EVOLVING LEGAL RESPONSES TO SOCIAL MEDIA: LITIGATION, LEGISLATION AND SYSTEM ARCHITECTURE

MARCUS SMITH AND GREGOR URBAS*

I INTRODUCTION

Social media is an important form of technology that facilitates online social interaction, enabling users to post self-generated content such as text or photographs, create profiles, engage with others and form networks with people holding similar interests and views.[1] Online profiles and communication have become a central aspect of modern life, and their influence has grown over the COVID-19 period, when physical interaction became constrained during extended periods of lockdown in many countries. Social media is now also a major advertising medium for business.[2]

Social media companies are repositories of information, recording details of interactions, what individuals are interested in and the places they go—raising privacy and other regulatory concerns. Many social media companies are large and increasingly powerful. Applications such as Facebook, Twitter, Instagram, YouTube and WhatsApp are capable of transmitting information to millions of people instantly; Facebook alone has approximately three billion active monthly users worldwide.[3] These companies have become highly profitable and powerful due to the vast amount of data generated by their users and the value of this data for advertising purposes. Social media providers use 'attention based' business models, generating revenue through the sale of advertising opportunities to other businesses and trading in consumer data, including detailed user profiles and data-mining analysis, often using artificial intelligence (AI).[4] There have been a number of recent

* Marcus Smith (LLM, MPhil, PhD) is a graduate of the Australian National University and Cambridge University and an Associate Professor of Law at Charles Sturt University, Centre for Law and Justice.
Gregor Urbas (BA (Hons), LLB (Hons), PHD) is a graduate of the Australian National University and an Adjunct Associate Professor of Law at the Australian National University, College of Law.

[1] Jonathan Obar and Steven Wildman, 'Social Media Definition and the Governance Challenge: An Introduction to the Special Issue' (2015) 39 *Telecommunications Policy* 745.

[2] Ibid.

[3] Statistica, *Number of Monthly Active Facebook Users Worldwide* (Web Page, 2021) <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide>.

[4] See generally Jacob Lund Orquin and Michel Wedel, 'Contributions to Attention Based Marketing: Foundations, Insights, and Challenges' (2020) 111 *Journal of Business Research* 85.

legal developments in relation to the regulation of social media in Australia, including the High Court decision in *Fairfax Media Publications Pty Ltd v Voller* ('*Voller*'),[5] the Social Media (Anti-Trolling) Bill 2022 (Cth) and recommendations in relation to a registration system for social media account holders.[6]

This article examines the rapidly evolving law relating to social media in Australia, considering present regulation, laws and the prospect of future system architecture to verify the identity of social media account holders as a means of better addressing the complex issues associated with social media. The article discusses whether combining legislation and regulatory constraints imposed through system architecture should be considered in light of the scale, global impact and technical sophistication of social media technology. The article begins with a consideration of self-regulation, followed by litigation and legislative reforms, before finally reflecting on whether social media could be regulated through a future identity verification requirement.

II SELF-REGULATION

As with many new forms of technology, the legal status and regulation of social media is comparatively underdeveloped, and its scale and complexity mean that regulation is challenging. The efficiency with which individuals can anonymously disseminate information using social media infrastructure has become an increasingly significant phenomenon. The influence and power of social media companies to affect people's lives, businesses and governments, along with the associated potential for social media sites to be used to disseminate misinformation, has prompted calls for more stringent regulation.

The term 'fake news' has gained traction as a description of deliberately false or heavily biased media reports, especially in the context of political discussion. Fake news has been defined as:

> [F]abricated information that mimics news media content in form but not in organizational process or intent. Fake-news outlets, in turn, lack the news media's editorial norms and processes for

[5] Fairfax Media Publications Pty Ltd v Voller; Nationwide News Pty Limited v Voller; Australian News Channel Pty Ltd v Voller [2021] HCA 27 ('Voller').
[6] House Standing Committee on Social Policy and Legal Affairs, Parliament of Australia, *Inquiry into Family, Domestic and Sexual Violence* (Report, March 2021) ('Inquiry into Domestic Violence Report').

