The Impacts of Social Media in Australia

Research Brief | December 2021

Centre for
Digital Wellbeing

Table of Contents

Executive Summary

Social media has fundamentally shifted the way Australians maintain connection, consume content and share information. The tools of e-commerce and the digitalisation of advertising are driving the creation of platforms organised by algorithms that are designed to change human behaviour and our methods of social connection. The influence of these platforms, such as Facebook, Instagram and Twitter, are the subject of intense scrutiny as governments grapple with the rise of misinformation, polarisation and threats to democracy. Ongoing research examining the impacts of social media on mental health and wellbeing corroborates what many have long suspected to be true: social media use is linked to adverse outcomes, in particular for youth. The issue of safety online and offline is also a cause for concern, as social media is related to the rise of technology-facilitated abuse, especially gender-based violence. This research brief considers the impacts of social media on mental health and wellbeing, safety and social cohesion, based on the latest findings from Australian and international research.

There is a significant body of emerging research on the link between social media and mental health. Some studies indicate that social media use may lead to increased anxiety and depression, while other studies identify positive outcomes for particular cohorts, such as feelings of increased social support and connectedness. Negative online interactions, such as critical comments or lack of 'likes', may contribute to perceptions of exclusion and disconnection, while unfavourable social comparison with other users can adversely affect self-esteem through exposure to highly curated lives. Adolescents in particular are more vulnerable to the negative impacts of social media use, given the development and maintenance of social connections are crucial during this life stage. Research shows that younger individuals are more likely to engage in social comparison and approval-seeking online, which can contribute to loneliness, isolation and increased negative self-rumination. Additionally, emerging research shows that social media addiction can lead to increased psychological distress, showing similar adverse outcomes as children addicted to gaming. While gaming addiction has been a focus of policy making, the impact of heavy social media usage on mental health has been given little attention in an Australian context.

In relation to safety, in particular women's safety online and offline, there is evidence to suggest that social media has facilitated the rise of new forms of abuse and harassment. Technology-facilitated abuse includes behaviours such as online harassment, stalking, impersonation and threats, with much of the research focusing on abuse within the context of domestic and family violence. The anonymity of social media has created an online disinhibition effect where individuals may feel less constrained in perpetrating abuse and harassment. Recent studies suggest that social media has been weaponised by perpetrators of domestic violence to continue their campaign of coercive control against victims. The potential for social media access to become a space of vulnerability for victim-survivors has been highlighted as a key concern by researchers and professionals working in the domestic violence space.

The impact of social media on social cohesion is a growing priority for governments. Social media has shifted the way Australians consume news and share information, with research highlighting consequent impacts on trust in government and trust in media. 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. Further, social media has reduced confidence in mainstream media reporting and opened space for the spread of misinformation. Research indicates that optimisation algorithms can help spread misinformation, with fake news and images promoted on platforms due to their high level of engagement, such as number of likes, comments and shares, regardless of the authenticity or veracity of the information provided. Digitisation has also given foreign malicious entities new tools in the promotion of misinformation, creating a point of vulnerability for foreign interference in democratic processes.

On the other hand, there is research to suggest that some forms of engagement with social media can have positive effects on civic engagement and political participation as users are exposed to political information online that activates their sense of collective and personal agency, increasing their likelihood of participation in civic activities.

In response to the myriad concerns raised by citizens, governments have attempted to address some of the issues outlined above by incentivising social media platforms to increase transparency and uphold user privacy and safety. Australia's regulatory responses to social media have to date focused on reducing cyberbullying, minimising the spread of terrorist and extremist content and addressing power imbalances in the media marketplace. An emerging priority of social policy will lie in mitigating the negative outcomes that are a by-product of the digitalisation of human experience. While the growth of technological reach into the social and behavioural realm is inevitable, the emphasis must be on creating a human centred future.

Introduction

Where digital infrastructure intersects with social infrastructure, a public space is created where individuals can create networks, stay connected, access information and engage in political and civic life.[1] Social media platforms, such as Facebook, Twitter, TikTok and Instagram, exist at the intersection between digital and social infrastructure, and have a presence in the lives of most Australians.

In recent years, the impact of social media platforms on shared values and institutions has gained significant attention in both research and policymaking. Governments worldwide have become increasingly cognisant of the role of social media platforms in disseminating disinformation, with potential negative flow-on effects on democracy, polarisation and social cohesion. The influence of social media use on the mental health and wellbeing of individuals has also been flagged as a cause for concern. Further, the role of social media platforms in facilitating online abuse and harassment, particularly in the context of gender-based violence, remains a pressing issue.

This research brief examines the impacts of social media in Australia, focusing on the areas of mental health and wellbeing, safety and social cohesion. Policy responses to some of the challenges arising from the use (and misuse) of social media platforms, including hate speech, disinformation and harmful content, are also outlined. As social media is an evolving phenomenon, research and policy responses to its impacts are dynamic and likely to change rapidly. This brief is designed to summarise the existing body of evidence on the role and influence of social media in Australian society and beyond.

Methodology

This research brief draws on peer-reviewed research and grey literature. As social media is a relatively new and shape-shifting phenomenon, the brief is also informed by relevant commentary published on blogs and online media outlets. Where available, research from the Australian context has been cited, although insights from international literature – in particular the United States (US), Canada and the United Kingdom (UK) – have also been included. Studies were identified through Google Scholar with additional sources gathered by snowballing.

Defining social media

While various definitions of social media exist, this brief adopts the definition proposed by the Australian Bureau of Statistics, which refers to '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.'[2]

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 compared to 2015.[3] Other platforms like Instagram (which is owned by Facebook) have also seen a dramatic increase in usage, with more than 8 million unique visitors in an average four weeks between March 2018 and March 2019.[3] Twitter and Pinterest have also experienced growth in visitors, although Facebook remains the most widely used social network in Australia.

Overall, women tend to spend more time on social media than men.[4] The most prolific users were women aged between 14 and 24 years.[4] Young Australians spend an average of 7 hours and 22 minutes on their phones per day, predominantly on social media and gaming.[5]

Today, children are exposed on social media before they are born, with parents sharing ultrasound pictures on social media, leading to technology companies starting their collection and use of children's data from an early age. One estimate suggests over 72 million data points are collected about children by technology companies by the time they reach 13.[6]

Mental health and wellbeing impact of social media

While social media can aid social connection, there is a growing body of evidence devoted to the damaging impacts of social media on mental health and wellbeing, in particular research demonstrating increased social media use leads to social isolation, stress, depression, and anxiety.[7]

This section provides an overview of existing research on the relationships between social media use and mental health, the specific experiences and impacts of social media in relation to adolescents, as well as the current evidence on social media addiction. As will become clear, the research is not conclusive.[8] While there is a growing correlation between social media and depression, research has not kept pace with the rapid uptake in usage. Further, only very recently have discussions begun to focus on algorithms designed to drive engagement and extend time on the platforms, leading to addictive properties and creating dependency.

Social media is still relatively new and, although research indicates that heavy social media usage can have harmful implications on wellbeing, the long-term effects of significant usage are still unknown. Nonetheless, the design features of social media platforms are modifying our patterns of human behaviour, pointing to a further escalation of trends. With limited attention given in the social policy arena to date, this chapter will showcase the current research, gaps and limitations, which may inform decision makers and legislators on how to mitigate negative outcomes of social media use.

Understanding social media's impact on mental health should be central to social policy

The link between social media use and an increase in mental illnesses has received significant attention in recent years. Several studies report a correlation between the use of social media and the presence of depressive symptoms.[9] Other studies have demonstrated that age, pre-existing mental health conditions, usage rate and style of engagement all determine whether social media use will impact positively or negatively on a user's mental health.

Seabrook et al. (2016) found that for individuals with depression or anxiety, social media use may negatively impact mood and mental health, due to the potential for increased exposure to negative online social interactions, such as cyberbullying. However, the authors also noted that for some cohorts, the positive aspects of social media use, such as perceived access to social support and enhanced social capital, may be a protective factor against depression and anxiety.[10] Similarly, Baker et al. (2016) found, through a systematic review of 30 quantitative studies, a range of factors that may influence the relationship between social media use and mental health.[9] In particular, individual differences such as gender and personality play a key role, with one study highlighting that participants who scored higher on measures of neuroticism had more problematic social media use and displayed more depressive symptoms, compared to those lower in neuroticism.[9]

Researchers have proposed various explanations behind the link between depressive symptoms and time spent on social media. Pantic (2014) highlights that one key reason is that online communication may create altered (and often incorrect) impressions of other users' physical and personality traits, contributing to perceptions that other users are happier and more successful, and exacerbating pre-existing dispositions towards depression.[11] This finding is supported by Feinstein et al. (2013) in a US-based study of college-age young adults, which found that in the context of social media, negative social comparisons with other users may place individuals at risk of rumination and depressive symptoms.[12] An increase in Facebook usage is associated with a risk of increased unfavourable social comparison, where one assesses another person to be 'better off' than them.[13,14] Another study found that as little as 20 minutes on Facebook can negatively impact a person's self-esteem due to exposure to highly curated lives.[13] Further, social media usage can cause feelings of exclusion, isolation, and disconnection through negative online interactions such as critical comments or a lack of favourable engagement.[10]

Communication via social media may also replace in-person interactions, contributing to increased loneliness and worsening of existing mental health conditions.[7] One longitudinal study suggested that Facebook use is negatively associated with overall wellbeing while real-world social networks were positively associated with overall wellbeing.[15] The authors used data from over 5,000 participants to show that while having more Facebook friends was associated with better mental health, using Facebook to like posts, click links and share status updates was negatively associated with wellbeing.[15] One possible explanation for this negative association, as identified by the authors, is that large quantities of social media use may detract from more meaningful in-person experiences.

Research on the association between anxiety disorders and social media use is less explored compared to the literature available on depression and social media. However, a study of 563 young adults in the US found that more time spent on social media was significantly associated with anxiety symptoms.[16] The authors hypothesised that social media platforms may be a source of stress contributing to elevated anxiety symptoms and related impairment among users. However, it was also noted that individuals with anxiety may engage in more frequent social media use as a form of validation to enhance their self-worth. Therefore, while there is a strong relationship between anxiety disorders and social media use, the study suggests that the association may be complex and not solely causal.

Another study from the US suggests there is a link between the use of multiple social media platforms and self-reported depression and anxiety. Primack et al. (2016) surveyed 1,787 young adults on their use of social media and measured their depression and anxiety symptoms.[17] They found that participants who reported using between 7 to 11 different social media platforms were substantially more likely to exhibit increased levels of depression and anxiety, compared to participants who used up to two platforms.[17] The authors posited that this association could be attributed to a range of factors, including increased multitasking between platforms, which is related to poor mental health outcomes; different rules of engagement between platforms, which users may experience difficulty navigating, potentially leading to negative moods and emotions; and heightened potential for social gaffes, as the increased number of platforms used creates more opportunities for miscommunication and embarrassment.

Overall, the literature suggests that both the type and frequency of social media use influences mental health outcomes. For example, Verduyn et al. (2015) examined the different impacts on wellbeing stemming from active versus passive Facebook use, with active use referring to activities that involved direct exchanges with other users (e.g. commenting on posts) and passive use being the consumption of information without direct exchanges (e.g. scrolling through newsfeeds).[18] The authors found that passive Facebook usage led to declines in affective wellbeing over time and theorised that this was because individuals were exposed to more positive information about other users, inducing envy. The results are in line with a study by Nereim et al. (2020), which found that reading posts on social media is more strongly associated with depression than making posts.[19]

Social media can aid social connection

While much of the literature reviewed focuses on the negative impacts of social media on mental health and wellbeing, some recent studies have also linked social media use with feelings of increased social support and connectedness, particularly in certain sub-groups such as older cohorts. Feelings of connectedness and belonging are associated with higher life satisfaction[20] as well as lower levels of depression and anxiety.[10] Social media can support individuals to maintain relationships, regardless of geographical location, either by solidifying pre-existing relationships or creating new ones.[7,10,20,21] In a systematic review of 65 studies, Frost and Rickwood (2017) identified a number of positive psychological effects stemming from Facebook use including reduced feelings of isolation, enhanced wellbeing, greater health and wellness satisfaction and perceived social and emotional support from others.[22] Together, these benefits act as protective factors for mental health.

Other benefits of social media include ease of communication,[7] particularly for individuals who may experience difficulties with face-to-face communication, including people with certain mental health conditions, like social anxiety, or individuals with behavioural or developmental disorders such as autism. As social media platforms provide an environment where recognition of body language and socially appropriate eye contact is not necessary, individuals can exercise more control over their communication with other users.[7]

Another aspect of social media use on psychological wellbeing is what Erfani and Abedin (2018) term 'authentic self-presentation and self-disclosure'.[20] In a systematic review of 22 studies, the authors found that when social media facilitated the communication of personal thoughts and feelings with other users, this self-disclosure was positively correlated to life satisfaction, subjective wellbeing and happiness.[20] This finding is supported by Bailey et al. (2020) in their study on the tension between self-idealisation and authentic self-expression on Facebook based on data from over 10,000 users.[23] The authors compared users' self-reported personality with predictions of personality based on their Facebook likes and status updates, finding that individuals who were more authentic in their self-expression on Facebook reported greater life satisfaction. Again, this relationship may not be directly causal, but resultant.

