Disabling access to illegal online content by way of takedowns — [2021] NZLJ 341

New Zealand Law Journal

Disabling access to illegal online content by way of takedowns

Rachel Sue Yin Tan, University of Waikato, with a strategy for online hate speech regulation

INTRODUCTION

Social media usage has proliferated over the past decade with an approximate 2.95 billion reported users worldwide ("Number of social media users worldwide 2010–2021" <www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>). To many, social media brings a form of escapism, and especially in times of a pandemic, it serves as a means to communicate to others and a method of forming connections. ***Online*** hate speech is on the rise especially after the 15 March 2019 terror attacks on the Muslim community in the city of Christchurch. Netsafe received approximately 600 complaints and enquiries in relation to hateful digital communications in the days after the attack (Neil Melhuish and Edgar Pacheco "Measuring trends in ***online*** hate speech victimisation and exposure, and attitudes in New Zealand" (2019) at 15 <dx.doi.org/10.2139/ssrn.3501977>). The Christchurch terror attacks made people think about the fine line between freedom of speech and hate speech. ***Takedowns*** have been introduced as part of a Bill in Parliament following the Christchurch Call ("Christchurch Call" <www.christchurchcall.com/call.html>). While the effects of ***takedowns*** are undisputed, more refinement is needed for their application. Other measures exist, namely ***content*** censorship, geo-blocking and web-filtering which also play important roles in combatting ***illegal content***.

The Christchurch terrorist planned his attack using social media as a tool to livestream the shootings for 17 minutes on social media ("Christchurch Call" <www.christchurchcall.com/call.html>). The Christchurch events spurred people to deliberate on whether there are sufficient measures to regulate ***online*** hate speech. As a result, the Government introduced a Bill in early 2020 that seeks to amend the current *Films, Videos, and Publications Classification Act 1993* (hereinafter "the Act"); the Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of ***Online*** Harm) Amendment Bill (hereinafter "Amendment Bill"). Along with that the *Royal Commission of Inquiry into the Terrorist Attack on Christchurch Mosques on 15 March 2019* that was established to investigate matters of national priority ***by way*** of examining the facts and findings, finally presented its recommendations to Parliament (Questions about the Royal Commission of Inquiry).

The European Union's e-Commerce Directive particularly, arts 12-14 of Directive 2000/31/EC stipulates that there is a limited liability imposed onto service providers for ***content*** posted by a third party (Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce') 2000 (European Union) <eur-lex.europa.eu/legal-***content***/EN/TXT/?uri=uriserv%3AOJ.L_.2000.178.01.0001.01.ENG>). Member States of the European Union such as France and Germany are actively refining their legal frameworks.

Methods of disabling online access — censorship

Censorship is a measure used to curtail ***online*** hate speech. Throughout history, censorship has been used by European rulers in imposing a system to obtain government licenses for the printing and publication of scientific and artistic expressions. The reasoning for this was purely in order to restrict the publication of ***content*** that was thought to threaten morality and public order (Mette Newth "The Long History of Censorship" (2010) <www.beaconforfreedom.org/liste.html?tid=415&art_id=475>).

A speech called *Areopagitica* by John Milton in the 1640s is arguably the tipping point in English censorship history (Newth, above). It addressed the English Parliament and challenged the government over the Licensing Order 1643 which controlled publication and censorship at the time. It can be said that Milton is an important personality in the concept of freedom of speech and the free press, as articulated in his four arguments put forward to Parliament below:

1. With respect to the prepublication censorship with the Catholic Church eradicate "all those authorities, civil as well as religious, which refused to recognize its dominion" (Hilary Gatti "The Humanities as the Stronghold of Freedom John Milton's Areopagitica and John Stuart Mill's On Liberty" in Rens Bod, Jaap

Maat, and Thijs Weststeijn (eds) *The Making of the Humanities* (Amsterdam University Press, 2012) at 167–182).

2. With respect to moral values, readers should be exposed to both morally acceptable and incorrect books (Kevin R Davis "John Milton" The First Amendment Encyclopedia — Presented By The John Seigenthaler Chair of Excellence In First Amendment Studies <www.mtsu.edu/first-amendment/article/1259/john-milton>).
3. It is an ineffective method of protecting public morality and religion — to censor is by all means impractical (Davis, above).
4. With respect to licensing, it has an effect of diminishing the public's ability to reason as they only accept authority (Davis, above).