> ensuring the accuracy and credibility of information ... It is particularly pernicious in that it is parasitic on standard news outlets, simultaneously benefiting from and undermining their credibility.[7]

Such reports may seem credible, indeed indistinguishable from reports published by reputable news sources, but consumers are rarely in a position to establish the provenance of material that flashes across their social media pages to establish its veracity. Social media platforms and smartphones have made it cheap and simple to create content, and for that content to be effectively and widely disseminated. The spread of misleading or inaccurate information or theories can be extremely pervasive when combined with social media platforms, particularly through automated dissemination and AI. The misinformation extends to conspiracy theories about governments and their programs, false rumours about public figures and pseudoscientific therapies for health issues (including in relation to vaccination), and is often a vehicle for social engineering to achieve political objectives. The efficiency with which social media can disseminate information is now well understood, following its use in association with big data analytics by the former consultancy firm Cambridge Analytica to inform online advertising strategies for the Republican Party during the 2016 presidential election campaign in the United States. In association with poll results and other intelligence, the strategy sought to identify and understand individuals in key electorates and then use social media advertisements specifically tailored to target their personality and social views to influence their vote.[8]

Subsequent analysis has shown that fake news on social media was a significant, though not overwhelming, source of information for voters on both sides of the United States political divide, with readers more likely to believe stories aligning with their own political views.[9] A further example of social media being used to disseminate misinformation occurred in relation to the COVID-19 pandemic. The World Health Organization Director-General referred to 'fighting an infodemic' in

[7] David Lazer et al, 'The Science of Fake News' (2018) 359 *Science* 1094, 1094.

[8] The firm was later dissolved after criticism about the legality of hiring the firm for the presidential campaign in light of prohibitions on the involvement of foreign citizens in United States election campaigns and whether the scale of the activity had compromised the integrity of the election itself. In 2019, Facebook was fined US$5 billion over its management of user data following inquiries into the arrangement: Julia Carrie Wong, 'The Cambridge Analytica Scandal Changed the World – But it Didn't Change Facebook', *The Guardian* (online, 18 March 2019), <https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook>.

[9] Hunt Allcott and Matthew Gentzkow, 'Social Media and Fake News in the 2016 Election' (2017) 31(2) *Journal of Economic Perspectives* 211.

relation to the nature of the virus, how the pandemic began and the risks associated with vaccines.[10] Misleading information in relation to the COVID-19 pandemic, which has involved automated dissemination using botnets, includes that it was engineered as a bioweapon by China and is transmitted by 5G technology.[11]

The largest social media companies are based in the United States but operate globally and have historically been self-regulating. Platforms such as Facebook publish policies outlining the kinds of content or use that are considered unacceptable, such as violence and incitement, promotion of criminal groups or activities, and fraud or deception.[12] Given the complexities of Internet regulation, due to its international accessibility across traditional jurisdictional boundaries, and the absence of significant external regulation in general, evaluation and removal of offending content is largely left to social media providers and the community of social media users. However, Facebook has received increasing scrutiny from regulators in several countries over its management of user data, following inquiries into its relationship with data-mining operations such as Cambridge Analytica.[13]

Social media companies have developed in-house capabilities to remove content that is considered harmful in response to this scrutiny. In 2020, Facebook updated its Coordinating Harm and Publicizing Crime policy, committing to prohibiting 'people from facilitating, organizing, promoting, or admitting to certain criminal or harmful activities targeted at people, businesses, property or animals'.[14] YouTube's Transparency Report provides data on the number and breakdown of videos removed after being flagged for attention:

> YouTube relies on teams around the world to review flagged videos and remove content that violates our Community Guidelines; restrict videos (e.g., age-restrict content that may not be appropriate for all audiences); or leave the content live when it doesn't violate our guidelines.[15]

[10] Salman Bin Naeem, Rubina Bhatti and Aqsa Khan, 'An Exploration of How Fake News is Taking Over Social Media and Putting Public Health at Risk' (2020) 38 *Health Information and Libraries Journal* 1.
[11] I Ullah et al, 'Myths and Conspiracy Theories on Vaccines and COVID-19: Potential Effect on Global Vaccine Refusals' (2021) 22 *Vacunas* 93.
[12] Facebook, *Community Standards* (Web Page) <https://www.facebook.com/communitystandards/>.
[13] Wong (n 8).
[14] Facebook (n 12).
[15] YouTube, *Community Guidelines Enforcement* (Web Page) <https://transparencyreport.google.com/youtube-policy/removals?hl=en>.