Existing research indicates that social media is most likely to have a net positive impact on mental health in older cohorts, in cohorts who experience other barriers to social engagement

and when used in conjunction with face-to-face interactions. To better understand the context and environmental factors that underlie the growth of mental health conditions in Australia, acknowledging the connection between social media usage and depression and anxiety is paramount. Forward looking social policy should consider the impacts of new technologies on mental health and when and how their effects can be mitigated.

Adolescents are more vulnerable to the negative impacts of social media use

There is considerable research on the impacts of social media on adolescent mental health and wellbeing. Most social media platforms were launched after 2003, with today's adolescents and young adults being the first generation to grow up with social media.[24] In Australia, YouTube and Facebook remain the most popular social media platforms used by 86 per cent and 75 per cent of 13–17-year-olds respectively, followed closely by Instagram at 70 per cent and Snapchat at 67 per cent.[25]

According to the World Health Organization (WHO), an estimated 10 to 20 per cent of adolescents globally experience mental health challenges.[26] Depression and anxiety disorders are among the most common mental health conditions and are known to impact adolescent development negatively.[27,28] As such, understanding the influence of social media use during adolescence is of particular importance.

As with studies focused on adult cohorts, the impacts of social media use on mental health by adolescents is nuanced depending on the nature and level of use.[24] Early reviews, such as Best et al. (2014) tended to report either mixed or no effects of social media use on adolescent wellbeing.[24] However, more recent research suggests that younger individuals may be more vulnerable to potential negative impacts of social media during adolescence, given the emphasis on social connectedness during this period.[29] For instance, adolescents in particular may be more prone to engage in behaviours like social comparison and approval-seeking online, which may contribute to loneliness, isolation and increased rumination.[30] One study found that youth who reported seeking approval and peer status online faced a higher risk of substance use and risky sexual online behaviour one year later.[13] Another study indicated that adolescents may be highly sensitive to acceptance and rejection through social media, and that their heightened emotional sensitivity may make them specifically reactive to emotion-arousing media.[31]

As with adult cohorts, the type of social media use has a strong influence on mental health and wellbeing outcomes. A study by Frison and Eggermont (2020) echoed earlier findings in relation to the negative impacts of passive Facebook use (e.g. monitoring other peoples' profiles and content without direct exchanges between users).[32] Drawing on data from two surveys of 1,612 adolescents in Belgium, the authors found that passive Facebook use decreased adolescents' perceptions of social support, possibly due to negative social comparison with peers, contributing to increased loneliness. However, the authors also noted that the relationship between loneliness and passive Facebook use may be bidirectional, as lonely individuals may use Facebook in a passive manner to distract from daily distress and to fulfil specific coping needs.

Similarly, in a US-based study, Barry et al. (2017) surveyed 113 pairs of parents and adolescents, finding that social media activity – such as the number of accounts adolescents had and self-reported frequency of checking social media – was positively associated with loneliness and fear of missing out (FOMO).[29] Further, the authors found that anxiety and depressive symptoms were highest among adolescents with a relatively high number of social media accounts. However, a number of limitations were noted with this study, including the small sample size and relative homogeneity of the cohort, with the majority of participants (over 80 per cent) identifying as White/Caucasian.[29] Therefore, the generalisability of these findings may be limited.

Adolescents who engage in self-objectifying behaviour on social media, such as taking selfies or asking other users to rate their looks, may also be at risk of developing body shame. Salomon and Brown (2019) surveyed 142 adolescents on their social media use and found that greater levels of self-objectifying behaviour on social media predicted greater body surveillance (excessive monitoring of appearance), which may lead to negative feelings about one's body or feeling weak about failing to achieve a certain body type.[33] Internal research from Facebook shows that 18 per cent of survey respondents think that Instagram made body image issues worse.[34] For teen girls, one in three say that Instagram make them feel worse about themselves. The research also showed that social comparison and body image issues are more prevalent among teenagers than adults.

While acknowledging the potentially harmful effects of social media use, adolescents may, in some instances, experience positive impacts on mental health and wellbeing. Michikyan and Suárez-Orozco (2016) identify two main drivers behind adolescent social media use, namely identity development and intimacy.[35] The authors highlight that social media platforms provide youth with opportunities to reveal and express various characteristics of themselves, particularly if their identities may not be accepted in the offline world, using the example of gender diverse young people. Adolescents may also turn to social media to share their personal experiences, facilitating intimacy with other users.[35] For example, adolescents struggling with mental health conditions may turn to social media to find others with similar experiences with whom they can share insights and advice.[27] When social support and connectedness are present, adolescents and young adults may feel less isolated,[7] feel more confident[24] and generally experience higher life satisfaction.[36]

What COVID-19 taught us about social media's impact on wellbeing

The COVID-19 pandemic and resulting global public health responses have included stay at home orders and lockdowns, minimising face-to-face contact with people outside immediate household members. This has significantly impacted mental health and wellbeing. Additionally, screen time and time spent on social media and gaming has increased.[37] While technology has helped us to remain connected throughout this process, it has also resulted in excessive online engagement and concerns that a rise in screen time is impacting on children and adolescents' wellbeing.[38]

Some 18 months into the pandemic, research is slowly emerging on how we have responded to and been affected by the changing nature of social engagement. Early evidence from China found that increased social media exposure heightened the likelihood of anxiety and

depression during lockdowns.[39] In a cross-country study by Geirdal et al. (2021) 63 per cent of Australians surveyed reported emotional distress during April and May 2020, when Australia was in its first nationwide lockdown.[40] Emotional distress was significantly more prevalent among the respondents who reported they used social media several times a day (76 per cent), compared to those using social media less frequently (24 per cent). Although both studies support trends prevalent before the pandemic, both studies have limitations in terms of accurately elucidating causal relationships between social media use and mental health.

Social media has for many been a critical source of access to COVID-19 information and resources. The easy distribution of public health information has been highlighted as one of the key benefits of social media in the pandemic.[41] However, the easy access to COVID-19 content on social media may create new stressors and fears, especially for children and adolescents. Social media has also been used to spread misinformation about the virus and vaccination options, leading to increased treatment and vaccine hesitancy.[42]

The outcomes and drivers of social media addiction need to be understood

The topic of internet addiction is an established area of research, with an emerging focus on social media, gaming and general screen time given the move to mobile platforms.[43] With over three billion users world-wide, the addictive nature of social media use is an evolving field of research. Recent studies focusing on the addictive properties of social media have provided various insights into its drivers and outcomes.[44]

The Netflix documentary, The Social Dilemma, released in 2020, detailed how technology companies use algorithms as a tool to intentionally keep their users engaged. Design techniques, such as push notifications, refreshing, infinite scrolling, newsfeeds, likes, comments and shares have created feedback loops aimed at capturing our attention for longer. Social media is adaptive. It adjusts based on our preferences and behaviours. This makes it more engaging and interesting — and more addictive. While the idea of social media addiction is still relatively new, experts have begun to compare social media with other dopamine-induced addictions, as the platforms produce the same neural circuitry caused by gambling and recreational drugs to ensure consumers remain engaged.[45] An internal Facebook study of 22,410 users found that 30 per cent of users felt that Instagram made problematic use worse. Problematic use refers to difficulty managing the amount of time spend on social media.[34]

Perceptions of social networking and the use of social media differ significantly across generations. What may seem like excessive use to the older generation may feel normal to the younger generation without creating any negative daily repercussions.[46] A recent study established that teenagers and young adults may spend up to nine hours a day on social media without necessarily feeling addicted to the platforms. [44] A single determinant, such as time spent on social media, does not necessarily constitute addiction; however, it may still be severely disruptive to living a fulfilling life.[44] Therefore, recent research has used six components traditionally related to behavioural and chemical addiction (salience, tolerance, mood modification, withdrawal, relapse, conflict) to identify and study social media addiction.[47] For example, social media use could be considered problematic when an

individual is preoccupied with social media, attempts to create ways to free up more time for usage, and systematically increases time spent online in order to reduce feelings of guilt, anxiety and restlessness. Should the individual also become stressed and irritable when they cannot use social media or abide by personal resolutions to reduce usage, impacting relationships, sleep quality, physical and mental health, then their behaviour could be considered addictive.[47]

A study of 179 German university students found that Facebook addiction was associated with negative mental health outcomes, including anxiety. The authors found that anxiety symptoms increased over time, particularly when experiencing withdrawal, as users became nervous without the possibility of using Facebook.[48] Similarly, a study by Hou et al. (2019) of 232 Chinese students found that social media addiction was associated with negative mental health and academic performance.[49] A systematic review of 132 articles by Ahmed et al. (2021) found most studies concluded that excessive social media use negatively impacts an individual's wellbeing through increased stress, depression, reduced sleep quality, reduced quality of social relationships, or problems at work or in school settings.[44] However, researchers have started to theorise that the relationship between causes, effects and social media addiction is more complex. Outcomes of social media addiction such as anxiety, low self-esteem and depression have been simultaneously identified as predictors of social media addiction, leading researchers to hypothesise about the existence of a bidirectional relationship.[44] For instance, findings from a study of 329 students in Afghanistan indicated that an increase in social media addiction resulted in an increase in levels of depression, while depression at the same time predicted social media addiction, making social media usage a self-reinforcing cycle.[50]

Overall, explanations for social media addiction in the literature reviewed can be grouped into the following interlinked categories: dispositional factors, socio-environmental factors, motivational or behavioural factors, and technological factors.[47,51] Dispositional factors group together personality factors (neuroticism, extroversion, agreeableness, etc.), innate psychological needs (feeling of belongingness, social contact), and basic cognitions (core beliefs, attributions, expectations, etc.).[47] In relation to socio-environmental or socio-cultural factors, family dynamics,feelings of social isolation, loneliness and lack of support are most frequently cited as predictors of social media addiction.[44] Researchers have observed that social media offers immediate rewards, with features such as likes, shares and friend counts presenting a highly visible way of obtaining rewards and recognition.[52] These seemingly positive outcomes can foster certain types of behaviour, in this instance excessive social media use. If the behaviour has additionally led to avoidance of negative experiences, such as boredom, it is more likely to repeat itself and could possibly increase in frequency which may lead to social media addiction.[46,47,51]

However, it is important to note that research on social media addiction is still in its infancy. Systematic reviews by Sun (2021) and Andreassen (2015) indicate that theories on the development of social media addiction are inconsistent.[47] For instance, a study of 807 Malaysian university students found that using YouTube for entertainment predicted addiction while its usage for information purposes did not. That same study found no support for hypotheses that extraversion and openness – two personality traits often cited as predictors of social media addiction – lead to compulsive YouTube use.[53] Contrary to that

finding, Blackwell et al. (2017) demonstrated that extraversion was a significant predictor of social media addiction, confirming previous research by Wilson et al. (2010).[54] Other comparable studies focusing on dispositional factors like neurobiological explanations, as well as socio-environmental factors, behavioural factors and technological factors all yield varying or inconclusive outcomes.[47,51]

The multitude of explanatory factors and reported outcomes can be attributed to the wide variety of theories, frameworks and methodologies used by researchers to examine the determinants and outcomes of social media addiction. Andreassen (2015) points to the use of different measurement tools like the Bergen Facebook Addiction Scale and the Facebook Dependence Questionnaire, among others, which all define addiction differently, as one of the major challenges for current and future research.[47] Additionally, Griffiths et al. (2014) critiques the small subject samples used in studies, along with the narrow focus on youth as the predominant study cohort.[55] Even for youth, statistics on the prevalence of social media addiction are lacking.[46,47] In general, the literature reviewed indicates a myriad of gaps in current research, with further work needed to support the development of a comprehensive and substantial body of evidence on social media addiction.[44]

Linking gaming addiction to social media usage

Studies suggest that some 78,000 Australian children are now addicted to games.[56] Gaming addiction, a medical condition recognised by the World Health Organisation as gaming disorder,[57] has startling mental and physical health implications for both children and their parents. As video and computer games have been around longer than social media, they provide a guide to the impact of screen based behavioural addictions on human behaviour. While this research brief does not review the literature on gaming addiction in children and adolescents, it does deserve some mentioning as, unlike social media addiction, gaming addiction has been a topic of discussion and concern amongst parents and decision makers in Australia for some time.[58] Like social media, more sophisticated games use algorithms to encourage prolonged usage.

Additionally, while some researchers argue that social media addiction and gaming addiction should be treated as separate addictions[59] instead of one overall internet addiction,[60] a study by Wong et al. (2019) showed that both social media addiction and internet gaming addiction were similarly associated with increased psychological distress.[61]This is in line with arguments made by Professor Adam Alter. In his book, *Irresistible*, Alter discusses the rise of addictive technology by applying examples with reference to gaming addiction works and social media addiction. While social media addiction is still a new phenomenon, Alter argues that we can use what we know about other behaviour addictions to understand social media addiction.[62]

This supports the argument that gaming and social media share similar factors impacting mental health in adolescence.[63]

Safety behind and beyond the screen

Social media has emerged in recent years as a mechanism to perpetrate abuse and harassment, particularly in the context of gender-based violence. While the literature on the relationship between social media and abusive behaviour is relatively nascent, emerging research from Australia, Canada and the US suggests that social media is a platform and tool for perpetrators to monitor and abuse victim-survivors. Technology-facilitated abuse has become a key part of domestic and family violence, as an easy way to continue to abuse, control, stalk and harass while physically apart.

