectual due to the core technologies that spread misinformation. Tripodi undertook a study of conservative Christians in the US, who sought to critically interrogate media, news and other information, by fact checking source material against multiple internet sources.

---

1 Ghosh, D., & Scott, B. (2018). #digitaldeceit: The Technologies Behind Precision Propaganda on the Internet. Retrieved from <https://www.newamerica.org/public-interest-technology/policy-papers/digitaldeceit/>, Srnicek, N. (2017). Platform capitalism: Cambridge, UK: Polity Press.

2 Guo, L., Rohde, J. A., & Wu, H. D. (2018). Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S. election networks. *Information, Communication & Society*, 1-18., Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *Journal of Computer-Mediated Communication*, 18(2), 40-60

3 Ghosh, D., & Scott, B. (2018).

4 Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*(8), 56. doi:10.1109/MC.2018.3191268

5 Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.

The author found that “fact checking” returned further false information because 1) search terms used by people are biased, 2) search engines are returning results that are far-right or alt-right to mainline conservatives without them knowing, and 3) bad faith actors are exploiting the intellectual exploration of conservatives.<sup>6</sup>

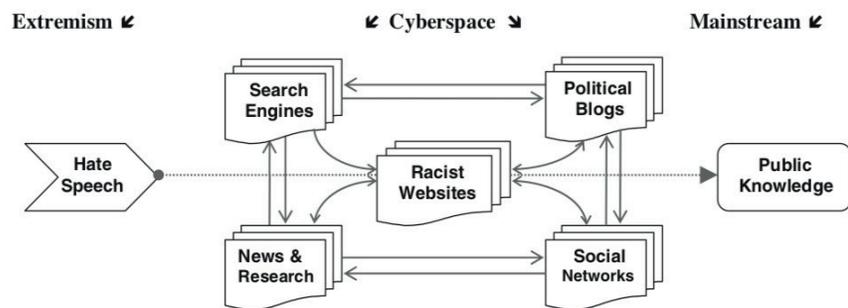
Alongside the democratisation of information there has been no commensurate democratisation in the structures, tools, or an individual’s ability to accurately identify (or assess) quality information.

---

<sup>6</sup> Tripodi F. Searching for Alternative Facts. Analysing Scriptural Inference in Conservative NewsPractices. 2018. Data & Society. Accessed May 2018 <https://datasociety.net/output/searching-for-alternative-facts/>

# 3. THE SPREAD OF RACIST IDEAS

Research shows that the loose and networked effects of the internet have acted to “launder” racist, and notably white supremacist, information into the mainstream. Dominant online structures do not expose these ideas to the sunlight of an open space, rather research shows they sanitise them and hide them in plain sight, so people cannot recognise them for what they are. Figure 1 below describes the process of “information laundering” of hate speech.



**Figure 1** Model of information laundering.

Communication Theory 22 (2012) 427–448 © 2012 International Communication Association

435

*“Through the Internet, hate groups are entering into mainstream culture by attaining legitimacy from the established media currencies of the cyber community, primarily search engines and interlinking social networks. These conventional pathways can unwittingly lead an online information seeker to extremist content that...has already been designed for them to appear as educational, political, scientific, and even spiritual in nature. More importantly, however, this network can also allow content from these websites to travel outside of their domain, to merge with mainstream spaces like YouTube, Facebook, political blogs, and even occasionally news cycles.” Klein, 2012*

For example multimedia convergence adds legitimacy to the white power websites, especially for young people who are familiar with and trusting of specific internet sources and brands. A study of 26 white supremacist websites in 2010 showed that YouTube videos appeared frequently, for example the creativityalliance YouTube channel.<sup>2</sup> Cognitive research shows that people have a poor memory for source material,

1 Klein, Adam. “Slipping racism into the mainstream: A theory of information laundering.” *Communication Theory* 22.4 (2012): 427-448.

2 Klein, A. G. (2010). *A space for hate: The White Power movement’s adaptation into cyberspace*. Duluth, MN: Litwin Books, LLC.

often misattributing bad information to trusted sources.<sup>3</sup> When those sources converge with malicious sources the problem is exacerbated.

