

Inquiry into Social Media and Online Safety

Submission

12 January 2022



Table of Contents

Submission to the Inquiry into Social Media and Online Safety	3
Introduction	3
Executive summary	4
Recommendations	4
Defining social media and online platforms	5
Social media usage trends	6
Understanding the design of social media platforms	7
Impacts of social media on individuals	9
Mental health	9
Online bullying and trolling	11
Technology-facilitated abuse	12
Impacts of social media on society	12
Increasing polarisation	13
Driving extremism	15
Spreading misinformation and disinformation and leaving our society vulnerable	16
Conclusion	20

Committee Secretary
Select Committee on Social Media and Online Safety
PO Box 6021
Parliament House
Canberra ACT 2600

12 January 2022

Submission to the Inquiry into Social Media and Online Safety

Introduction

The Centre for Digital Wellbeing (CDW) welcomes the opportunity to make a submission to the inquiry into social media and online safety.

CDW is a research centre focusing on social media's impact on mental health and wellbeing, safety, and social cohesion in the Australian community. By bringing together a network of health, mental health, digital technology and policy experts, the Centre seeks to increase awareness within Australia's policy domain on the impact of social media on Australian society. CDW simultaneously facilitates critical discussions on social media's impact on social cohesion and mental health to inform users and policy makers in Australian society.

For further information on any of the points raised in our submission, please contact CDW on

[REDACTED]

Executive summary

Social media has fundamentally shifted the way Australians maintain connections, consume content and share information. Australians are increasingly conducting more of their lives online. This trend, already well established prior to the COVID-19 pandemic, has accelerated over the past two years. Our shift online has transpired with limited consideration of the ramifications of doing so and with limited oversight of the tech companies running the online platforms on which we now conduct so much of our lives. There are benefits to our increasing connectedness online, but there are also costs. Until recently, limited attention has been paid to the costs and harms of social media – both to the individual and to society. Although there is growing public awareness and concern about the negative effects of social media, they remain contested and poorly understood.

This inquiry is a welcome and important step in identifying the harms associated with social media and other online platforms and establishing a world-leading regulatory framework to mitigate these harms. A regulatory and policy response will need to consider and mitigate the impact of social media at an individual level as well as at a societal level. At an individual level, social media can lead to addiction, depression and increased suicide ideation. At a societal level, social media usage is associated with increased polarisation and reduced social cohesion. Regulatory responses will also need to address issues that arise from both the algorithms that platforms are built on, as well as the design features that are embedded to enhance platform engagement (such as likes and filters). While algorithms dictate the display of content and are the engine rooms of profitability for social media companies, design features can significantly increase the level of harm caused to individual users.

These harms are not exclusive to Australia. They present a global challenge. Reflecting this, there has been increasing public scrutiny on the influence, power and practices of big tech companies around the world in recent years. The power of these companies and the extent to which social media has become ingrained in people's lives can give rise to the perception that the trend is irreversible and there is little that can be done to mitigate social media's negative impacts. This is false. Governments and civil society have a role to play in responding to the challenges presented by social media and other online platforms. They have a duty to do so. Recognising this, a range of regulatory approaches have been proposed, considered and implemented in recent years. Just as governments have succeeded in reining in big tobacco, who knowingly understood the harms of their products on society, so too can they succeed in responding to the harms of social media.

Recommendations

- That the Government increase transparency and accountability of algorithms by mandating social media companies publish key features of their algorithms. This would provide greater information to users and policymakers around the impact of algorithms on their feeds. As part of transparency, measures should be considered such as requiring public reporting on the number of times misinformation is shared prior to removal. Other

regulation that enhances accountability should be considered as part of a regulatory package.

- That the Government consider regulating a duty of care to users as part of a licensing scheme for social media providers to operate a digital space. Over time, this would likely impact on design features and algorithmic functions that are harmful to all users, but particularly youth. A clear legislative requirement such as this would compel social media companies to act on evidence that any of its features are significantly damaging.
- That the Government address the need for education, guidance, tools and resources in relation to online safety and harm, specifically with parents and carers, children and teenagers, educators, migrant and other communities, the LGBTIQ+ community and other vulnerable groups.
- That the Government reform the Voluntary Australian Code for Disinformation and Misinformation administered by the Australian Communications and Media Authority (ACMA) to provide more adequate protection against disinformation and misinformation, including making the code mandatory, creating material penalties for breaches and better intervention strategies.
- That the Government establish an office for disinformation. The role of this office would be to monitor the spread of disinformation, inform the public of disinformation strategies and bring to light campaigns designed to erode trust in public institutions, destabilise social cohesion and distort public opinion.
- That the Government invest in digital literacy education, including in schools and through the settlement process, with a focus on disinformation and critical thinking in online information consumption.

It is likely that as new forms of technology (such as blockchain) emerge, there will be an increased capacity to utilise intermediary technology to change the asymmetry of power between users and social companies. Such technology may also enable the pricing of individual users' data. Policy responses will need to be adaptive to the development of such technologies.

An analysis of international regulatory frameworks can be found at Annexure 1.

Defining social media and online platforms

While various definitions of social media exist, the Australian Bureau of Statistics defines social media as 'web and mobile-based technologies which are used to turn communication into interactive dialogue among organisations, communities and individuals. These include blogs and

micro-blogs such as Twitter; content communities such as YouTube; and social networking sites, such as Facebook, Instagram and LinkedIn'.¹

'Online platform' is used to describe a range of services available on the Internet such as social networks, online marketplaces, search engines, app stores, communications services, and payment systems. An online platform can be defined as a digital service that facilitates interactions between two or more distinct but interdependent sets of users (organisations or individuals) who interact through the service via the Internet.²

Social media usage trends

In recent years, the number of Australians using social media has increased significantly. Australia follows global social media trends, with the most used platforms including Facebook, YouTube, Instagram, TikTok and Twitter.³ Social media has become an integral and, for many, an unavoidable part of our lives, particularly during the COVID-19 pandemic, with more than one in three Australians increasing their use of social media apps following the introduction of COVID-19 restrictions.⁴

- **Facebook:** With approximately 75 per cent of Australians being active Facebook users, Facebook is the most used platform in Australia.⁵ In December 2021, there were almost 20 million Australian users, two million more than the previous year and 7 million more than a decade ago.⁶ Despite this increase in overall usage, in recent years there has been a decline in Facebook usage among younger people worldwide, including in Australia (17.2 per cent to 16.9 per cent).⁷
- **Instagram:** Instagram use has increased, with 39% of Australians using the platform in 2018⁸ and 42% in 2021.⁹ The largest age group of users are aged 25-34 (approximately a third of users), with women using the platform more than men.¹⁰ Compared to Facebook,

¹ Australian Bureau of Statistics, 'Summary of IT Use and Innovation in Australian Business, 2014-15', (2016) <<https://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/8166.0Main%20Features52%2014-15?opendocument&tabname=Summary&prodno=8166.0&issue=2014-15&num=&view>>.

² OECD, 'What is an "online platform"?' in *An Introduction to Online Platforms and Their Role in the Digital Transformation*, (OECD Publishing, Paris, 2019).

³ Simon Kemp, 'Digital 2021: Australia', (9 February 2021) *DataReportal* <<https://datareportal.com/reports/digital-2021-australia>>.

⁴ Australian Communications and Media Authority, 'COVID restrictions helped increase digital communication use for older Australians', (22 April 2021) <<https://www.acma.gov.au/articles/2021-04/covid-restrictions-helped-increase-digital-communication-use-older-australians>>.

⁵ App Annie, 'The state of mobile 2021', (2021) <<https://www.appannie.com/en/go/state-of-mobile-2021/>>.

⁶ Napoleon Cat, 'Facebook users in Australia: December 2021', (2022) <<https://napoleoncat.com/stats/facebook-users-in-australia/2021/12/>>.

⁷ Andrew Hutchinson, 'Facebook's Daily Active Usage Has Stalled in the US - A Sign of Concern for The Social Network?', (31 January 2021) *Social Media Today* <<https://www.socialmediatoday.com/news/facebook-daily-active-usage-has-stalled-in-the-us-a-sign-of-concern-for/594253/>>.

⁸ Napoleon Cat, 'Instagram users in Australia: September 2018', (2018) <<https://napoleoncat.com/stats/instagram-users-in-australia/2018/09/>>.

⁹ Ibid.

¹⁰ Ibid.

where users are relatively evenly spread amongst gender, Instagram has a slightly younger and more female skew.¹¹ While Instagram attracts a younger population, its user group of 18-24 year olds has declined by 3% points in the last three years.

- **TikTok:** In almost every country, TikTok outperforms other social media apps in hours per user.¹² In 2019, average Australian TikTok users spent nearly 8 hours per month on the app. In 2020, this had increased to 17 hours per month.¹³ TikTok had almost 2.5 million Australian users in 2021, and is expected to keep growing exponentially.¹⁴ TikTok is most popular amongst young Australians, and almost 60% of the users are women and girls.¹⁵
- **YouTube:** YouTube has nearly 17 million unique visitors from Australia every month.¹⁶ Most YouTube users fall in the group of 15-35 with a male-female ratio of 11:9. YouTube is considered the second largest search engine after Google and the second largest social media platform.¹⁷
- **Twitter:** 30% of internet users aged 16-64 in Australia use Twitter. However, as almost half of the created accounts on the platform have never posted, Twitter reports that the potential audience for advertising is 2.9 million people in Australia (nearly 12% of the population). Twitter remains the most gender imbalanced platforms, where nearly 70% of the audience defines as male, and just 30% as female.¹⁸

Understanding the design of social media platforms

Many Australians see social media as a tool that helps them stay connected with friends and family, access information and content and browse products online. While this is true, it is not the complete picture. Social media users are not just individuals who are using a service, they – or more specifically, their attention – are a product that is being sold. Social media companies are commercial entities reliant on advertising revenue. For example, in 2020, nearly all of Facebook's

¹¹ L&A Social and Pureprofile, 'Social Media Census 2020', (2020) <<https://landasocial.com/wp-content/uploads/2020/09/SocialMediaCensus.pdf>>.

¹² Marissa Dellatto, 'TikTok Hits 1 Billion Monthly Active Users', (27 September 2021) *Forbes* <<https://www.forbes.com/sites/marissadellatto/2021/09/27/tiktok-hits-1-billion-monthly-active-users/?sh=50f2972844b6>>.

¹³ SMPPerth, 'Facts & Figures // TikTok statistics for 2021', (3 August 2021) <<https://www.smp Perth.com/resources/tiktok/tiktok-statistics/>>.

¹⁴ Roy Morgan, 'Nearly 2.5 million Australians using TikTok – up over 850,000 (+52.4%) during first half of 2020', <<http://www.roymorgan.com/findings/8538-launch-of-tiktok-in-australia-june-2020-202010120023>>.

¹⁵ Ibid.

¹⁶ David Correll, 'Social Media Statistics Australia – December 2021', (1 January 2022) *Social Media News* <<https://www.socialmedianews.com.au/social-media-statistics-australia-december-2021/>>.

¹⁷ GMI Blogger, 'YouTube user statistics 2022', (3 January 2022) *Global Media Insight* <<https://www.globalmediainsight.com/blog/youtube-users-statistics/>>.

¹⁸ Simon Kemp, 'Digital 2021: Australia', (9 February 2021), *DataReportal* <<https://datareportal.com/reports/digital-2021-australia>>

\$86 billion revenue came from advertising on Facebook and Instagram.¹⁹ Companies are incentivised to maximise a user's time spent on the platform and their level of engagement by maintaining their attention for as long as possible. This allows the companies to show more advertisements and other paid content to users and maximise their profits. Further, by tracking everything that a user does online, social media and other online platforms build up detailed data profiles of every user, enabling them to display more targeted content in future. This deep profiling of individuals enables further monetisation.

Early social media platforms sorted content chronologically. However, to manage growing volumes of content, and maintain and increase engagement, platforms began using increasingly sophisticated algorithms and design tools to determine what content to show users, in what order and how. The introduction of algorithm-based feeds on social media platforms has resulted in chronologically arranged posts substituted for tailored content that the algorithm deems to be the most engaging for each individual user, in turn increasing engagement.²⁰ The algorithms used by big tech companies are valuable and closely protected commercial secrets. There is no oversight of how they are developed and used. However, we do know that they take into account factors such as who users are friends with, who they interact with most (e.g. comments and likes), the content of those interactions, what pages they like, what content they read or watch or save to read or watch later, how popular content is among other users, groups they are in, which advertisers have paid to target them and what other similar users like.²¹

In addition, design features of social media platforms such as push notifications, refreshing, infinite scrolling, likes, comments and shares contribute to the creation of feedback loops aimed at capturing our attention for longer.²² Social media platforms are designed to be addictive, as users will then spend more time on them.²³ Related to the addictive nature of social media, fear of missing out (FOMO) is a psychological phenomenon that plays on people's anxiety and fear on not being part of the group. It motivates people to get back online and increase their engagement.²⁴ The more engagement certain posts receive online, the bigger the fear of missing out grows. This leads

¹⁹ Keach Hagey and Jeff Horwitz, 'Facebook tried to make its platform a healthier place. It got angrier instead', (15 September 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215>>.

²⁰ See, eg, Ella Kipling, 'What is 'chronological feed' on Instagram and why did it leave?', (9 December 2021) *HITC* <<https://www.hitc.com/en-gb/2021/12/09/what-is-chronological-feed-on-instagram-and-why-did-it-leave/>>.

²¹ See, eg, Keach Hagey and Jeff Horwitz, 'Facebook tried to make its platform a healthier place. It got angrier instead', (15 September 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215>>; Maria Alessandra Golino, 'Algorithms in Social Media Platforms', (24 April 2021) *Institute for Internet & Just Society* <<https://www.internetjustsociety.org/algorithms-in-social-media-platforms>>; Adam Mosseri, 'Shedding more light on how Instagram works', (8 June 2021) *Instagram* <<https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>>.

²² *The Social Dilemma* (Netflix, 2020).

²³ See, eg, *The Social Dilemma* (Netflix, 2020); Hannah Schwär, 'How Instagram and Facebook are intentionally designed to mimic addictive painkillers', (12 August 2021) *Business Insider* <<https://www.businessinsider.com/facebook-has-been-deliberately-designed-to-mimic-addictive-painkillers-2018-12?r=AU&IR=T>>.

²⁴ King University Online, 'Scared to stay in: the psychology of FOMO', (4 September 2019) <<https://online.king.edu/news/psychology-of-fomo/>>.

to an obsessiveness about keeping-up-to-date, which can in turn contribute to increased anxiety and the outrage cycle.

The adaptive nature of social media platforms due to their optimisation algorithms, adjusting users' feeds based on their preferences and behaviours, makes them more engaging and interesting to users, and hence more addictive. High levels of social media use are increasingly compared to dopamine-induced addictions, as design features of the platforms such as notifications and likes induce dopamine hits, causing the user to feel pleasure which they continue to seek through ongoing use of the platform. This is similar to the reaction seen with gambling and recreational drugs.²⁵

Impacts of social media on individuals

Mental health

Social media can aid in increasing connectedness and social support, particularly in our current era of public health restrictions.²⁶ However, as noted, it can be incredibly addictive. We do not yet have a full understanding of the risks of high levels of prolonged social media usage, especially for children and adolescents. While it is not conclusive, the emerging research is concerning, particularly with respect to young girls.²⁷ Today's youth are the first generation to grow up with social media. The evidence suggests that we are undertaking a mass experiment on them that may have serious consequences for their mental health.

In September 2021, the Wall Street Journal published a series of articles, commonly referred to as the 'Facebook Files' based on damning internal Facebook documents, released by whistle-blower, Frances Haugen. The series included reports on internal studies, demonstrating that Facebook was aware that Instagram has a negative impact on teenage users, prioritises profit over public safety, and that its design features amplify hate, political unrest and misinformation. The files outline internal research comprised of findings from focus groups, online surveys and diary studies in 2019 and 2020, which confirmed serious issues with Instagram including that it has a significant impact on the mental health of young users, particularly teenage girls.

Facebook's internal research found that teenagers blame Instagram for increases in anxiety and depression rates. Three per cent of teenage girls said that when they felt bad about their bodies, Instagram made them feel worse, and among users who reported suicidal thoughts, 13 per cent in

²⁵ Addiction Centre, 'Social Media Addiction', (2021) <<https://www.addictioncenter.com/drugs/social-media-addiction/>>; Jamie Waters, 'Constant craving: how digital media turned us all into dopamine addicts', (22 August 2021) <<https://www.theguardian.com/global/2021/aug/22/how-digital-media-turned-us-all-into-dopamine-addicts-and-what-we-can-do-to-break-the-cycle>>.

²⁶ Elizabeth M Seabrook, Margaret L Kern and Nikki S Rickard, 'Social Networking Sites, Depression, and Anxiety: A systematic Review', (2016) 3(4) *JMIR Mental Health* 50.

²⁷ Jonathan Haidt and Jean Twenge, *Is there an increase in adolescent mood disorders, self-harm, and suicide since 2010 in the USA and UK? A review*, (Unpublished manuscript, New York University 2021)

the UK and 6 per cent in the US linked the thoughts to Instagram. Researchers found that those struggling with the platform's psychological effects were not logging off from the platform. While teenagers frequently reported not liking the amount of time spent on the platform and wanting to spend less time on Instagram, they lacked the self-control to do so and felt they had to be present on the platform. The research noted that teenagers often feel addicted, and despite knowing that it is bad for their mental health, they feel unable to stop.²⁸ Facebook has taken minimal efforts to address these issues and has consistently downplayed the app's negative consequences on teenagers.²⁹

The evidence outlined in the Facebook Files reflected pre-existing concerns and emerging research about social media's impact on the mental health of adolescents. Rates of teen depression, anxiety, self-harm and suicide among teenagers have substantially and significantly increased since the early 2010s, correlating with increased use of social media and smartphone ownership.³⁰ The impact of social media is particularly profound for teenage girls, with those who spend significant time on social media being two to three times more likely than girls who use it less frequently to feel depressed.³¹

Children and teenagers are more vulnerable to the negative impacts of social media use, given the importance of social connections in those formative years. When all their friends are engaging with each other online, children and teenagers must effectively choose between using social media or social isolation. If they miss out on online conversations, trends, games or challenges, they risk social exclusion. Having been drawn onto the platforms, children and teenagers must then create and engage with content, fuelling the cycle of addiction. Further, the need to perform for their audience creates pressure and a desire for social approval, while also contributing to anxiety, as users wait anxiously after posting for likes and positive comments. These performative aspects of social media create inherent risks. Negative online interactions, including critical comments or lack of 'likes', may contribute to perceptions of exclusion and disconnection. Further, unfavourable social comparison to other users through exposure to highly curated 'perfect' lives can adversely affect the self-esteem of children and teenagers.

Parents and carers are in an unenviable situation when it comes to regulating their children's social media use. If they do not allow their children to use social media, their children may be excluded or

²⁸ Georgia Wells, Jeff Horwitz and Deepa Seetharaman, 'Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show', (14 September 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739>>.

²⁹ Damien Gayle, 'Facebook aware of Instagram's harmful effect on teenage girls, leak reveals', (15 September 2021) *The Guardian* <<https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals>>; Georgia Wells, Jeff Horwitz and Deepa Seetharaman, 'Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show', (14 September 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739>>.

³⁰ Jonathan Haidt, 'The Dangerous Experiment on Teen Girls', (21 November 2021) *The Atlantic*, <<https://www.theatlantic.com/ideas/archive/2021/11/facebook-dangerous-experiment-teen-girls/620767/>>.

³¹ Ibid.

socially isolated. However, allowing social media use may negatively affect their child's mental health. Parents, carers and educators have no assistance to navigate their concerns. Australia needs a comprehensive national policy to address social media's harms and promote the health and wellbeing of children and teenagers.

Online bullying and trolling

Social media platforms' features and tools facilitate and perpetuate online bullying and trolling, with serious consequences for victims. While online bullying shares similar characteristics with face-to-face bullying (offline bullying), including hostile or aggressive acts and repeated negative behaviour, the ability for abusers to remain anonymous exacerbates the issue. Anonymity shields perpetrators from the possible consequences of their behaviour that they would have experienced in 'offline bullying', and may enlarge the potential pool of perpetrators. Perpetrators do not witness firsthand the impacts of their bullying on the victim, which reduces opportunities for remorse and empathy.³² People are more likely to continue abusive behaviour when they do not see the victim's immediate reaction.