Recent activities by the companies behind the platforms show action taken to make platforms safer to use, especially for women and marginalised groups. These changes have been implemented as a response to demands by users. They demonstrate that technology companies are not immune to demands for safer usage of their products by their users.

The literature reviewed has adopted various terms to describe this online behaviour. Nuanced differences between the terms exist; however, the common denominator is the use of technology, in particular social media, to stalk, isolate, manipulate, threaten and harass victim-survivors. For ease of reference, this research brief adopts the term 'technology-facilitated abuse' to capture the broad spectrum of relationships between victim-survivors and perpetrators – be they strangers, relatives, friends, current or former partners.

Online disinhibition creates conditions for digital harassment and abuse

Researchers have attributed the proliferation of abusive behaviours in the digital space to a variety of factors, the most influential being the online disinhibition effect. The effect, as conceptualised by Suler (2005), provides useful context for some of the behaviours experienced on social media networks.

The online disinhibition effect encompasses several factors which can lead to uninhibited behaviours and communications online.[64] Anonymity is a key contributor to the disinhibition effect, as it may spur users to separate their online actions from their real identity, leading to compartmentalisation. Similarly, physical invisibility – where the identity of users is known but there is no interaction in the offline world – may instil courage in individuals to act differently online in ways outside of their normal behaviour. Asynchronicity refers to the time lapse between some online interactions, which Suler argues can disinhibit users as they do not have to cope with someone's immediate reaction.[64] Separately, these factors may not necessarily induce significant changes in online behaviour compared to the offline world but collectively, they can provide an explanation for why individuals may feel less constrained in perpetrating abusive behaviour against other users on social media.

Social media has facilitated the rise of new forms of abuse and harassment

The use of social media to perpetrate abusive behaviour has received significant attention in the past decade. Cyberbullying in particular has emerged as a key area of concern in relation to the wellbeing of children and adolescents. Cyberbullying, also known as electronic aggression, shares a number of characteristics with face-to-face bullying ('offline bullying') including hostile or aggressive acts, repeated negative behaviour, intent to harm and a power differential between the victim and perpetrator.[65] However, distinctive elements of cyberbullying include the possibility of a wider audience for the abuse, potential for the perpetrator to remain anonymous and challenges for the victim in disconnecting from the abuse, due to constant access to the Internet.[66] Anonymity is an important factor identified in the research, as it allows the pool of potential perpetrators to be expanded and it means that perpetrators do not necessarily witness the impacts of their bullying on the victim, reducing opportunities for remorse and empathy.[66] Further, it is noted that the nature of social media and other digital platforms means that abusive content can be viewed, saved and re-posted repeatedly, re-traumatising the victim before a larger potential audience.[67]

The range of behaviours that could be considered cyberbullying is extensive, with studies describing various activities such as flaming, which involves posting hostile messages and insults; imitating others online; name-calling; and sending or posting humiliating photos or videos.[67] While cyberbullying can be perpetrated across a range of mediums, one 2014 study of 384 Australian primary and secondary schools found that Facebook was reported as the most common social media platform used for cyberbullying, with over 55 per cent of case examples reporting the use of Facebook in creating 'hate' pages and fake accounts.[68] Respondents reported that Instagram was also commonly used for cyberbullying, including posting inappropriate images and videos as well as tagging – which involves mentioning a user in conjunction to an image or video – in offensive ways.[68]

The prevalence of cyberbullying in Australia ranges significantly between studies, although existing research is predominantly focused on the prevalence of cyberbullying among children and adolescents, with fewer studies devoted to adult cohorts. According to the eSafety Commissioner, between June 2016 and June 2017, around one in five young people in Australia reported being socially excluded, threatened or abused online.[69] This figure has likely increased significantly in the intervening years as social media use has increased. In contrast, based on a systematic review of 46 studies, Jadambaa et al. (2019) found that the lifetime prevalence of cyberbullying victimisation among children and adolescents in Australia was around 7 per cent, compared to approximately 25 per cent for traditional (face-to-face) bullying.[70] Variation in reported rates of cyberbullying can be attributed to a range of factors, such as the lack of a consistent definition of cyberbullying as well as social desirability bias, with perpetrators reluctant to self-report engagement in cyberbullying.[70,71] Further, victims of cyberbullying may underreport due to fears of losing access to technology and retaliation from the perpetrator.[72]

The link between traditional bullying and cyberbullying is supported by many studies, with research indicating both forms of bullying may occur together.[72] While studies on the drivers of cyberbullying in the Australian context are relatively scarce, one qualitative study of high school students in New South Wales found that young people engage in cyberbullying to gain social benefits over peers, as well as to manage anxiety and social pressures.[73] Further, girls were more likely to engage in sexualised abuse and harassment, as well as harassment in

relation to body size and appearance.[73] However, the generalisability of the study's findings are limited by the small sample size.

One influential study from the US on cyberbullying perpetration and victimisation sheds more light on predictors of both. Kowalski et al. (2014) found that being a victim of cyberbullying is strongly associated with being a perpetrator, with other predictors of cyberbullying perpetration being aggression, moral disengagement, risky online behaviour and narcissism.[74] In relation to cyberbullying victimisation, the authors found that the strongest predictor of being a victim of cyberbullying was also being a victim of face-to-face bullying, echoing the findings of other studies. Other risk factors for experiencing cyberbullying victimisation included anger, risky online behaviour, frequency of Internet use and social anxiety.[74]

The impacts of cyberbullying on victims can be severe. In their systematic review of 26 studies, Aboujaoude et al. (2015) found that reported pathology in cyberbullying victims included suicidal ideation, depression, emotional distress, insomnia and substance use.[72] Kowalski et al. (2014) note that perpetrators of cyberbullying were more likely to report low life satisfaction and higher levels of loneliness, although the associations between cyberbullying and these outcomes was weak.[74]

Image-based abuse is a subset of technology-facilitated abuse that has received particular attention in the literature. Image-based abuse refers to the non-consensual sharing of images with the intention of coercing, threatening, objectifying, harassing or abusing the victim.[75] Image-based abuse, perhaps more widely known as 'revenge porn', has received mainstream attention as a sexually motivated act of retaliation. However, researchers have identified a wide range of other drivers for image-based abuse, including monetary gain, social status enhancements, control and intimidation.[75] 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.[75] A substantial proportion (40 per cent) of victims of image-based abuse reported their images had been distributed across multiple platforms.[75] However, an important caveat noted by the researchers is that many victims of image-based abuse may not be aware that their images have been distributed.[75]

Recent reporting by the Australian Federal Police (AFP) indicates paedophiles are exploiting increased unattended screen time by children during COVID-19 lockdowns, with child abuse material shared online doubling since the start of the pandemic.[76] Increased screen time has left children vulnerable to online abuse; examples include being tricked into sending sexualised images by predators and then being blackmailed for even more explicit material.

Online abuse also happens to adults and sexuality and gender diverse individuals

Globally, 38 per cent of women have directly experienced abuse online, rising to 45 per cent for Gen Zs and Millennials.[77] The abuse is often far worse for marginalised groups such as the LGBTIQ+ community and women of colour.[78] While we know marginalised groups are more likely to experience violence and abuse online, research on adult experiences with technology-facilitated abuse in Australia is limited. A study by Amnesty International and Element AIused machine learning to survey millions of tweets received by 778 journalists and

politicians from the UK and the US throughout 2017, finding that 7.1 per cent of tweets sent to women in the study were abusive. This amounted to 1.1 million tweets mentioning 778 women across the year, or one every 30 seconds.[78] Black women were disproportionally targeted, being 84 per cent more likely than white women to be mentioned in abusive tweets.

A 2015 study of 3,000 Australian adults aged 18 to 54 years found that over 60 per cent of respondents reported experiencing some form of digital harassment and abuse in their lifetime.[79] Researchers also surveyed respondents on their relationship or connection, if any, to the people who had harassed or abused them online. The most common perpetrators were reported to be strangers (28.2 per cent), followed by friends (21.8 per cent), with current, past or potential intimate partners representing 14.4 per cent of perpetrators.[79] However, it should also be noted that in some instances (16 per cent) the identity of perpetrators was not known by the victim.[79]

Compared to other age groups, young adults aged 18 to 24 were more likely to experience digital harassment and abuse.[79] Another study found that individuals who identify as sexuality and gender diverse were more likely to experience higher rates of digital harassment and abuse. A survey of over 300 adults who identify as sexuality and gender diverse in Australia and Britain found individuals who are transgender experience higher rates of sexual, sexuality and gender-based harassment and abuse on the internet, compared to heterosexual cisgender individuals.[80] This type of abuse is described by the researchers as 'harmful and unwanted behaviours either of a sexual nature, or directed at a person on the basis of their sexuality or gender identity.'[80]

Social media is weaponised in the context of gender-based violence

There is a growing body of literature on technology-facilitated abuse in the context of gender-based violence. Dragiewicz et al. (2018) propose the use of the term 'technology facilitated coercive control' (TFCC) to describe abuse and violence facilitated by digital media and perpetrated by current or former intimate partners.[81] TFCC encompasses a range of abusive behaviours, such as threats delivered via social media and stalking using 'check-ins' where the location of the user can be revealed by their posts.[81]

Concerns about the role of media platforms, particularly Facebook, in facilitating deviant behaviour such as stalking, are not new. Chaulk and Jones (2011) surveyed over 200 Canadian university students by mapping their behaviour and activities conducted on Facebook. They compared participants leaving unwanted messages and monitoring user profiles for updates, against behaviours commonly associated with obsessive relational intrusion, which refers to behaviours that may not constitute stalking in a legal sense, but share similar characteristics.[82] The authors found social networking sites could be avenues for obsessive and intrusive behaviours, further noting that some forms of monitoring and surveillance – such as constantly monitoring someone's Facebook profile for updates – can be conducted in relative anonymity, with most users unaware that this behaviour is occurring.[82] Similarly, a survey of 365 college students in the US found that common digital dating behaviours among participants included the use of social media to monitor a dating partner's whereabouts or checking a dating partner's friends and contacts.[83]

Key features of social media, such as mutual networks of contacts, may facilitate unwanted engagement between perpetrators and victim-survivors of domestic and family violence. In a US-based qualitative study of 40 intimate partner violence professionals and 32 survivors, Freed et al. (2018) found that shared online social circles made it difficult for victim-survivors to 'delete' the perpetrator from their online life, providing an example of a perpetrator's family members using Facebook to harass the victim-survivor.[84] The authors further highlighted that a lack of digital savviness may make it more challenging for victim-survivors to recognise forms of technology-facilitated abuse.

One study involving 546 Australian domestic violence practitioners found that almost all survey respondents (98 per cent) reported working with clients who had experienced digital coercive control.[85] Digital coercive control, an alternative term to TFCC, refers to the use of digital technology, including social media to harass, threaten, stalk and abuse partners, ex-partners and children.[85] Facebook was identified as the most common social media platform used by perpetrators with respondents highlighting how 'technology enabled the perpetrator to create a sense of omnipresence … women felt they had no space that was free from the perpetrator's invasive contact or monitoring'.[85] Domestic violence practitioners also expressed that advice for women to limit or delete their social media presence was not always helpful, as for some women, remaining in contact with the perpetrator online was a means to gauge his mental state and the level of risk posed.[85] Further, as highlighted in a 2019 study of 55 domestic and family violence survivors in Queensland, pressure to disconnect from technology in response to online abuse places a disproportionate burden on survivors, while the perpetrators' online activity remains unrestricted.[86]

Research on teenagers and young people's experiences with technology-facilitated abuse is limited but offers useful insight into emerging forms of abuse made possible through the advent of social media. For example, Hellevik (2019) highlights examples of controlling behaviour on social media in his study of teenage victims of digital intimate partner violence and abuse in Norway. The most common form of digital controlling behaviour, as described by participants, was being pressured by their partner into blocking or deleting contacts from their social networking sites, with most cases involving girls pressured into deleting boys from their list of contacts.[87] While isolation from friends and family in real life is a common tactic used by perpetrators of domestic and family violence[88], social media provides another domain for perpetrators to exert control over their victim-survivors' social connections.

Perceptions of the impact of technology-facilitated abuse versus in-person abuse may differ between genders. In their UK-based study of technology use in adolescent romantic relationships, Stonard et al. (2017) found that adolescent males were more likely than females to minimise the perceived harm and seriousness of technology-facilitated dating abuse, compared to abuse experienced in person.[89] According to some participants, the ease for abuse to be stopped or prevented by blocking the perpetrator was advantageous. However, some female participants highlighted the pervasive nature of online abuse, noting that technology facilitated more opportunities for abusive behaviour to continue, even after the relationship had ended.