The emergence of bias in the media is another effect of information washing. A meta analysis in 2017 found that “that Muslims tend to be negatively framed, while Islam is dominantly portrayed as a violent religion”.<sup>4</sup>

The case of VDARE.com shows how information laundering works in reality. VDARE.com emerged in 1999 as an anti-immigrant website in the US, and by 2003 was added by the Southern Poverty Law Center as a hate group. It both attracted pro-white communities and published anti-Hispanic and anti-Semitic essays. These online essays take the form of “scholarly texts” and become certified in the public domain, research that then becomes cited in media by commentators, influencing the public discourse and thinking. By 2012 VDARE had reached a mainstream audience, Its founder invited to speak at Conservative Political Action Conference (CPAC).

*“In this objective, the Internet becomes the key to activating the process of information laundering, turning hateful rhetoric into public knowledge.” Klein, 2012*

---

<sup>3</sup> Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, (2012) 13(3), 106-131.

<sup>4</sup> Ahmed, Saifuddin, and Jörg Matthes. “Media representation of Muslims and Islam from 2000 to 2015: A meta-analysis.” *International Communication Gazette* 79.3 (2017): 219-244.

# 4. IMPACTS ON PEOPLE

Online structures have the potential to both facilitate positive experiences for people of minority ethnic and racial groups and drive harm against them. The balance is currently tipped far towards harm.

## 4.1 DIRECT BENEFITS FROM THE INTERNET

The examination of several case-studies in the Australian context shows digital media can facilitate the formation of both 'ad-hoc' and longer-term, group-based online communities focused on fighting racism, for example IndigenousX.<sup>1</sup> Social media platforms, such as Twitter have been found to play a key role in the organisation of protest movements, like that of Arab Spring, Los Indignados and Occupy movements.<sup>2</sup> Social media may act to reduce political inequality, through encouraging political engagement.

The Workshop's own research in New Zealand shows social media is used by minority ethnic groups to access politicians and political action (unpublished survey research).

However, while evidence suggests the internet has given minority groups greater access to similar voices, collective action and political actors, it has also exposed them to more hate and bigotry. In 2017, one in ten New Zealanders experienced hate speech online and three in ten encountered hateful content.<sup>3</sup> One in three New Zealand women experienced online abuse and harassment. Of those women, three in four (75%) said they had not been able to sleep well, one in two (49%) feared for their physical safety and one in three (32%) feared for the physical safety of their families as a result.<sup>4</sup> In 2018, one in three Māori (32%), and one in five Asian (22%) and Pacific (21%) people experienced racial abuse and harassment online.<sup>5, 6</sup>

---

1 Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A.-M., Bahfen, N., Connolly, K. (2017). *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*: Springer.

2 Bennett, W. L., Segerberg, A., & Walker, S. (2014). Organization in the crowd: peer production in large-scale networked protests. *Information, Communication & Society*, 17(2), 232-260.

3 Netsafe, (2018) *Online Hate Speech A survey on personal experiences and exposure among adult New Zealanders*.

4 Amnesty International New Zealand (2018), *Amnesty reveals alarming impact of online abuse against women*

5 ActionStation (2019), *The People's Report on Online Hate, Harassment and Abuse*.

6 Netsafe has published two additional reports which reflect the findings from the Amnesty and ActionStation reports, and in addition include people in the 14-17 age group, and provide a measure of the distress and harm caused. They are: *Netsafe (2018) New Zealand Teens and Digital Harm: Statistical Insights into Experiences, Impact and Response* and *Netsafe (2017) Harmful Digital Communications in New Zealand: Annual Population Survey 2017*. A Netsafe third report included findings about the experience of non-heterosexual adults (e.g., young adults aged 16-29) and those who identify as non-heterosexual were more likely to be targeted by Image-Based Sexual abuse). *Netsafe (2019) New Zealand Teens and Digital Harm: Statistical Insights into Experiences, Impact and Response*.