In order to get a good grasp of censorship, a distinction is made between ex-ante censorship ('prior restraint') with ex-post measures (post-publication measures) to ensure it does not spread to the wider community (Andreas J Wiesand and others *Culture and Human Rights: The Wroclaw Commentaries:* (De Gruyter, Berlin, Boston 2016)). The Internet is ever-changing and can be used to disseminate information and ***content*** easily with a press of or click of a button (Wiesand, above). While its origins encompass the censorship on books, it has now evolved and shifted its primary focus to other literary expressions such as caricatures, computer games and other digital media.

In February 2021, social media giant Facebook decided to block all Australian news outlets on the platform. The Australian government tabled to Parliament a media legislation which is a world first. The new media legislation would effectively force a negotiation between Google and Facebook with media and news organisations for the usage of their ***content*** on the respective social media platforms' newsfeed and Google searches (Amanda Meade "Australia is making Google and Facebook pay for news: what difference will the code make?" *The Guardian* (***online*** ed, 9 December 2020)). The rationale behind this media legislation is to safeguard the livelihood of news media businesses to ensure they are fairly paid for their ***content*** (Meade, above). This comes as a result of the closure of many news media businesses in the country in 2020.

The issues with censorship can be illustrated in the case of *Google Inc v Equustek Solutions Inc* ([2017] 1 SCR 824 Supreme Court of Canada 36602 2001-01-01) which took place in British Columbia. The case involved a small tech company, E, launching a legal action against D, who while acting as E's product distributor had re-labelled the products to pass them off as its own (at 3). To make matters worse, D also made away with trade secrets and confidential information belonging to E. D initially filed a statement of defence. However, it was dropped as he absconded (at 52). The issues in that case were "(i) whether Google can be ordered, pending trial of action, to globally de-index websites of a distributor which, in breach of several court orders, is using those websites to unlawfully sell intellectual property of another company; (ii) whether the Supreme Court of British Columbia had jurisdiction to grant injunction with extraterritorial effect; (iii) whether, if it did, it was just and equitable to do so" (*Google Inc v Equustek Solutions Inc,* above). In that case, the majority judgment failed to recognise the fact that it was at a juncture to (i) step away from hyper-regulation of ***illegal content***; (ii) refer to and apply geo-location technologies instead; and therefore, (iii) failed to address the issues pertaining to jurisdiction (Dan Svantesson "Supreme Court of Canada challenges the idea of state sovereignty" (29 August 2017) Oxford University Press's Academic Insights for the Thinking World <blog.oup.com/2017/08/supreme-court-canada-state-sovereignty/>).

Flowing from that, the usage of censorship to restrict people from their basic freedom of speech and expression is a human rights issue (Liza Negriff "The Past, Present, and Future of Freedom of Speech and Expression in the People's Republic of China" (2009) Topical Research Digest: Human Rights in China at 130 <www.du.edu/korbel/hrhw/researchdigest/china/China.pdf>). These examples demonstrate that censorship is a potent method of ***disabling access*** thus it makes one contemplate its effectiveness against a wide range of problems and issues. While it can be used to combat hate speech, it can be used to hinder people from enjoying civil liberties.

Methods of disabling online access — geo-blocking

Every computer or device has a unique numerical identifier or Internet Protocol (hereinafter "IP") address that can be used to determine a user's geographical location. These IPs are distributed to Internet Service Providers (hereinafter "ISP") who then distributes them to its customers (Karl Schaffarczyk "Explainer: what is geoblocking?" (2021) The Conversation <theconversation.com/explainer-what-is-geoblocking-13057>). When a computer sends a request to the server for ***access*** to ***content***, the server will know where to send the request based on the IP address (Schaffarczyk, above). For example, when you attempt to watch an episode of a programme, for