More so, the perpetrator's ability to save, re-post and re-share abusive content an unlimited number of times may trap the victim in a perpetual loop of trauma. Constant access to the Internet and notifications makes it difficult to disconnect from abuse. The permanence of abusive content posted online coupled with the large audience who may view the content exacerbate challenges for victims of bullying. The consequences of online bullying are felt particularly by women, with 38 per cent of women directly experiencing abuse online, rising to 45 percent among teenagers and young adults.³³ The rates of abuse are significantly worse for marginalised individuals and groups, such as LGBTIQ+ and women of colour.³⁴

The tools created by platforms to address cyberbullying are largely designed for English speaking users, to address abuse in English. Mechanisms to address abuse in other languages have been either poorly developed or not developed at all.³⁵ While Facebook is used in more than 110 languages, its community standards – which ban users from posting offensive material – have been translated only into 41 languages. Accordingly, more than 60 per cent of Facebook users do

³² Aashish Srivastava, Roger Gamble and Janice Yean Mei Boey, 'Cyberbullying in Australia: Clarifying the problem, considering the solutions' (2013) 21(1) *International Journal of Children's Rights* 25.

³³ Web Foundation, 'Facebook, Google, TikTok and Twitter make unprecedented commitments to tackle the abuse of women on their platforms', (1 July 2021) <<https://webfoundation.org/2021/07/generation-equality-commitments/>>.

³⁴ Ibid.

³⁵ Billy Perrigo, 'Facebook says it's removing more hate speech than ever before. But there's a catch', (27 November 2019) *Time* <<https://time.com/5739688/facebook-hate-speech-languages/>>; Daniel Howley, 'Facebook exec on moderating hate speech outside the US: 'Language is a challenge'', (7 October 2021) *Yahoo! Finance* <<https://finance.yahoo.com/news/facebook-exec-on-moderating-hate-speech-outside-of-us-language-a-challenge-160356598.html>>.

not have accessible tools to address and alleviate cyberbullying.³⁶ This issue is pervasive even in English-speaking countries such as Australia, which has an increasingly diverse population.

Technology-facilitated abuse

Perpetrators of family and domestic violence have increased capacity to perpetuate abuse through technology and social media, with technology-facilitated abuse increasingly evident in domestic and family violence situations. A 2015 survey of domestic and family violence frontline workers found that 98 per cent of clients had experienced abuse through technology.³⁷ By controlling access to and monitoring social media accounts, and using the platforms' tools to stalk and harass, perpetrators gained additional control over victims' lives. This abuse extends to image-based abuse, the non-consensual sharing of images to coerce, threaten, objectify, harass or abuse the victim.³⁸ A survey of over 4,000 participants found that around one in five Australians have experienced image-based abuse, with men and women equally likely to report being a victim.³⁹

Due to limited knowledge regarding the impacts of social media, adolescents and children are also subject to image-based harassment. Examples include minors coerced into sending sexualised images by predators and then blackmailed for more explicit material. The Australian Federal Police (AFP) recently reported that paedophiles have been exploiting increased unattended screen time by children during COVID-19 lockdowns, with child abuse material shared online doubling since the beginning of the pandemic.⁴⁰

Impacts of social media on society

Social media has become integral to everyday life and society. Social media and online platforms have integrated themselves into news, work, homelife, politics, and continue to evolve at an ever-increasing speed. Social media has made positive contributions to society through fostering and building social networks, increasing access and communication including across borders, and providing public platforms to those traditionally excluded from public debates. However, social media platforms are also doing immense damage to society by increasing polarisation, facilitating the proliferation of misinformation and disinformation and enabling trolls and bots to manipulate

³⁶ Maggie Fick and Paresh Dave, 'Facebook's flood of languages leave it struggling to monitor content', (23 April 2019) *Reuters* <<https://www.reuters.com/article/us-facebook-languages-insight-idUSKCN1RZ0DW>>.

³⁷ eSafety Commissioner, 'Protecting Voices at risk online', (2020) <https://www.esafety.gov.au/sites/default/files/2020-08/Protecting%20voices%20at%20risk%20online_0.pdf>.

³⁸ Harmony Alliance, 'National Consultation Report February 2021, (2021) <<http://harmonyalliance.org.au/wp-content/uploads/2021/02/HA-Membership-Forum-Report.pdf>>.

³⁹ Nicola Henry, Anastasia Powell and Asher Leigh Gevaux Flynn, 'Not Just 'Revenge Pornography': Australians' Experiences of Image-Based Abuse: A Summary Report' (2017) RMIT University 9.

⁴⁰ N Bita 'Sick messages of sex predators targeting Aussies kids online', (2021) *The Daily Telegraph* <<https://www.dailytelegraph.com.au/technology/sick-messages-of-sex-predators-targeting-aussie-kids-online/news-story/71312a16dc36a672b47dc11950617bf2>>.

public opinion. Overall, the picture that is emerging is that social media undermines trust in institutions and threatens Australia's social cohesion.

Increasing polarisation

Social media has disrupted our traditional sources of news and information. Over 12.7 million Australians (60.8 per cent) now cite the internet as their main source of news, including nearly 7.9 million Australians (37.7 per cent) who nominate social media as their primary source.⁴¹ While previously centralised news services vetted and curated information and facts, and people were exposed to a range of news sources, social media has led to the decentralisation of information sources through the rise of peer-to-peer content and the proliferation of information through digital channels.⁴² While offering some benefits, decentralisation of information has led to an erosion of trust in information and a reduced trust in vetted fact-based services.⁴³

Social media works to amplify status-seeking behaviours and to promote identity projection. In doing so, social media motivates people to engage with and share content. Humans have a strong preference towards information that confirms existing biases and beliefs.⁴⁴ Confirmation bias leads us to not only preference, but to seek out information that confirms our existing assumptions and ideas. Contradictory or differing viewpoints tend to reinforce our beliefs, rather than moderate them. As people are more likely to engage with content they agree with, the algorithms underpinning social media platforms will continue to show them more of that content.

It is generally understood that exposure to conflicting viewpoints leads to high levels of political tolerance towards others and awareness of the legitimacy of oppositional viewpoints.⁴⁵ Accordingly, the algorithmic design of social media poses significant challenges for social cohesion as it means that people are exposed to less diverse perspectives. When social media is used as a primary source for information and newsgathering, there is a significant risk that people are exposed mainly to ideologically compatible news and information, creating an echo chamber and contributing to further polarisation in society. These echo chambers can make us insular, less curious and less open-minded towards different ideas, which can fuel animosity towards 'the other'.

⁴¹ Roy Morgan, 'It's official: Internet is Australia's main source of news; TV remains most trusted', (21 August 2020) <<http://www.roymorgan.com/findings/8492-main-sources-news-trust-june-2020-202008170619>>.

⁴² Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report* (June 2019) <<https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf>> 279.

⁴³ Caroline Fisher, Sora Park, Jee Young Lee, Glen Fuller and Yoonmo Sang (News & Media Research Centre, University of Canberra), *Digital News Report: Australia 2019*, (17 June 2019) <<https://apo.org.au/node/240786>>.

⁴⁴ Raymond S Nickerson, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises', (1998) 2(2) *Review of General Psychology* 175.

⁴⁵ Diana C. Mutz, 'Cross-cutting Social Networks: Testing Democratic Theory in Practice', (2002) 96(1) *American Political Science Review* 111.

Further, the gradual deterioration of behaviour and attitudes online and offline has created a vicious cycle of anger and outrage.⁴⁶

Social media encourages and optimises content that generates clicks and is shared.⁴⁷ As emotive content that plays to people's underlying concerns and fears is more likely to be viewed and shared, social media preferences that content.⁴⁸ There is ample evidence that content that feeds negative emotions, such as anger, fear or uncertainty is more likely to be viewed and passed on.⁴⁹ Information that is highly sensational or likely fake but confirms our views will often be given preference over content that is factual but counter to our beliefs. As a result, there is an incentive for publishers to go viral through the creation of more divisive, negative and sensationalised content.

As an example, Facebook dramatically changed its News Feed algorithm in 2018 to boost 'meaningful social interactions'. Its stated objective was to strengthen bonds between users and improve their wellbeing, through encouraging more interaction with friends and family, and less time interacting with professionally produced content. However, the changes had a contradictory effect, and the algorithm's heavy weighting of reshared material that made people angry increased misinformation, outrage, violent content and sensationalism on the platform.⁵⁰ Internal Facebook documents reveal that many political parties reported that the change to the algorithms shifted the way they approach their target audience, with them creating more negative content. One political party in Poland has reported shifting from 50/50 negative/positive posts to almost 80% negative.⁵¹ Although Facebook has now identified this as an issue, it has not acted in a coordinated manner to restrict the features of its platform that amplify harmful posts, as this would impede its usage and growth. Instead, it has chosen to adopt ad hoc responses to crises such as the January 6 Capitol riot.⁵²

⁴⁶ Charles Arthur, *Social Warming: The Dangerous and Polarising Effects of Social Media*, (Oneworld publications, 2021).

⁴⁷ Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an interdisciplinary framework for research and policy making* (Council of Europe, 27 September 2017) <<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>> 13.

⁴⁸ Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report* (June 2019) <<https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf>> 342-3.

⁴⁹ See, eg, Aaron Mak, 'The Most Important Answer from the Facebook Whistleblower' (5 October 2021) *Slate* <<https://slate.com/technology/2021/10/facebook-whistleblower-hearing-haugen-misinformation-anger.html>>.

⁵⁰ Keach Hagey and Jeff Horwitz, 'Facebook tries to make its platform a healthier place. It got angrier instead.', (15 September 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215>>; Christiano Lima, 'A whistleblower's power: Key takeaways from the Facebook Papers', (26 October 2021) *The Washington Post* <<https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/>>.

⁵¹ Keach Hagey and Jeff Horwitz, 'Facebook tries to make its platform a healthier place. It got angrier instead.', (15 September 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215>>

⁵² Jeff Horwitz and Justin Scheck, 'Facebook increasingly suppresses political movement it deems dangerous', (22 October 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-suppresses-political-movements-patriot-party-11634937358>>.

Driving extremism

Social media not only facilitates increased polarisation in society by entrenching us in our own echo chambers, it also pushes users towards more extreme content. There is increasing evidence to suggest that social media has primed us to expect more emotive and sensational news content.⁵³ In part, this is a function of a coevolving system. As we have an in-built bias to content that affirms our views, identity and sense of belonging, we preference content that is evocative and spend more time engaging with it.⁵⁴ In doing so, users are training an algorithm and an information ecosystem that rewards sensationalism. Simultaneously we are also training ourselves to normalise, accept and expect content and information that is presented as sensationalist, affirming or outrageous.⁵⁵

An example of how social media platforms direct us to ever extreme content is evidenced by various pieces of research. In one internal Facebook study, a researcher created a Facebook account for a fictional 41-year-old conservative mother with an interest in 'young children, parenting, Christianity, Civics and Community'. After this fictional account liked memes and joined conservative groups on the first day, Facebook began recommending almost exclusively right-wing content on the second day. By the fifth day, it was recommending QAnon content and right-wing conspiracy theories. Facebook's internal research found similar effects for a fictional liberal user.⁵⁶ In 2019, the New York Times documented this effect, reporting on the experience of Caleb Cain, who 'fell down the alt-right rabbit hole' when YouTube's algorithms began recommending him progressively more extreme far-right content.⁵⁷ Similarly, the New Zealand Royal Commission on Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019 noted how right-wing extremists have converged on social media platforms and how they play an 'instrumental role in spreading right-wing ideology across different online communities, including more mainstream ones'.⁵⁸ In particular, it highlighted the role that YouTube played in radicalising the Christchurch shooter and the platform's broader role in promoting far right content and radicalisation, although it also noted YouTube's amendments to its recommendation system make it less likely to recommend and access extreme content.⁵⁹

⁵³ Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an interdisciplinary framework for research and policy making*, (27 September 2017) Council of Europe <<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>> 13.

⁵⁴ Ibid.

⁵⁵ Adam Alter, *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked* (Penguin Group, 2018).

⁵⁶ Jeff Horwitz and Justin Scheck, 'Facebook increasingly suppresses political movement it deems dangerous', (22 October 2021) *The Wall Street Journal* <<https://www.wsj.com/articles/facebook-suppresses-political-movements-patriot-party-11634937358>>.

⁵⁷ Kevin Rose, 'The making of a YouTube Radical', (8 June 2019), *NY Times* <<https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html>>.

⁵⁸ Royal Commission, *Report of the Royal Commission of Inquiry into the terrorist attack on the Christchurch masjidain on 15 March 2019* (December 2020) <<https://christchurchattack.royalcommission.nz/the-report/download-report/download-the-report/>>.

⁵⁹ Ibid.

Spreading misinformation and disinformation and leaving our society vulnerable

The decentralisation of information through social media has led to the proliferation of content that aims to mislead, deceive and/or polarise. Social media's preference for angry, emotional and sensational content is inextricably tied with its key role in spreading misinformation, disinformation and extremist views online.

Misinformation is a term used to describe content or information that is false but was created or shared without the intent to cause harm. Misinformation is false or out-of-context information that is presented as fact, and can include made-up news articles, false information shared on social media platforms, doctored images and videos, and scam advertisements.⁶⁰

Conversely, disinformation is the purposeful or deliberate creation and dissemination of false information with the intention to mislead or cause harm.⁶¹ Disinformation can take many forms. It can include false or fake news content or fake news sites, images or text that are altered or distorted, or videos or commentary that include elements of fact mixed with elements of falsehood or exaggeration. Disinformation can also include real material used within a context that presents a distorted view of reality, such as a clip of a speech that is given a new and false attribution of meaning. The amalgamation of false information with truth is a common tool used in disinformation campaigns and is highly effective as a tactic of influence.⁶²

Disinformation, or the amplification of misinformation, is a strategy employed by a growing number of state and non-state actors, in part as it is a low-cost high-yield means to spread narratives, distort opinion and undermine trust in the public institutions of the target state.⁶³ While disinformation campaigns are a longstanding tactic, social media networks and the growing connection between social media, infotainment channels and influencers has created new and unprecedented avenues to pursue disinformation strategies. While offering some benefits, the decentralisation of information has simultaneously led to an erosion of trust in information and a reduced trust in vetted fact-based services.⁶⁴ Disinformation on social media is difficult to

⁶⁰ Australian Communications and Media Authority, *Online Misinformation* (2021) <<https://www.acma.gov.au/online-misinformation>>; Meira Gebel, 'Misinformation vs. disinformation: What to know about each form of false information, and how to spot them online, (16 January 2021) *Business Insider* <<https://www.businessinsider.com/misinformation-vs-disinformation?r=AU&IR=T>>.

⁶¹ Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report* (June 2019) <<https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf>> 616-617.

⁶² See, eg, Australian Communications and Media Authority, *Misinformation and news quality on digital platforms in Australia: A position paper to guide code development* (June 2020) <<https://www.acma.gov.au/australian-voluntary-codes-practice-online-disinformation>> 40-41; Claire Wardle and Hossein Derakhshan, Council of Europe Report, *Information Disorder: Toward an interdisciplinary framework for research and policy making* (27 September 2017) <<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>> 17.

⁶³ Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report* (June 2019) <<https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf>> 353.

⁶⁴ Caroline Fisher, Sora Park, Jee Young Lee, Glen Fuller and Yoonmo Sang (News & Media Research Centre, University of Canberra), *Digital News Report: Australia 2019* (17 June 2019) <<https://apo.org.au/node/240786>>.

identify.⁶⁵ This decreases the capacity of governments to combat disinformation and increases the susceptibility of Australian citizens to disinformation campaigns. As content that is semi-factual becomes normalised, campaigns that intertwine fact with fiction are more effective and harder to combat. Our tendency to seek out information that confirms our existing beliefs and social media's amplification of particular content can be manipulated by malicious actors—directly and indirectly—to destabilise, erode trust, confuse messaging or flood public opinion.

For example, machine-driven communications tools can write news articles and reports that are very hard to distinguish from human-generated content. With big data and machine learning availability, these tools can simulate human accounts (known as 'bots'). These accounts appear to users as real people; making comments, liking posts, sharing news and interacting with other accounts as though they are human. Large numbers of bots operating in a coordinated fashion are known as a bot farm and can be used to boost engagement on posts. Bot farms are significant because the endorsement of a post by a large number of users, regardless of their credibility, boosts persuasiveness. Bot farms can generate a high volume of likes and follows for selected content, the mass generation of engagement on posts creating the perception of mass support,⁶⁶ while simultaneously boosting the content in other user's feeds due to its popularity. Machine-driven communication tools also have the advantage of time on their side because they operate 24/7, 365 days per year, constantly analysing data and being the first to generate content, giving more credibility to false information and fuelling the cycle of outrage.⁶⁷

In contrast to bots, troll farms are groups who work together to post content—often inflammatory, divisive and false—to social media platforms.⁶⁸ Trolls may use bots to amplify some of their messages.⁶⁹ Troll farms are increasingly being used as a tool of foreign interference by state and non-state actors to sow seeds of division in society and encourage discontent. A Facebook internal report written in October 2019 and obtained by MIT Technology Review revealed that troll farms operating out of Eastern Europe were running Facebook's most popular Christian American and Black American pages, its second largest Native American page and the fifth largest women's page. The pages reached 140 million US users per month and 360 million global users per week. Between 85 and 95 per cent of the people being reached by each individual page had never followed the pages.⁷⁰ The report noted that while troll farms primarily affect the US, Australia is also a popular

⁶⁵ Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report* (June 2019) <<https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf>> 353.

⁶⁶ Matt Chessen, 'Understanding the psychology behind computational propaganda' in Shawn Powers and Markos Kounalakis *Can public diplomacy survive the internet? Bots, echo chambers and disinformation*, (May 2017) 19 <<https://www.state.gov/wp-content/uploads/2019/05/2017-ACPD-Internet.pdf>>.

⁶⁷ Ibid.

⁶⁸ Karen Hao, 'Troll farms reached 140 million Americans a month on Facebook before 2020 election, internal report shows', (16 September 2021) *MIT Technology Review* <<https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election/>>.

⁶⁹ Donara Barojan, 'Understanding bots, botnets and trolls', (13 November 2018) *International Journalists' Network* <<https://ijnet.org/en/story/understanding-bots-botnets-and-trolls>>.

⁷⁰ Karen Hao, 'Troll farms reached 140 million Americans a month on Facebook before 2020 election, internal report shows', (16 September 2021) *MIT Technology Review* <<https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election/>>.

target.⁷¹ The incredible reach of pages being run by malicious actors, including to people who never followed the page in the first instance, poses great risks to social cohesion as misinformation, disinformation and divisive, hateful and inflammatory content can be easily projected on to Australian users.

Indeed, the damage caused by misinformation and disinformation has been evidenced throughout the COVID-19 pandemic, both in Australia and globally. Trust in authorities has been disrupted and undermined through the spread of anti-vaccination and conspiracy-infused misinformation across social media.⁷² There is no cohesive ideology or coherent message underpinning anti-vaccination and anti-lockdown groups; instead, they bring together threads of different conspiracy narratives.⁷³ In so doing, they broaden their appeal and draws more individuals in as they seek an explanation for their own individual grievances. As in much of the rest of the world, this movement has led to threats of violence against public figures and heated, often violent, protests, which are organised on social media platforms and have links to overseas conspiracy groups.⁷⁴ Further, conspiracy theorist groups, including those with overseas links, have targeted remote Indigenous communities with COVID-19 misinformation.⁷⁵ Marginalised culturally and linguistically diverse communities have also been targeted with tailored messaging designed to undermine trust in government and media messaging. This has contributed to anger, distrust, confusion, refusal to test and vaccine hesitancy among some communities.⁷⁶

This latter point highlights Australia's unique risk factors when it comes to the spread of misinformation and disinformation on social media and other online platforms. Our vulnerability is heightened by the size and diversity of our diaspora communities. There are currently more than 300 languages spoken in Australia today, with the vast majority spoken by migrant and refugee

⁷¹ Karen Hao, 'Troll farms reached 140 million Americans a month on Facebook before 2020 election, internal report shows', (16 September 2021) *MIT Technology Review* <<https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election/>>.

⁷² Elise Thomas, 'The threat of conspiratorial COVID-sceptic extremism', (16 November 2021) *Institute for Strategic Dialogue* <https://www.isdglobal.org/digital_dispatches/the-threat-of-conspiratorial-covid-sceptic-extremism/>.

⁷³ Ibid.