## 4.2 DIRECT HARM FROM THE INTERNET

Hate speech pre-dates the internet. However, alongside the rise of digital technology, there has been a 66% increase in hate groups in the US since 2000 (602 groups to 1002).<sup>7</sup> In 2011 there were over 14,000 documented hate sites, blogs, and social networks operating across the web.<sup>8</sup>

The structures and features of dominant online tools and platforms, including capitalization of people's attention, the design of search engines and social media algorithms, networking effects across websites, media and social media, allow for increasingly coordinated and targeted hate towards minority-ethnic groups. Research also found the use of Encryption, the Darknet and Cryptocurrencies to engage in activities of hate and extremism.<sup>9</sup>

Racialised hate speech (otherwise known as cyber racism) is specifically targeted towards ethnic minority groups, and has become increasingly coordinated in recent years, through the rise of the "alt-right". It has become a global phenomenon, affecting "refugees and ethnic minorities in Europe, Muslim Blacks and Jews in the United States, Indigenous Australians" and others. The impact on affected groups is to retreat to safe locations to find solidarity, instead of engaging in public discourse and institutions, where thinking may be changed.<sup>10</sup> Spirals of silence have been found in large surveys, where majority groups act to suppression of minority groups dissenting views and force withdrawal into safe spaces.<sup>11</sup>

Two social phenomenon should be mentioned here - pluralistic ignorance. Where the amplifying effects of the internet have been found to push those with more moderate views to polarised extremes due to a belief extreme views are more common than they are. And the false consensus effect, where by those with hateful and racist views believe their views to be more widely shared.

Anne Pedersen has carried out extensive research into the beliefs and attitudes about minority groups in Australia.<sup>12</sup> In one of her studies people were asked about their attitudes to Aboriginal Australians and asylum seekers. They were also asked to estimate what percentage of people they believed held the same attitudes as them. Whether an attitude was positive or negative, people tended to overestimate the community support for their own beliefs. However, those who held the most fringe and negative beliefs about these two groups of people, vastly overestimated the support for their beliefs in the community compared to those with positive attitudes. Those with positive beliefs estimated 46% of the community shared their attitudes, those with negative beliefs

---

7 U.S. Hate Groups Top 1000 (2011). Retrieved from <http://www.splcenter.org/get-informed/news/us-hate-groups-top-1000>

8 Simon Wiesenthal Center. (2011). 2011 digital terrorism & hate report launched at Museum of Tolerance New York. Retrieved from <http://www.wiesenthal.com/site/apps/nlnet/content2.aspx?c=lsKWLBpJLnF&b=6478433&ct=9141065>

9 <http://henryjacksonsociety.org/wp-content/uploads/2018/04/Terror-in-the-Dark.pdf>

10 Jakubowicz, A., et al. (2017). *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*: Springer.

11 Chen, H.-T. (2018). Spiral of silence on social media and the moderating role of disagreement and publicness in the network: Analyzing expressive and withdrawal behaviors. *New media & society*, 20(10), 3917-3936

12 Pedersen, Anne, Brian Griffiths, and Susan E. Watt. "Attitudes toward Out groups and the perception of consensus: All feet do not wear one shoe." *Journal of Community & Applied Social Psychology* 18, no. 6 (2008): 543-557.

estimated 71% of the wider community shared their beliefs. The actual figure is 1.9% for the most extreme negative views. The more negative an attitude towards Aboriginal Australians or asylum seekers, the greater the estimate of community consensus became.