example, Saturday Night Live on NBC's platform app or its website that is based in the United States of America while you are located in New Zealand, you might be receive a message indicating that 'this video is not available in your location'. This is an example of geo-blocking whereby ***content*** is geographically blocked, or ***content*** is accessible only from a specified geographic location. This comes with several ramifications such as circumventing a blocked site or platform; privacy implications which depend on the range within which an internet intermediary is able to link a user's location; and the over-utilising of geo-blocking can have severe effects (Dan Jerker B Svantesson "Delineating the Reach of Internet Intermediaries' ***Content*** Blocking — ccTLD Blocking, Strict Geo-Location Blocking or a Country Lens Approach" 2014 11 SCRIPTed). To circumvent geo-blockers, users can turn to a Virtual Private Network (hereinafter "VPN") to ***access*** ***content*** that they otherwise would not be able to. A VPN functions as an intermediary. Instead of altering or changing a user's IP address, it sends a request to the server on behalf and then, delivers the ***content*** at the end of that process (Schaffarczyk, above). In addition to VPNs, there is The Onion Router (hereinafter "TOR") which was originally created for users to enhance their privacy and ***online*** security. It has in recent times been used by users to circumvent geo-blockers ("The Tor Project" <www.torproject.org/about/history/>). It is, in other words, a circumvention tool used to remedy censorship, geo-blockers, or to ***access*** other ***content***. It works also to ensure that users are not tracked so their information and data is not exploited by organizations (The Tor Project, above). But VPNs and TOR may also be utilised by individuals with criminal intentions such as child pornography, scammers and ***content*** relating to ***online*** hate speech, in particular distribution. Where there is a *will* there is a *way.* No matter how much regulation, determined people will still find a way around it by the aforementioned means.

The importance of mentioning geo-blockers as a means of circumvention is that it is essential to be aware of the other ways of obtaining ***content*** and information ***online***. This demonstrates that the cons outweigh the pros particularly in using geo-blocking to combat hate speech. Hateful material ***online*** can be found by anyone who is adamant in locating it and this can be done anonymously.

Methods of disabling online access — web-filter

Web-filters are another method of combatting ***online*** hate speech. As mentioned prior, the Film, Videos and Publications Classification Bill looks to update the current Act. The Bill intends to amend and to introduce five key changes to the Act, one of them being a web-filter. This introduction elicited the attention and to an extent, disapproval of legal experts, tech experts and civil libertarians opposing the Bill (Andrew Chen and New Zealand Council for Civil Liberties "Submission to the Governance and Administration Committee on the Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of ***Online*** Harm) Amendment Bill"). This is mainly because the Bill features a *future* (not developed yet) web-filter that will generally block and filter objectionable ***content*** on the internet. Since no further information as to its mechanisms were mentioned in the Bill, the worst was assumed by those opposing the Bill in that it is potentially dangerous should *future* New Zealand government be more inclined to being totalitarian. Currently, there is one web-filter operating, but for tackling child exploitation. This filter is government-backed and is called the Digital Child Exploitation Filtering System which blocks sexual and exploitative material (***Rachel Tan*** "Submission to the Governance and Administration Committee on the Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of ***Online*** Harm) Amendment Bill").

The Film, Videos and Publications Classification Bill also provides for ***takedown*** notices from the web (Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of ***Online*** Harm) Amendment Bill (2020) 268-1, pt 2 and see Amendment Bill Digest 2626 <www.parliament.nz/en/pb/bills-and-laws/bills-digests/document/52PLLaw262611/films-videos-and-publications-classification-urgentharm>). The world of social media is akin to a free and open space for users to escape, interact and is also seen as a venue to exchange views within its limits. However, in an instance in which a user conducts themselves contrary to respective community guidelines and rules, then a ***takedown*** notice is issued and a social media platform then removes ***content***. In extreme situations in which not only the post is contentious, but its account holder is essentially producing an inciteful and hateful post, then a social media platform may deplatform the account in question. This idea is not novel as it has been discussed since about 2017 as a method of addressing cyberbullying using the notice and ***takedown*** model (Brian O'Shea "A New Method to Address Cyberbullying in the United States: The Application of a Notice-and-***Takedown*** Model as a Restriction on Cyberbullying Speech Notes" (2017) 69 Fed Comm LJ).