⁷⁴ Elise Thomas, 'The threat of conspiratorial COVID-sceptic extremism', (16 November 2021) *Institute for Strategic Dialogue* <https://www.isdglobal.org/digital_dispatches/the-threat-of-conspiratorial-covid-sceptic-extremism/>; Elise Thomas, 'Australia's fragmented, conspiracy-focused anti-lockdown movement', (14 September 2021) *Institute for Strategic Dialogue* <https://www.isdglobal.org/digital_dispatches/australias-fragmented-conspiracy-focused-anti-lockdown-movement/>; Christopher Knaus and Michael McGowan, 'Who's behind Australia's anti-lockdown protests? The German conspiracy group driving marches', (27 July 2021) *The Guardian* <<https://www.theguardian.com/australia-news/2021/jul/27/who-behind-australia-anti-covid-lockdown-protest-march-rallies-sydney-melbourne-far-right-and-german-conspiracy-groups-driving-protests>>.

⁷⁵ Michael McGowan, 'Old Parliament House fire protesters linked to anti-vaccine and conspiracy groups', (31 December 2021) *The Guardian* <<https://www.theguardian.com/australia-news/2021/dec/31/old-parliament-house-fire-protesters-linked-to-anti-vaccine-and-conspiracy-groups>>.

⁷⁶ Ahmed Yussuf, 'Anti-vax 'fearmongers' spreading misinformation are targeting Australia's diverse communities, leaders, experts warn', (6 September 2021) *ABC News* <<https://www.abc.net.au/news/2021-09-06/fears-misinformation-targeting-australias-diverse-communities-/100405706>>.

communities.⁷⁷ Members of these communities often depend on social media to maintain contact with loved ones overseas; however, their use of online platforms in languages other than English leaves them vulnerable to misinformation and disinformation, something that is very difficult for authorities to detect and respond to.

As Australia's diversity increases, in-language articles and content on Australian-based social media accounts and platforms has flourished. A feature of disinformation campaigns within an Australian context has been in-language content creation. These campaigns are highly effective as they remain hidden and are difficult to combat. In-language content often picks up elements of factual news and distorts or represents information to not only sway opinion, but to undermine confidence around conventional or mainstream news institutions. The development of in-language disinformation tactics is concerning, and significant consideration needs to be given to strategies to combat such campaigns, particularly in the lead up to the next federal election. It is notable that some campaigns appear to combine in-language content creation with efforts to influence through the broader information ecosystems, such as infotainment channels and influencers.

In the 2016 Census 5.6 per cent of respondents identified as having Chinese ancestry with 509,555 born in China.⁷⁸ Between the 2011 Census and the 2016 Census the number of Australian-born Chinese increased by nearly 60 per cent.⁷⁹ During this period, platforms such as WeChat and Weibo have grown in use and are now the main source of information for a significant number of Chinese born Australians. In the lead up to the 2019 election, an in-language disinformation campaign circulated in Chinese on WeChat and Facebook. The campaign warned that if Labor won the election, they would introduce an inheritance tax or 'death tax'. Memes and articles specific to this misinformation this were promoted through private chat groups and peer-to-peer conversations. While it is difficult to ascertain the origin of the campaign, it appeared to be highly organised and sustained. As WeChat enables groups of up to 500 people to post content, the information ecosystem for Chinese Australians is largely hidden from public view. However, content is still widely understood to be censored and monitored by Chinese state authorities and Chinese Australians have had posts or accounts deleted. Outside groups, users can post on Moments and official accounts promote information and interpretations of Australian news. While official posts did eventually emerge describing the 'death tax' as fake news, the promotion of content through groups and through Facebook was highly effective and spread rapidly.

While social media platforms have begun investing in combating misinformation and disinformation (although not nearly to the extent required), the majority of this is directed towards

⁷⁷ Australian Bureau of Statistics. Cultural Diversity in Australia: 2016 Census Data Summary. Census of Population and Housing: Reflecting Australia – Stories from the Census, 2016. Australian Bureau of Statistics, Canberra (2016) <<https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2071.0main+features302016>>.

⁷⁸ Department of Home Affairs and Australian Bureau of Statistics, *China-born Community Information Summary* (2018) <<https://www.homeaffairs.gov.au/mca/files/2016-cis-china.PDF>>.

⁷⁹ Ibid.

misinformation and disinformation occurring in English.⁸⁰ Reporting by the Washington Post has demonstrated how misinformation shared on social media platforms in languages other than English is left on those platforms for much longer than the same content is left on in English.⁸¹ 87 per cent of Facebook's misinformation spending is in English, but only about nine per cent of its users speak English.⁸² Consultations by the Migration Council Australia have revealed concerns among migrant and refugee communities about the spread of COVID-19 misinformation on social media in languages other than English and the challenges faced by governments in identifying and countering this information.⁸³

The Australian Government has recognised the need to combat disinformation. In December 2019, as part of its response to the ACCC's Digital Platforms Inquiry Report, the Government asked the digital industry to develop a voluntary code of conduct for disinformation and news quality.⁸⁴ The *Australian Code of Practice on Disinformation and Misinformation*, drafted by Digital Industry Group Inc (DiGi) was released in February 2021. The voluntary code commits a diverse set of technology companies, including Facebook, Twitter, and Google, to reducing the risk of online misinformation causing harm to Australians. The signatories committed to safeguards to protect Australians and must publicly report their efforts in response to disinformation and misinformation. The Australian Communications and Media Authority reports on the efficacy of the code. However, the code is hampered by its self-regulatory and opt-in approach.⁸⁵ To adequately address disinformation and misinformation, reforms to the code, creating material penalties and better intervention strategies are needed.

Conclusion

Social media has fundamentally impacted the way Australians engage, connect, and consume. Social media continues to become a more integral and unavoidable part of our lives. However, this shift online has occurred without considered attention to the significant online harms stemming from online platforms.

Social media positively contributes to individuals and Australian society through developing social networks, increasing communication and offering public platforms to all users. Simultaneously, for

⁸⁰ Stephanie Valencia, 'Misinformation online is bad in English. But it's far worse in Spanish', (28 October 2021) *The Washington Post* <<https://www.washingtonpost.com/outlook/2021/10/28/misinformation-spanish-facebook-social-media/>>.

⁸¹ Ibid.

⁸² Ibid.

⁸³ Report forthcoming.

⁸⁴ Australian Government, *Regulating in the Digital Age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry* (2019) <<https://treasury.gov.au/sites/default/files/2019-12/Government-Response-p2019-41708.pdf>>.

⁸⁵ DiGi, 'Disinformation code', (22 February 2021) <<https://digi.org.au/disinformation-code/>>; Asha Barbaschow, 'Facebook, Google, Microsoft, TikTok, and Twitter adopt Aussie misinformation code', (22 February 2021) *ZDNet* <<https://www.zdnet.com/article/facebook-google-microsoft-tiktok-and-twitter-adopt-aussie-misinformation-code/>>; Josh Taylor, 'What is the Australian government doing to crack down on big tech, and why?', (30 October 2021) *The Guardian* <<https://www.theguardian.com/australia-news/2021/oct/30/what-is-the-australian-government-doing-to-crack-down-on-big-tech-and-why>>.

individuals, social media is addictive, can have detrimental effects on mental health, especially for children and teenagers, and critically increases the capacity of online bullying and technology facilitated abuse. On a broader societal scale, social media is increasing polarisation, enabling the proliferation of misinformation and disinformation, and facilitating the manipulation of public opinion by trolls and bots. While social media and other online platforms can bring us together, they can also deepen divisions and threaten social cohesion.

Accordingly, targeted, holistic and effective regulation is needed to counteract the negative effects of social media on Australians and Australian society as a whole. The impact of social media on individuals and Australian culture and society is profound, and solutions must go beyond regulation and into social policy spheres.

Annexure 1

International Regulation of Social Media

December 2021



Table of Contents

Introduction	3
About us	3
Background	4
The impact of social media on mental health, safety, and social cohesion	4
The need for a regulatory response	6
Australian Regulatory Response	7
International Regulatory Responses	12
The United States	12
European Union	18
EU Member States	25
Norway	29
United Kingdom	29
New Zealand	32
Canada	33
China	36
Singapore	37
India	39
Conclusion	41

Introduction

The role of social media and the influence and power of big tech companies has increasingly come under public scrutiny. There is mounting evidence about the negative effects of social media on mental health and wellbeing, particularly among young people, as well as significant concern about the role of social media platforms in spreading disinformation and misinformation, undermining trust in institutions and threatening social cohesion.

In this context, there is a growing recognition of the need for effective regulation of social media platforms in order to ensure the development and use of this technology is human-centred. Global discussion have centred on how countries can protect their citizens from various online threats. At the time of writing, the Australian Government has recognised these challenges, establishing a parliamentary inquiry into social media and online safety, which will “examine the practices of these companies—and whether more needs to be done”.⁸⁶ This inquiry is a welcome and important step in establishing an Australian regulatory framework that draws on international best practice and experience.

The purpose of this review is to identify what approaches other countries have taken to regulate social media platforms and understand the context in which those regulations have been implemented. In doing so, this paper will identify experiences that Australian policymakers can learn from and draw upon in developing Australia’s regulatory framework. It will form the foundation of recommendations best suited to an Australian context.

About us

The Centre for Digital Wellbeing is a policy research centre focusing on social media’s impact on mental health and wellbeing, safety and social cohesion in the Australian community. The Centre seeks to collate research and increase awareness within Australia’s policy domain on the effects of social media and to provide advice to government on policy and regulatory responses, including international best practice. The Centre has an Advisory Council comprising a network of health, mental health, digital technology and social policy experts who inform the Centre’s work direction and policy development.

⁸⁶ Australian Government, Prime Minister of Australia, *Parliamentary committee to put big tech under the microscope* (1 December 2021) <https://www.pm.gov.au/media/parliamentary-committee-put-big-tech-under-microscope>

Background

Social media has fundamentally shifted the way Australians maintain connections, consume content, and share information. In recent years, the number of Australians using social media has increased significantly. As of March 2019, over 17 million Australians aged over 14 years used Facebook, representing an increase of nearly 4.2 million users since 2015.⁸⁷ Other platforms such as Instagram, TikTok, WeChat, YouTube, Pinterest, and Twitter also experienced significant growth in that period.⁸⁸ Social media platforms have seen further increases in usage during the COVID-19 pandemic, with more than one in three Australians increasing their use of social networking apps following the introduction of COVID-19 restrictions.⁸⁹ While there are many benefits of digitisation, there is increasing evidence that social media, gaming, and excessive screen time can severely impact our mental health and wellbeing, heighten the risk of online abuse and harassment, deepen societal divisions, and challenge social cohesion.⁹⁰

Despite social networking platforms providing age restrictions, more children are online than ever before. A recent U.S. study found that for children aged 10-12, 49% of parents report the use of social media in the first six months of 2021. For children aged 7-9, 32% of parents report their child using social media.⁹¹ In the same survey, parents reported finding it challenging to monitor children's behaviour online.

The impact of social media on mental health, safety, and social cohesion

Social media can aid increase connectedness and social support and build skills appropriate for the digital age. However, increased use of social media also carries risks, especially for children and adolescents. Social media can be incredibly addictive, with platforms intentionally designed to maximise the time users spend on them, for example through the use of algorithms to display content users will find engaging based on previous behaviour online. Like gambling and gaming, users can become captivated and obsessed, spending more time than anticipated on the platforms.

Research on the impact of social media on mental health and wellbeing is far from conclusive. However, growing evidence indicates that social media leads to social

⁸⁷ Roy Morgan, 'Facebook on top but Instagram and Pinterest growing fastest' (17 May 2019) <http://www.roymorgan.com/findings/7979-social-media-trends-march-2019-201905170731>.

⁸⁸ Ibid.

⁸⁹ Australian Communications and Media Authority, 'COVID restrictions helped increase digital communication use for older Australians' (22 April 2021) <https://www.acma.gov.au/articles/2021-04/covid-restrictions-helped-increase-digital-communication-use-older-australians>.

⁹⁰ John A. Naslund, Ameya Bondre, John Torous, & Kelly A. Aschbrenner, 'Social Media and Mental Health: Benefits, Risks and Opportunities for Research and Practice' (2020) 5 *Journal of Technology in Behavioral Science*, 245-257; Elly Robinson, 'Parental involvement in preventing and responding to cyberbullying' (2013) 92 *Family Matters*, 68-76; Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, & Brendan Nyhan 'Social Media, Political Polarization, and Political Disinformation: A review of the Scientific Literature' (2018) *SSRN*.

⁹¹ Mott Poll Report. 'Sharing too soon? Children and social media apps' (2018) 39 (4) *C.S. Mott Children's Hospital University of Michigan*.

isolation, stress, depression, and anxiety. Specifically, adolescents are more vulnerable to the consequences of social media use.⁹²

In recent years social media has also emerged as a mechanism to perpetrate abuse and harassment, including cyberbullying and the promotion of violence and self-harm, age-inappropriate content, and gender-based violence. Social media platforms have started responding to concerns raised by individuals and organisations, implementing measures such as artificial intelligence to identify and block fake accounts, and providing access to safety tools. However, these measures remain limited, and social media platforms' overall self-regulation is weak and ineffective.

Social media can also deepen divisions and challenge social cohesion, including through the spread of misinformation and disinformation and the creation of echo chambers. The platforms are built to encourage and optimise content that generates clicks and is shared.⁹³ It therefore preferences viral content that is more sensational and more emotive.⁹⁴ As disinformation often acts on emotive plays and sensational claims, the content created is more likely to go viral and be shared, hence making it more difficult to contain. Anti-vaccine misinformation during the pandemic is a recent example of how easily false content can spread and the damaging effects that can have.

Social media has become a place where incorrect information spreads quickly. It has changed our traditional consumption of news and information. Over 12.7 million Australians (60.8 per cent) now cite the Internet as their primary source of news, including nearly 7.9 million Australians (37.7 per cent) who nominate social media as their primary source.⁹⁵ Traditionally, centralised news has been perceived as vetted and curated information and facts. In contrast, the decentralisation of information on social media has led to an erosion of trust in information and institutions and reduces overall trust in vetted fact-based channels.⁹⁶

Further, with the use of algorithms, social media platforms expose their users to similar content they have previously engaged in. This creates echo chambers, where users are repeatedly exposed to similar views and opinions with minimal exposure to opposing views about current affairs. Echo chambers can make people insular, less curious, and less open-minded towards different ideas, which can fuel animosity towards 'the other'. The spread of disinformation, misinformation, and inflammatory content can have severe consequences for social cohesion.

Additionally, social media has opened new opportunities for foreign actors to undermine Australia's institutions. Social media provides an easy and accessible avenue for

⁹² Jenna Palermo Christofferson, 'How is Social Networking Sites Effecting Teen's Social and Emotional Development: A Systemic Review' (2016) *Social Work Master's Clinical Research Papers*.

⁹³ Claire Wardle and Hossein Derakhshan, Council of Europe Report, *Information Disorder: Toward an interdisciplinary framework for research and policy making* (2017) <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>

⁹⁴ Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report* (June 2019) 342-3

⁹⁵ Roy Morgan. 'It's official: Internet is Australia's main source of news; TV remains most trusted', (21 August 2020), <http://www.roymorgan.com/findings/8492-main-sources-news-trust-june-2020-202008170619>.

⁹⁶ Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report* (June 2019) Chapter 6.

malicious actors to interfere in the democratic process by communicating directly with citizens and spreading disinformation to erode social cohesion. To combat disinformation online, users must first learn to recognise false information online, highlighting the need for further education and awareness.

The need for a regulatory response

Governments around the world have in recent years begun considering what regulatory response is required to mitigate the negative effects of social media. At the same time as recognising the challenges posed by social media, it is also important to recognise its benefits including its capacity to facilitate connections and communications across borders and give voice to those who may have traditionally been excluded from public debate. Any regulation needs to ensure that these benefits are maintained.

This paper provides an overview of the different regulatory frameworks adopted internationally, with a view to informing Australian policymakers about current trends and debates and options for regulation in Australia.

Australian Regulatory Response

Australia's regulatory response to social media to date has primarily focused on cyberbullying, terrorist and extremist content, and the media marketplace. Less consideration has been given to impacts on mental health (specifically relating to children, teenagers, and vulnerable groups) and the platforms' addictive properties. Further, while there is growing awareness of social media's role in spreading misinformation and disinformation, particularly in the context of COVID-19, and the corresponding negative effects on social cohesion, there has been a limited policy response to date. Communications Minister Paul Fletcher recently suggested Australia's defamation laws needed to ensure social media companies faced the same rules as traditional media. Deputy Prime Minister Barnaby Joyce has stated it is "essential" the Government pushes tech companies to clamp down on misinformation, noting Australia should take the US's lead. Prime Minister Scott Morrison has echoed Joyce's sentiment, urging tech companies to take more responsibility for content published on their platforms.⁹⁷

Responding to social media's threat to social cohesion and impact on mental health should be a policy priority of government. To date, Australia has taken a piecemeal and reactive approach to the regulation of harmful online practices.⁹⁸ A better coordinated and proactive approach across sectors is needed to ensure social media platforms are appropriately and adequately regulated in Australia to minimise their negative effects.

Cyberbullying and Online Content

The *Enhancing Online Safety Act 2015* (Cth) establishes a two-tiered scheme for social media services to remove cyberbullying material targeted towards Australian children in response to cyberbullying on social media platforms.⁹⁹ Tier 1 social media platforms, including Twitter, TikTok, and Snapchat, participate in the scheme voluntarily. If a complaint is made to these platforms about cyberbullying material and the material is not removed within a specific period (currently 48 hours), the eSafety Commissioner may issue a request to have the material removed from the service. The Minister of Communications declares Tier 2 social media services following a recommendation by the eSafety Commissioner. Facebook, Instagram, and YouTube have been declared to be Tier 2 social media services. Tier 2 social media services may be subject to civil penalties and legally binding notices if they do not comply with requests to remove cyberbullying material.

The *Online Safety Act 2021* (Cth) builds on the existing regulatory framework established in the *Enhancing Online Safety Act* and will take effect on 23 January 2022. The new Act

⁹⁷ Stephanie Bory, 'Social media a 'coward's palace', says Prime Minister, as he promises more action to hold online abusers responsible', (7 October 2021), *ABC News* <https://www.abc.net.au/news/2021-10-07/prime-minister-defends-dutton-twitter-defamation-action/100522002>

⁹⁸ Katharine Gelber, 'A better way to regulate online hate speech: require social media companies to bear a duty of care to users' (14 July 2021), *The Conversation* <https://theconversation.com/a-better-way-to-regulate-online-hate-speech-require-social-media-companies-to-bear-a-duty-of-care-to-users-163808>

⁹⁹ *Enhancing Online Safety Act 2015*; eSafety Commissioner, *Working with social media*, Australian Government. <https://www.esafety.gov.au/about-us/consultation-cooperation/working-with-social-media>.

introduces additional compliance obligations, including an online content scheme for removing specific material and a complaints-based removal notice scheme. The reform broadens the scheme to capture harms occurring on services other than social media.¹⁰⁰ A new set of industry codes are expected to be developed within 12 months of Royal Assent to guide industry compliance with their new obligations and to promote the adoption of responsible processes for dealing with online content and safety issues. The desired approach would see the codes developed by industry and then reviewed and endorsed by the eSafety Commissioner. The Commissioner has the authority to impose industry-wide standards if the codes cannot be agreed or do not meet the desired safety outcomes.¹⁰¹

The eSafety Commissioner has released general guidelines on social media use for parents, children, and young people. There is no recommended time limit for screen time; instead, the guidelines identify warning signs, such as reduced personal hygiene or becoming withdrawn from friends and family, that suggest online activity is becoming problematic for children and young people.¹⁰²

Extremist content

In response to the Christchurch terror attack on 15 March 2019 where the perpetrator live-streamed footage of the event on social media platforms, the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth)¹⁰³ was passed, requiring Internet, content, and hosting providers to report abhorrent violent conduct occurring in Australia on their services to the Australian Federal Police.¹⁰⁴ Failure to report violent material may result in fines of up to \$888,000 for corporations,¹⁰⁵ and failure to remove the material from their services may result in fines of up to \$11.1 million or 10 per cent of annual turnover, whichever is higher.¹⁰⁶

Australia also signed the Christchurch Call, a voluntary commitment from governments and online service-providers aimed at addressing terrorist and extremist content online, established by the New Zealand and French Governments.¹⁰⁷ Government signatories have committed to considering appropriate action to prevent the use of online services to disseminate terrorist and violent extremist content through actions such as the development of industry standards or voluntary frameworks, as well as regulatory or policy measures that are consistent with international human rights law and the principle of a free, open and secure internet. Online service providers have committed to

¹⁰⁰ *Online Safety Act 2021* (Cth)

¹⁰¹ eSafety Commissioner, *Online Safety Act 2021 Factsheet*, Australian Government.