Another important finding is that victims of technology-facilitated abuse may include children. In 2019-20, a survey of 515 professionals working with domestic violence cases in

Australia found that children were involved in technology-facilitated abuse in approximately 27 per cent of domestic violence cases.[90] According to participants, perpetrators used technology to try to learn children's locations, to ask children about adult victims' locations or to request the adult victims' phone numbers. The most commonly used social media platforms for these behaviours included Facebook (59 per cent) and Snapchat (43 per cent).[90] Snapchat is a messaging application where pictures and messages shared by recipients are typically available for a limited time span before they automatically disappear; the temporary nature of these messages feeds into the online disinhibition effect, where perpetrators do not expect to be held accountable for these posts once they vanish.

While there is some research on the association between technological and in-person intimate partner violence, more studies are needed to explore the nature of the relationship, including longitudinal data on the lifetime prevalence of technology-facilitated abuse. A study of 278 Canadian university students found that social media use, but not texting, significantly predicted the perpetration of in-person and technological intimate partner violence.[91] One explanation for this correlation, as identified by the researchers, is that social media may provide a more public platform, and therefore create more visible opportunities, for abuse to take place. Further, social media provides young people with constant access to their dating partners, including the ability to monitor their partner's activity. Another Canadian study with over 1,500 participants examined the link between Facebook use, Facebook jealousy and intimate partner violence.[92] Researchers found that Facebook jealousy – which was measured through a Likert scale with a range of items such as 'I feel jealous when my partner posts a provocative picture (e.g., bikini, naked chest) on his/her Facebook wall'– was positively correlated with intimate partner violence. The authors theorised that feelings of jealousy and powerlessness in response to imagined or real threats to the relationship may trigger negative feelings and behaviours aimed at regaining control, which underpin the association between Facebook use and intimate partner violence.

There is limited research to date on the help-seeking behaviour of victim-survivors of technology-facilitated abuse. However, one study of 152 domestic violence advocates and 46 victim-survivors in Victoria found that over half of victim-survivors indicated they had not sought assistance in response to technology-facilitated stalking, with embarrassment being the main reason for not seeking help.[93] Further consideration will need to be given to safety campaigns and awareness raising efforts around such behaviours.

Digital social networks can also support survivors of gender-based violence

While the literature reviewed generally highlights negative aspects of the relationship between social media and gender-based violence, there is potential for online social platforms to be harnessed as a force for good. Social media platforms are increasingly being used as mechanisms to contribute to public discourse around gender-based violence.

McCauley et al. (2018) highlight an example of the hashtag #MaybeHeDoesntHitYou, which circulated Twitter in 2016, with users tweeting about non-physical forms of intimate partner violence.[94] The authors analysed 1,229 original tweets – for instance '#MaybeHeDoesntHitYou but he will sit with his friends and crack cruel jokes about you while

you're in earshot' – and identified the potential for social media platforms to provide a powerful platform to air the realities and dispel the myths of intimate partner violence.

One US-based study notes that Pinterest, a social media platform that is traditionally associated with crafts, recipes and home inspiration, may be a useful target for public health messaging around gender-based violence. Pinterest allows users to 'pin' content such as visuals and messages onto a virtual board, which other users can then view and share. Carlyle et al. (2018) analysed 750 posts on Pinterest in relation to intimate partner violence, noting that as a traditionally female-dominated platform – with over 80 per cent of users being women – Pinterest could be used by public health professionals as a channel to provide support and resources.[95]

The effectiveness of self-regulation by social media platforms is weak

The perception of social media platforms as 'neutral intermediaries' initially shielded companies from taking action against online abuse and harassment perpetrated by their users.[96] Social media platforms were initially adamant that user-generated content and the results curated by their algorithms were not to be interfered with, in the interests of free speech and the concept of the public forum. However, the real-world consequences of such platforms being used and misused by perpetrators of antisocial behaviours prompted action. In recent years, platforms such as Twitter and Facebook have responded to the misuse of their networks by introducing a range of measures, some of which are highlighted below.

Twitter has implemented features designed to support victims of abuse and harassment in the digital and 'real' world. One example is the #ThereIsHelp notification service, which was originally launched to provide mental health resources to vulnerable users.[97] The service displays to users searching for terms associated with self-harm or suicide on Twitter a notification encouraging them to reach out for help and provides contact details of available support services. In Australia, for instance, the contact details of Lifeline Australia are provided in the notification.[97] In 2020, Twitter expanded the feature to include terms associated with gender-based violence, with users searching related terms receiving a notification with contact details for local helplines and resources designed to encourage them to seek help. As of November 2020, the feature is available in 24 countries, including Australia.[98]

In 2017, Facebook introduced a number of tools in response to concerns raised by domestic violence support organisations. The tools were developed in consultation with organisations as well as victim-survivors' advocates. One feature uses artificial intelligence to proactively identify and block fake accounts, in order to prevent situations where users perpetrating online abuse and harassment may create multiple accounts to continue contacting victim-survivors, after their initial account is blocked.[99] Another tool is designed to support survivors who wish to remain in contact with perpetrators online, as blocking may lead to further harassment offline. The feature allows users to ignore messages, which are sent to a separate folder, as well as read messages without the sender seeing that they have been read.[99] These features, according to the founder of a domestic violence support organisation and adviser to Facebook, are intended to 'respond to what victims need — a means to track abusive

behaviour without escalation and to protect themselves from the emotional toll that harassment can take.'[100]

More recently, at the United Nations Generation Equality Forum in Paris in 2021, Facebook, Google, TikTok and Twitter announced a package of commitments to tackle online abuse and improve women's safety on their platforms.[77] This announcement came after concerted advocacy by high profile women, including former Australian prime minister Julia Gillard, globally calling for social media companies to tackle the abuse of women on their platforms.[101] Some of the commitments include offering settings controlling who can see, share, comment or reply to posts and providing easy access to safety tools, thus reducing the amount of abuse women see.

As the majority of these initiatives are in their infancy, their effectiveness in combatting online abuse and providing support to victims is yet to be determined. However, it shows that the technology companies behind social media platforms can be responsive to demands from their users and can implement measures aiming to improve online safety.

Social media and social cohesion

The impact of social media on social cohesion is also a growing source of interest in both research and policymaking. As already mentioned, social media can bring us together, however, as will be highlighted in this chapter, it can also deepen division and challenge social cohesion. The combination of easy access to and spreading of information, and the echo chamber enabled by algorithms exposing us to similar views and opinions repeatedly, can lead to polarisation and insulate people from opposing views about current affairs. Additionally, it has become easier for foreigner malicious actors to affect public discourse by spreading misinformation. Malicious entities can be challenging to identify, meaning educating the broader population on the impact of misinformation on social cohesion is critical.

Social media use impacts political trust and can lead to polarisation

Social cohesion remains a highly fluid concept, with no universally agreed definition. The Scanlon Foundation Research Institute, which has produced an annual national survey on social cohesion in Australia since 2007, considers five domains of relevance: belonging, worth, social justice, political participation and acceptance or rejection.[102] Further, shared values and a sense of community are considered critical components of social cohesion.[103] While the literature on the relationship between social media and social cohesion as a whole is limited, there is some research on the impacts of social media on trust in government, democracy and participation. This is outlined below.

Trust in government and trust in media are critical for a well-functioning democracy.[104] There are two strands of argument in relation to social media use and trust in government identified in the literature reviewed. Firstly, social media use may lead to increased exposure to political information that solidifies partisan biases, contributing to further polarisation. Secondly, in diversifying sources of news and in diluting the authority of information, social media has reduced confidence in mainstream media reporting and opened space for the concept of false or misleading news content.

Australians' usage of social media as a news source is similar to that of other developed countries, where between 40 and 60 per cent of adults rely on social media for news.[105] As of August 2020, almost 7.9 million Australians (38 percent) reported using social media as their main source of news.[106] However, only a very small proportion (5 per cent) considered social media a trusted source, with TV, radio and print media more frequently identified as trustworthy.[106] Trust in social media as a news source is influenced by different factors. For example, Sterrett et al. (2019) conducted an experiment where respondents were shown a health article shared via Facebook, with the source of the article manipulated to reflect either a reputable source (The Associated Press) or a fictional news website.[107] The authors found there was no significant difference in trust between a fictional or reputable news source; instead, respondents were more likely to trust an article if it was shared by a public figure they rated as trustworthy. The authors theorised that having a trusted public figure share the article may compensate for the fact that respondents may have limited or non-existent knowledge of the news source.

In analysing data from 27 European countries, Ceron (2015) found that consumption of news from online media websites, that is, online versions of traditional media outlets, was associated with higher trust in political institutions, whereas newsgathering from social media was linked to lower levels of political trust.[108] The author hypothesised that traditional media outlets were more likely to reinforce established political systems, while the bottom-up nature of social media facilitated the dissemination of alternative information undermining political trust. However, the author noted that the direction of causality is likely mixed, as citizens with higher levels of political distrust may be more likely to switch from mainstream media outlets to social media platforms.

Social media use can also shape citizen evaluations of politicians' personalities, which may influence the development or withdrawal of trust in government. In a survey of 1,117 respondents in Germany, Starke et al. (2020) found that interactions with politicians on social media increased the perceived likeability of candidates.[109] The authors theorised that interactions on social media may reduce the social distance between users and political leaders, contributing to the perception of leaders as individuals rather than abstract figures. However, social media use was not found to influence evaluations of other traits including leadership, benevolence and responsiveness.[109] On a similar note, research from Norway suggests that voters perceive politicians as more honest in social media, compared to talk shows and news interviews. [104] This finding applied in particular to young people (aged under 30 years), which the authors attributed to the virtuous circle theory, where voters tend to consider politicians more honest in the media formats they consume more frequently.[104]

Governments may also use social media as a tool to nudge public opinion and spread information. Bradshaw and Howard (2018) examined the behaviour of 'cyber troops', state-sponsored organisations responsible for distributing fake or manipulated content online in 28 countries, which were categorised as either democracies, authoritarian regimes or crisis states.[110] The authors argued that democracies have the highest level of capacity to conduct disinformation campaigns due to existing levels of investment in research and innovation in this area. For example, as suggested by the authors, political parties in democracies may use political bots to generate high follower counts on social media or amplify certain narratives or hashtags over others, thereby distorting partisan conversations. While coordinated and large actors may have greater capacity to spread 'fake news', individual users too can play a role. In a study of social media users in Chile, Valenzuela et al. (2019) found that politically engaged users were more likely to spread misinformation in relation to politics, science and natural disasters that matched their bias.[111] The authors proposed a number of reasons for this association; for example, politically engaged users may participate in disinformation campaigns to advance their own agendas or they may be more exposed to misinformation in general online, with exposure being a key predictor of sharing false claims.

There is a growing body of evidence on the relationship between social media use, misinformation and polarisation. Tucker et al. (2018) highlight that social media platforms are particularly vulnerable to disinformation campaigns, which can further polarisation.[105] First, the authors argue that a business model focused on ad revenue means that social media platforms are less stringent regarding political advertising, leading to the proliferation of bot accounts and the purchasing of followers to spread misinformation in relation to politics. Second, optimisation algorithms can help spread disinformation, with fake news and images

promoted on platforms due to their high level of engagement, such as number of likes, comments and shares, regardless of the degree of truth. Overall, the authors found that the online spread of misinformation, negative and uncivil exchanges, and in some instances, propaganda, negatively impacts social cohesion. In particular, polarisation has supercharged political misinformation, which has increasingly infiltrated public debate.[105] The flow-on impacts on democracy are considerable, with the authors highlighting the state of politics in the US as a prime example.[105] The authors argue that misinformation can distort the views of individual citizens, which in turn shapes collective public opinion and may impact policy and election outcomes. Negative partisanship and increased polarisation can lead to distorted perceptions about political parties that deepen political divisions in society. While the impacts of polarisation on democracy are more readily apparent in US politics, there are lessons that can be drawn for the Australian context in relation to the power of misinformation.

Social media as an echo chamber of information

It is generally held that cross-cutting exposure, or the exposure to conflicting political viewpoints, leads to high levels of political tolerance towards others and awareness of the legitimacy of oppositional viewpoints.[112] However, algorithms customising and personalising online experiences ensure that users are only exposed to information that aligns with previous consumption patterns, which minimises expose to diverse opinions.[113] Initial research, including a study by Messing and Westwood (2014), suggested that social media users will often read news articles shared by their friends, even if they are contrary to their own views.[114] However such safeguards are dependent on a diversification within the user's social group. Further, when social media is used as a primary source for information and newsgathering, there is a significant and demonstrated risk that people are exposed mainly to ideologically compatible news and information.[114] The idea of social media as an 'echo chamber' is supported by the theory of selective exposure, where an individual is only exposed to information corresponding with their views through a selective process.[114] For example, in their analysis of social media use during the 2016 US presidential election, Klein and Robinson (2020) found that social media use was associated with the reaffirmation of partisan allegiances. For example, respondents with highly favourable attitudes towards the Democratic Party reported greater trust in the party as their level of social media use increased.[115] These echo chambers, supported by the algorithArchivems used on social media platforms, can make us insular, less curious and less open-minded towards different ideas, which can fuel animosity towards 'the other'.