The landmark case of *Google Spain v González* (*Google Spain SL, Google Inc v Agencia Española de Protección de Datos (AEPD), Mario Costeja González* 2014 European Court of Justice) lead to Google setting up an Advisory Council that comprises of its legal division and external communities who collectively decide to grant ***takedown*** requests.

This would involve a court order requesting a search engine, for example Google, to remove specific data such as personal information from search results.

The caveat here is that the data and information in question may still be accessible and searchable in other search engines or in archives. Borrowing the concept of the right to be forgotten from the privacy perspective, deplatforming may be a method of regulating ***online*** hate speech in social media. The concept of deplatforming is defined as a removal of a user's account on social media upon breaking respective platform rules and standards (Richard Rogers "Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media" (2020) 35 European Journal of Communication 213–229). This means that social media takes further steps to ensure that the user's account in question ceases to exist in its platform and along with it, all ***illegal*** hate speech.

It is argued that the method of deplatforming could work; however, it is not a long-term remedy.

An example of deplatforming was the removal of Gab (a less moderated social media platform) resulting from the Pittsburgh Synagogue and high school shooting in the United States. Gab was used to spread and incite hateful sentiments to the Jewish community and anti-immigrant sentiments by alt-right groups in the city of Pittsburgh (Jeremy Blackburn, Robert W Gehl and Ugochukwu Etudo "Does 'deplatforming' work to curb hate speech and calls for violence? 3 experts in ***online*** communications weigh in" (16 January 2021) The Conversation <*theconversation.com/does-deplatforming-work-to-curb-hate-speech-and-calls-for-violence-3-experts-in-**online**-communications-weigh-in-153177*>. Gab was not efficient in policing hateful expression unlike its conventional rivals, Facebook or Twitter, because it took a passive approach (Abby Vesoulis "How Gab Became the Social Media Site Where the Pittsburgh Suspect's Anti-Semitism Thrived" Time (Washington, 27 October 2018)). After the unfortunate events, app stores and cloud infrastructure providers removed Gab from its platforms. Recently, as a result of the United States Capitol Hill riot on 6 January 2021 (Khavin Dmitriy and others "US Capitol Riot" *The New York Times* (***online*** ed, 15 August 2021), Twitter took action to permanently suspend Donald Trump's Twitter account (other social media accounts did so too such as Apple, Google) (Blackburn, Gehl and Etudo, above). Deplatforming could indicate a paradigm shift in which social media platforms have powers and should also be responsible for combatting ***online*** hate speech.

Facebook has a big role to play in relation to regulating hate speech as a result of its number of users. The onus set on social media companies is greater than of governmental or legislative enforcement. It is reported that Facebook deletes nearly 288,000 hate speech posts per month (Richard Ashby Wilson "*Hate: Why We Should Resist it with Free Speech, Not Censorship* by Nadine Strossen (review)" 2019 41(1) Human Rights Quarterly).

The concept of *flagging* is deployed as a tool to regulate ***illegal*** ***content*** and ***online*** hate speech. To *flag* means to report offensive or ***illegal*** ***content*** to a social media platform; it gives users the capability to express any concern that contravenes the social media platform's Community Guidelines. Not only is this a technical feature of some social media platforms, but a complex interplay between "the users, platforms, human and algorithms, and the social norms and regulatory structures of social media" (Kate Crawford and Tarleton Gillespie "What is a flag for? Social media reporting tools and the vocabulary of complaint" (2016) 18(3) New Media & Society 410). In terms of Facebook and Instagram's Oversight Board, on the one hand, its existence indicates that it has poor policies at its Community Standards level, while on the other, it shows that it is an independent appellate course of action (Evelyn Douek "Facebook's 'Oversight Board:' Move Fast with Stable Infrastructure and Humility" (2019) 21 North Carolina Journal of Law & Tech 1).

It has been recommended that instead of looking at a legislative lens, a more holistic approach ought to be applied (Evelyn Douek "Governing ***Online*** Speech: From "Posts-As-Trumps" To Proportionality And Probability" (2021) 121(3) Columbia Law Review 759). It is observed that there is a paradigm shift of ***content*** moderation since the beginning of the COVID-19 pandemic; in that we find that social media platforms are now "balancing societal interest and choosing between error costs on a systemic basis" (Douek, "Governing ***Online*** Speech", above). As a result of this shift, their regulating or governing speech would need to be adapted — failure to do so would mean that platform rules and guidelines that govern hate speech will be viewed as illegitimate simply because there is no direct answer to the questions in writing the rules surrounding ***online*** hate speech. It is vital that the rule formation process obtain public acceptance in order to be viewed as legitimate and acceptable (Douek, "Governing ***Online*** Speech", above).