<https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf>

¹⁰² eSafety Commissioner, *Time online*. Australian Government. <https://www.esafety.gov.au/parents/big-issues/time-online>.

¹⁰³ Attorney-General's Department, *Abhorrent violent material*. Australian Government. <https://www.ag.gov.au/crime/abhorrent-violent-material>.

¹⁰⁴ *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, s474.33

¹⁰⁵ *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, s474.33

¹⁰⁶ *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, s474.34

¹⁰⁷ Christchurch Call, *The Christchurch Call to Action to Eliminate Terrorist and Violent Extremist Content Online*, (2019). <https://www.christchurchcall.com/christchurch-call.pdf>.

implementing measures to prevent the upload of this content, with the Christchurch Call supported by social media platforms including Facebook and Twitter.¹⁰⁸

Media

Earlier this year, Australia introduced a mandatory code of conduct, the *News Media and Digital Platforms Mandatory Bargaining Code*,¹⁰⁹ to address bargaining power imbalances between digital platforms, such as Google and Facebook, and Australian news media businesses. The Code enables eligible news businesses to bargain individually or collectively with digital platforms over payment for the inclusion of news on their platforms. The Code established a negotiation framework for news businesses and digital platforms to reach binding agreements and provides for an independent arbiter to determine the remuneration if parties cannot reach an agreement. While the Code was supported by both major parties in the Australian Parliament,¹¹⁰ it was met with significant opposition by Facebook and Google. In response to the development of the legislation, in February 2021 Facebook temporarily blocked Australian users from viewing and sharing news on its platform, even blocking information and government pages, including health and emergency services.¹¹¹ The draft legislation was amended (and subsequently passed in parliament in February 2021) to include a mediation period to allow digital platforms and news businesses to attempt to reach agreement before entering into arbitration, and to consider platforms' existing agreements with publishers before deciding on the application of the Code.¹¹²

Blocking illegal online services

The *Telecommunications Act 1997* (Cth) allows Australian Government agencies to block illegal online services.¹¹³ Following a review, the Department of Communications and the Arts published guidelines in 2017 on the use of the provision that entails "good practice measures" to be followed, including obtaining authorisation before disrupting online services, and limiting disruptions to instances of serious offenses or national security threats.

¹⁰⁸ Christchurch Call, *The Christchurch Call to Action to Eliminate Terrorist and Violent Extremist Content Online*, (2019) <https://www.christchurchcall.com/christchurch-call.pdf>

¹⁰⁹ *Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Act 2021*

¹¹⁰ Lisa Valentin, 'Landmark media code set to become law with Labor's backing', (16 February 2021), *The Sydney Morning Herald*, <https://www.smh.com.au/politics/federal/landmark-media-code-set-to-become-law-with-labor-s-backing-20210216-p572wv.html>

¹¹¹ Reuters, 'Facebook news ban stops Australians from sharing or viewing Australian and international news content', (18 February 2021), *ABC News*, <https://www.abc.net.au/news/2021-02-18/facebook-to-restrict-sharing-or-viewing-news-in-australia/13166208>; Amanda Meade, Josh Taylor & Daniel Hurst, 'Facebook reverses Australia news ban after government makes media code amendments', (23 February 2021), *The Guardian* <https://www.theguardian.com/media/2021/feb/23/facebook-reverses-australia-news-ban-after-government-makes-media-code-amendments>

¹¹² Josh Frydenberg & Paul Fletcher, *News Media and Digital Platforms Mandatory Bargaining Code*. (8 December 2020) <https://ministers.treasury.gov.au/ministers/josh-frydenberg-2018/media-releases/news-media-and-digital-platforms-mandatory-bargaining#:~:text=The%20News%20Media%20and%20Digital public%20interest%20journalism%20in%20Australia>

¹¹³ *Telecommunications Act 1997*, s313(3)

Privacy

In October 2021, the Australian Government released an Exposure Draft of the Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill (*Online Privacy Bill*)¹¹⁴ for submissions and feedback by 6 December 2021. The Bill proposes amendments to the *Privacy Act 1988* (Cth), including establishing a framework to develop, implement and enforce a binding online privacy code to regulate large online platforms, social media platforms, and data brokers. The introduction of an Online Privacy Code is part of the response to the ACCC's Digital Platforms Inquiry Report,¹¹⁵ which made extensive recommendations to strengthen privacy protections for individuals and improve transparency and accountability in data handling practices. The Bill would prevent social media platforms from accessing a child's data without a parent or guardian's permission and require companies to make all reasonable attempts to verify the age of users. The Bill would introduce stricter penalties and enforcement powers to enable the Office of the Australian Information Commissioner to resolve matters more effectively.

Disinformation and misinformation

In December 2019, as part of its response to the ACCC's Digital Platforms Inquiry Report, the Australian Government asked the digital industry to develop a voluntary code of conduct for disinformation and news quality.¹¹⁶ The *Australian Code of Practice on Disinformation and Misinformation* was released in February 2021. It was drafted by Digital Industry Group Inc (DiGi), a non-profit industry association advocating for the digital industry in Australia. The voluntary code commits a diverse set of technology companies, including Facebook, Twitter, and Google, to reducing the risk of online misinformation causing harm to Australians. The signatories committed to safeguards to protect Australians and must publicly report their efforts in response to disinformation and misinformation. The Australian Communications and Media Authority reports on the efficacy of the code. The code has been criticised for its self-regulatory and opt-in approach, which may hinder its effectiveness.¹¹⁷

Defamation

On 25 October 2021, Nationals MP Anne Webster introduced a private member's Bill (the *Social Media (Basic Expectations and Defamation) Bill 2021*), which would enable the

¹¹⁴ Attorney-General's Department, *Online Privacy Bill Exposure Draft*, Australian Government <https://consultations.ag.gov.au/rights-and-protections/online-privacy-bill-exposure-draft/>

¹¹⁵ Australian Competition and Consumer Commission, *Digital platforms inquiry - final report* (2019) <https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report>

¹¹⁶ Australian Government, *Regulating in the Digital Age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry* (2019) <https://treasury.gov.au/sites/default/files/2019-12/Government-Response-p2019-41708.pdf>

¹¹⁷ DiGi, *Disinformation code*, <https://digi.org.au/disinformation-code/>; Asha Barbaschow, 'Facebook, Google, Microsoft, TikTok, and Twitter adopt Aussie misinformation code', (22 February 2021), ZDNet <https://www.zdnet.com/article/facebook-google-microsoft-tiktok-and-twitter-adopt-aussie-misinformation-code/>; Josh Taylor, 'What is the Australian government doing to crack down on big tech, and why?', (30 October 2021), *The Guardian* <https://www.theguardian.com/australia-news/2021/oct/30/what-is-the-australian-government-doing-to-crack-down-on-big-tech-and-why>

Communications Minister to set basic expectations of social media service providers regarding the hosting of defamatory material on social media platforms. The proposed legislation would ensure that service providers are liable for defamatory material hosted on their platforms and not removed within a reasonable timeframe after notice from the eSafety Commissioner.¹¹⁸ It remains unclear whether the government will adopt the Bill.¹¹⁹

On 28 November 2021, Prime Minister Scott Morrison and Attorney-General Michaelia Cash announced proposed new legislation which would include the introduction of new court powers to force social media platforms to unmask anonymous online trolls, with the aim of better protecting Australians online. The reforms, described by the Government as ‘world-leading’, will ensure social media companies are considered publishers and can be held liable for defamatory comments posted on their platforms. and the legislation is expected to be introduced into parliament in early 2022.¹²⁰ While reserving the Opposition’s position, the Federal Opposition leader Anthony Albanese questioned how effective the imposition of domestic controls would be on a global industry and whether they could easily be avoided, for example by the use of foreign IP-addresses.¹²¹

Foreign Interference

Other legislation relevant to foreign interference includes the *Telecommunications Legislation Amendment (International Production Orders) Bill 2020*¹²² and the *Surveillance Legislation Amendment (Identify and Disrupt) Act 2021*.¹²³

¹¹⁸ *Social Media (Basic Expectations and Defamation) Bill 2021*

¹¹⁹ Josh Taylor, ‘What is the Australian government doing to crack down on big tech, and why?’, (30 October 2021), *The Guardian* <https://www.theguardian.com/australia-news/2021/oct/30/what-is-the-australian-government-doing-to-crack-down-on-big-tech-and-why>; Paul Karp, ‘Social media giants face \$10m fines for privacy breaches under proposed government reform’, (25 October 2021), *The Guardian* <https://www.theguardian.com/australia-news/2021/oct/25/social-media-giants-face-10m-fines-for-privacy-breaches-under-proposed-government-reform>

¹²⁰ Australian Government, Prime Minister of Australia, *Combatting online trolls and strengthening defamation laws*, (28 November 2021) <https://www.pm.gov.au/media/combating-online-trolls-and-strengthening-defamation-laws>

¹²¹ Tom Lowrey, ‘Social media companies could be forced to give out names and contact details under new anti-troll laws’, (28 November 2021), *ABC News* <https://www.abc.net.au/news/2021-11-28/social-media-laws-online-trolls/100657004>

¹²² Establishes a new legal framework to access overseas communication data for law enforcement and national security purposes, facilitating access to encrypted communications provided by non-Australian companies

¹²³ Grants the Australian Federal Police and Australian Criminal Intelligence Commission (ACIC) the ability to request new types of warrants to investigate and disrupt “serious” crime.

International Regulatory Responses

The United States

A surge in false, misleading and inflammatory content surrounding the November 2020 United States elections led to the violent attack on the U.S. Capitol on 6 January 2021. Following the storming of the Capitol, the U.S. Congress held a congressional hearing on 25 March 2021 interrogating the CEOs of Facebook, Google and Twitter about how their social media platforms spread extremism and misinformation, and the role of these platforms in the attack.¹²⁴ Executives from Facebook, YouTube and Twitter testified before a Senate Judiciary subcommittee on 27 April 2021 on the ways their platforms' algorithms influence users. Senators from both political sides criticised the negative effects of the advertising-supported business models and questioned the serving of harmful misinformation on the platforms.¹²⁵

In mid-September 2021, the Wall Street Journal published a series of articles, commonly referred to as the 'Facebook Files', based on internal Facebook documents, released by a whistle-blower. The series included reports on internal studies, demonstrating that Facebook was aware of the negative impact of Instagram on teenage users, how the platform prioritises profit over public safety, and how its design features amplify hate, political unrest and misinformation.¹²⁶ On 5 October 2021, Frances Haugen came forward as the anonymous Facebook whistle-blower who released the internal documents.¹²⁷ Following the widespread media attention, she was invited to testify before the Senate Committee on Commerce, Science and Transportation about how the social network knowingly harms people (especially teenagers) with toxic content and how the company is failing to adequately protect against threats emerging from foreign entities including Russia, China and Iran.¹²⁸

Social media has been at the forefront of Congressional and Senate hearings and oversight efforts over the past few years, including an October 2019 hearing on content

¹²⁴ US Congress, *Hearing on "Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation"*, (2021) <https://www.congress.gov/event/117th-congress/house-event/111407>; House Committee on Energy and Commerce, *Hearing on "Disinformation Nation: Social media's role in promoting extremism and misinformation"*, (2021) <https://energycommerce.house.gov/committee-activity/hearings/hearing-on-disinformation-nation-social-medias-role-in-promoting>

¹²⁵ US Senate Hearing, *Algorithms and Amplification: How Social Media Platforms' Design Choices Shape Our Discourse and Our Minds* <https://www.judiciary.senate.gov/meetings/algorithms-and-amplification-how-social-media-platforms-design-choices-shape-our-discourse-and-our-minds>

¹²⁶ Reed Albergotti, 'Frances Haugen took thousands of Facebook documents: This is how she did it', (26 October 2021), *The Washington Post* <https://www.washingtonpost.com/technology/2021/10/26/frances-haugen-facebook-whistleblower-documents/>; Wall Street Journal Investigation, 'The Facebook Files', *The Wall Street Journal*, [wsj.com/articles/the-facebook-files-11631713039](https://www.wsj.com/articles/the-facebook-files-11631713039)

¹²⁷ Daniel E. Slotnik, 'Whistle-Blower Unites Democrats and Republicans in Calling for Regulation of Facebook', (5 October 2021), *The New York Times* <https://www.nytimes.com/live/2021/10/05/technology/facebook-whistleblower-frances-haugen>

¹²⁸ US Senate Committee on Commerce, Science & Transportation, *Hearings on "Protecting Kids Online: Testimony from a Facebook Whistleblower"*, (5 October 2021) <https://www.commerce.senate.gov/2021/10/protecting%20kids%20online:%20testimony%20from%20a%20facebook%20whistleblower>

moderation¹²⁹ and a June 2020 hearing on the rise of disinformation and extremism online.¹³⁰ These are part of a broader policy discussion for reform currently taking place within the United States.¹³¹

Discussion on regulating social media platforms has largely centred on the First Amendment to the U.S. Constitution which prohibits censorship by the government, and on proposed amendments to Section 230 of the *Communications Decency Act*¹³² which provides that an internet provider (such as a social media platform) cannot be treated as the publisher or speaker of third-party content. Section 230 gives social media platforms broad protection from liability for defamatory content, and broad scope to moderate discussions and remove or not remove posts.¹³³

Section 230(c)(1) provides "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider".¹³⁴ Section 230(c)(2) provides "Good Samaritan" protection from civil liability for operators of interactive computer services who remove or moderate third-party material they deem obscene or offensive, including constitutionally protected speech.¹³⁵

Both Republicans and Democrats have advocated for the repeal of Section 230. As President, Donald Trump asserted online platforms were editing his and other right-wing sources' content, and they should no longer be protected from proceedings charging them with discrimination.¹³⁶ Conversely, President Joseph Biden has argued for the repeal of Section 230 on the basis that online platforms should be held responsible for disseminating false or misleading content.¹³⁷ Concerns have been raised regarding implications of reform on the First Amendment, for example that Section 230 encourages the moderation of content, and that the First Amendment protects social media platforms from hate speech liability.¹³⁸

¹²⁹ House on Energy & Commerce, *Hearing on "fostering a healthier internet to protect consumers"* (16 October 2019) <https://energycommerce.house.gov/committee-activity/hearings/hearing-on-fostering-a-healthier-internet-to-protect-consumers>

¹³⁰ House on Energy & Commerce, *Joint hearing on "A country in crisis: How disinformation online is dividing the nation"* (24 June 2020) <https://energycommerce.house.gov/committee-activity/hearings/joint-hearing-on-a-country-in-crisis-how-disinformation-online-is>

¹³¹ Gerrit De Vynck, Cat Zakrzewski & Elizabeth Dwoskin, 'Big tech CEOs face lawmakers in House hearing on social media's role in extremism, misinformation', (10 April 2021), *The Washington Post*

<https://www.washingtonpost.com/technology/2021/03/25/facebook-google-twitter-house-hearing-live-updates/>

¹³² 47 U.S.C. s230 [https://uscode.house.gov/view.xhtml?req=\(title:47%20section:230%20edition:prelim\)](https://uscode.house.gov/view.xhtml?req=(title:47%20section:230%20edition:prelim))

¹³³ Section 230 was developed in response to lawsuits against Internet service providers in the early 1990s that resulted in different interpretations of whether the service providers should be treated as publishers or, alternatively, as distributors of content created by its users. It was enacted as part of the Communications Decency Act, and the Act was challenged and ruled by the Supreme Court in *Reno v. American Civil Liberties Union* (1997) to be unconstitutional. However, Section 230 was determined to be severable from the rest of the legislation and remained.

¹³⁴ 47 U.S.C. s230 [https://uscode.house.gov/view.xhtml?req=\(title:47%20section:230%20edition:prelim\)](https://uscode.house.gov/view.xhtml?req=(title:47%20section:230%20edition:prelim))

¹³⁵ 47 U.S.C. s230 [https://uscode.house.gov/view.xhtml?req=\(title:47%20section:230%20edition:prelim\)](https://uscode.house.gov/view.xhtml?req=(title:47%20section:230%20edition:prelim))

¹³⁶ Abram Brown, 'What is section 230 - and why does Trump want to change it?' (28 May 2020), *Forbes* <https://www.forbes.com/sites/abrambrown/2020/05/28/what-is-section-230-and-why-does-trump-want-to-change-it/?sh=7b540c33389d>

¹³⁷ Sinan Aral, *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health—And How We Must Adapt*. Currency, New York, 2020.

¹³⁸ Ellen Goodman & Ryan Whittington, 'Section 230 of the Communications Decency Act and the Future of Online Speech', (9 August 2019), *The German Marshall Fund of the United States* <https://www.gmfus.org/news/section-230-communications-decency-act-and-future-online-speech>.

There have been a number of proposed reforms to regulate social media platforms and while there appears to be wide-ranging support for reforming Section 230, there is little agreement among political leaders about how this should occur. Unlike Australia where the introduction of legislation is tightly controlled by the government and the party system and Members of Parliament and Senators vote along party lines, the U.S. political system allows any member of the House of Representatives or Senate to propose legislation, with or without party support. Consequently, while many reforms have been proposed, they have not gained enough momentum and consensus to be implemented.

There have been numerous proposed reforms to Section 230, including many introduced in the 117th congressional session (2020-2021), several of which have been outlined below. These proposed Bills look to reform social media through amendments to Section 230 in four key ways: by repealing section 230 in whole, limiting the scope of Section 230, imposing new obligations or altering the 'Good Samaritan' part of Section 230.¹³⁹ While there are unique aspects to the regulatory debate and legislative reform in the US, it plays a leading role in setting the global policy agenda around social media, particularly given that the key technology companies are headquartered there. Accordingly, it is important to monitor and understand their proposals.

Regulating Online Platforms

Following the Facebook Files revelations, two major new Bills were put forward targeting big tech and social media. The most recent Bill put forward by Energy and Commerce Chair Frank Pallone and leading House Democrats¹⁴⁰, the *Justice Against Malicious Algorithms Act*¹⁴¹ would amend Section 230 to remove absolute immunity in certain instances, specifically when an online platform knowingly or recklessly uses an algorithm to recommend harmful content that contributes to physical or severe emotional injury.¹⁴² The Bill targets algorithms that materially contribute to a physical or severe emotional injury to a person. However, the proposed reform only applies to algorithms or search features that rely on personalisation.

The *Platform Accountability and Consumer Transparency Act (PACT Act)* is a bipartisan bill introduced to Congress in March 2021, aimed at imposing new obligations on internet companies. The PACT Act was originally introduced in the 2019-2020 Congressional session. The updated version of the Bill seeks to make content moderation for social

¹³⁹ Meghan Enand et al, 'All the ways congress wants to change section 230' (23 March 2021), *Slate* <https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html>

¹⁴⁰ Including Energy and Commerce Committee Chairman Frank Pallone, Jr. (D-NJ), Communications and Technology Subcommittee Chairman Mike Doyle (D-PA), Consumer Protection and Commerce Subcommittee Chair Jan Schakowsky (D-IL), and Health Subcommittee Chair Anna Eshoo (D-CA)

¹⁴¹ *Justice Against Malicious Algorithms Act*, <https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/101421%20EC%20Section%20230%20Text.pdf>

¹⁴² House Committee on Energy & Commerce, *E&C Leaders announce legislation to reform Section 230* <https://energycommerce.house.gov/newsroom/press-releases/ec-leaders-announce-legislation-to-reform-section-230>

media platforms more transparent and increase consumer protections.¹⁴³ The PACT Act requires platforms to issue public statements on their policies regarding moderation, demonisation and the removal of user content, in addition to publishing transparency reports summarising their actions and statistics. The PACT Act additionally gives State Attorneys General the authority to bring legal action against platforms that violate federal civil law.¹⁴⁴ The Bill has been referred to the Senate Committee on Commerce, Science, and Transportation.¹⁴⁵

The proposed *Safeguarding Against Fraud, Exploitation, Threats, Extremism and Consumer Harms Act* (SAFE TECH Act) was introduced in May 2021 and limits the scope of section 230 immunity. The SAFE TECH Act removes the legal protections for platform providers in situations where they have accepted payment to either make the speech available or have created or funded the speech.¹⁴⁶ The Bill also creates new exceptions to the liability protections in cases involving civil rights laws, antitrust laws, stalking, harassment or intimidation laws, international human rights laws and wrongful death action.¹⁴⁷ The objective of the SAFE TECH Act is to hold social media companies accountable for enabling cyber-stalking, targeted harassment, and discrimination.¹⁴⁸ The bill was referred to the Subcommittee on Communications and Technology in May 2021.¹⁴⁹

The *Protecting Americans from Dangerous Algorithms Act*, introduced in March 2021, similarly removes liability immunity for a platform, focusing on the algorithmic promotion of harmful, radicalising content interfering with civil rights.¹⁵⁰ Under the proposed legislation, companies may still use Section 230 as a defence in cases if they distribute

¹⁴³ U.S. Senator for Hawai'i Brian Schatz (17 March 2021). Schatz, Thune Reintroduce Legislation To Update Section 230, Strengthen Rules, Transparency on Online Content Moderation, Hold Internet Companies Accountable For Moderation Practices. <https://www.schatz.senate.gov/news/press-releases/schatz-thune-reintroduce-legislation-to-update-section-230-strengthen-rules-transparency-on-online-content-moderation-hold-internet-companies-accountable-for-moderation-practices>.