#### **4.3 INDIRECT HARM FROM RACIST DISCOURSE ON POLICY TO ADDRESS THE PROBLEM**

Washing information into the mainstream helps build and maintain a public environment less likely to support policies that protect the freedoms and rights and well-being of minority groups. It furthers shallow understanding of minorities lives and experiences, focussing instead on explanations that include individual and moral failings. This causes genuine harm - policies that fail to address long term negative outcomes, and reinforce them. A large scale analysis of empirical research in health Williams and Mohammed show how different aspects of racism affect wellbeing.<sup>13</sup>

First, Williams and Mohammed found evidence that racism at an institutional level leads to policies and actions that limit Black Americans' access to important resources and opportunities; for example, by restricting access to certain neighbourhoods, education, employment and other community resources (also observed in New Zealand research<sup>14</sup>). Second, racism embedded in cultural narratives (such as internet, media and popular culture) shapes negative emotions about Black people, and leads to stereotypes and prejudice that damage people's wellbeing. Finally,

they found a large body of evidence showing that "experiences of racial discrimination are an important type of psychosocial stressor that can lead to adverse changes in health status and altered behavioural patterns that increase health risks."

The authors draw our attention to the role that the absence of positive feelings towards stigmatised groups has in shaping policy preferences of wider society. The cultural narratives create a social and political environment hostile to policies that attempt to change the social and economic conditions that drive harm. Altered behaviours and health status that stem from wider social conditions are interpreted by others through the narratives of discrimination. These narratives posit that it is inferiority, biology and/or individual weakness that determine poorer outcomes, not social, economic, cultural or environmental conditions. This interpretation prevents support for action, such as the introduction of policies and regulations, to change those conditions.

---

<sup>13</sup> Williams, D. R., & Mohammed, S. A. (2013). Racism and health I: Pathways and scientific evidence. *American Behavioral Scientist*, 57(8), pp. 1152-1173.

<sup>14</sup> Harris, Ricci, et al. "Racism and health: The relationship between experience of racial discrimination and health in New Zealand." *Social science & medicine* 63.6 (2006): 1428-1441.

# TE TIRITI O WAITANGI

While out of the scope of this paper, we recommend that an assessment be made of the impact of online hatred and abuse on the rights and protections set out in Te Tiriti o Waitangi, as a foundational document of New Zealand.

# CONCLUSIONS

The evidence for the connection between online hate and 'real-life' harm is clear: there is a connection, albeit a more complex one than some have hypothesised. And although we need more research to better understand the details of how that connection works, sufficient evidence already exists to make the case for action to, as Sir Tim Berners-Lee has put it 'save the web'.

*As The Case for the Web report concludes, 'The web is under threat - but the web we want is not out of reach. It's up to us to overcome these threats and ensure the web remains an open platform that is truly a force of good for everyone.'*

Specifically in relation to online hatred and abuse, The Case for the Web report recommends that governments and companies must enact policies and enforce regulations to protect the right to safety alongside the right to freedom of speech.

In New Zealand, the public conversation about protecting the right to safety alongside the right to freedom of speech has too often been overly simplistic and has ignored the growing evidence base on the clear yet complex role that online hatred plays in creating, spreading and sustaining 'real life' harm. 'Real life harm' is too often understood only in the form of acts of violent extremism. Obviously such acts are a very serious concern, and have a particular relevance to public discourse in New Zealand at the moment.

However, research shows that online hatred and bigotry also do real-life harm in other, more complex and interrelated ways including - amongst others:

- through the 'laundering' effect of hiding extreme views plain site in mainstream public discourse,
- through the direct psychological harm of encountering hatred online, and
- through the influence of racist and bigoted discourse on support for policies to protect the freedoms, rights and well-being of minority groups.

The current debate about the role of harmful online content in perpetrating real-life harm provides a critical opportunity to refocus the debate away from a false dichotomy between the right to safety and freedom from discrimination, on the one hand, and the the right to freedom of expression on the other.

New Zealand is currently well-placed to contribute to a broader, more well-informed and evidence-based conversation about the relationship between online hatred and real-life harm. There is an important role for New Zealand to play in reclaiming the values of the open web, and reasserting the critical components of an open internet.

If the open web is to be saved, it's salvation will lie in a broader public understanding of the critical role of open and diverse internet governance, open architecture and open technology rather than the current focus on a narrowly defined, and widely misunderstood and misrepresented, concept of freedom of expression.

Broadening this conversation, and ensuring that it is informed by the best available evidence, is critical work.