Figures for the January to March 2021 period indicate that six million videos were removed by YouTube, the majority through automated flagging and smaller numbers through user or non-government or government agency flagging.[16] Controversially, Twitter permanently suspended the account of then–US President Donald Trump in response to allegations that he had used the platform to incite the violent protests at the Capitol Building on 6 December 2021, leading to heated debate about whether his free speech rights had been violated by a private corporation.[17] Generally, self-regulation by social media platforms does not provide any guarantee that material will be removed in a transparent and systematic way, free from personal or corporate political biases. The purchase and associated privatisation of Twitter by Elon Musk in 2022 followed public debate about free speech online and the decisions of social media companies to censor posts and de-platform users, including political figures. This development is likely to further disrupt the sector and increase public debate on the issue (at the time of writing, the sale of Twitter had been approved by the company's board but remained subject to shareholder and regulatory agency approval).[18]

Automated flagging involves the use of AI-based filters that incorporate machine learning to detect posts that breach prohibitions against harmful or illegal content, often referred to as 'hate speech'. In the second half of 2020, it was reported that Facebook's use of AI for flagging such content had increased significantly:

> From July to September, Facebook's AI tools proactively detected 94.7% of the hate speech removed by the company, up from 80.5% in the same period last year, Facebook said. The social network attributed the uptick to improvement in its automated tools, including better training of the machines. In the third quarter, Facebook took action against 22.1 million pieces of content for hate speech. The company's photo service, Instagram, took action against 6.5 million pieces of hate speech content.[19]

Coupled with increased pressure on social media platforms to retain and disclose identifying information of users (eg, when required by court orders), this use of system architecture represents

[16] Ibid.
[17] Patrick Ganninger, 'Freedom of Tweets: The Role of Social Media in a Marketplace of Ideas' (2021) *SLU Law Journal Online* 63.
[18] Mike Isaac and Lauren Hirsch, 'Elon Musk and Twitter Reach Deal for Sale', *New York Times* (online, 25 April 2022) <https://www.nytimes.com/live/2022/04/25/business/elon-musk-twitter>.
[19] Queenie Wong, 'Facebook's AI is Flagging More Hate Speech Before You Report It', *CNet Tech* (online, 20 November 2020) <https://www.cnet.com/tech/services-and-software/facebooks-ai-is-flagging-more-hate-speech-before-you-report-it/>.

a significant response to the proliferation of harmful or objectionable content. Despite such measures, it is unlikely that self-regulation alone will be an adequate response in the longer term, given the scale of social media in the dissemination of news and other information globally.[20] Attempting to fit social media within a more traditional governance regime, developed over many decades for media such as newspapers, radio and television, is unlikely to be effective and may have unintended consequences.[21] Without committing to either of the labels of 'platform' or 'publisher', the role of social media companies in providing a linkage between users and content makes it more accurate to describe them as 'intermediaries':

> By calling them intermediaries, let's recognise that social media platforms are fundamentally in the middle – that is, they mediate between users who produce content and users who might want it. That makes them similar to not only search engines and ISPs, but also traditional media. They too face a regulatory framework designed to oversee how they mediate between producers and audiences, between speakers and listeners. Social media platforms are not only in the middle between user and user, and user and public but between citizens and law enforcement, policymakers, and regulators charged with governing their behaviour.[22]

The developments discussed above have also led to increased scrutiny of social media companies, and both the public and government regulators are becoming more aware of how these companies utilise the vast amount of personal information they generate. Some companies claim to be reducing the amount of data they collect. For example, in early 2020, Google announced that it had begun blocking third-party cookies (data from websites that track users' viewing history) in its web browser, which will limit the ability of companies like Facebook to track users and offer tailored advertising. This is likely to affect Facebook's business model and impact its earnings.[23]

In countries such as the United States and Australia, an array of statutory obligations imposes specific reporting and monitoring obligations. However, there remains a lack of clarity regarding to what extent online providers are required to actively seek out prohibited content so as to block and

[20] See generally Phillip Napoli, 'Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers' (2015) 39 *Telecommunications Policy* 751.
[21] Terry Flew et al, 'Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance' (2019) 10 *Journal of Digital Media and Policy* 33.
[22] Tarleton Gillespie, 'Regulation of and by Platforms' in Jean Burgess (ed), *The Sage Handbook of Social Media* (Sage, 2018).
[23] Marcus Smith and Gregor Urbas, *Technology Law: Australian and International Perspectives* (Cambridge University Press, 2021).

report it. Australian regulators have highlighted existing legislation, which provides that carriage service providers must 'do the provider's best to prevent telecommunications networks and facilities from being used in, or in relation to, the commission of offences against the laws of the Commonwealth or of the States and Territories'.[24] In practice, social media companies are required to work with law enforcement, either on request or under their own policies. Over time, as social media has become more widely used, influential and understood, there has been a shift away from self-regulation as a regulatory model, with courts and legislatures increasingly willing to take action using traditional legal approaches.