¹⁴⁴ Frank Konkel, 'Bipartisan Bill Would Hold Tech Companies Responsible for Moderating Content', (17 March 2021), Nextgov. <https://www.nextgov.com/policy/2021/03/bipartisan-bill-would-hold-tech-companies-responsible-moderating-content/172739/>; Ashley Johnson & Daniel Castro, 'PACT Act Would Increase Platform Transparency, But Undercut Intermediary Liability', (7 August 2020), Information Technology & Innovation Foundation <https://itif.org/publications/2020/08/07/pact-act-would-increase-platform-transparency-undercut-intermediary>

¹⁴⁵ PACT Act, s797 <https://www.congress.gov/bill/117th-congress/senate-bill/797/text?q=%7B%22search%22%3A%5B%22PACT+Act%22%2C%22PACT%22%2C%22Act%22%5D%7D&r=5&s=1>

¹⁴⁶ Mark R. Warner, US Senator from the Commonwealth of Virginia, (5 February 2021), Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230. <https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230>.

¹⁴⁷ Taylor Hatmaker, 'The SAFE TECH ACT offers Section 230 reform, but the law's defenders warn of major side effects' (6 February 2021), TechCrunch. https://techcrunch.com/2021/02/05/safe-tech-act-section-230-warner/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2x1LnNvbS8&guce_referrer_sig=AQAAAK_1kWpdJ_ZrtSdGq4U1bPw3cTXtx7mSHfZtpkVUbdVRUBvylrITpop-GtBDjd9xYur9rKclH_zXPMfjuSE72yaHNJdtXxPhztFH0X1JEWhHoyxP2alepdEMUAqT5eBuiK0dVLBXi01pOj8ESvgRhM_hbjAfZJu2vvQ05KCeZv8_Lf.

¹⁴⁸ Mark R. Warner, US Senator from the Commonwealth of Virginia, Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230, (5 February 2021) <https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230>.

¹⁴⁹ SAFE TECH Act <https://www.congress.gov/bill/117th-congress/house-bill/3421/actions?r=3&s=3>

¹⁵⁰ *Protecting Americans from Dangerous Algorithms Act* <https://www.congress.gov/bill/117th-congress/house-bill/2154/text?q=%7B%22search%22%3A%5B%22Malinowski%22%5D%7D&r=2&s=3>; Congressman Tom Malinowski, Representing the 7th District of New Jersey, Reps. Malinowski and Eshoo reintroduce bill to hold tech platforms accountable for algorithmic promotion of extremism (24 March 2021) <https://malinowski.house.gov/media/press-releases/reps-malinowski-and-eshoo-reintroduce-bill-hold-tech-platforms-accountable>.

content using methods that are "obvious, understandable, and transparent" to a reasonable user.¹⁵¹ The Bill was referred to the Subcommittee on Communications and Technology in March 2021.

Other examples of legislation under consideration include Republican Senator Rick Scott's *Safe Social Media Act*, introduced in May 2021, which would require the Federal Trade Commission, in coordination with the Centers for Disease Control, to conduct a study on social media use among American teenagers and children including the use of personal information in algorithms, the mental health effects and the long-term impact of extended usage.¹⁵² The *Abandoning Online Censorship (AOC) Act* was introduced in February 2021 and would repeal section 230.¹⁵³ The *Health Misinformation Act of 2021* was introduced in July 2021 and would create an exception to liability protections for platforms using algorithms to promote health misinformation, and would take effect for the remainder of the public health emergency.¹⁵⁴ The *Disincentivizing Internet Service Censorship of Online Users and Restrictions on Speech and Expression Act (DISCOURSE)* was introduced in June 2021 and would amend the Good Samaritan provision so that platforms would only receive liability protections when content that is extremist, obscene or unlawful was moderated. The Bill would also add a clause making it more difficult to be protected under section 230 when content is moderated so that it "burdens" religious exercise.¹⁵⁵

The *Protect Speech Act*, introduced in June 2021 would narrow a platform's ability to use Section 230 as a defence for content removal.¹⁵⁶ The *21st Century Foundation for the Right to Express and Engage in Speech Act (21st Century FREE Speech Act)* would repeal section 230, and replace it with a reclassification of platforms as common carriers, required to provide their services to everyone, imposing additional responsibilities to platforms. Liability protections would only apply to platforms removing content in accordance with their content moderation policies, and the bill would also establish a private right of action.¹⁵⁷ The *Stop Shielding Culpable Platforms Act*, introduced in March 2021 would amend Section 230 to note it does not prevent a provider or user of a platform from being treated as the distributor of information provided by a third party.¹⁵⁸ The *See Something, Say Something Online Act of 2021*, introduced in January 2021 would require platforms to report suspicious transmissions they detect, and platforms would have to take reasonable steps to prevent and address these transmissions. Section 230 would could not be used as a defence if the provider should have been reasonably aware of the suspicious

¹⁵¹ *Protecting Americans from Dangerous Algorithms Act* <https://www.congress.gov/bill/117th-congress/house-bill/2154/text?q=%7B%22search%22%3A%5B%22Malinowski%22%5D%7D&r=2&s=3>

¹⁵² A Bill to require the Federal Trade Commission to conduct a study regarding social media use by teenagers, S. 1630, 117th Congress. (2021). <https://www.congress.gov/bill/117th-congress/senate-bill/1630/text?r=4&s=1>.

¹⁵³ AOC Act <https://www.congress.gov/bill/117th-congress/house-bill/874/text?q=%7B%22search%22%3A%5B%22section+230%22%5D%7D&r=1&s=3>

¹⁵⁴ AOC Act <https://www.congress.gov/bill/117th-congress/house-bill/874/text?q=%7B%22search%22%3A%5B%22section+230%22%5D%7D&r=1&s=3>

¹⁵⁵ DISCOURSE Act, s2228 <https://www.congress.gov/bill/117th-congress/senate-bill/2228>

¹⁵⁶ *Protect Speech Act* <https://www.congress.gov/bill/117th-congress/house-bill/3827/text?q=%7B%22search%22%3A%5B%22protect+speech%22%5D%7D&r=1&s=2>

¹⁵⁷ *Century FREE Speech Act*, s1384 <https://www.congress.gov/bill/117th-congress/senate-bill/1384?s=9&r=2>

¹⁵⁸ *Stop Shielding Culpable Platforms Act* <https://www.congress.gov/bill/117th-congress/house-bill/2000/text?q=%7B%22search%22%3A%5B%22jim+banks%22%5D%7D&r=4&s=5>

transmission.¹⁵⁹ The *Curbing Abuse and Saving Expression in Technology (CASE-IT) Act*, introduced in January 2021 would prevent a platform from using Section 230 as a defence for one year if the company creates, posts, materially contributes to, or induces another person to contribute to illegal online content.¹⁶⁰ The *Protecting Constitutional Rights From Online Platform Censorship Act*, introduced in January 2021 would remove the Good Samaritan provision and make it unlawful for any internet platform to restrict access or availability to content.¹⁶¹

While there appears to be some level of bipartisan consensus that there is an issue, the number and breadth of proposed legislation demonstrates the difficulty in agreeing on a regulatory solution.

Antitrust

The *American Innovation and Choice Online Act*,¹⁶² put forward by a bipartisan group of Senators is an antitrust measure, is another high profile Bill recently proposed. This Bill does not look at Section 230; instead, it would prevent U.S. technology giants from giving an advantage to their own products over those of competitors. This comes after a various high-profile hearings in which the Senate Judiciary Subcommittee on Competition Policy, Antitrust and Consumer Rights investigated the conduct of big tech companies including Apple and Google.¹⁶³ The Bill would help restore competition online by establishing 'commonsense' rules for dominant digital platforms to prevent them from abusing their market power to harm competition, online businesses and consumers and from reducing incentives to innovate. The proposed Act outlines clear rules to protect competition, and gives enforcers strong and flexible tools to deter violations and hold platforms to account.¹⁶⁴ The *American Innovation and Choice Online Act* follows on from the proposed House of Representatives Bill, the *American Choice and Innovation Online Act* which was introduced on 11 June 2021, and ordered to be amended by the House on 24 June 2021.¹⁶⁵

While the *American Innovation and Choice Online Act* is one of the most widely supported bills on antitrust measures, there are other similar proposed reforms, several of which are listed below. The House of Representatives Antitrust Subcommittee has proposed the *Augmenting Compatibility and Competition by Enabling Service Switching (ACCESS) Act*, which would require covered platforms to "maintain a set of transparent, third-party-

¹⁵⁹ See *Something, Say Something Online Act of 2021*, s27 <https://www.congress.gov/bill/117th-congress/senate-bill/27/text?q=%7B%22search%22%3A%5B%22see+something+say+something%22%5D%7D&r=1&s=1>

¹⁶⁰ CASE-IT Act <https://www.congress.gov/bill/117th-congress/house-bill/285/text?q=%7B%22search%22%3A%5B%22section+230%22%5D%7D&r=2&s=3>

¹⁶¹ *Protecting Constitutional Rights from Online Platform Censorship Act* <https://www.congress.gov/bill/117th-congress/house-bill/83/text?q=%7B%22search%22%3A%5B%22section+230%22%5D%7D&r=3&s=2>

¹⁶² *American Innovation and Choice Online Act* <https://www.klobuchar.senate.gov/public/cache/files/7/d/7d176f5c-c84b-4207-8d96-77469fe1db44/903C851389B04EA66A4D2133A3EA18CF.sil21b56.pdf>

¹⁶³ Michael W. Scarborough & M Kevin Castello, 'Senate Zeros in on Big Tech with Latest Antitrust Reform Bill' (2021) 11 (334) *National Law Review* <https://www.natlawreview.com/article/senate-zeros-big-tech-latest-antitrust-reform-bill>

¹⁶⁴ US Senator Amy Klobuchar, *Support Builds for Bipartisan Legislation From Klobuchar, Grassley, and Colleagues to Rein in Big Tech* (18 October 2021) <https://www.klobuchar.senate.gov/public/index.cfm/2021/10/support-builds-for-bipartisan-legislation-from-klobuchar-grassley-and-colleagues-to-rein-in-big-tech>

¹⁶⁵ *American Choice and Innovation Online Act* <https://www.congress.gov/bill/117th-congress/house-bill/3816/actions>

accessible interfaces . . . to enable the secure transfer of data to a user",¹⁶⁶ and the *Ending Platform Monopolies Act*, which would prohibit technology platforms with at least 50,000,000 active monthly U.S. users and a market capitalisation of over \$600 billion from selling products or services that they own and control.¹⁶⁷

Protecting Children

U.S. Democratic Representative Kathy Castor introduced an updated *Protecting the Information of our Vulnerable Children and Youth Act (Kids PRIVCY Act)* on 29 July 2021 to strengthen the *Children's Online Privacy Protection Act (COPPA)*. The Bill expands privacy protections for children and teenagers, and incorporates key elements of the U.K.'s Age-Appropriate Design Code. This includes banning companies from providing targeted advertisements to children and teenagers, requiring opt-in consent for individuals under 18, creating a right to access, correct and delete personal information, expanding coverage of companies, and strengthening enforcement.¹⁶⁸

European Union

The European Union (EU) is comprised of 27 sovereign and independent countries (and their citizens), known as Member States, who have pooled some of their 'sovereignty', delegating some of their decision-making powers to shared institutions such as the European Council, the European Parliament and the European Commission. Generally, the European Commission proposes new laws, the European Parliament and European Council adopt them, and then Member States and the European Commission then implement them. Regulations passed are applicable and binding in all Member States directly, and while they do not have to be passed into national law by Member States, national laws may need to be amended to avoid conflict with the regulation.¹⁶⁹

The European Union is striving to be global role model for the digital economy and internationally promote its digital standards. As such a large and influential market, their regulation of social media contributes to the setting of global norms. The legislation proposed by the EU will have significant impacts on European democracy, and how they are able to balance the market, state, civil society and speech.¹⁷⁰ The EU and its member

¹⁶⁶ ACCESS ACT <https://cicilline.house.gov/sites/cicilline.house.gov/files/documents/ACCESS%20Act%20-%20Bill%20Text%20%281%29.pdf>

¹⁶⁷ *The Ending Platform Monopolies Act* <https://cicilline.house.gov/sites/cicilline.house.gov/files/documents/Ending%20Platform%20Monopolies%20-%20Bill%20Text.pdf>

¹⁶⁸ U.S. Representative Kathy Castor, (29 July 2021), *Rep. Castor Reintroduces Landmark Kids PRIVCY Act to Strengthen COPPA, Keep Children Safe Online* <https://www.congress.gov/bill/117th-congress/house-bill/4801?r=12&s=1> <https://castor.house.gov/news/documentsingle.aspx?DocumentID=403677>; <https://www.congress.gov/bill/117th-congress/house-bill/4801?r=12&s=1>

¹⁶⁹ https://eeas.europa.eu/archives/delegations/singapore/documents/more_info/eu_publications/how_the_european_union_works_en.pdf

¹⁷⁰ Damian Tambini, 'Media Policy in 2021. As the EU takes on the tech giants, will the UK' (12 January 2021) *London School of Economics* <https://blogs.lse.ac.uk/medialse/2021/01/12/media-policy-in-2021-as-the-eu-takes-on-the-tech-giants-will-the-uk/>

states are pursuing regulation in various areas, including artificial intelligence, digital markets and services, connectivity, cybersecurity, data and digital identity.¹⁷¹

The European Commission's approach focuses on digital transformation, making technology work for people and fostering a cohesive democratic society through investment in digital skills. Europe's overarching digital strategy, *Shaping Europe's Digital Future*,¹⁷² released on 19 February 2020, invests in digital skills for all Europeans and protects against cyber threats. The Strategy focuses on ensuring that technology – in particular artificial intelligence – is developed in a way that respects individuals rights, and maintains their trust. The digital strategy targets a fair and competitive digital economy, and an open, democratic and sustainable society.¹⁷³

Digital Services and Markets

The European approach to regulating the digital environment centres around the *Digital Markets Act*¹⁷⁴ and *Digital Services Act*,¹⁷⁵ which were both proposed on 15 December 2020. These Acts are currently proposals. In order for them to become binding on EU member states, they require approval by the European Council and the European Parliament. This is expected to take 18 months from when they were proposed by the European Commission. These Acts aim to provide users with access to a wide range of safe products and services online. Upon implementation, these Acts would impose an EU-wide obligation on Member States to ensure that digital services connecting consumers to goods, services and content also protect user's fundamental rights. The regulations would be binding in their entirety and directly applicable in all Member States, as harmonisation and cooperation cannot be achieved by Member States acting in silos.

The *Digital Markets Act* is based on the understanding that social media and digital platforms have strong network effects, particularly as increased use by business and end-users (the consumers of the goods and services) drives further demand. In mediating the connection between businesses and end-users, platforms have the potential to create lock-in effects (making the user dependent on them for products and services, unable to use another service without substantial switching costs) and a significant level of dependence of both businesses and users. In turn, this enhances the bargaining power of the platforms, creates unequal relationships between market actors, and impedes innovation.

¹⁷¹ European Commission, *Priorities 2019-2024, A Europe fit for Digital Age* https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en

¹⁷² European Commission, *Shaping Europe's Digital Future* https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_en

¹⁷³ European Commission, *Shaping Europe's Digital Future* https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_en

¹⁷⁴ Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608116887159&uri=COM%3A2020%3A842%3AFIN>

¹⁷⁵ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services>

The *Digital Markets Act* proposes that the collection of large amounts of data from end-users by digital market gatekeepers should be regulated and transparent to protect the privacy interests of end-users. The Act requires data sharing across platforms to be a voluntary choice by end-users rather than the only available and accessible option. Further, social media and other digital platforms would be required to offer end-users the ability to opt out of processes that require access through a gateway controlled by a single gatekeeper. Under the Act, platforms would be restricted in their the deep profiling of end-users – where large volumes of information about a user are combined. Where profiling processes are in play, users would need to be informed of the profiling’s purpose and impact. Service providers would need to demonstrate the steps taken to ensure user awareness of the use of profiling, and their consent. Platforms will not be able to combine personal data from core platform services with any other service offered by the gatekeeper unless user consent is provided.¹⁷⁶

Other measures may complement this Act; for example, a proposal by the Committee on Economic and Monetary Affairs has called for banning platforms from displaying micro-targeted advertisements that rely on deep profiling.¹⁷⁷

The *Digital Services Act* seeks to improve users' online safety, and better protect their fundamental rights and online anonymity where possible.¹⁷⁸ This proposal tackles core operations of platforms, namely how information is prioritised and presented on its online interface. Significant online platforms (with more than 45 million end-users, or an equivalent of 10% of the European Union population)¹⁷⁹ would be required to ensure recipients are appropriately informed of the information presented to them. The Act defines the responsibilities of digital services providers, specifically online platforms, social media, and online marketplaces. Further, it outlines obligations and procedures to tackle illegal content and disinformation, and offers the opportunity to challenge content moderation decisions. The proposal introduces safeguards protecting fundamental rights, allowing citizens to freely express themselves while maintaining rights to effective remedies, non-discrimination, the rights of the child, and personal data and privacy protection.¹⁸⁰ Platforms would be required to ensure recipients of online advertisements know what information has been used to personalise advertising content, and platforms will have to obtain user consent prior to processing data for targeted advertising.

¹⁷⁶ Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608116887159&uri=COM%3A2020%3A842%3AFIN>

¹⁷⁷ European Parliament resolution of 18 June 2020 on competition policy – annual report 2019 (2019/2131(INI)) https://www.europarl.europa.eu/doceo/document/TA-9-2020-0158_EN.html

¹⁷⁸ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services>

¹⁷⁹ Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608116887159&uri=COM%3A2020%3A842%3AFIN>

¹⁸⁰ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services>

These regulated platforms would also have to ensure public access to repositories of advertisements displayed on their online interfaces, to facilitate supervision and research into emerging risks brought about by the distribution of advertising online. These risks include illegal advertisements or manipulative techniques and disinformation with a real and foreseeable negative impact on public health, public security, civil discourse, political participation and equality.¹⁸¹

The Act would allow for protocols in response to extraordinary situations affecting public security or public health. In such a crisis, the Commission would initiate the drafting of a protocol to coordinate a rapid, collective, and cross-border response in the online environment.⁵ Further, the Commission encourages these platforms to participate in the drafting, testing and application of crisis protocols. This may include displaying prominent information on the crisis provided by Member States authorities on the Union level, initiating or adjusting cooperation between online platforms, facilitating a faster response to removing access to illegal or harmful content, and termination of services providing such content to their recipients. The Commission is aware of a possible conflict between the protection of human rights and freedom of expression and hence seeks collaboration in the drafting and implementation of appropriate protocols.¹⁸²

Platforms would be required to respond without delay and inform the issuing authority of the actions taken when an order is made against a specific item of illegal content. The Digital Services Coordinator is required to transmit the order to all other Digital Services Coordinators. Member States will be required to appoint a Digital Services Coordinator (DSC), an authority to supervise, investigate, and enforce the regulation in the Member State. The DSCs have powers to require information from providers, carry-out on-site inspections, ask any staff member or representative of the providers for explanations regarding investigation cases. They are also required to publish annual reports on their activities. DSC's can impose fines and penalties for failure to comply with the regulation.¹⁸³ DSCs from each member State will form the European Board of Directors.¹⁸⁴

The European Union has also developed an initiative, EU4Digital, to extend the Digital Single Market to the Eastern Partnership (which includes Armenia, Azerbaijan, Belarus, Georgia, Republic of Moldova, and Ukraine). EU4Digital promotes key digital economy and society concerns in line with EU norms and practices.¹⁸⁵ It supports the reduction of roaming tariffs, developing high-speed broadband to boost economies and expand e-

¹⁸¹ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services>

¹⁸² Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services>

¹⁸³ Penalties shall not exceed 6% of the annual income or turnover of the provider for the failure to comply with regulations, and no more than 1% of the annual profit or turnover if the provider is providing misleading information to the coordinators.