Social media can strengthen civic engagement and political participation

The evidence on the effect of social media use on political participation and civic engagement is mixed. One argument is that social media use, and internet use in general, diverts people's time and attention from civic engagement and political participation, as time spent on social media can lead to withdrawal from community activities.[116] However, an opposing perspective highlights that social media platforms can contribute to strengthening relationships and social capital building, thereby enhancing participation in civic and political life among users.[116] In an analysis of modern social movements, digital platforms have been extensively used to coordinate activities across the political spectrum.

The use of social media platforms to express political views and share personal stories through the widespread use of hashtags has been accompanied by scholarly debate on whether digitally networked participation can be considered a new mode of political participation. Digitally networked participation is undertaken by individual users to activate their own social networks in order to raise awareness about, or amplify political and social pressures for the solution of, a political or social problem.[117] While this form of activism is sometimes derided as 'slacktivism', with critics highlighting the low costs and impacts associated with activities such as liking a political post on Facebook or sharing images and videos with hashtags associated with specific causes, Theocharis (2015) argues that as long as the act has political motivations or intentions, it should be considered a form of political participation.[117] For example, #IAmNotAfraidToSayIt was a social media campaign initiated by a Ukrainian activist to highlight the prevalence of sexual violence and harassment in Ukrainian society, with users contributing their own experiences of sexual harassment and gender-based violence more broadly via the hashtag.[118] Lokot (2018) analysed around 3,500 Facebook posts containing the hashtag, concluding that while the campaign did not result in legislative reform, it succeeded in shifting the tone of public dialogue around sexual harassment from 'abstract and shameful' to 'personal and persistent'.[118]

The purpose of social media use by an individual is a moderating factor in determining whether its usage will lead to a decrease or increase in political and civic participation. In particular, the use of social media for entertainment purposes as opposed to political information gathering may lead to a decrease in cross-cutting exposure to information, which has a negative impact on civic and political engagement, political moderation and the quality of democratic politics.[105] However, in contrast to entertainment motivated use, informational uses of social media are generally found to increase participation in political activities.[111] In a meta-analysis of 22 studies, Skoric et al. (2016) found that social media use generally has a positive association with citizen engagement and political participation. In particular, the authors identified a significant positive relationship between informational social media use (where users seek, gather and share various kinds of information on social media platforms) and participation, in line with previous research suggesting that exposure to political information online can increase political expression and action.[119] For example, in a study of 502 participants in Chile, Halpern et al. (2017) found that sharing political content on social media, namely Facebook and Twitter, activated users' sense of collective and personal agency, which increased their likelihood of political participation.[120] The authors argued that political discussion on social media empowered users to feel more informed and capable of affecting politics, thereby spurring political participation. Similarly, Lane et al. (2017) surveyed 950 respondents in the US on their political information sharing on social media and found that even for individuals whose primary motivation for using social media was maintaining relationships, as opposed to political engagement, the presence of political information shared online was linked to increased political expression.[121] The authors theorised that individuals may be incentivised to engage with their social contacts on political matters to maintain connection.

In contrast, Theocharis and Lowe (2016) found that social media users from Western countries generally demonstrated an increased detachment from politics online which could explain a decrease in political engagement overall.[122] Nevertheless, where political

engagement is observed, it encourages the formation of socio-political groups. Through the creation of these groups, the willingness of group-members to accept positions different from their own decreases.[111] Similarly, Cappellini et al. (2019) theorised that social media allows for socio-ideological sorting and the classifying of people and ideas into 'us and them' groups to take place. However, it should be noted that in online spaces, the lines between these groups can in some instances be quite firm, while in others they are interchanging.[123] Additionally, in some of these groups, violent media content is shared, which can undermine social norms and values and, in some instances, translate into offline violent behaviour. The impact of online extremist views on the offline world can be significant.[124] A Belgian study found that exposure to extremism through social media was related to political violence.[124] Similarly, Williams et al. (2020) found that hate speech focused on race and religion was related to London's offline hate crimes.[125]

A recent Facebook-funded report by researchers from the University of Sydney and the University of Queensland found that LGBTIQ+ groups in the Asia Pacific experiencing hate speech on Facebook are less likely to receive support from the platform when they are targeted by hate speech. The study found that Facebook does not effectively capture the language and context dependent nature of hate speech, meaning that the platform is still enabling vilification and discrimination on public pages. The researchers call for Facebook to regularly consult with minority groups to understand evolving hate speech forms and support users targeted by hate speech. They also call for mandatory training in how to identify and report hate speech for anyone who runs a public Facebook page.[126]

In addition to political participation, research suggests that social media use can facilitate forms of civic engagement such as volunteering. In a study of over 2,000 millennials in the US, Lee (2020) found that a moderate level of Facebook use (between one to four times a week) was positively associated with volunteering.[116] However, an important caveat should be noted as the author found that moderate Facebook use did not predict increased participation in regular volunteering, which suggests that social media use alone does not necessarily cause increased civic engagement. Nevertheless, social media platforms can be an effective tool in increasing the ease of volunteering. Kaun and Uldam (2017) examined the role of social media, in particular Facebook, in facilitating volunteer-led initiatives in response to the migration crisis in Sweden in 2015.[127] The authors found that, in the context of a humanitarian crisis, Facebook was the primary platform where volunteers could coordinate their efforts and financial donations in a way that was both cost- and time- effective.

While the benefits of online volunteering can be numerous, there are also a number of associated challenges in relation to the engagement, recruitment, and retention of virtual volunteers. Retention is a specific problem as the tasks given to virtual volunteers are often short-term in nature, with few organisations trying to create long-term opportunities online.[128] Further, managers of online volunteers may have limited e-leadership skills.[129] While research suggests that virtual volunteering may reduce inequalities experienced in offline volunteering by mobilising different types of people,[130] individuals still require access to technology to participate.

Foreign interference through social media

Social media provides the opportunity to engage with other people across the globe. While it has enhanced connectedness across borders, social media has made it cheaper, faster and less risky for foreign actors to engage in activities contrary to Australia's sovereignty, values and national interest.

Digital foreign interference allows foreign actors to communicate directly with a country's citizens, making foreign interference on social media more difficult to contain. Once virtually connected, the malicious actor can create and disseminate targeted propaganda and misinformation, contributing to the distortion of democratic processes or undermining legitimate information, such as public health information. The information can match the narrative and political objectives of the foreign actor's government, leading to animosity between political or social groups, weakening public discourse and disrupting elections.

One of the most cited examples is Russia's interference in the 2016 US Presidential Election, where the Russian Government hacked campaigns, released politically damaging information on social media and spread propaganda online to damage the Clinton campaign and sow distrust in American democracy and democratic processes.[131]

The Australian Government established a Counter Foreign Interference Taskforce in 2019.[132] With the digitisation of our democratic participation, such as e-voting, e-petitioning and digital debate, and easy access to the spread of misinformation, we are opening up new opportunities for foreign actors to impact the democratic foundations of our society and erode social cohesion.[133]

Australia is developing a robust regulatory framework to protect us from digital foreign interference However, the effectiveness of legislative initiatives remain to be seen.[133] Digital foreign interference is, by its nature, difficult to identify, which can impact the punitive effects of legislation. A public education campaign and a continued focus on building and maintaining strong regulations is necessary to protect our democratic processes and minimise the impact of misinformation on social cohesion.

Regulating social media around the world: policy responses to date

Governments worldwide have attempted to respond to some of the issues raised in this research brief by incentivising social media platforms to increase transparency and holding companies accountable for user content, particularly in relation to privacy and safety. Some of the different approaches adopted by states are outlined below, noting that some of these legislative and policy instruments are still in development.

Australia

Australia's regulatory response to social media has to date focused on cyberbullying, terrorist and extremist content, and the media marketplace.

In response to cyberbullying on social media platforms, in 2015, the *Enhancing Online Safety Act 2015* (Cth) came into force, establishing a two-tiered scheme for social media services to remove cyberbullying material targeted at Australian children.[134] Tier 1 social media platforms, including Twitter, TikTok and Snapchat, participate in the scheme on a voluntary basis. 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. Tier 2 social media services are declared by the relevant Minister 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 will lift industry standards, and introduces additional compliance obligations.

The eSafety Commissioner has also released general guidelines on social media use for various groups including parents, children and young people. There is no recommended time limit for screen time or social media user; instead, the guidelines identify warning signs, such as reduced personal hygiene or becoming withdrawn from friends and family, that could suggest online activity is becoming problematic for children and young people.[135]

To combat the sharing of violent material on social media platforms, the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) was passed, requiring internet, content and hosting providers to report abhorrent violent conduct occurring in Australia on their services to the Australian Federal Police.[136] Failure to report the violent material may result in fines of up to $888,000 for corporations (at the time of writing). Likewise, failure to remove abhorrent violent material from their services may result in fines of up to $11.1 million or 10 per cent of annual turnover, whichever is higher.

The Act was passed in response to the Christchurch terror attack on 15 March 2019, where footage of the events was streamed online by the perpetrator. Australia also signed the Christchurch Call, which is a voluntary commitment from governments and online service providers aimed at addressing terrorist and extremist content online.[137] The Christchurch Call was established by the New Zealand Government and the French Government in 2019.

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 also committed to implementing measures to prevent the upload of terrorist and violent extremist content, with the Christchurch Call supported by social media platforms such as Facebook and Twitter.

Australia has recently introduced a mandatory code of conduct, the *News Media and Digital Platforms Mandatory Bargaining Code*, 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 business and digital platforms to reach binding agreements and ensures that an independent arbiter will determine the remuneration if parties cannot reach an agreement. While the Code achieved broad support in the Australian Parliament, 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. The draft legislation was amended (and subsequently passed in parliament in February 2021) to include a mediation period to allow digital platforms to agree before being entering into arbitration, and to take into consideration platforms existing agreements with publishers before deciding on the application of the code. [138]

Australia is currently considering regulations that could make it illegal for social media companies to direct children to harmful content. While work on the new regulations had already begun, Wall Street Journal's *The Facebook Files* series provided further impetus to protect children online. In October 2021, the Australian government released draft legislation that would enable the creation of a binding online-privacy code for tech companies, ensuring that the best interest of a child is the primary consideration during data collection, use and disclosure. For children under 16 years old, parental consent must be obtained by social media platforms before collecting, using or disclosing personal information.

The United States

Following the storming of the US Capitol on 6 January 2021, the United States Congress held a congressional hearing interrogating the CEOs of Facebook, Google and Twitter about the part their social media platforms played in the attack and the spread of misinformation on their platforms.[139] Additionally, the Congress held meetings with the policy leads from Facebook, YouTube and Twitter in April 2021 about social media echo chambers and the use of algorithms.[140] These hearings are part of a wider policy discussion taking place within the United States.[141]

Discussion on reforms to platform regulation in the US have included a focus on amendments to section 230 of the Communications Decency Act, passed in 1996. Section 230 creates a broad protection for social media platforms and other online intermediaries from being held liable for transmitting third-party content (i.e. their users' posts), while also allowing them to

remove the content. This section is often referred to and seen as a shield from liability, protecting online platforms. Moreover, concerns have been raised regarding implications of reform on the First Amendment, for example Goodman E. et al. (2019) argue that Section 230 encourages the moderation of content, and that the First Amendment protects social media platforms from hate speech liability.[142] Nevertheless, bills including the Platform Accountability and Consumer Transparency Act (PACT Act), the Safeguarding Against Fraud, Exploitation, Threats, Extremism and Consumer Harms Act (SAFE TECH Act) and the Protecting Americans from Dangerous Algorithms Act are proposing reforms to section 230, while still maintaining its core elements.

The PACT Act, a bipartisan bill, was introduced to Congress in March 2021. The Bill seeks to make content moderation for social 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 power to bring legal action against platforms that violate federal civil law.[144] The Bill was referred to the Senate Committee on Commerce, Science, and Transportation and at the time of writing has not progressed.

The proposed SAFE TECH Act was announced in February 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.[145] 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. [146] The bill was referred to the Senate Committee on Commerce, Science and Transportation in February 2021. The objective of the SAFE TECH Act is to hold social media companies accountable for enabling cyber-stalking, targeted harassment, and discrimination.[147]

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.[148] The Bill was referred to the Subcommittee on Communications and Techology 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[149] – and the Kids Internet Design and Safety Act, which aims to stop online practices such as manipulative marketing, amplification of harmful content and damaging design features.[150]

While there is no guarantee that these bills will be enacted into legislation, they demonstrate the growing concern among policymakers in the US that algorithmic functions are reshaping human behaviour and regulation of social media platforms is required.