III LITIGATION

There has been a number of recent examples of litigation involving social media companies in recent years. The Australian Information Commissioner commenced proceedings against Facebook in 2020 in the Federal Court, and this litigation is ongoing.[25] The key allegation is that Facebook has breached Australian privacy law by allowing third parties to harvest the personal information of hundreds of thousands of Australian Facebook users without their consent, along with those in other countries, which was subsequently disclosed to groups (including Cambridge Analytica).

There have also been several defamation cases addressing the issue of anonymous individuals posting information in the form of comments and reviews online. The key issue here is who should be held accountable: the individuals posting the comments, the individual that created the social media page that the comments are posted on, or the social media company itself. Where the identity of a defamatory poster is unknown to the potential plaintiff, discovery proceedings against a social media company may be initiated. In a Federal Court ruling in February 2020, Google was ordered to disclose information that could identify a poster using the online handle 'CBsm23' and who allegedly defamed a Melbourne dentist.[26] In another proceeding involving alleged defamatory

[24] *Telecommunications Act 1997* (Cth) s 313(1). Subsection 313(3), which allows blocking of Internet content at the request of law enforcement agencies, was the subject of a parliamentary inquiry that reported in 2015: Standing Committee on Infrastructure and Communications, Parliament of Australia, *Inquiry into the Use of Subsection 313(3) of the Telecommunications Act 1997 by Government Agencies to Disrupt the Operation of Illegal Online Services* (Report, 1 June 2015).

[25] *Australian Information Commission v Facebook Inc* [2020] FCA 531. See also Office of the Australian Information Commissioner, 'Statement on Facebook Proceedings' (Media Release, 22 April 2020) <https://www.oaic.gov.au/updates/news-and-media/statement-on-facebook-proceedings/>.

[26] *Kabbabe v Google LLC* [2020] FCA 126 (12 February 2020). See also *Smith v Jones* [2020] NSWDC 262, a case involving malicious and defamatory Google reviews posted about a solicitor.

reviews about an Australian lawyer, legal action was taken against Google to force disclosure of the identity of an online reviewer who the plaintiff suspected was a competitor in the legal sector.[27]

A noteworthy defamation proceeding in mid-2020 culminated in the award of A$875,000 to a federal parliamentarian, Anne Webster, her general practitioner husband and their charity supporting new mothers.[28] The plaintiffs sued the author of a number of highly defamatory Facebook posts, which included bizarre allegations that the politician was a member of a paedophile cabal. The Federal Court judge noted the role of social media in spreading the malicious allegations:

> Fortunately for the Websters, their long lives of decency and good deeds, coupled with the incoherence of much of Ms Brewer's messages, make it reasonably unlikely that any but the most suggestible individuals would think the less of them as a result of Ms Brewer's publications. However, social media has provided Ms Brewer with a platform by which she is able to reach suggestible individuals who may believe her claims. As the applicants observed, not all community members know the Websters personally or by their positive reputation. Ms Webster's Facebook page has several thousands of followers. Consistent with her stated aim, Ms Brewer's defamatory publications have spread along the grapevine into the Mildura community.[29]

The proceedings also involved preliminary injunctions against the defendant, restraining her from continuing her Facebook publications.[30] However, Webster was also dissatisfied with Facebook's delay in taking down the offending posts, prompting the parliamentarian to push for new legislation setting out minimum expected standards of response to complaints about defamatory material.[31]

In late 2021, the High Court published its decision in the *Voller* case.[32] This case was concerned with whether media companies, such as *The Australian* newspaper, that posted their news content on their Facebook page, and allowed comments about those articles, were liable for defamatory posts made by third parties on the Facebook page that they controlled. The decision,

[27] Danny Tran, 'Gangland Lawyer Zarah Garde-Wilson Launches Court Action to Unmask Google Reviewer', *Australian Broadcasting Corporation* (online, 20 February 2020) <https://www.abc.net.au/news/2020-02-20/gangland-lawyer-zarah-garde-wilson-court-action-against-google/11982866>.
[28] Webster v Brewer (No 3) [2020] FCA 1343.
[29] Ibid [39] (Gleeson J).
[30] Webster v Brewer [2020] FCA 622; Webster v Brewer (No 2) [2020] FCA 727.
[31] Social Media (Basic Expectations and Defamation) Bill 2021 (Cth), introduced into the Australian Parliament on 25 October 2021, discussed further below.
[32] *Voller* (n 5).