¹⁸⁴ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services>

¹⁸⁵ EU4Digital Website, <https://eufordigital.eu/>

services, coordinated cyber-security, and harmonizing digital frameworks in areas ranging from logistics to health, skills, and jobs creation in the digital industry.¹⁸⁶

Concern has been voiced regarding the effect of the proposed legislation on digital services and digital markets. Specifically, there is concern that it could allow repressive governments to suppress speech.¹⁸⁷ Amnesty International welcomed the *Digital Services Act*, although stated that it does not go far enough to protect people's human rights.¹⁸⁸ Their position is that companies should not be granted more responsibility regarding the adjudication of the legality of the content and should not bear liability for failure to remove it if they are not aware of its presence. Amnesty calls for stricter limits in the targeting of online advertisements and that deep-profiling should be an opt-in option rather than an opt-out.¹⁸⁹

Civil societies around the globe formed *Digital Services Act Human Rights Alliance* in May 2021, calling on the EU to focus on the protection of human rights.¹⁹⁰ Recognising that the European Union approach will have global influence, they call for some changes to the *Digital Services Act*. Their recommendations include not legally imposing automated content moderation tools and to focus on protecting human rights, especially on very large online platforms. In their joint statement, they state that decreasing the response time for removal requests will result in more pressure on small platforms and the removal of legitimate content.¹⁹¹ They also urge legislators to prevent public authorities from becoming trusted flaggers (entities approved by the EU as having expertise and competence in identifying illegal content) and that the conditions for instituting trusted flagger status should not be determined solely by private platforms.¹⁹²

Digital Education

The European Union has drafted the *Digital Education Action Plan 2021-2027* to support the sustainable and effective modernisation of education and training systems. The plan has two key priorities, to foster the development of a high-performing digital education

¹⁸⁶ EU4Digital Website, <https://eufordigital.eu/>

¹⁸⁷ See, eg, European plans to regulate internet will have major impacts on civic space at home and abroad' (10 May 2021), *OpenGlobalRights*, <https://www.openglobalrights.org/european-plans-to-regulate-internet-will-have-major-impacts-on-civic-space-at-home-and-abroad/>

¹⁸⁸ Amnesty International, *Amnesty International Position of the proposals for a Digital Services Act and a Digital Markets Act* (2020) <https://www.amnesty.eu/news/amnesty-international-position-on-the-proposals-for-a-digital-services-act-and-a-digital-markets-act/>

¹⁸⁹ Amnesty International, *Position Paper on the Proposal for a Digital Services Act and Digital Markets Act* (2021) https://www.amnesty.eu/wp-content/uploads/2021/04/Amnesty-International-Position-Paper-Digital-Services-Act-Package_March2021_Updated.pdf

¹⁹⁰ Acces Now, *Digital Services Act: Bad decisions can lead to global consequences* (22 October 2021) <https://www.scoop.co.nz/stories/WO2110/S00210/digital-services-act-bad-decisions-can-lead-to-global-consequences.htm>

¹⁹¹ Digital Services Act Human Rights Alliance, *Joint Statement of the Digital Services Act Human Rights Alliance* (21 October 2021) https://www.eff.org/files/2021/10/21/digital_services_act_human_rights_alliance_statement_upd.pdf

¹⁹² Ibid.

eco-system (priority 1), and to enhance digital skills and competencies for digital transformation (priority 2).¹⁹³

The plan encompasses a variety of actions which target key challenges of the digital era, including:

- Creating the *European Digital Content Framework* by 2023, with an understanding of the underlying 'supply side' and 'demand side' issues relating to digital education content in response to the problems algorithms pose for educational resources (Action 3);
- Creating common guidelines for teachers and educators to foster digital literacy and tackle disinformation through education and training, with a planned finalisation date of September 2022 (Action 7);
- Developing a European Digital Skills Certificate (EDSC), which will enhance the transparency and mutual recognition of digital skills certification by governments, employers, and other stakeholders across Europe (Action 9). The EDSC is expected to be fully operational 2023;
- Improving the provision of digital skills in education and training, to empower Europeans with basic and advanced digital skills (Action 10). The aim is for at least 65% of Europeans to have at least basic digital skills by 2025, with the proposal to be finalised by the end of 2022;
- Collecting cross-national data on student digital skills and reducing the share of low-achieving 13-14 year olds in computer and information literacy to below 15% by 2030 (Action 11);
- Providing Digital Opportunity Traineeships, to provide higher education students the opportunity to gain professional experience in digital fields demanded by the labour market (Action 12).¹⁹⁴

The EU Commission has also funded the Digital Wellbeing Educators Project, which focuses on increasing lecturers and teachers' capacity to integrate the promotion of students digital wellbeing into education.¹⁹⁵ The project helps students critically assess the media they consume and create, to become responsible and confident digital citizens.¹⁹⁶ The project provides resources to introduce practical strategies on digital competency. Participating institutions, universities, and colleges have strengthened their commitment to support their staff and students digital wellbeing. Further, an App has been developed with short courses on digital wellbeing for students. Feedback on the App noted its assistance in assessing social media usage and improving critical thinking skills.¹⁹⁷

¹⁹³ European Commission, *Digital Education Action Plan 2021-2027* https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en

¹⁹⁴ European Commission, *Digital Education Action Plan 2021-2027* https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en

¹⁹⁵ European Commission, *Digital Wellbeing Educators Project* <https://ec.europa.eu/programmes/erasmus-plus/projects/eplu-project-details/#project/2018-1-UK01-KA203-048214>

¹⁹⁶ European Commission, *Joint Communication to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions, Action Plan Against Disinformation* (2018) https://eeas.europa.eu/sites/default/files/action_plan_against_disinformation.pdf

¹⁹⁷ European Commission, *Digital Wellbeing Educators Project* <https://ec.europa.eu/programmes/erasmus-plus/projects/eplu-project-details/#project/2018-1-UK01-KA203-048214>

Disinformation

In 2018 the EU Commission issued an Action Plan Against Disinformation (the Plan) in response to the threat of online disinformation, specifically by Russia. The Plan outlines the EU's coordinated response to disinformation, and is based on the cooperation of EU institutions, Member States, civil society, and the private sector.¹⁹⁸

The Plan aims to improve EU capabilities to detect, analyse and expose disinformation by: using data mining; and increasing the number of analysts; and investing in relevant analytical tools. Further, the plan centres on strengthening joint responses to disinformation through the creation of the *Rapid Alert System* which provides alerts on disinformation in real-time. This system would improve information-sharing and awareness among the Member States. The plan also raises awareness and increases the level of digital literacy of platform users, through campaigns including the European Week of Media Literacy. These campaigns encourage independent quality journalism, promote media freedom, and support pluralism.

The European External Action Service (the EU's diplomatic wing), created the East StratCom Task Force to support the plan. The Task Force's mandate is to expose disinformation in countries within and neighbouring the EU, the three priority regions being to the EU's East, South, and Western Balkans.

The Plan mobilises the private sector to tackle disinformation through the Code of Practice. This Code is the first time globally that industry has voluntarily agreed to self-regulatory standards to combat disinformation.¹⁹⁹ Signatories to the Code committed to respond to disinformation, invest in detection technologies, and address verifiable false or misleading information.²⁰⁰ Initial signatories included Facebook, Google, Twitter, and Mozilla.²⁰¹ The platforms have agreed to invest in products, technologies and programs to assist users make informed decisions when they encounter online news that might be false; invest in technological means to prioritise relevant, authentic and authoritative information in search and feed features; and invest in features making diverse perspectives about public interest topics easier to locate.²⁰² The implementation assessment on the Code was generally positive, and indicated that the Code had set the foundation for further activities. While the Code has improved awareness on disinformation, and led to the implementation of policies by platforms to increase

¹⁹⁸ European Commission, *Joint Communication to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions, Action Plan Against Disinformation* (2018) https://eeas.europa.eu/sites/default/files/action_plan_against_disinformation.pdf

¹⁹⁹ European Commission, *Code of Practice on Disinformation* <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>

²⁰⁰ European Commission, *Code of Practice on Disinformation* <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>

²⁰¹ European Commission, *Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and The Committee of the Regions Tackling online disinformation: a European Approach* (2018) <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN>

²⁰² European Commission, *Joint Communication to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions, Action Plan Against Disinformation* (2018) https://eeas.europa.eu/sites/default/files/action_plan_against_disinformation.pdf

collaboration with researchers and fact-checkers, the voluntary nature of the Code and lack of sufficient communication between Signatories and researchers poses issues for enforceability and effectiveness.²⁰³

A multi-disciplinary committee, the Social Observatory for Disinformation and Social Media Analysis (SOMA) has been established to support the plan.²⁰⁴ Through SOMA, fact-checkers collaborate with tech specialists, data collection experts, and applied machine learning to exploit existing verification platforms and provide experts with necessary resources and tools to fight disinformation. SOMA is negotiating with major social media networks to access their content and data. SOMA also conducts investigations into disinformation narratives in the European Union.²⁰⁵

The Plan is complemented by other EU efforts to regulate online platforms. In 2021, the European Council implemented a regulation to address the online dissemination of terrorist content. This regulation gives the competent authority of each Member State the power to issue an order to hosting service providers to remove terrorist content or disable access to terrorist content in all Member States. The hosting provider must remove content or disable access in all member States within one hour of receipt of the removal order.²⁰⁶

EU Member States

In parallel with the EU actions, Member States have taken legislative and non-legislative measures to tackle disinformation and social media concerns more broadly.²⁰⁷

Germany

The *Network Enforcement Act* (NetzDG) was implemented in 2017, to combat hate speech and misleading information on social media. While the legislation did not enforce new requirements for social media platforms, it imposed large fines for noncompliance with existing legal obligations. Under NetzDG, platforms are required to respond to complaints of unlawful content and determine whether the content is illegal in accordance with the German Criminal Code. If illegal, the content must be removed within 24 hours, or in some cases, within seven days. Illegal content may include the incitement of violence or hatred against national, religious, ethnic, or racial groups. Penalties for non-compliance include fines of up to €50 million (\$79 million AUD) per violation. On 28 June 2021, NetzDG was

²⁰³ Iva Plasilova et al., 'Study for the assessment of the implementation of the Code of Practice on Disinformation', (2020), European Commission <https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation>

²⁰⁴ European Commission, *Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and The Committee of the Regions Tackling online disinformation: a European Approach* (2018)

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN>

²⁰⁵ SOMA, <https://www.disinfobservatory.org/>

²⁰⁶ Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online (Text with EEA relevance) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2021.172.01.0079.01.ENG

²⁰⁷ Iva Plasilova et al., 'Study for the assessment of the implementation of the Code of Practice on Disinformation', (2020), European Commission <https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation>

amended, to increase the content and comparability of social media providers' transparency reports and improve the user-friendliness of reporting channels for complaints. The amendment introduced an appeals procedure for measures taken by the social network platform.²⁰⁸

Austria

The *Communications Platforms Act* came into force in 2021, in response to an increase in hate speech, harassment, and the spreading of false information on online platforms. Under the legislation, providers (domestic and foreign providers of for-profit communication platforms that have more than 100,000 users in Austria or revenues exceeding EUR 500,000) are required to establish effective and transparent procedures for reporting and deleting. Deletion must occur within 24 hours if the illegality is "obvious to a legal layman", or within 7 days if a detailed examination is necessary. Platforms are required to store deleted postings for at least ten weeks for any possible prosecution. Providers are required to submit an annual review on illegal content handling, or a quarterly review for platforms exceeding one million registered users. If non-compliant with the legislation, fines of up to EUR 10 million can be imposed on the platform, and fines of up to EUR 1 million can be imposed on members of the managing board.²⁰⁹

The EU Commission noted the Act may impede the freedom to provide services and may lead to unnecessary additional costs and administrative burdens. The EU Commission also questioned why Austria implemented its own legislation while the Digital Services Act is being formulated. Nevertheless, the Austrian government's position is that the urgency of the issue required immediate implementation of national measures, before EU wide regulations are implemented.²¹⁰

Sweden

Sweden has sought to increase the misinformation literacy of its citizens, with the Civil Contingencies Agency (MSB) releasing the *Countering Information Influence Activities: A Handbook for Communicators* in 2018. The publication provides communicators working in public administration with resources in the event of an actual or anticipated disinformation influence campaign.²¹¹ The Swedish Minister for Digital Development Anders Ygeman noted in early 2020 that he wanted to introduce legislation to increase accountability for social media platforms in removing offensive content.²¹² Additionally,

²⁰⁸ *Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)* https://www.bmjbv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?jsessionid=AD99C47B2608D

²⁰⁹ European Commission, *Draft Federal Act on measures to protect users on communication platforms (Communication Platforms Act)* <https://ec.europa.eu/growth/tools-databases/tris/en/search/?trisaction=search.detail&year=2020&num=544>

²¹⁰ Gabriela Staber, *Communication platforms face new obligations and high fines in Austria* <https://www.lexology.com/library/detail.aspx?q=fcf46df4-4694-4f10-b11b-67564a824470>

²¹¹ Swedish Civil Contingencies Agency: *Countering information influence activities: A handbook for communicators* (2018) <https://www.msb.se/RibData/Filer/pdf/28698.pdf>

²¹² Fanny Svärd, 'Government wants to legislate against illegal content in social media', (18 February 2020), *Radio Sverige* <https://sverigesradio.se/artikel/7410502>

Sweden has also advocated for a tougher stance in the EU against platforms such as Google and Facebook and fraudulent and illegal material posted on their platforms.²¹³

Spain

The Spanish Government's policy – *Spain Digital 2025* – includes nearly 50 measures to promote the country's digital transformation process over 5 years, aligned to the EU Digital Strategy. *Spain Digital 2025* targets digital connectivity, cybersecurity and strengthening the digital skills of the general public.²¹⁴ In July 2021, the Spanish Government adopted the *Charter on Digital Rights*, fulfilling a mandate in *Spain Digital 2025*, to reinforce and extend citizens' rights, generate certainty in the new digital age and increase people's confidence in the face of the disruption that technology represents. The Charter includes rights on freedom, the right to identity in the digital environment, data protection, pseudonymisation, the right to not be traced and profiled, the right to cybersecurity and to digital inheritance.²¹⁵

Denmark

Denmark's primary focus on social media is on ensuring clear guidelines around product advertisement on platforms. As businesses are relying more on "influencers" and "bloggers" to sell their products online, clear and transparent advertising has become a priority.²¹⁶ Section 4 of the *Danish Marketing Practices Act* dictates that advertising through social media should be clearly distinguishable, to ensure target groups recognise content as an advertisement and are able to judge the content accordingly.²¹⁷

In March 2021, Danish lawmakers proposed legislation to make tech giants, such as Facebook and Google, pay Danish media for using content on their platforms.²¹⁸ The legislation was implemented in June 2021 and builds on an EU directive giving individual media outlets the right to agree deals with tech giants.²¹⁹ The legislation was inspired by the Australian *News Media and Digital Platforms Mandatory Bargaining Code*.

Denmark has also adopted an EU Code of Practice on Disinformation that applies to Denmark as an EU Member State. Further, since 2018 the Danish Government has

²¹³ TricksFast, 'Ygeman: Indecent of Google and Facebook for not stopping posts', (12 January 2020), <https://tricksfast.com/sweden/ygeman-indecent-of-google-and-facebook-for-not-stopping-posts/>

²¹⁴ Government of Spain, *Digital Spain 2025*

https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/210204_Digital_Spain_2025.pdf

²¹⁵ Carta Derechos Digitales https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf; Government of Spain, 'The Government adopts the Digital Rights Charter to articulate a reference framework to guarantee citizens' rights in the new digital age' (14 July 2021)

https://www.lamoncloa.gob.es/lang/en/gobierno/news/Paginas/2021/20210713_rights-charter.aspx

²¹⁶ *Danish Marketing Practices Act* s 4. 'Covert advertising'. <https://www.consumerombudsman.dk/marketing-practices-act/covert-advertising/>

²¹⁷ *Danish Marketing Practices Act* s 4. 'Covert advertising'. <https://www.consumerombudsman.dk/marketing-practices-act/covert-advertising/>

²¹⁸ Ritzau, 'Denmark proposes new law to make Facebook pay for news and music', (26 March 2021), <https://www.thelocal.dk/20210326/denmark-proposes-new-law-to-make-facebook-pay-for-news-and-music/>

²¹⁹ European Parliament, *Agreement reached on digital copyright rules*, (13 February 2019), <https://www.europarl.europa.eu/news/en/press-room/20190212IPR26152/agreement-reached-on-digital-copyright-rules>

increased media literacy among government and defence employees to combat disinformation.²²⁰ Digital literacy and technology is also built into early years educational settings. Most kindergartens use digital technologies in their pedagogical practices as well.²²¹ Screen time and the use of digital technologies continue to be a paradox as parents are torn between their inclination to limit their child's screen use and to ensure children are digitally literate.²²²

France

Legislation was passed in December 2018 cracking down on the dissemination of false information. The law allows election candidates to sue for the removal of contested news reports during election periods, and requires social media platforms to disclose the source of funding for sponsored content. The legislation imposes a quick-response judicial review of potentially "manipulative" information shared during electoral periods.²²³ Online platforms are required to establish a mechanism for users to flag false information in an easily accessible and visible way. The legislation also outlines that French public schools should teach children how to navigate online information. Critics have noted that it could jeopardise democracy and censor the press.²²⁴

Other Member States

Other member states have implemented a range of measures to tackle disinformation, including:

- Lithuania: implemented the *Law on Provision on Information to the Public*, making it illegal to spread disinformation and enabling the Radio and Television Commission to block channels spreading disinformation.²²⁵ Lithuania has also implemented European Initiatives such as "Debunk.eu", that unites the media, society and the state to fight against disinformation;²²⁶
- Greece: established the website Ellinica Hoaxes in 2013 to debunk disinformation items;²²⁷
- Sweden: In collaboration with researchers from Lund University, the Swedish Civil Contingencies Agency published a handbook describing different techniques used by malicious actors and the methods they can use to spread disinformation;²²⁸

²²⁰ Council of Europe, *Mapping of Media Literacy Practices and Actions in EU-28*, (2016), <https://rm.coe.int/1680783500>.

²²¹ Media & Learning, 'Early Years Education and Digital media in Denmark', (1 October 2020), <https://media-and-learning.eu/type/featured-articles/early-years-education-and-digital-media-in-denmark/>

²²² Media & Learning, 'Early Years Education and Digital media in Denmark', (1 October 2020), <https://media-and-learning.eu/type/featured-articles/early-years-education-and-digital-media-in-denmark/>

²²³ Politico, 'French Parliament passes law against 'fake news'', (4 July 2018), <https://www.politico.eu/article/french-parliament-passes-law-against-fake-news/>

²²⁴ Michael-Ross Fiorentino, 'France passes controversial 'fake news' law', (22 November 2018), [EuroNews https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law](https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law)

²²⁵ Republic of Lithuania, *Law on Provision of Information to the Public* https://www.legislationline.org/download/id/5542/file/Lithuania_law_provision_information_public_am2006_en.pdf

²²⁶ <https://debunk.eu>

²²⁷ Iva Plasilova et al., 'Study for the assessment of the implementation of the Code of Practice on Disinformation', (2020), *European Commission* <https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation>

²²⁸ Iva Plasilova et al., 'Study for the assessment of the implementation of the Code of Practice on Disinformation', (2020), *European Commission* <https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation>

- Latvia: implemented school workshops educating teachers and students how to differentiate fact from fiction;²²⁹
- Luxemburg: the BEE SECURE initiative includes *"Share Respect – Stop Online Hate Speech"*, within which files about false online information are published. The initiative also includes advice to parents on coping with their children's media consumption;²³⁰
- Cyprus: designed and implemented media literacy programs to educate high school students on how to recognise disinformation;
- Finland: invested in strengthening media literacy through partnerships between the schools and fact-checker agencies;²³¹
- The Netherlands: launched a public awareness campaign aimed at informing people about disinformation.