Lessons from other jurisdictions

France and Germany are two countries that have developed ***online*** hate speech regulations in recent years. These

countries have faced lots of controversy with regards to hate speech and in recent times, have enacted (or attempted, in France's case) to enact laws to protect the public from ***online*** hate speech. France attempted to enact hate speech laws, only to have its Constitutional Court strike most provisions out, and Germany enacted its hate speech laws successfully albeit with some controversy.

In 2019, France drafted a piece of legislation to regulate hate speech called "Avia Law". Specifically, the laws gave the powers to social media platforms to regulate their spaces and ***takedown*** (within a very brief period of time, 24-hours) any hateful ***content*** found to be discriminatory pursuant to race, gender, disability, sexual orientation and religious background (Chloe Hadavas "France's New ***Online*** Hate Speech Law Is Fundamentally Flawed" (26 May 2020) Slate <slate.com/technology/2020/05/france-hate-speech-law-lutte-contre-haine-sur-internet.html>). For other ***online*** harms, primarily in relation to child pornography and terrorist activities, the laws established that these platforms should remove the ***content*** pursuant to the aforementioned within one hour of flagging (Hadavas, above). Failure to remove harmful within the stipulated time frames would mean a substantial fine; approximately EU 1,250,000 which is equivalent to NZD 2,085,687.50 (Hadavas, above).

The French National Assembly took a very narrow stance potentially making the country authoritarian and akin to George Orwell's 1984. As a result, the French Constitutional Council struck down several provisions in the draft legislation. The reason was because it went against freedom of speech and was, therefore, unconstitutional. The Constitutional Council also felt that the response times for ***takedowns*** were too short thereby increasing the risk of over-censorship or over-blocking.

To eradicate ***online*** hate speech, Germany enacted the "network enforcement act" officially called Netzwerkdurchsetzungsgesetz (hereinafter "NetzDG"). The NetzDG was enacted by the Bundestag in 2017 which had comparable objectives like France's Avia Law. Its objectives include combatting hate crime and regulating social media networks effectively (Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act, NetzDG) — Basic Information (2017) 2021, art 1, s 1). Article 1 of the NetzDG stipulates that for an effective framework, there is a requirement for transparency to handling complaints; it also developed the following obligations for social media platforms to adhere to:

- Reporting procedure — social networks must develop a reporting system for users to report ***illegal content*** and it must be easy, direct, permanent and accessible for its user (art 1, above).
- Assessment of ***illegal content*** in question — respective social media platforms must assess reported ***content*** and determine if the ***content*** in question does in fact violate the law or not (art 1, above).
- ***Takedown*** or block ***access*** — social networks must remove the ***content*** within 24-hours (upon receiving the complaint) (art 1, above).
- Other criminal ***content*** — ***takedown*** or blocking ***access*** of such ***content*** must be within 7 days upon receiving the complaint. If the ***content*** in question comes from "a recognised institution of regulated self-governance" it too will have 7 days to make a decision on the ***content*** (art 1, above).

The scope of the NetzDG lies within s 1 and it crosses over to the German Criminal Code (Strafgesetzbuch — StGB) to provide the meaning of unlawful ***content*** by virtue of the country's criminal code (German Criminal Code (Strafgesetzbuch — StGB) 2021 (15 May 1871), s 130 Incitement of masses). According to s 1(3) of the NetzDG, there are several sections in the German Criminal Code that defined unlawful ***content***. As an overview, unlawful ***content*** is classified into these categories: "endangering democratic state under rule of law; treason and endangering external security; resistance to state authority; offences against public order; offences relating to religion and ideology; offences against sexual self-determination; insult; offences against personal liberty; forgery of documents" (ss 86, 86(a), 89(a), 91, 100(a), 111, 126, 129 to 129(b), 130, 131, 140, 166, 184(b) in connection with 184(d), 185 to 187, 241 or 269, above). To put it into context, the relevant category for hate speech is offences against public order and s 130 of the German Criminal Code defines the provisions for the incitement of masses. That is to say, anyone who (1) incites hatred *against* a group of people (based on race, religion, background, section of the population or individuals belonging to one of the categories) and (2) violates human dignity will face jail term from between the minimum of three months to the maximum of five years (s 130(1), above).