which has potential implications for all individuals and organisations that maintain websites and social media pages, maintained that they were. An extract from Gageler and Gordon JJ (in the majority) observes that the 'advent of the Internet has resulted in a "disaggregation" of the process of publication and has facilitated a shift from "one-to-many" publication to "many-to-many" publication'.[33] In rejecting the publishers' argument, they referred to the benefit gained by the publishers in using Facebook to disseminate their news product:

> The primary judge found that over 15 million Australians are Facebook users. The appellants chose to operate public Facebook pages in order to engage commercially with that significant segment of the population.[34] ... Having regard to those findings, the appellants' attempt to portray themselves as passive and unwitting victims of Facebook's functionality has an air of unreality. Having taken action to secure the commercial benefit of the Facebook functionality, the appellants bear the legal consequences.[35]

The High Court decided that it was the news organisations, in posting the articles on Facebook and allowing the defamatory comments of the third parties, that were liable for defamation, rather than the social media company, Facebook:

> Each appellant became a publisher of each comment posted on its public Facebook page by a Facebook user as and when that comment was accessed in a comprehensible form by another Facebook user. Each appellant became a publisher at that time by reason of its intentional participation in the process by which the posted comment had become available to be accessed by the other Facebook user. In each case, the intentional participation in that process was sufficiently constituted by the appellant, having contracted with Facebook for the creation and ongoing provision of its public Facebook page, posting content on the page the effect of which was automatically to give Facebook users the option (in addition to 'Like' or 'Share') to 'Comment' on the content by posting a comment which (if not 'filtered' so as to be automatically 'hidden' if it contained 'moderated words') was automatically accessible in a comprehensible form by other Facebook users.[36]

[33] Ibid 86.
[34] Ibid 100.
[35] Ibid 102.
[36] Ibid 98.

This outcome may be fair in the context of this case and in light of factors such as the resources of large media companies and the benefit they gain in using Facebook's services, but the anonymity of individuals posting comments, the scale of dissemination afforded by social media platforms and Internet-based discussion forums, and the internationalisation of the Internet means that this approach may not be feasible in addressing the issue more broadly and in the longer term. Anonymity is a major contributing factor—individuals are more likely to be disrespectful, abusive and defamatory and promulgate false assertions if their comments are not attributable. These examples highlight the challenges for liberal democracies and the maintenance of effective legal systems, in light of the changes social media has made to human communication in the modern world. While online anonymity provides for greater personal freedom and rapid exchange of ideas, over time, it will likely be necessary for governments to provide greater regulation in this area.

IV LEGISLATION

A recent development that is likely to be viewed as more positive for traditional media organisations in Australia is the recently enacted legislation that seeks to support the business models of these companies in the face of declining revenues resulting from an increasing number of people obtaining their news from social media, including news that was originally researched and reported by traditional media companies. Social media has disrupted the way information is disseminated, contributing to a decline in the profitability of traditional news media organisations, and Australia is now at the forefront of social media regulation internationally. In 2020, the federal government introduced laws limiting the extent to which social media companies can freely distribute news reporting through the *Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Act 2021* (Cth). Under a code of conduct developed by the Australian Competition and Consumer Commission, companies such as Facebook and Google are required to pay traditional news media companies to publish their content on their platforms.[37] The government stated that the code 'reflects the importance of a diverse and well-resourced news media sector to our democracy and the Australian people' and was implemented to:

[37] Georgia Hitch, 'Facebook, Google to be Forced to Pay for News as Part of New Mandatory Code of Conduct to Support Traditional News Media', *Australian Broadcasting Corporation* (online, 31 July 2020) [1] <https://www.abc.net.au/news/2020-07-31/draft-mandatory-code-conduct-facebook-google-pay-for-news/12510776>.