Norway

In June 2021, in response to increasing concerns around the impact of social media on the Norwegian population's mental health and body image insecurity, the Norwegian Government passed legislation requiring content creators to disclose when they have retouched or added a filter to photos.²³²

United Kingdom

In the UK, there is increasing concern about online activity and harmful content. Technology firms have been accused of not addressing online abuse, with soccer clubs and other sporting authorities boycotting social media platforms in April 2021 to shine a spotlight on the increasing problem.²³³ British schoolgirl Molly Russell's suicide in 2017 after viewing graphic self-harm images on Instagram ignited the calls for regulation.²³⁴ There have also been calls to enact 'David's law' to clamp down on social media abuse of public figures and end anonymity online after the murder of M.P. Sir David Amess in October 2021,²³⁵ which followed on from significant online threats and abuse directed

²²⁹ Latvian Academy of Culture, *Professional Development Conference for teachers* https://lka.edu.lv/en/international-cooperation/international-projects/nordplus-projects/film-and-media-education/professional-development-conference-teachers/?edit_off

²³⁰ Iva Plasilova et al., 'Study for the assessment of the implementation of the Code of Practice on Disinformation', (2020), *European Commission* <https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation>

²³¹ Iva Plasilova et al., 'Study for the assessment of the implementation of the Code of Practice on Disinformation', (2020), *European Commission* <https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation>

²³² Vedtak til lov om endringer i markedsføringsloven mv. (merking av retusjert reklame) (Legislation amending the Swedish Marketing Practices Act) <https://www.stortinget.no/globalassets/pdf/lovvedtak/2020-2021/vedtak-202021-146.pdf>

²³³ Michael Holden, 'UK unveils law to fine social media firms which fail to remove online abuse', (12 May 2021), *Reuters* <https://www.reuters.com/technology/uk-unveils-law-fine-social-media-firms-which-fail-remove-online-abuse-2021-05-11/>

²³⁴ BBC News, 'Molly Russell social media material 'too difficult to look at'', (26 September 2020), <https://www.bbc.com/news/uk-england-london-54307976>

²³⁵ Jessica Elgot, 'PM urged to enact 'David's law' against social media abuse after Amess's death', (19 October 2021), *The Guardian* <https://www.theguardian.com/uk-news/2021/oct/18/pm-urged-to-enact-davids-law-against-social-media-abuse-after-amesss-death>

towards politicians in recent years,²³⁶ including the murder of MP Jo Cox in 2016.²³⁷ The proposed reforms to the regulation of social media and online safety take a similar approach to the European Union's proposals, looking simultaneously at human welfare and free speech, treating social media platforms as public environments, not as publishers.²³⁸

Online Harms and Online Safety

The UK Government released an *Online Harms White Paper* (the Paper) in April 2019, claiming that the existing patchwork of regulation and voluntary initiatives had not gone far or fast enough to keep online users safe. The Paper proposed a single regulatory framework to tackle the issue, which centres on a duty of care for internet companies, including social media platforms. Compliance with duty of care obligations was to be overseen and enforced by an independent regulator.²³⁹ The Paper received a varied reaction, including concerns that harms were insufficiently defined and it may threaten freedom of expression.²⁴⁰ The UK Government consulted on the White Paper, and subsequently a draft *Online Safety Bill* was included in the Queen's speech of 11 May 2021, and published the following day. A Joint Committee has been established to consider the draft legislation, with a report deadline of 10 December 2021.²⁴¹

The draft Bill would impose duties of care on providers of online content-sharing platforms and search services, to address illegal content on their services. This illegal content includes terrorism offences, child sexual exploitation and abuse offences, offences directed at an individual as the victim, and offences set out in secondary legislation.²⁴² Companies within the scope of the legislation would need to take "robust action to tackle illegal abuse, including swift and effective action against hate crimes, harassment and threats directed at individuals and keep their promises to users about their standards". The definition of harm would be that which may cause significant adverse physical or psychological impact on individuals.²⁴³

²³⁶ Jennifer Scott, 'Can Online Safety Bill tackle social media abuse of MPs?', (20 October 2021), *BBC News* <https://www.bbc.com/news/uk-politics-58958244>

²³⁷ BBC News, 'Labour MP Jo Cox 'murdered for political cause'', (14 November 2016), *BBC News* <https://www.bbc.com/news/uk-37978582>

²³⁸ Parmy Olsen, 'The appeal of British efforts to keep social media under watch', (3 November 2021), *Mint* <https://www.livemint.com/opinion/columns/the-appeal-of-british-efforts-to-keep-social-media-under-watch-11635872121309.html>

²³⁹ UK Government. (2020). *Fact sheet – Online Harms Full Government Response*. <https://researchbriefings.files.parliament.uk/documents/CBP-8743/CBP-8743.pdf>

²⁴⁰ Claudine Tinsman, 'Will the government's online safety laws for social media come at the cost of free speech?', (24 December 2020), *The Conversation* <https://theconversation.com/will-the-governments-online-safety-laws-for-social-media-come-at-the-cost-of-free-speech-152352>

²⁴¹ UK Government. (2020). *Fact sheet – Online Harms Full Government Response*. <https://researchbriefings.files.parliament.uk/documents/CBP-8743/CBP-8743.pdf>

²⁴² Clause 41(3)(c) of the *Online Harms Bill*

²⁴³ Part 1 contains definitions of the services to which the Bill would apply.

Part 2 sets out the duties of care that would apply to providers of user-to-user and search services – i.e. duties to undertake risk assessments, and duties with regards to content that is illegal, harmful to children and harmful to adults; Part 4 sets out Ofcom's powers and duties, including duties to carry out risk assessments and to maintain a register of categories of services. Part 4 also establishes Ofcom's functions and powers with respect to the use of technology in relation to terrorism content and child sexual exploitation and abuse content, information-gathering, enforcement, research, and media literacy; Part 5 provides for the grounds and avenues for appeals against Ofcom's decisions, and for designated bodies to make super complaints.

Ofcom, the UK's independent communications regulator would be appointed as the online harms regulator, their remit broadening to include setting codes of practice, establishing a transparency, trust and accountability framework, and requiring all in-scope companies to have effective and accessible mechanisms for users to report concerns. Ofcom's powers would include the ability to fine companies up to £18 million or 10% of annual global turnover (whichever is higher) and have the power to block access to sites if they are non-compliant. However, social media platforms will set their own definitions of risk assessment, which may lend itself to less diligent reporting.²⁴⁴

The proposed legislation has been heavily critiqued by civil liberties organisations, however it was welcomed by children's safety organisations.²⁴⁵ Critique centres on freedom of expression and privacy concerns, including private messaging, legal but harmful content and journalistic material.²⁴⁶

Counter Terrorism

The *Counter-Terrorism and Border Security Act 2019* amended the *Terrorism Act 2000* in response to attacks in London and Manchester in 2017, including provisions related to online activity for individuals.²⁴⁷ The legislation states it is an offense to view terrorist material over the Internet,²⁴⁸ and individuals face up to 15 years in prison for viewing or accessing material that is useful or likely to be useful in preparing or committing a terrorist act, even if there is no demonstrated intent to commit such acts.²⁴⁹ The legislation includes exceptions for journalists and academics accessing materials in the course of their work.²⁵⁰ Critique of the legislation centres on threats to freedom of expression.

Prosecution of Digital Offences

The Crown Prosecution Service publishes guidelines for the prosecution of crimes committed by social media users.²⁵¹ The guidelines inform decisions on whether criminal charges should be pursued against individual social media users for a range of offences. These guidelines were updated in 2014 to include digital harassment offences committed under the *Sexual Offences Act 2003*. The guidelines were updated in 2016 to include more abusive online behaviours, including online harassment, trolling, threats, disclosure of

²⁴⁴ UK Government. (2020). *Fact sheet – Online Harms Full Government Response*.

<https://researchbriefings.files.parliament.uk/documents/CBP-8743/CBP-8743.pdf>

²⁴⁵ Alex Hern, 'Online safety bill 'a recipe for censorship', say campaigners', (13 May 2021), *The Guardian*

<https://www.theguardian.com/media/2021/may/12/uk-to-require-social-media-to-protect-democratically-important-content>

²⁴⁶ John Woodhouse, 'Regulating online harms', (12 August 2021), *House of Commons Library*

<https://researchbriefings.files.parliament.uk/documents/CBP-8743/CBP-8743.pdf>; Heather Burns, 'Online harms plans threaten the future of freedom of expression', (15 December 2020), *Open Rights Group*

<https://www.openrightsgroup.org/blog/online-harms-freedom-of-expression-remains-under-threat/>

²⁴⁷ *Counter-Terrorism and Border Security Act 2019* <https://bills.parliament.uk/bills/2255>

²⁴⁸ Amends s58 of the *Terrorism Act 2000*

²⁴⁹ Chapter 2 – Punishment and management of terrorist offenders. S7-11 Sentencing

²⁵⁰ Freedom House, *United Kingdom* <https://freedomhouse.org/country/united-kingdom/freedom-net/2021>

²⁵¹ The Crown Prosecution Service, *Guidelines on prosecuting cases involving communications sent via social media* http://www.cps.gov.uk/legal/a_to_c/communications_sent_via_social_media/

sexual images without consent, grooming, stalking online, and online mobbing.²⁵² There have been recent calls for closer monitoring.

Cyberbullying

Cyberbullying is not explicitly covered by UK regulation, however, there are several acts under which it may be deemed a criminal offence. Under section 127 *Communications Act 2003*, it is an offence to send via any electronic communication network a message deemed grossly offensive or of an indecent or menacing manner.²⁵³ Under section 1 of the *Malicious Communications Act 1988*²⁵⁴ it is an offence to send communication that is indecent or grossly offensive, for the purpose of causing distress to the recipient. The Act encompasses threats and information which are false.

New Zealand

The evolution of social media has resulted in increasing potential for exposure to harmful content, explicitly evident by the livestreaming of the Christchurch terror attack. The existing New Zealand regulatory system was designed around analogue publication such as books and free-to-air TV, and does not have the capacity to respond to digital media types, including content made available online.²⁵⁵ Further, the Christchurch Call, the international effort to curb violent extremism and terrorist content spread through tech companies' algorithms, was initiated by NZ Prime Minister Jacinda Ardern and French President Emmanuel Macron in May 2021.

Harmful Content

A comprehensive review of content regulation was announced by the Hon Jan Tinetti, Minister of Internal Affairs, on 10 June 2021, to design and create a modern, flexible and coherent regulatory framework to mitigate the harmful impacts of content. Content includes any publicly available communicated material (video, audio, images and text), regardless of how it is communicated. Specifically, harmful content ranges from child sexual exploitation material, adult content that children can access and violent extremist content.²⁵⁶ The review has a broad scope, encompassing areas not covered by existing legislation such as misinformation and disinformation, in addition to broadcasting and advertising standards, the *Harmful Digital Communications Act*, the classification system

²⁵² Crown Prosecution Service, *CPS publishes new social media guidance and launches Hate Crime consultation* (10 October 2016)

https://web.archive.org/web/20161013201133/www.cps.gov.uk/news/latest_news/cps_publishes_new_social_media_guidance_and_launches_hate_crime_consultation/

²⁵³ *Communications Act 2003* <https://www.legislation.gov.uk/ukpga/2003/21/section/127>

²⁵⁴ *Malicious Communications Act 1988* <https://www.legislation.gov.uk/ukpga/1988/27/section/1>

²⁵⁵ The Department of Internal Affairs, *The Content Regulatory System Review* <https://www.dia.govt.nz/media-and-online-content-regulation>

²⁵⁶ The Department of Internal Affairs, *The Content Regulatory System Review* <https://www.dia.govt.nz/media-and-online-content-regulation>; Hon Jan Tinetti, *Govt acts to protect NZers from harmful content* (10 June 2021)

<https://www.beehive.govt.nz/release/govt-acts-protect-nz-ers-harmful-content>; Hon Jan Tinetti Proactive release of Cabinet material about the initiation of the media content regulatory review (2 July 2021) [https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/\\$file/Cabinet-material-about-the-initiation-of-the-media-content-regulatory-review.pdf](https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/$file/Cabinet-material-about-the-initiation-of-the-media-content-regulatory-review.pdf)

and Chief Censor's office. The consultation is twofold, with targeted stakeholder engagement in mid-to-late 2021, and public consultation anticipated for early 2022.²⁵⁷

The *Harmful Digital Communications Act 2015* assists people dealing with serious or repeated harmful digital communications and provides 10 communication principles that guide how to communicate online. The Act covers any harmful digital communications which include racist, sexist and religiously intolerant comments, cyberbullying and comments about disabilities or sexual orientation.²⁵⁸ Netsafe has responsibility to resolve reports on alleged breaches, however they are not an enforcement agency. The District Court handles cases of harmful digital communications that Netsafe has not been able to resolve. Criminal penalties include a fine of up to \$50,000 for an individual or up to \$200,000 for a body corporate, or up to two years jail for posting or sending a digital communication with intent to cause harm.²⁵⁹

Canada

Prime Minister Justin Trudeau secured a third election victory in September 2021 after calling a snap election, resulting in a minority government and necessitating a negotiated position with smaller parties to govern and pass legislation.²⁶⁰ The dissolution of Parliament on 15 August 2021, which paved the way for the snap election, put a pause on several Bills to regulate social media platforms, that were considered both ambitious and controversial.²⁶¹ As the Government commences its third term, and as a minority government, the Canadian regulatory space will likely change.

Proposed Legislation

The Canadian Government's proposed legislative and regulatory framework creates rules for how social media platforms and other online services must address harmful content.²⁶² The framework sets out: the entities subject to the new rules; the types of harmful content that would be regulated; new rules and obligations for regulated entities; and two new regulatory bodies and an Advisory Board to administer and oversee the new

²⁵⁷ The Department of Internal Affairs, *The Content Regulatory System Review* <https://www.dia.govt.nz/media-and-online-content-regulation>; Hon Jan Tinetti, *Govt acts to protect NZers from harmful content* (10 June 2021) <https://www.beehive.govt.nz/release/govt-acts-protect-nzers-harmful-content>; Hon Jan Tinetti Proactive release of Cabinet material about the initiation of the media content regulatory review (2 July 2021) [https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/\\$file/Cabinet-material-about-the-initiation-of-the-media-content-regulatory-review.pdf](https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/$file/Cabinet-material-about-the-initiation-of-the-media-content-regulatory-review.pdf)

²⁵⁸ Harmful Digital Communications Act 2015 <https://www.legislation.govt.nz/act/public/2015/0063/latest/whole.html>; <https://www.netsafe.org.nz/what-is-the-hdca/>

²⁵⁹ Netsafe, 'What is the NDCA?', (1 September 2021), *Netsafe* <https://www.legislation.govt.nz/act/public/2015/0063/latest/whole.html>; <https://www.netsafe.org.nz/what-is-the-hdca/>

²⁶⁰ Leylan Cecco, 'Justin Trudeau secures a third victory in an election 'nobody wanted'', (22 September 2021), *The Guardian* <https://www.theguardian.com/world/2021/sep/21/justin-trudeau-wins-third-election-victory>

²⁶¹ Blayne Haggart & Natasha Tusikov, 'Resetting the Debate on Regulating Social Media: Part One', (8 September 2021), *Centre for International Governance Innovation* <https://www.cigionline.org/articles/resetting-the-debate-on-regulating-social-media/>

²⁶² Government of Canada, *Consultation closed: The Government's proposed approach to address harmful content online* <https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content.html>

framework and enforce its rules and obligations.²⁶³ The legislative framework would apply to online communication service providers, which is intended to capture major platforms such as Facebook, Twitter and YouTube, and exclude products and services such as fitness applications or travel review websites. Further, the legislation would target five categories of harmful content: terrorist content, content that incites violence, hate speech, non-consensual sharing of intimate images, and child sexual exploitation content.

While the definitions would draw on existing law, they would be modified to tailor them to a regulatory context. Obligations in the legislation would require regulated entities to do whatever is reasonable and within their power to monitor for harmful content on their platforms, including through automated systems based on algorithms. Once platform users flag content, regulated entities would have to respond by assessing whether it should be made inaccessible in Canada, and if the content meets the legislated definitions, the entity would have to make the content inaccessible within 24 hours.²⁶⁴

The Government presented a discussion guide summarising the approach and a technical paper proposing instructions to inform the legislation for public consultation. Consultation closed on 25 September 2021, and no further information on the “proposed approach” is yet available.

In addition, several proposed Bills were introduced in the 43rd Canadian Parliament, 2nd Session, which ended in August 2021, and were not passed before the election was called. They will need to be reintroduced by the new government, and it is likely they will be amended in some capacity.²⁶⁵ Bill C-36 proposes amendments to several Canadian laws, including the *Canadian Human Rights Act* to make it a discriminatory practice to communicate (or cause the communication of) hate speech on the internet where the hate speech is likely to encourage the vilification of an individual or group of individuals on a prohibited ground of discrimination.²⁶⁶ The Bill would make online hate speech punishable by a fine of up to \$700,000 Canadian dollars, and imprisonment. The Bill was read in Parliament in June 2021 and has not yet been passed.²⁶⁷ The Bill is intended to complement the proposed legislation on combatting online harms.²⁶⁸ There have been concerns voiced regarding censorship and the government's authority to determine what is hate speech.²⁶⁹

²⁶³ Government of Canada, *Consultation closed: The Government's proposed approach to address harmful content online* <https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content.html>

²⁶⁴ Government of Canada, *Discussion guide* <https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html>

²⁶⁵ Bill C-36, An Act to amend the Criminal Code and the Canadian Human Rights Act <https://openparliament.ca/bills/43-2/https://openparliament.ca/bills/43-2/C-36/>; Bills for the 43rd Parliament, 2nd Session <https://openparliament.ca/bills/43-2/>

²⁶⁶ Bill C-36 First Reading <https://parl.ca/DocumentViewer/en/43-2/bill/C-36/first-reading#ID1RB>

²⁶⁷ Bill C-36, An Act to amend the Criminal Code and the Canadian Human Rights Act <https://openparliament.ca/bills/43-2/C-36/>

²⁶⁸ Dale Smith, 'Here's what died on the order paper', (17 August 2021), *National Magazine* <https://nationalmagazine.ca/en-ca/articles/law/hot-topics-in-law/2021/here-s-what-died-on-the-order-paper>

²⁶⁹ Standing for Freedom Center Staff, 'Canada proposes another 'hate speech' law and this one is just as threatening to free speech', (28 June 2021), *Standing for Freedom* <https://www.standingforfreedom.com/2021/06/28/canada-proposes-another-hate-speech-law-and-this-one-is-just-as-threatening-to-free-speech/>

Another proposed piece of legislation is Bill C-10, an Act to amend the *Broadcasting Act*. The Bill passed the House of Commons but did not receive Senate endorsement before the election was called. The proposed legislation would allow the federal government to regulate video content on social media the same way it regulates national broadcasting, through the Canadian Radio-television and Telecommunications Commission, protecting domestic cultural industries as Canadians turn to internet platforms for music and videos.²⁷⁰ This would regulate social media platforms by requiring them to provide information, pay Canadian content contributions and put in place discoverability of Canadian content rules. This proposed legislation was controversial, as some considered that it would implement censorship of social media and control the content Canadians view.²⁷¹

Bill C-11 would enact the *Consumer Privacy Protection Act*, which would update Canadian privacy legislation to address online activities, by protecting the personal information of individuals while recognising the need of organisations to collect, use or disclose personal information in the course of commercial activities.²⁷² This Bill was first read in November 2020, however did not reach committee study.²⁷³

Electoral Integrity

In the lead-up to the 2019 Canadian federal election,²⁷⁴ the Canadian Government released the *Canada Declaration on Electoral Integrity Online*.²⁷⁵ The non-binding declaration establishes a set of common commitments with online platforms to safeguard federal elections from malicious interference and build a healthier online ecosystem. The declaration contains initiatives aimed at enhancing integrity, transparency and authenticity which include assisting users to better understand the sources of information they are viewing; removing fake accounts and inauthentic content on their platforms; and ensuring transparency for regulated political advertising.²⁷⁶ The declaration was updated on 11 August 2021 to include a clearer focus on inauthentic

²⁷⁰ Bill C-10, An Act to amend the Broadcasting Act and to make related and consequential amendments to other Acts <https://parl.ca/DocumentViewer/en/43-2/bill/C-10/third-reading>