In September 2021, journalists from Wall Street Journal released a series of articles and podcast episodes, *The Facebook Files*, where former Facebook employee Frances Haugen

revealed how Facebook's services are toxic for especially teen girls, how its in-built algorithms promote inflammatory content and how it puts profit over people.[151] Following the release of the Facebook Files, Facebook's Global Head of Safety, Antigone Davis gave evidence to the Subcommittee on Consumer Protection, Product Safety and Data Security.[152] Facebook received substantial criticism from the Senators regarding how the platform has neglected to share their internal findings on the harms their platforms have on mental health and children. Additionally, Senators expressed concern about children under the age of 13 being on Facebook's different platforms and questioned what Facebook did to prevent this. As a result, Facebook has paused the development of Instagram Kids, a platform tailored for children under 13 years old.

Canada

In the lead-up to the 2019 Canadian federal election, the Canadian Government released the *Canada Declaration on Electoral Integrity Online*.[153] The non-binding declaration was designed to increase the accountability of social media platforms in responding to disinformation online, particularly information disseminated with the intent of undermining free and fair elections. The declaration contains twelve initiatives aimed at enhancing integrity, transparency and authenticity. The initiatives include assisting users to better understand the sources of information they are seeing; removing fake accounts and inauthentic content on their platforms; and ensuring transparency for regulated political advertising. The declaration was signed by Facebook, Twitter, Google and Microsoft.[154]

European Union

In response to the dissemination of disinformation and its associated challenges, the European Commission established the *Code of Practice on Disinformation* in 2018.[155] The code aims to combat the spread of misleading information that may cause public harm, including threats to democratic political processes and EU citizens' health, the environment and security. The code tasks relevant signatories with various commitments, including improving the scrutiny of advertisements to reduce the revenues of purveyors of disinformation; enabling public disclosure of political advertising; diluting the visibility of disinformation by improving the findability of trustworthy content; and providing annual reports on their efforts to combat disinformation. The code has been signed by Facebook, Twitter, Google, Mozilla, Microsoft and TikTok.[155]

The European Union (EU) also passed the *General Data Protection Regulation* (GDPR) in 2018, which has been described as the 'toughest privacy and security law in the world'.[156] The GDPR requires organisations to adopt various measures to protect any personal data that may target or collect in relation to people in the EU. Organisations are required to handle data securely by implementing appropriate technical and organisational measures; in the event of a data breach, data subjects must be notified within 72 hours or penalties may result. There are two tiers of GDPR fines, with some violations potentially resulting in fines of up to €10 million ($16 million AUD) or 2 per cent of the firm's worldwide annual revenue, whichever is higher, and more serious violations incurring fines of up to €20 million ($32 million AUD) or 4 per cent of the firm's worldwide annual revenue, whichever is higher.

Germany in particular has adopted stringent measures to address the dissemination of hate speech and other illegal content on social media. Germany passed the Network Enforcement Act (NetzDG) in 2018 which targets social networks with more than two million registered users in Germany.[157] Under NetzDG, platforms are required to respond to complaints of unlawful content, determine whether the content is illegal according to the German Criminal Code and, if so, remove it from their services within 24 hours, or in some cases, within seven days. Unlawful content may include, for example, incitement of violence or hatred against national, religious, ethnic or racial groups. Penalties for failing to comply with NetzDG may include fines of up to €50 million ($79 million AUD) per violation. Platforms are also required to publish their handling of complaints twice a year.[158]

Singapore

In order to combat fake news, Singapore passed the *Protection from Online Falsehoods and Manipulation Act* in 2019. The legislation requires online platforms, including social media sites, search engines and news aggregation services, to issue corrections or remove content that the Singapore Government deems false. According to the Act, false statements of fact are considered particularly serious if they are prejudicial to the security and bilateral relations of Singapore, or incite feelings of hatred between different groups.[159] Failure to comply with the act may result in fines of up to $1 million Singapore dollars.[159] This legislation is important to the greater region, as the Asian headquarters of Facebook and Twitter are both located in Singapore.[160]

United Kingdom

In 2019, the UK Government released its Online Harms White Paper, outlining proposed measures to enhance user safety online by requiring companies within scope, including social media platforms and search engines, to address illegal content on their services.[161] The proposed laws impose a duty of care on companies to protect their users from harmful content – that is, material which may cause significant physical or psychological impacts on individuals – by requiring companies to designate what content is not acceptable in their terms and conditions, and enforce this effectively. Ofcom, the UK communications regulator, has been appointed as the new online harms regulator and will be able to enforce penalties of up to £18 million ($32 million AUD) or 10 per cent of global annual turnover, whichever is higher, or stop services from operating, if they are non-compliant.[161] The UK Government has also released interim codes of practice designed to support companies to take action against terrorism and child sexual exploitation and abuse on their services. The codes of practice are voluntary and non-binding. The UK Government published a draft Online Safety Bill in May 2021, and a Joint Committee has been established to consider the draft legislation, with a report deadline of December 2021.

Conclusion

Social media is having a growing impact on the lives of Australians and has fundamentally impacted and changed the way we engage. While there are benefits to social media, such as aiding social connection for certain cohorts, there are significant negative and damaging aspects to social media, which have been given limited attention in the social policy arena to date.

This research brief has provided an overview of existing research on the impacts of social media in Australia, focusing on mental health and wellbeing, safety and social cohesion.

The research literature in this paper indicates a strong link between social media use and an increase in mental illnesses, especially the presence of depressive symptoms, with factors such as age, pre-existing mental health conditions, usage activity and frequency of use all having an impact to varying degrees. Children and adolescents are particularly vulnerable to the negative impacts of social media use, given the importance of social connections in those formative years. With social media still being relatively new, its long-term effects are unknown, although the trends are highly concerning.

Further, there is increasing evidence of social media being used as a mechanism to perpetrate abuse and harassment, particularly in the context of gender-based violence. Traditionally, policymaking has focused on cyberbullying, particularly in relation to children and adolescents. This remains an important focus. However, technology-facilitated abuse has become an increasingly prevalent feature in domestic and family violence, with perpetrators using social media platforms to monitor, abuse, control, stalk, isolate and harass victim-survivors.

Finally, while social media can bring us together, it can also deepen divisions and challenge social cohesion. Australians are increasingly relying on social media as their main source of news. This is problematic when the algorithms used by social media platforms are designed to show people what they want to see, meaning people are less likely to be exposed to content that challenges their worldview. Further, social media helps spread misinformation and disinformation, as information is promoted on platforms in response to high levels of engagement, regardless of its truth. Social media has also exposed Australia to a growing risk from foreign interference, as has been evidenced in other countries such as the US.

The risks posed by social media are increasingly being recognised, as demonstrated by the literature reviewed in this paper, and there are a range of regulatory responses being implemented by governments in Australia and internationally. However, it is clear that more needs to be done to understand its effects on mental health and wellbeing, safety and cohesion, to mitigate its negative impacts and to better regulate it.

References

[1] Zuckerman, E. (2020). 'What is Digital Public Infrastructure? An essay, in the form of an FAQ, about the possibility of digital social spaces built with taxpayer dollars.' *Centre for Journalism & Liberty.* https://www.journalismliberty.org/publications/what-is-digital-public-infrastructure.

[2] Australian Bureau of Statistics. (2016). *Summary of IT Use and Innovation in Australian Business, 2014-15.* https://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/8166.0Main%20Features52014-15?opendocument&tabname=Summary&prodno=8166.0&issue=2014-15&num=&view.

[3] Morgan, R.. (2019). *Facebook on top but Instagram and Pinterest growing fastest.* http://www.roymorgan.com/findings/7979-social-media-trends-march-2019-201905170731.

[4] Morgan, R. (2018). *Young Women the Queens of social media in Australia.* http://www.roymorgan.com/findings/7584-social-media-minutes-by-gender-age-march-2018-201805110812.

[5] EY Sweeney. (2017). *Digital Australia: State of the Nation: 2017 Edition*. https://digitalaustralia.ey.com/Documents/Digital_Australia_2017%20edition.pdf.

[6] Williams, D., Farthing, R. & McIntosh, A. (2021). *Surveilling young people online: An investigation into TikTok's data processing practices*. Reset Australia. https://au.reset.tech/uploads/resettechaustralia_policymemo_tiktok_final_online.pdf

[7] Naslund, J.A., Bondre, A., Torous, J. & Aschbrenner, K.A. (2020). 'Social Media and Mental Health: Benefits, Risks and Opportunities for Research and Practice'. *Journal of Technology in Behavioral Science,* 5, 245-257.

[8] Karim, F., Oyewande, A. A., Abdalla, L. F., Chaudhry Ehsanullah, R. & Khan, S. (2020). 'Social Media Use and Its Connection to Mental Health: A Systematic Review'. *Cureus*, 12(6).

[9] Baker, D. & Algorta, G.P. (2016). 'The Relationship Between Online Social Networking and Depression: A Systematic Review of Quantitative Studies'. *Cyberpsychology, Behavior, and Social Networking,* 19(11), 638-648.

[10] Seabrook, E.M., Kern, M.L. & Rickard, N.S. (2016). 'Social Networking Sites, Depression, and Anxiety: A systematic Review'. *JMIR Mental Health*, 3(4):e50.

[11] Pantic, I. (2014). 'Online Social Networking and Mental Health'. *Cyberpsychology, Behavior, and Social Networking,* 17(10), 652-657.

[12] Feinstein, B.A., Hershenberg, R., Bhatia, V., Latack, J.A., Meuwly, N. & Davila, J. (2013). 'Negative Social Comparison on Facebook and Depressive Symptoms: Rumination as a Mechanism'. *Psychology of Popular Media Culture,* 2(3), 161-170.

[13] Hussain, Z. & Griffiths, M.D. (2018). 'Problematic Social Networking Site Use and Comorbid Psychiatric Disorders: A Systematic Review of Recent Large-Scale Studies'. *Frontiers in Psychiatry,* 9(686).

[14] de Vries, D.A. & Kühne, R. (2015). 'Facebook and self-perception: Individual susceptibility to negative social comparison on Facebook'. *Personality and Individual Differences*, 86, 217-221.

[15] Shakya, H.B. & Christakis, N.A. (2017). 'Association of Facebook Use With Compromised Well-Being: A Longitudinal Study'. *American Journal of Epidemiology,* 185(3), 203-211.

[16] Vannucci, A., Flannery, K.M. & Ohannessian, C.M. (2017). 'Social media use and anxiety in emerging adults'. *Journal of Affective Disorders,* 207, 163-166.

[17] Primack, B.A., Shensa, A., Escobar-Viera, C.G., Barrett, E.L., Sidani, J.E., Colditz, J.B. & James, A.E. (2017). 'Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults'. *Computers in Human Behaviour,* 69, 1-9.

[18] Verduyn, P., Lee, D.S., Park, J., Shablack, H., Orvell, A., Bayer, J., Ybarra, O., Jonides, J. & Kross, E. (2015). 'Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence'. *Journal of Experimental Psychology,* 144(2), 480-488.

[19] Nereim, C., Bickham D. & Rich M.J. (2020). 'Social media and adolescent mental health: who you are and what you do matter'. *Journal of Adolescent Health.* 66(2), 118-119.

[20] Erfani, S.S. & Abedin, B. (2018). 'Impacts of the use of social network sites on users' psychological well-being: A systematic review'. *Journal of the Association for Information Science and Technology,* 69(7), 900-912.

[21] Huang, C. (2017). 'Time Spent on Social Network Sites and Psychological Well-Being: A Meta-Analysis'. *Cyberpsychology, Behavior, and Social Networking,* 20(6), 346-354.

[22] Frost, R.L. & Rickwood, D.J. (2017). 'A systematic review of the mental health outcomes associated with Facebook use'. *Computers in Human Behaviour,* 76, 576-600.

[23] Bailey, E.R., Matz, S.C., Youyou, W. & Iyengar, S.S. (2020). 'Authentic self-expression on social media is associated with greater subjective well-being'. *Nature Communications,* 11(4889).

[24] Best, P., Manktelow, R. & Taylor, B. (2014). 'Online communication, social media and adolescent wellbeing: A systematic narrative review'. *Children and Youth Services Review,* 41, 27-36.

[25] eSafety Commissioner. (n.d.) *Young people and social media usage.* Australian Government. https://www.esafety.gov.au/about-us/research/youth-digital-dangers/social-media-usage.

[26] World Health Organization. *Adolescent mental health.* https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

[27] Keles, B., McCrae, N. & Grealish, A. (2020). 'A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents'. *International Journal of Adolescence and Youth,* 25(1), 79-93.

[28] McCrae, N. G. (2017). 'Social Media and Depressive Symptoms in Childhood and Adolescents: A systematic Review'. *Adolescent Research Review,* 2, 315-330.

[29] Barry, C.T., Sidoti, C.L., Briggs, S.M., Reiter, S.R. & Lindsey, R.A. (2017). 'Adolescent social media use and mental health from adolescent and parent perspectives'. *Journal of Adolescence,* 61, 1-11.

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

[31] Crone, E.A. & Konijn, E.A. 2018. Media use and brain development during adolescence. *Nature Communications*, *9*(588), 1-10.

[32] Frison, E., & Eggermont, S. (2020). Toward an Integrated and Differential Approach to the Relationships Between Loneliness, Different Types of Facebook Use, and Adolescents' Depressed Mood. *Communication Research, 47*(5), 701-728.