> ensure that news media businesses are fairly remunerated for the content they generate, helping to sustain public interest journalism in Australia … [and] provide a framework for good faith negotiations between the parties and a fair and balanced arbitration process to resolve outstanding disputes.[38]

Most of the laws that have been enacted in relation to content posted on social media sites to date are criminal law legislation seeking to addressing online bullying and harassment facilitated by social media platforms. Recent survey data indicates that one in three Australians has experienced some form of online harassment, affecting their health, safety and productivity.[39] An enormous amount of vitriol is likely attributable to online anonymity, and this content can be widely observed on the comments pages of public websites. In the social media context, the term 'trolling' is used to describe these types of comments, ranging from seemingly genuine contributions that subvert the flow of interactions to outright abuse.[40]

Online harassment may target an individual, a group or an entire race, ethnicity, religion or sexual orientation. Legal recourse is often difficult as the offenders are anonymous, but where an offender can be identified and is within jurisdiction, then there are laws that can be used to prosecute them in Australia. For instance, under Commonwealth law, the following section relating to online harassment is applicable:

> Using a carriage service to menace, harass or cause offence
>
> (1) A person commits an offence if:
>
> (a) the person uses a carriage service; and
>
> (b) the person does so in a way (whether by the method of use or the content of a communication, or both) that reasonable persons would regard as being, in all the circumstances, menacing, harassing or offensive.

[38] Treasurer of Australia, 'Parliament Passes News Media and Digital Platforms Mandatory Bargaining Code' (Media Release, 25 February 2021) [3]–[4] <https://ministers.treasury.gov.au/ministers/josh-frydenberg-2018/media-releases/parliament-passes-news-media-and-digital-platforms>.

[39] The Australia Institute, 'Online Harassment and Cyberhate Costs Australians $3.7b' (Media Release, 28 January 2019) [2] <https://australiainstitute.org.au/post/online-harassment-and-cyberhate-costs-australians-3-7b/>.

[40] See generally Hannah Barton, 'The Dark Side of the Internet' in Irene Connoly et al (eds), *An Introduction to Cyberpsychology* (Routledge, 2016) 58.

Penalty: Imprisonment for 3 years.[41]

The meaning of 'offensive' is explicated in an earlier provision:

Determining whether material is offensive

> (1) The matters to be taken into account in deciding for the purposes of this Part whether reasonable persons would regard particular material, or a particular use of a carriage service, as being, in all the circumstances, offensive, include:
>
> > (a) the standards of morality, decency and propriety generally accepted by reasonable adults; and
> >
> > (b) the literary, artistic or educational merit (if any) of the material; and
> >
> > (c) the general character of the material (including whether it is of a medical, legal or scientific character).[42]

An example of these provisions being applied to online trolling is the case of ***R v Hampson***, in which the defendant had posted a range of highly offensive material, some of which included child sexual exploitation, on social media pages commemorating deceased children. He was sentenced to imprisonment and ordered to undergo psychiatric assessment.[43] Remarks by the Court of Appeal judges noted not only the depraved nature of the offender's Facebook trolling but also that 'it interfered with the legitimate use of the Internet by members of the public'.[44]

While there are some legislated requirements for online service providers to report suspected illegal content, such as child exploitation material, to the police, there is no general requirement to do so in relation to offensive material.[45] Offences relating to the use of online services for illegal or offensive purposes usually provide a 'safe harbour' for Internet service providers and other intermediaries so long as they operate purely in the technical capacity of a provider rather than

[41] *Criminal Code Act 1995* (Cth) s 474.17(1) ('*Criminal Code Act*'). In 2018, an aggravated form of the offences was added as s 474.17A, with a five year maximum where the commission of the offence involved the misuse of 'private sexual material', and a three year maximum if this was after civil penalty orders had already been made. See also *Enhancing Online Safety (Non-consensual Sharing of Intimate Images) Act 2018* (Cth).

[42] *Criminal Code Act* (n 41) s 473.4. Note that this standard also applies to the definition of 'child abuse material' under the Code, so that not every explicit image is prohibited (eg, content in an online medical course).

[43] ***R v Hampson*** [2011] QCA 132, 1–2. The initial three-year sentence for two s 474.17 offences was reduced on appeal.

[44] Ibid [36] (Muir JA).

[45] *Criminal Code Act* (n 41) s 474.25.

content originator or moderator.[46] Arguments against the imposition of broader monitoring and reporting obligations rely on their characterisation as intermediaries, analogous to a postal service, rather than as publishers. However, this position of neutrality comes into conflict with demands that those who facilitate harmful speech by others bear some responsibility for its harmful effects.