²⁷¹ Dale Smith, 'Here's what died on the order paper', (17 August 2021), *National Magazine* <https://nationalmagazine.ca/en-ca/articles/law/hot-topics-in-law/2021/here-s-what-died-on-the-order-paper>; Timothy Gindi, Marty Rabinovitch & Angela Papeo, 'Canada: Bill C-10: The Future of Regulated Canadian Content' (22 October 2021) *Mondaq* <https://www.mondaq.com/canada/social-media/1118340/bill-c-10-the-future-of-regulated-canadian-content>

²⁷² Bill C-11 First Reading <https://parl.ca/DocumentViewer/en/43-2/bill/C-11/first-reading>

²⁷³ Bill C-11: An Act to enact the Consumer Privacy Protection Act and the Personal Information and Data Protection Tribunal Act and to make consequential and related amendments to other Acts <https://www.parl.ca/LegisInfo/en/bill/43-2/C-11>

²⁷⁴ Further in the electoral space, *The Elections Modernization Act* (Bill C-76) received Royal Assent in December 2018. The legislation prohibits the use of foreign funds by third parties for partisan advertising and activities. It heightens transparency measures and clarifies offences related to false statements and foreign interference. <https://www.canada.ca/en/democratic-institutions/news/2019/01/combating-foreign-interference.html>

²⁷⁵ Government of Canada, *Canada Declaration on Electoral Integrity Online* (2021) <https://www.canada.ca/en/democratic-institutions/services/protecting-democracy/declaration-electoral-integrity.html>

²⁷⁶ Joan Bryden, 'Several tech giants sign onto Canadian declaration on electoral integrity', (27 May 2019), *Global News Canada* <https://globalnews.ca/news/5323084/tech-giants-electoral-integrity/>

behaviour online, providing further protection of free expression. Facebook, Google, Microsoft, Twitter and TikTok have endorsed the Declaration.²⁷⁷

China

China has one of the world's most restrictive media environments, utilising censorship to regulate and control information online, in the news and on social media. Chinese authorities blocked platforms including Facebook, Twitter and Google in July 2009 following riots in Xinjiang, to restrict communication among independence activists.²⁷⁸ Commonly referred to as the 'Great Firewall', various methods are utilised to control online expression, including blocking websites, filtering keywords and censoring social media. In 2014 the Cyberspace Administration of China was established as the main body to censor the internet in China.²⁷⁹

At the same time, China has the world's largest social media market, with an estimated 927 million users in 2020.²⁸⁰ While many foreign social media companies are prohibited, Chinese companies are flourishing and include platforms such as Weibo, WeChat and Baidu.²⁸¹ Further, despite the ban, there are several ways to access blocked platforms in China, through virtual private networks (VPN) and proxy websites.²⁸²

Chinese platforms have evolved significantly into 'super apps', to encompass more of what people do online. For example, WeChat facilitates life online, and allows you to message friends, see updates in their feed, as well as take out loans, shop and arrange food delivery. It is anticipated that the major social networks such as Facebook will become super apps and will become increasingly important ways for people to stay connected, bank, shop and entertain themselves.²⁸³

Data protection and privacy

The rhetoric within China is that the prohibition of platforms such as Facebook, Twitter, and YouTube is critical to the protection of citizen's user data, as these platforms aggregate vast amounts of data on their users, which is stored and used. An archived article from MingPao News claims that Facebook is used as a channel for Western

²⁷⁷ Government of Canada, *The Government of Canada updates the Canada Declaration on Electoral Integrity Online* (2021) <https://www.canada.ca/en/democratic-institutions/news/2021/08/the-government-of-canada-updates-the-canada-declaration-on-electoral-integrity-online.html>

²⁷⁸ Techcrunch, 'China blocks access to Twitter, Facebook after riots', (7 July 2009), *Techcrunch* <https://techcrunch.com/2009/07/07/china-blocks-access-to-twitter-facebook-after-riots/>

²⁷⁹ KPMG China, *Overview of China's Cybersecurity Law* (2017) <https://assets.kpmg/content/dam/kpmg/cn/pdf/en/2017/02/overview-of-cybersecurity-law.pdf>

²⁸⁰ Statista, *Number of social network users in China from 2017 to 2020 with a forecast until 2026* (2021) <https://www.statista.com/statistics/277586/number-of-social-network-users-in-china/>

²⁸¹ Christina Lu, 'China's social media explosion' (11 November 2021), *Foreign Policy* <https://foreignpolicy.com/2021/11/11/china-social-media-tech-linkedin-wechat-censorship-privacy-regulation/>

²⁸² Sam Gaskin, 'A guide to digital security for reporters in Asia', (4 October 2019), *Asia Media Centre* <https://www.asiamediacentre.org.nz/features/a-guide-to-digital-security-for-reporters-in-asia/>

²⁸³ Alex Heath, 'The rise of the super app', (1 November 2021), *The Verge* <https://www.theverge.com/22738395/social-media-super-app-facebook-wechat-shopping>

Intelligence services to subvert other countries' regimes.²⁸⁴ Yin Yungong, the Director of News at the Chinese Academy of Social Sciences, has stated that while Chinese citizens love to engage in politics and discussions, there are issues on banned platforms (such as cyberbullying), from which Chinese citizens are being protected. Discussion also centres on how Facebook, Twitter, and Google threaten China's interests and Chinese customers' interests, with Facebook and Twitter considered especially dangerous as they have capacity to disseminate disinformation fast. Facebook's participation in spreading information on the Arab Spring has been used as a critical argument.²⁸⁵

Under China's new *Personal Information Protection Law*, which was announced in August 2021, and came into force on 1 November 2021, users are afforded greater protections from tech companies. The law contains provisions requiring any organisation or individual handling Chinese citizens' personal data to obtain prior consent and minimise data collection.²⁸⁶ The Chinese Government is expected to maintain broad access to the data. The framework comes in response to frustration in the government and Chinese society over online fraud, data theft and data collection by domestic technology giants. Previously, loose rules on data access allowed companies to develop new products and technology quickly, while simultaneously also stimulating a black market for consumer data. The new legislation unifies previously piecemeal law on personal information and protection, in addition to addressing increasingly pertinent issues such as the proliferation of facial recognition, and algorithmic discrimination.²⁸⁷

Social media addiction in children

TikTok (DouYin in China) has implemented rules regarding access to the app for children under 15 years old in response to gaming and media addiction in young people. Children can only use DouYin for 40 minutes a day, between 6 am to 10 pm. The content available is carefully curated, and consists of science experiments, museums and gallery exhibitions, and historical knowledge of China.²⁸⁸ While proof of age is not required upon registration, parents are advised to help children register their age in the app.

Singapore

²⁸⁴ MingPao News, *Chinese Academy of Social Sciences: Facebook becomes a destructive tool* (translated from '社科院：Facebook成颠覆工'), (8 July 2010),

<https://web.archive.org/web/20100711170314/http://hk.news.yahoo.com/article/100707/4/i1ol.html>

²⁸⁵ Tang QiWei, 'Chinese officials made it clear that Facebook will not be allowed to enter Chinese market'

(translated from '中国官员明确表态不允许FACEBOOK进入中国市场'), (12 September 2014), *Radio Free Asia*,

<https://www.rfa.org/mandarin/yataibaodao/meiti/vt-09122014132912.html>

²⁸⁶ Christina Lu, 'China's social media explosion' (11 November 2021), *Foreign Policy*

<https://foreignpolicy.com/2021/11/11/china-social-media-tech-linked-in-wechat-censorship-privacy-regulation/>

²⁸⁷ Eva Xiao, 'China passes one of the world's strictest data-privacy laws', (20 August 2021), *The Wall Street Journal* <https://www.wsj.com/articles/china-passes-one-of-the-worlds-strictest-data-privacy-laws-11629429138>; Li

Yuan, 'Personal-privacy concerns grip China', (31 August 2016), *The Wall Street Journal*

https://www.wsj.com/articles/personal-privacy-concerns-grip-china-1472665341?mod=article_inline

²⁸⁸ TengXun Net, 'DouYin pushes the strictest youth protection rules in history. Daily usage limited to 40 minutes'

(translated from '抖音推史上最严青少年保护措施 每天只能用40分钟'), (19 September 2021), *TengXun Net*

<https://new.qq.com/rain/a/20210919a0ckqc00>

Singapore's approach to social media regulation can be characterised as cautious and paternalistic.²⁸⁹ It has been criticised for curbing free speech and stifling political dissent.²⁹⁰ To better understand the Singapore context, the People's Action Party (PAP) has lead Singapore's parliamentary system since independence in the 1960s. The government allows for some political pluralism; however, it limits freedoms of expression, assembly, and association and restrains the growth of credible opposition parties.²⁹¹

Protection from Online Falsehoods

In response to the spread of deliberate online falsehoods, the *Protection from Online Falsehoods and Manipulation Act* (POFMA), known colloquially as Fake News Law, was passed in the Singapore Parliament on 8 May 2019, and came into effect on 2 October 2019.²⁹² POFMA's purpose is to prevent the communication of false statements and enable measures to counteract effects of this communication, to suppress the financing and promotion of online locations communicating these statements and to enable measures to detect and control this behaviour.²⁹³ Under the legislation, falsehoods are defined as statements of fact that are false or misleading,²⁹⁴ and these are determined by POFMA ministers. These false statements are considered particularly serious to the public interest if they are prejudicial to the security and bilateral relations of Singapore or if they incite feelings of hatred between different groups.²⁹⁵ Individuals in breach of POFMA can be liable for fines up to \$50,000 Singapore dollars and/or imprisonment of up to 5 years, and companies can be liable for up to \$1 million Singapore dollars.²⁹⁶

Critiques of this approach centre on how the law gives authorities excessive and broad powers to crack down on dissenting political views, with the first POFMA actions issued against individuals affiliated with the opposition political party.²⁹⁷ This legislation is significant to the greater region, as the Asian headquarters of Facebook and Twitter are

²⁸⁹ Jing Yi Tay, 'No news is good news, but "fake news" is bad news: A comparative analysis of Singapore's and Australia's measures to combat misinformation on social media', (2021) 33(2) *Singapore Academy of Law Journal*, 600–624

²⁹⁰ Agence France-Presse, 'Singapore passes foreign interference law allowing authorities to block internet content', (5 October 2021), *The Guardian* <https://www.theguardian.com/world/2021/oct/05/singapore-passes-foreign-interference-law-allowing-authorities-to-block-internet-content>

²⁹¹ Freedom House, *Singapore* <https://freedomhouse.org/country/singapore/freedom-net/2021>

²⁹² *Protection from Online Falsehoods and Manipulation Act 2019* <https://sso.agc.gov.sg/Act/POFMA2019?TransactionDate=20191001235959>; Tan Zhi Han, 'Protection from Online Falsehoods and Manipulation Act (POFMA): Regulating Fake News to Maintain Public Trust in Singapore' *Honrad Adenauer Stiftung* https://www.kas.de/documents/288143/11133938/Panorama_Trust_TanZhiHan.pdf/898f786c-229e-b2c6-a4d3-1b1e22128035?t=1608692256696

²⁹³ *Protection from Online Falsehoods and Manipulation Act 2019*, s5

²⁹⁴ *Protection from Online Falsehoods and Manipulation Act 2019*, s2(2)(b)

²⁹⁵ *Protection from Online Falsehoods and Manipulation Act 2019*, s4

²⁹⁶ Taylor Vinters, *5 things you need to know about Singapore's controversial new fake news law* <https://www.taylorvinters.com/article/5-things-you-need-to-know-about-singapores-controversial-new-fake-news-law>; Tech Law for Everyone, *POFMA: Singapore's anti-fake news law* <https://www.scl.org/articles/10541-pofma-singapore-s-anti-fake-news-law>; *Protection from Online Falsehoods and Manipulation Act* <https://www.mlaw.gov.sg/files/news/others/POFMABrochure.pdf>

²⁹⁷ Amnesty International, *Singapore: Social media companies forced to cooperate with abusive fake news law* (19 February 2020) <https://www.amnesty.org/en/latest/news/2020/02/singapore-social-media-abusive-fake-news-law/>; <https://sso.agc.gov.sg/Acts-Supp/18-2019>

both located in Singapore.²⁹⁸ Other countries, such as Sri Lanka, have implemented similar laws to control misleading and false statements online.²⁹⁹

Foreign Interference

The *Foreign Interference (Countermeasures) Act* (FICA) was passed in the Singapore Parliament on 4 October 2021, and seeks to prevent, detect and disrupt the use of hostile information campaigns and local proxies by foreign entities intending to interfere in domestic politics.³⁰⁰ FICA allows authorities to compel internet, social media service platforms and website operators to provide user information,³⁰¹ block content,³⁰² and remove applications if the information is harmful and is suspected as being carried out by foreign actors. People deemed "politically significant persons" under the law will have to comply with strict rules relating to donations³⁰³ and declare their links to foreign affiliations.³⁰⁴ Instead of a court, an independent tribunal chaired by a judge will hear appeals against government minister's decisions.³⁰⁵ The legislation does not apply to citizens airing their political opinions (unless they are agents of a foreign principal as defined in the Act), and it does not apply to foreign individuals and publications commenting and reporting on Singapore politics, even if the comments may be critical of the government.³⁰⁶

Critics note that because the law is vague and broadly worded it could be used to silence government critics. It also followed just weeks after independent media site The Online Citizen (a site for alternative political views) was shut down over alleged failures to identify its funding sources. This raised concerns that Singapore authorities are strengthening efforts to enforce greater state control over its citizens.³⁰⁷

India

²⁹⁸ Ashley Westerman, 'Fake News' Law Goes Into Effect In Singapore, Worrying Free Speech Advocates', (2 October 2019), *NPR* <https://www.npr.org/2019/10/02/766399689/fake-news-law-goes-into-effect-in-singapore-worrying-free-speech-advocates>.

²⁹⁹ Shreetesh Angwalkar, 'Sri Lanka Implements Singapore Style Law to Control Fake News', (23 April 2021), *Spherex* <https://www.spherex.com/regulation/sri-lanka-implements-singapore-style-law-to-control-fake-news>

³⁰⁰ *Foreign Interference (Countermeasures) Act 2021* [https://www.parliament.gov.sg/docs/default-source/default-document-library/foreign-interference-\(countermeasures\)-bill-24-2021.pdf](https://www.parliament.gov.sg/docs/default-source/default-document-library/foreign-interference-(countermeasures)-bill-24-2021.pdf)

³⁰¹ *Foreign Interference (Countermeasures) Act*, s108

³⁰² *Foreign Interference (Countermeasures) Act*, s33

³⁰³ *Foreign Interference (Countermeasures) Act*, Part 5, Division 3

³⁰⁴ *Foreign Interference (Countermeasures) Act*, Part 6, Division 1

³⁰⁵ *Foreign Interference (Countermeasures) Act*, Part 8, Division 2

³⁰⁶ Philip J. Heijmas, 'Singapore Proposes Law Combating Foreign Interference Online', (13 September 2021), <https://time.com/6097362/singapore-online-foreign-interference-bill/>; PR Week Staff, 'Singapore's new foreign interference law could impact social media, publishers', (6 October 2021), *PR Week* <https://www.prweek.com/article/1729525/singapores-new-foreign-interference-law-impact-social-media-publishers>

³⁰⁷ Amnesty International, *Singapore: Foreign interference law is a tool for crushing dissent* (5 October 2021) <https://www.amnesty.org/en/latest/news/2021/10/singapore-foreign-interference-law-dissent/>; Human Rights Watch, *Singapore: Withdraw Foreign Interference (Countermeasures) Bill* <https://www.hrw.org/news/2021/10/13/singapore-withdraw-foreign-interference-countermeasures-bill>; Agence France-Presse, 'Chilling': Singapore's 'fake news' law comes into effect' (2 October 2019), *The Guardian* <https://www.theguardian.com/world/2019/oct/02/chilling-singapores-fake-news-law-comes-into-effect>

India is the largest market of users of Facebook and WhatsApp.³⁰⁸ India has become increasingly less accommodating toward big tech companies, driven by a rise in India's homegrown platforms such as Reliance Jio, a rapid spread of misinformation on platforms, and the government's desire to have a greater level of control over social media.³⁰⁹ In July 2018, villagers in a rural Indian town beat five strangers to death over a rumour circulated on WhatsApp that the men had kidnapped children.³¹⁰ On 29 June 2020 India banned 59 apps developed by Chinese firms, including TikTok, over concerns that these apps were engaging in activities that threatened the "national security and defence of India, which ultimately impinges upon the sovereignty and integrity of India".³¹¹ In February 2021, India ordered Twitter to remove more than 1100 accounts and posts it alleged spread misinformation about farmers protesting agricultural reforms. The Indian government rebuked Twitter for not fully complying with the government order.³¹²

These are some of the events that have led to sweeping reforms to hold social media companies, streaming platforms and digital news publishers to account under direct government oversight. The legally enforceable *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021* targets misinformation and hate speech. The rules require social media companies to establish a grievance redressal mechanism, obliges platforms to remove content within 36 hours of receiving a legal order, and requires them to disable any post depicting an individual in a sexual act or conduct. Critics note that these rules were introduced and implemented without public consultation and may lead to outright censorship. Further, the rules may undermine user rights such as end-to-end encryption, and some consider that the Indian Government does not have the capacity to effectively enforce the rules.³¹³

³⁰⁸ Sankalp Phartiyal & Aditya Kalra, 'India tightens regulatory grip on Facebook, WhatsApp with new rules' (25 February 2021), Reuters <https://www.reuters.com/article/india-tech-regulation-idUSKBN2AP175>

³⁰⁹ Saheli Roy Choudhury, 'India wants to cut Big Tech down to size. Critics say the new rules may give the state too much power' (20 April 2021) CNBC <https://www.cnbc.com/2021/04/20/indias-social-media-law-puts-big-techs-power-into-states-hands-critics-say.html>; Sarah Frier, Naomi Nix & Sarah Kopit, 'How and why internet companies moderate speech online' (25 October 2-2021), *The Washington Post* https://www.washingtonpost.com/enterprise/how-and-why-internet-companies-moderate-speech-online/2021/10/21/e4c87baa-3293-11ec-8036-7db255bff176_story.html?utm_source=rss&utm_medium=referral&utm_campaign=wp_business

³¹⁰ Pranav Dixit & Ryan Mac, 'How WhatsApp destroyed a village' (9 September 2018), *Buzzfeed News* <https://www.buzzfeednews.com/article/pranavdixit/whatsapp-destroyed-village-lynchings-rainpada-india>

³¹¹ Government of India, *Government Bans 59 mobile apps which are prejudicial to sovereignty and integrity of India, defence of India, security of state and public order* (29 June 2020) <https://pib.gov.in/PressReleaseDetailm.aspx?PRID=1635206>; Manish Singh, 'Facebook, Twitter, WhatsApp face tougher rules in India', (25 February 2021), *Techcrunch* <https://techcrunch.com/2021/02/25/india-announces-sweeping-guidelines-for-social-media-on-demand-streaming-firms-and-digital-news-outlets/>

³¹² Saheli Roy Choudhury, 'India rebukes Twitter for not fully complying with government order to ban certain accounts', (11 February 2021), CNBC <https://www.cnbc.com/2021/02/11/india-rebukes-twitter-for-not-fully-complying-with-government-order.html>

³¹³ Saheli Roy Choudhury, 'India wants to cut Big Tech down to size. Critics say the new rules may give the state too much power', (20 April 2021), CNBC <https://www.cnbc.com/2021/04/20/indias-social-media-law-puts-big-techs-power-into-states-hands-critics-say.html>

Conclusion

Across jurisdictions there is significant debate about the role of social media and how best to regulate it. Growing concern over the negative impact of social media engagement on individuals and on social cohesion has created a consensus on the need for better regulations. However approaches have differed and many regulatory frameworks are still under construction or in consultation phase.

This paper has provided an overview of the different approaches adopted by countries around the world and the different contexts in which these responses have been implemented. The diversity of approaches outlined can be used to inform consideration of what can be adapted and adopted for regulation in the Australian context.

Targeted, holistic and effective regulation is needed to counteract the negative effects of social media on mental health and wellbeing and its threats to social cohesion. The impact of algorithmic culture is profound, and the solutions must go beyond regulation and into social policy spheres. As we race towards web 3.0, new forms of encrypted communications, the creation of the metaverse and quantum computing, the ability of regulators to keep pace and ensure a balance between public safety and civil liberties will present enormous challenges.