[33] Salomon, I., & Brown, C.S. (2019). 'The Selfie Generation: Examining the Relationship Between Social Media Use and Early Adolescent Body Image'. *Journal of Early Adolescence,* 39(4), 539-560.

[34] Facebook Research Deck. (2019). Hard life moments – mental health deep dive. https://about.fb.com/wp-content/uploads/2021/09/Instagram-Teen-Annotated-Research-Deck-1.pdf

[35] Michikyan, M. & Suárez-Orozco, C. (2016). 'Adolescent Media and Social Media Use: Implications for Development'. *Journal of Adolescent Research,* 31(4), 411-414.

[36] Piteo, E.M. & Ward, K. (2020). 'Review: Social networking sites and associations with depressive and anxiety symptoms in children and adolescents - a systematic review'. *Child and Adolescent Mental Health*, 25(4), 201-216.

[37] Norris, J. (2021). COVID-19 lockdowns saw increased screen time and sleep problems. Medical News Today. https://www.medicalnewstoday.com/articles/covid-19-lockdowns-saw-increased-screen-time-and-sleep-problems

[38] Gupta, T., Swami, M.K. & Nebhinani, N. (2020). 'Risk of digital addiction among children and adolescents during COVID-19 pandemic: Concerns, caution, and way out'. *Journal of Indian Association of Child Adolescents Mental Health,* 16(3)*,* 199-208.

[39] Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., Wang, Y., Fy, H. & Dai, J. (2020). 'Mental health problems and social media exposure during COVID-19 outbreak'. *PLOS ONE*, 15(4).

[40] Geirdal, A. Ø., Ruffolo, M., Leung, J., Thygesen, H., Price, P., Bonsaksen, T. & Schoultz, M. (2021). 'Mental health, quality of life, wellbeing, loneliness and use of social media in a time of social distancing during the COVID-19 outbreak. A cross-country comparative study'. *Journal of Mental Health,* 30(2), 148-155.

[41] Wong, A., Ho, S., Olusanya, O., Antonini, M. & Lyness, D. (2020). The use of social media and online communications in times of pandemic COVID-19. *Journal of the Intensive Care Society,* 22(3), 255-260.

[42] Susarla, A. (2021). Big tech has a vaccine misinformation problem – here's what a social media expert recommends. *The Conversation.* https://theconversation.com/big-tech-has-a-vaccine-misinformation-problem-heres-what-a-social-media-expert-recommends-164987

[43] Hawi, N. & Rupert, M.S. (2016). 'The Relations Among Social Media Addiction, Self-Esteem, and Life Satisfaction in University Students'. *Social Science Computer Review,* 35(5), 576-586.

[44] Ahmed, E. & Vaghefi, I. (January 2021). *Social Media Addiction: A Systematic Review through Cognitive-Behavior Model of Pathological Use.* Proceedings of the 54th Hawaii International Conference on System Sciences, Hawaii. https://www.researchgate.net/publication/348548390_Social_Media_Addiction_A_Systematic_Review_through_Cognitive-Behavior_Model_of_Pathological_Use

[45] Addiction Centre. (2021). *Social Media Addiction*. https://www.addictioncenter.com/drugs/social-media-addiction/

[46] Kuss, D. & Griffiths, M.D. (2017). 'Social Networking Sites and Addiction: Ten Lessons Learned'. *International Journal of Environmental Research and Public Health,* 14(3), 311.

[47] Andreassen, C. (2015). 'Online Social Network Site Addiction: A Comprehensive Review'. *Current Addiction Reports,* 2, 175-184.

[48] Brailovskaia, J. & Margraf, J. (2017). 'Facebook Addiction Disorder (FAD) among German students – A longitudinal approach'. *PLoS One,* 12(12), 1-15.

[49] Hou, Y., Xiong, D., Jiang, T., Song, L. & Wang, Q. (2019). 'Social media addiction: Its impact, mediation, and intervention'. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace,* 13(1).

[50] Haand, R. & Shuwang, Z. (2020). 'The relationship between social media addiction and depression: a quantative study among university students in Khost, Afghanistan'. *International Journal of Adolescence and Youth,* 25(1), 780-786.

[51] Sun, Y. & Zhang Y. (2021). 'A Review of theories and models applied in studies of social media addiction and implications for future research'. *Addictive Behaviors,* 114.

[52] Martinez-Pecino, R. & Garcia-Gavilán, M. (2019). 'Likes and Problematic Instagram Use: The Moderating Role of Self-Esteem'. *Cyberpsychology, Behavior and Social Networking,* 22(6), 412-416.

[53] Klobas, J.E., McGill, T.J., Moghavvemi, S. & Paramanathan, T. (2018). 'Compulsive YouTube usage: A comparison of use motivation and personality effects'. *Computers in Human Behavior,* 87, 129-139.

[54] Blackwell, D., Leaman, C., Tramposch, R., Osborne, C. & Liss, M. (2017). 'Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction'. *Personality and Individual Differences,* 116, 69-72.

[55] Griffiths, M., Kuss, D. J. & Demetrovics, Z. (2014). 'Social networking addiction: An overview of preliminary findings'. In K. Rosenberg & L. Feder (eds.), *Behavioral Addictions. Criteria, Evidence, and Treatment* (Academic Press, New York), 119-141.

[56] Biggs, H. (21 October 2019). 'SA Schools tackling child gaming addiction with new 'Unplugged' social program'. *9News*. https://www.9news.com.au/national/video-game-addiction-children-schools-program-unplugged-health-news-sa-australia/fd9fa5df-08cb-4dbf-99a4-3482f4463a05.

[57] World Health Organisation. (14 September 2018). *Addictive behaviours: Gaming disorder*. https://www.who.int/news-room/q-a-detail/addictive-behaviours-gaming-disorder.

[58] eSafety Commissioner. *Gaming*. Australian Government. https://www.esafety.gov.au/key-issues/staying-safe/gaming.

[59] Rehbein, R. & Mößle T. (2013). 'Video game and internet addiction: Is there a need for differentiation?' *Sucht*, 59(3), 129-42.

[60] Griffiths, M.D. & Pontes, H.M. (2014). 'Internet addiction disorder and internet gaming disorder are not the same'. *Journal of Addiction Research and Therapy,* 5(4).

[61] Wong, H.Y., Mo, H.Y., Potenza, M.N., Chan, M.N.M., Lau, W.M., Chui, T.K., Pakpour, A.H. & Lin, C. (2020). 'Relationships between severity of internet gaming disorder, severity of problematic social media use, sleep quality and psychological distress'. *International Journal of Environmental Research and Public Health*, 17(6), 1879.

[62] Alter, A. (2018). Irresistible – The rise of addictive technology and the business of keeping us hooked. *Penguin Books.*

[63] Young, K. (2009). 'Understanding online gaming addiction and treatment issues for adolescents'. *The American Journal of Family Therapy*, 37(5), 355-372.

[64] Suler, J. (2005). 'The online disinhibition effect'. *International Journal of Applied Psychoanalytic Studies,* 2(2), 184-188.

[65] Robinson, E. (2013). 'Parental involvement in preventing and responding to cyberbullying'. *Family Matters*, 92, 68-76.

[66] Ferrara, P., Ianniello, F., Villani, A. & Corsello, G. (2018). 'Cyberbullying a modern form of bullying: let's talk about this health and social problem'. *Italian Journal of Pediatrics,* 44(14), 1-3.

[67] Srivastava, A., Gamble, R. & Boey, J.Y.M. (2013). 'Cyberbullying in Australia: Clarifying the problem, considering the solutions'. *International Journal of Children's Rights,* 21(1), 25-45.

[68] Australian Government Department of Communications. (June 2014). *Estimates of cyber-bullying incidents dealt with by Australian schools: Final Report*. Prepared by IRIS Research.

[69] eSafety Commissioner. (n.d.). *Cyberbullying.* Australian Government. https://www.esafety.gov.au/key-issues/cyberbullying.

[70] Jadambaa, A., Thomas, H.J., Scott, J.G., Graves, N., Brain, D. & Pacella, R. (2019). 'Prevalence of traditional bullying and cyberbullying among children and adolescents in Australia: A systematic review and meta-analysis'. *Australian & New Zealand Journal of Psychiatry,* 53(9), 878-888.

[71] Carlson, B. & Frazer, R. (2018). *Cyberbullying and Indigenous Australians: A review of the literature.* Macquarie University.

[72] Aboujaoude, E., Savage, M.W., Starcevic, V. & Salame, W.O. (2015). 'Cyberbullying: Review of an Old Problem Gone Viral'. *Journal of Adolescent Health,* 57(1), 10-18.

[73] Nilan, P., Burgess, H., Hobbs, M., Threadgold, S. & Alexander, W. (2015). 'Youth, Social Media, and Cyberbullying Among Australian Youth: "Sick Friends"'. *Social Media + Society* 1(2*),* 1-12.

[74] Kowalski, R.M., Giumetti, G.W., Shroeder, A. & Lattanner, M.R. (2014). 'Bullying in the Digital Age: A Critical Review and Meta-Analysis of Cyberbullying Research Among Youth'. *Psychological Bulletin,* 140(4), 1073-1137.

[75] Henry, N., Powell, A. & Flynn, A. (2017). *Not Just 'Revenge Pornography': Australians' Experiences of Image-Based Abuse: A Summary Report.* RMIT University.

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

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

[78] Amnesty International. (n.d.) 'Troll patrol findings: Using Crowdsourcing, data science and machine learning to measure violence and abuse against women on Twitter'. https://decoders.amnesty.org/projects/troll-patrol/findings.

[79] Powell, A. & Henry, N. (2015). *Digital Harassment and Abuse of Adult Australians: A Summary Report*. RMIT University.

[80] Powell, A., Scott, A.J. & Henry, N. (2020). 'Digital harassment and abuse: Experiences of sexuality and gender minority adults'. *European Journal of Criminology,* 17(2), 199-223.

[81] Dragiewicz, M., Burgess, J., Matamoros-Fernández, A., Salter, M., Suzor, N.P., Woodlock, D. & Harris, B. (2018). 'Technology facilitated coercive control: domestic violence and the competing roles of digital media platforms'. *Feminist Media Studies,* 18(4), 609-625.

[82] Chaulk, K. & Jones, T. (2011). 'Online Obsessive Relational Intrusion: Further Concerns About Facebook'. *Journal of Family Violence,* 26(1), 245-254.

[83] Reed, L.A., Tolman, R.M. & Ward, L.M. (2016). 'Snooping and Sexting: Digital Media as a Context for Dating Aggression and Abuse Among College Students'. *Violence Against Women,* 22(13), 1556-1576.

[84] Freed, D., Palmer, J., Minchala, D., Levy, K., Ristenpart, T. & Dell, N. (2017). 'Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders'. *Proceedings of the ACM on Human-Computer Interaction* 1(46), 1-22.

[85] Woodlock, D., McKenzie, M., Western, D. & Harris, B. (2020). 'Technology as a Weapon in Domestic Violence: Responding to Digital Coercive Control'. *Australian Social Work,* 73(3), 368-380.

[86] Douglas, H., Harris, B.A. & Dragiewicz, M. (2019). 'Technology-Facilitated Domestic and Family Violence: Women's Experiences'. *The British Journal of Criminology,* 59(3), 551-570.

[87] Hellevik, P.M. (2019). 'Teenagers' personal accounts of experiences with digital intimate partner violence and abuse'. *Computers in Human Behaviour,* 92(1), 178-187.

[88] Stark, E. (2012). 'Looking Beyond Domestic Violence: Policing Coercive Control'. *Journal of Police Crisis Negotiations,* 12(2), 199-217.

[89] Stonard, K.E., Bowen, E., Walker, K. & Price, S.A. (2017). '"They'll Always Find a Way to Get to You": Technology Use in Adolescent Romantic Relationships and Its Role in Dating Violence and Abuse'. *Journal of Interpersonal Violence,* 32(14), 2083-2117.

[90] eSafety Commissioner. (2020). *Children and technology-facilitated abuse in domestic and family violence situations: Summary report.* Australian Government. https://www.esafety.gov.au/about-us/research/children-and-technology-facilitated-abuse-domestic-and-family-violence-situations.

[91] Duerksen, K.N. & Woodin, E.M. (2019). 'Technological intimate partner violence: Exploring technology-related perpetration factors and overlap with in-person intimate partner violence'. *Computers in Human Behaviour,* 98(1), 223-231.

[92] Daspe, M.E., Vaillancourt-Morel, M.P., Lussier, Y. & Sabourin, S. (2018). 'Facebook Use, Facebook Jealousy, and Intimate Partner Violence Perpetration'. *Cyberpsychology, Behavior, and Social Networking,* 21(9), 549-555.

[93] Woodlock, D. (2017). 'The Abuse of Technology in Domestic Violence and Stalking'. *Violence Against Women,* 23(5), 584-602.

[94] McCauley, H.L., Bonomi, A.E., Mass, M.K., Bogen, K.W. & O'Malley, T.L. #MaybeHeDoesntHitYou: Social Media Underscore the Realities of Intimate Partner Violence. *Journal of Women's Health*. https://vaw.msu.edu/wp-content/uploads/2016/03/1-McCauley-JWH-2018-MaybeHeDoesntHitYou.pdf

[95] Carlyle, K.E., Guidry, J.P.D. & Burton, C. (2018). 'Recipes for Prevention: An Analysis of Intimate Partner Violence Messages on Pinterest'. *Journal of Interpersonal Violence,* 36(11-12), NP6106.