From another point of view, there is an argument that the NetzDG essentially violates the International Covenant on Civil and Political Rights (hereinafter "ICCPR"). This is because art 19 of the ICCPR provides for individual rights and freedoms — that is to say that individuals have a "right to hold opinions without interference; right to freedom of expression that includes the freedom to seek, receive and impart information and ideas of all kinds; the right to exercise the rights aforementioned comes with responsibilities and duties which may be subject to restrictions provided by law eg, respecting the rights and reputations of others and when it involves national security or to

defend public order, public health or morals" (International Covenant on Civil and Political Rights (23 March 1976), art 19). This means that the challenge here is that the NetzDG goes against basic freedom of expressions in the ICCPR and is at the same time overly-regulating society (Rebecca Zipursky "Nuts About NETZ: The Network Enforcement Act and Freedom of Expression" (2019) 42(4) Fordham International Law Journal 1325).

An example of this relates to an incident that happened around 2016, involving a journalist and satirist, Jan Böhmermann who published a poem about Recep Tayyip Erdogan, Turkey's President. The poem was provocative as it was ridiculing and lampooning the President's genitalia (Deutsche Welle "German satirist Jan Böhmermann ***sues*** Angela Merkel over Erdogan poem remark" (2 April 2019) <www.dw.com/en/german-satirist-jan-böhmermann-***sues***-angela-merkel-over-erdogan-poem-remark/a-48158329>). As a result, Erdogan demanded that the German government act on the insults by approving criminal proceedings against Böhmermann enacted by Chancellor Angela Merkel, which was a questionable move at the time (Zipursky, above). It was considered controversial because the particular section used against Böhmermannwas s 103 of German Criminal Code (that has since been repealed) which was relied on by the Turkish President for the mere fact that it prohibits "insults against organs of representatives of foreign states" (Philip Oltermann "Obscure German law gives Angela Merkel a diplomatic headache" (2016) *The Guardian* <www.theguardian.com/world/2016/apr/14/obscure-german-law-angela-merkel-recep-tayyip-erdogan>). Germany faced immense pressure from the Human Rights Watch for the arrest and the activation of criminal proceedings against Böhmermann and as a result of this pressure, its Parliament reviewed and repealed that section from its criminal code (Zipursky, above). Prior to the repeal of this provision, it meant that German law could silence its citizens if it "was about a matter of public safety and wellbeing" (Zipursky, above). What we can learn from this is that even though Germany had initially been non-compliant with art 19 of the ICCPR, it consequently repealed s 130 of the German Criminal Code which is indicative of the nation's commitment to adhere to the general rule of international and human rights law (Zipursky, above).

Even though the legal issue is whether s 130 of the German Criminal Code is in fact proportionate to art 19 of the ICCPR that was designed to limit hate speech that incites violence, the fact that Germany repealed that provision in its criminal code speaks volumes of its aspiration to modernise its laws. By doing so, Germany demonstrates that it values the morality of the speech when it comes to censorship; and as a result, Germany has defended the NetzDG against critics by arguing that NetzDG protects both German civil rights and liberties (Zipursky, above).

One other important factor to consider is the notion that human dignity is inviolable, which is enshrined in ch 1 of the Fundamental Rights of the European Union. This is because Germany's Constitutional Courts, being the highest in the country, refers to human dignity as the core of the country's basic rights (Zipursky, above). In other words, the legal system in the country deems hate speech as a crime that should not receive any defense. This contrasts with the United States of America's First Amendment, where free speech is embedded into its core legislative framework (Zipursky, above).

In the process of effectively managing ***illegal*** ***content***, social networks are obliged by this Act to ensure that users are informed on decisions made along with justifications. It is worth noting that this new law applies to all social media platforms based outside of Germany (Bundesministerium der Justiz und für Verbraucherschutz "FAQ: Act to Improve Enforcement of the Law in Social Networks, 2017" (2021) <www.BMJV.de/SharedDocs/FAQ/EN/NetzDG/NetzDG.html>). This means that NetzDG imposes its laws and obligates social media networks to adhere to their framework even though they are based outside of Germany, most notably from the United States of America and China.

NetzDG functions to "encouraged accountability and transparency from large social media platforms"; however, critical questions about freedom of expression and the potential chilling effects of legislation come to mind (Heidi Tworek and Paddy Leerssen "An Analysis of Germany's NetzDG Law" (15 April 2019) A working paper of the Transatlantic High Level Working Group on ***Content*** Moderation ***Online*** and Freedom of Expression <www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf>). NetzDG canvasses most social media platforms such as Facebook, Instagram, Twitter, YouTube, Snapchat and Google, but not for messaging services like WhatsApp and professional network platforms like LinkedIn and Xing (Ben Knight "Germany implements new internet hate speech crackdown" (1 January 2018) DW <p.dw.com/p/2qBvi>).

NetzDG gave the federal government absolute powers parallel to censorship. It is in the view of the country's Federal government that social media platforms are not sufficient in their regulation of ***online*** hate crime. Aside from the imposition of a hefty fine of EU 50,000,000 (approximately NZD 84,000,000) (Knight, above) tech companies must produce an annual report containing the number of complaints and its decision-making practices along with details about the team responsible for generating the report (NetzDG, above). All this is to be released and accessible to the main public.

CONCLUSION

An overview of how the European Union law regulates ***illegal*** ***content*** is illustrated in their report (Sally Broughton Micova and Alexandre de Streel "Digital Services Act — deepening the internal market and clarifying responsibilities for digital services" (Centre on Regulation in Europe, Report, December 2020)) to the European Commission. The e-Commerce Directive is a *horizontal regulation* that is like an over-arching umbrella that covers and is applicable to all platforms and all types of ***illegal*** or harmful ***content***. On the other hand, *vertical regulation* is also used which consists of specific regulations adopted to regulate ***illegal*** ***content*** under EU law (Broughton-Micova and de Streel, above). It is at the *horizontal* level that the EU intends to update the current framework ***by way*** of having clear and definitive responsibilities in place pertaining to the obligations of digital services. This is demonstrated in the European Union's e-Commerce Directive (arts 12–14 of Directive 2000/31/EC) which stipulates that there be a limited liability imposed onto service providers for ***content*** posted by a third party as previously mentioned (Wolfgang Schulz "Regulating Intermediaries to Protect Privacy ***Online*** — The Case of the German NetzDG" (13 August 2018) <*ssrn.com/abstract=3216572* >). This is because NetzDG does not require social media platforms to search proactively for unlawful ***content*** (Bundesministerium der Justiz und für Verbraucherschutz "FAQ: Act to Improve Enforcement of the Law in Social Networks, 2017" (2021) <*www.BMJV.de/SharedDocs/FAQ/EN/NetzDG/NetzDG.html* 107>). This means that the onus is on the social media platforms to delete or block unlawful ***content*** once flagged or once they are notified. This makes for a completely reasonable request to safeguard the social media environments seeing as it is virtually impossible, time consuming and labour intensive (though artificial intelligence and machine learning could be employed here) to introduce this method of reporting. Regulating ***online*** hate speech is at a point of inflexion. As observed, there are ways to regulate ***online*** hate speech. However, the opinion is that we ought to employ a holistic approach ***by way*** of involving every aspect of society, the law, social media platforms and its users. Legal jurist Wolfgang Schulz of the University of Hamburg asserts that the German legal system does not have the capacity to regulate ***online*** harms in particular ***online*** hate speech (Schulz, above). As a result, NetzDG is looking at amending some provisions that have come under criticism and proven to be an erosion of civil liberties (Amélie Heldt "Germany is amending its ***online*** speech act NetzDG ... but not only that" (6 April 2020) Internet Policy Review, Journal on Internet Regulation). Censorship in particular became a major erosion of people's fundamental human right of free speech and the freedom of expression.□

End of Document