[96] Suzor, N., Dragiewicz, M., Harris, B., Gillett, R., Burgess, J. & Geelen, T.V. (2019). 'Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online'. *Policy & Internet,* 11(1), 84-103.

[97] Hinesley, K. (9 Semptember 2018). *Working together to prevent suicide*. Twitter. https://blog.twitter.com/en_au/topics/company/2018/working-together-to-prevent-suicide.html.

[98] Twitter Public Policy. (24 November 2020). *Our work to combat the 'shadow pandemic'*. Twitter.

https://blog.twitter.com/en_us/topics/company/2020/our-work-to-combat-the-shadow-pandemic.html.

[99] Davis, A. (19 December 2017). *New Tools to Prevent Harassment.* Facebook. https://about.fb.com/news/2017/12/new-tools-to-prevent-harassment/.

[100] Kahn, M. (19 December 2017). 'New Facebook Tools Give Women More Power Over Their Abusers'. *Elle*. https://www.elle.com/culture/a14461920/facebook-harassment-tools-domestic-violence/.

[101] ABC News. (2 July 2021). 'Julia Gillard joins prominent women in calling for Facebook, Google, Twitter and TikTok to stop social media abuse'. *ABC News*. https://www.abc.net.au/news/2021-07-02/julia-gillard-social-media-abuse-open-letter-facebook-twitter/100261524.

[102] Markus, A. (2019). *Mapping Social Cohesion: The Scanlon Foundation Surveys 2019*. Scanlon Foundation Research Institute.

[103] Scanlon Foundation Research Institute. (n.d.). *What is social cohesion?* https://scanloninstitute.org.au/what-social-cohesion.

[104] Gunn, E. & Roseberg, L.T. (2018). 'Trust in the Age of Social Media: Populist Politicians Seem More Authentic'. *Social Media + Society,* 4(1), 1-11.

[105] Tucker, J. G., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. & Nyhan, B. (March 2018). *Social Media, Political Polarization, and Political Disinformation: A review of the Scientific Literature.* William + Flora Hewlett Foundation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3144139.

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

[107] Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., Sonderman, J. & Loker, K. (2019). 'Who Shared It?: Deciding What News to Trust on Social Media'. *Digital Journalism,* 7(6), 783-801.

[108] Ceron, A. (2015). 'Internet, News, and Political Trust: The Difference Between Social Media and Online Media Outlets'. *Journal of Computer-Mediated Communication,* 20, 487-503.

[109] Starke, C., Marcinkowski, F. & Wintterlin, F. (2020). 'Social Networking Sites, Personalization, and Trust in Government: Empirical Evidence for a Mediation Model'. *Social Media + Society,* 6(2), 1-11.

[110] Bradshaw, S. & Howard, P.N. (2018). 'The Global Organization of Social Media Disinformation Campaigns'. *Journal of International Affairs,* 71(1.5), 23-31.

[111] Valenzuela, S., Halpern, D., Katz, J.E. & Miranda, J.P. (2019). 'The Paradox of Participation Versus Misinformation: Social Media, Political Engagement, and the Spread of Misinformation'. *Digital Journalism,* 7(6), 802-823.

[112] Mutz, D.C. (2002). 'Cross-cutting Social Networks: Testing Democratic Theory in Practice'. *American Political Science Review*, 96(1), 111-126.

[113] Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. Penguin Group; Rader, E. & Gray, R. (2015). 'Understanding user believes about algorithmic curation in the Facebook news feed'. In Begole, B. & Kim, J. (eds.) *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Republic of Korea.

[114] Nguyen, A. & Vu, H.T. (2019). 'Testing popular news discourse on the "echo chamber" effect: Does political polarisation occur among those relying on social media as their primary politics news source?'. *First Monday,* 24(5), 6.

[115] Klein, E. & Robinson, J. (2020). 'Like, Post, and Distrust? How Social Media Use Affects Trust in Government'. *Political Communication,* 37(1), 46-64.

[116] Lee, Y. (2020). 'Facebooking Alone? Millennials' Use of Social Network Sites and Volunteering'. *Nonprofit and Voluntary Sector Quarterly,* 49(1), 203-217.

[117] Theocharis, Y. (2015). 'The Conceptualization of Digital Networked Participation'. *Social Media + Society,* 1(2), 1-14.

[118] Lokot, T. (2018). '#IAmNotAfraidToSayIt: stories of sexual violence as everyday political speech on Facebook'. *Information, Communication & Society,* 21(6), 802-817.

[119] Skoric, M.M., Zhu, Q., Goh, D. & Pang, N. (2016). 'Social media and citizen engagement: A meta-analytic review'. *New Media & Society,* 18(9), 1817-1839.

[120] Halpern, D., Valenzuela, S. & Katz, J.E. (2017). 'We Face, I Tweet: How Different Social Media Influence Political Participation through Collective and Internal Efficacy'. *Journal of Computer-Mediated Communication*, 22(6), 320-336.

[121] Lane, D.S., Kim, D.H., Lee, S.S., Weeks, B.E. & Kwak, N. (2017). 'From Online Disagreement to Offline Action: How Diverse Motivations for Using Social Media Can Increase Political Information Sharing and Catalyze Offline Political Participation'. *Social Media + Society,* 3(3), 1-14.

[122] Theocharis, Y. & Lowe, W. (2016). 'Does Facebook increase political participation? Evidence from a field experiment'. *Information, Communication & Society,* 19(10), 1465-1486.

[123] Cappellini, B., Kravets, O. & Reppel, A. (2019). 'Shouting on social media? A borderscapes perspective on a contentious hashtag'. *Technological Forecasting and social change,* 145, 428-437.

[124] Pauwels, L., Brion, F., Schils, N., Laffineur, J., Verhage, A., De Ruyver, B. & Easton, M. (2014). *Explaining and understanding the role of exposure to new social media on violent extremism. An integrative quantitative and qualitative approach.* Academic Press.

[125] Williams, M., Burnap, P., Javed, A., Liu, H. & Ozalp, S. (2020). 'Hate in the machine: anti-black and anti-muslim social media posts as predictors of offline racially and religiously aggravated crime'. *British Journal of Criminology,* 60(1), 93-117.

[126] Sinpeng, A., Martin, F., Gelber, K. & Shields, K. (2021) *Facebook: Regulating hate speech in the Asia Pacific*. Department of Media and Communications, The University of Sydney & The School of Political Science and International Studies, The University of Queensland. https://ses.library.usyd.edu.au/bitstream/handle/2123/25116.3/Facebook_hate_speech_Asia_report_final_5July2021.pdf?sequence=3&isAllowed=y.

[127] Kaun, A. & Uldam, J. (2018). ''Volunteering is like any other business': Civic participation and social media'. *New Media & Society,* 20(6), 2186-2207.

[128] Lachance, E. (2020). 'COVID-19 and its impact on Volunteering: Moving Towards Virtual Volunteering'. *Leisure Sciences*, 43(1-2), 104-110.

[129] Silva, F., Proença, T. & Ferreira, M. (2018). 'Volunteers' perspective on online volunteering - a qualitative approach'. *International Review on Public and Nonprofit Marketing,* 15(4), 531-552.

130 Ackermann, K. & Manatschal, A. (2018). 'Online Volunteering as a means to overcome unequal participation? The profiles of online and offline volunteers compared'. *New Media & Society,* 20(12), 4453-4472.

131 Marineau, S. (2020). Fact check US: What is the impact of Russian interference in the US presidential election? *The Conversation*. https://theconversation.com/fact-check-us-what-is-the-impact-of-russian-interference-in-the-us-presidential-election-146711

132 Department of Home Affairs. (2020). *Countering foreign interference*. Australian Government. https://www.homeaffairs.gov.au/about-us/our-portfolios/national-security/countering-foreign-interference.

133 Dowling, M. (2020). 'The digital dilemma: Countering foreign interference'. *Stretton Institute, The University of Adelaide*. https://www.adelaide.edu.au/stretton/news/list/2020/11/03/the-digital-dilemma-countering-foreign-interference.

134 eSafety Commissioner. (n.d.). *Working with social media.* Australian Government. https://www.esafety.gov.au/about-us/consultation-cooperation/working-with-social-media.

135 eSafety Commissioner. (n.d.). *Time online.* Australian Government. https://www.esafety.gov.au/parents/big-issues/time-online.

136 Attorney-General's Department. (n.d.). *Abhorrent violent material.* Australian Government. https://www.ag.gov.au/crime/abhorrent-violent-material.

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

138 Frydenberg, J. & Fletcher, P. (2020). *News Media and Digital Platforms Mandatory Bargaining Code*. The Hon Josh Frydenberg MP: Treasurer of the Commonwealth of Australia. 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.

139 US Congress Hearing. (2021). Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation. https://www.congress.gov/event/117th-congress/house-event/111407

140 Committee on the Judiciary, Subcomittee on Privacy, Technology, and the Law, US senate. (2021). 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

141 De Vynck, G., Zakrzewski, C. & Dwoskin, E. (10 April 2021). *Big tech CEOs face lawmakers in House hearing on social media's role in extremism, misinformation.* The Washington Post. https://www.washingtonpost.com/technology/2021/03/25/facebook-google-twitter-house-hearing-live-updates/.

142 Goodman, E.P. & Whittington, R. (9 August 2019). *Section 230 of the Communications Decency Act and the Future of Online Speech* (Report No.20)*. The German Marshall Fun of

the United States. https://www.gmfus.org/news/section-230-communications-decency-act-and-future-online-speech.

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

[145] 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.

[146] Hatmaker, T. (6 February 2021). 'The SAFE TECH ACT offers Section 230 reform, but the law's defenders warn of major side effects'. *TechCrunch*. https://techcrunch.com/2021/02/05/safe-tech-act-section-230-warner/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS88&guce_referrer_sig=AQAAAK_1kWPdJZrtSdGq4U1bPw3cTXtx7mSHfZtpkVUbdVRUBvyIrITpop-GtBDjd9xYur9rKcIH_zXPMfjuSE72yaHNJdtXxPhztFH0X1JEWhHoyxP2aIepdEMUAqT5eBuiKOdVLBXiO1pOj8ESvgRhMhbjAfZJu2vvQ05KCeZv8_Lf.

[147] 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.

[148] Congressman Tom Malinowski, Representing the 7th District of New Jersey. (24 March 2021). *Reps. Malinowski and Eshoo reintroduce bill to hold tech platforms accountable for algorithmic promotion of extremism*. https://malinowski.house.gov/media/press-releases/reps-malinowski-and-eshoo-reintroduce-bill-hold-tech-platforms-accountable.

[149] 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.

[150] Senator Ed Markey. (2021). Sentors Markey and Blumenthal, Rep. Castor reintroduce legislation to protect children and teens from online manipulation and harm. https://www.markey.senate.gov/news/press-releases/senators-markey-and-blumenthal-rep-castor-reintroduce-legislation-to-protect-children-and-teens-from-online-manipulation-and-harm

[151] A Wall Street Journal Investigation (2021). The Facebook Files. *The Wall Street Journal.* https://www.wsj.com/articles/the-facebook-files-11631713039

[152] Subcommittee on Consumer Protection, Product Safety, and Data Security Hearing. (2021). Protecting Kids Online: Facebook, Instagram, and Mental Health Harms https://www.commerce.senate.gov/2021/9/protecting-kids-online-facebook-instagram-and-mental-health-harms

[153] Government of Canada. (2019). *Canada Declaration on Electoral Integrity Online.* https://www.canada.ca/en/democratic-institutions/services/protecting-democracy/declaration-electoral-integrity.html.

[154] Bryden, J. (27 May 2019). 'Several tech giants sign onto Canadian declaration on electoral integrity'. *Global News Canada.* https://globalnews.ca/news/5323084/tech-giants-electoral-integrity/.

[155] European Commission. (2021). *Code of Practice on Disinformation.* https://ec.europa.eu/digital-single-market/en/code-practice-disinformation.

[156] Wolford, B. (n.d.). 'What is GDPR, EU's new data protection law?' *GDPR.EU.* https://gdpr.eu/what-is-gdpr/.

[157] He, D. (2020). 'Governing Hate Content Online: How the *Rechtsstaat* Shaped the Policy Discourse on the NetzDG in Germany'. *International Journal of Communication,* 14(1), 3746-3768.

[158] Twitter Transparency. (2021). *Germany: Network Enforcement Act*. https://transparency.twitter.com/en/reports/countries/de.html.

[159] Republic of Singapore.. (2019). *Protection from Online Falsehoods and Manipulation Act 2019.* Singapore Statutes Online. https://sso.agc.gov.sg/Acts-Supp/18-2019.

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

[161] UK Government. (2020). *Fact sheet — Online Harms Full Government Response.* https://www.gov.uk/government/publications/fact-sheet-online-harms-full-government-response.