Unfortunately, abuse on the basis of ethnicity, religion, gender and sexuality is prevalent on some websites and in social media. Online 'hate speech' contributes to real-world crimes, ranging from violent assaults to acts of terrorism.[47] Balancing legitimate freedom of expression against the protection of minorities within society is a difficult task. Contemporary debates tend to focus on what kind and level of harm is sufficient to merit interference by the state, particularly in the form of criminal sanction. While some criminal offences for vilification exist, other laws impose aggravated sentences for hate-motivated crimes or include such acts within statutory definitions of unlawful discrimination.[48]

Incitement to violence is a criminal act in most jurisdictions, whether this occurs online or otherwise, but fewer countries have laws prohibiting content that depicts or describes violent acts. Content regulation in countries such as Australia is generally achieved through classification systems and conditions on broadcasting and other telecommunications licensing schemes, rather than criminal offences. However, some more extreme material has come to be prohibited, with responsibility increasingly falling on social media providers to remove the material expeditiously or face substantial fines.

The mass shooting of dozens of people in two mosques in New Zealand in early 2019, live-streamed on Facebook by the offender, who also posted his intentions and beliefs on online forums before the attacks, highlighted the problem:

> The world got a terrible reminder of how flawed existing social-media policies and algorithms are for policing violent and offensive content. In the days before the shooting, the perpetrator apparently boasted of his plans and posted an online manifesto. He then broadcast the horrific act live on Facebook. The attack left 49 people dead and dozens more injured. Over the past 18

[46] Ibid s 473.5.
[47] Abbee Corb, 'Online Hate and Cyber-Bigotry: A Glance at our Radicalized Online World' in Nathan Hall et al (eds), *Routledge International Handbook on Hate Crime* (Routledge, 2015).
[48] In Australia, for example, websites denying the Holocaust have been found to be in breach of the *Racial Discrimination Act 1975* (Cth). See, eg, *Jones v Toben* [2002] FCA 1150.

> months, following harassment and fake-news scandals, social-media companies have invested heavily in content moderators. But this did little to stop video of the shooting from spreading. Not only was the live stream reportedly up for 20 minutes, but the resulting video was then reposted on YouTube, with some clips remaining up for over an hour.[49]

The Australian Government subsequently announced laws relating to 'abhorrent violent material', defined as audio and/or visual material that records or streams acts and is produced by a person involved in the acts, being acts of terrorism, murder, torture, rape or kidnap.[50] The offences that were enacted do not target these acts directly (as they are already criminalised under existing laws) but impose criminal liability on content and hosting services, including social media providers, for failure to notify and expeditiously remove such content.[51] This approach signifies a shift in the legal position of Internet and social media providers.

The provisions set out the obligations of Internet service providers, content service providers and hosting service providers, where 'content service' includes social media providers.[52] Obligations include reporting abhorrent violent material to police and expeditiously removing or ceasing to host the material if it is reasonably capable of being accessed within Australia. Penalties for failure to comply with these obligations can amount to millions of dollars or a sum of 10% of annual turnover for corporate entities.[53] Defences are provided for law enforcement and other official dealings in such information, for journalistic and research uses, and for ensuring implied freedom of political communication is preserved.[54]

In association with this type of legislation, the Australian Government has established the Office of the eSafety Commissioner, which also has a role in relation to regulating bullying, harassment, and the posting of abhorrent or violent content on social media sites and the Internet

[49] Will Knight, 'The Mass Shooting in New Zealand Shows How Broken Social Media Is', *MIT Technology Review* (online, 15 Match 2019) [1]–[3] <https://www.technologyreview.com/2019/03/15/65970/the-mass-shooting-in-new-zealand-shows-how-broken-social-media-is/>.

[50] *Criminal Code Act* (n 41) ss 474.30–474.45, added by the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) with effect from 5 April 2019. Further provisions, ss 474.46–474.48, were added with effect from 19 September 2019 to prohibit online incitement of trespass to agricultural land (eg, during protests) under the *Criminal Code Amendment (Agricultural Protection) Act 2019* (Cth).

[51] Criminal Code Act (n 41) ss 474.30–474.45.

[52] Ibid s 474.30, defining 'content service' and 'hosting service'.

[53] *Criminal Code Act* (n 41) ss 474.33, 474.34.

[54] Ibid ss 474.37, 474.38.

more broadly.[55] Legislative powers have been provided to facilitate removal of this form of content from the Internet and impose financial penalties on social media companies for non-compliance, for example, in relation to 'cyber-bullying material', defined as material likely to have the effect of 'seriously threatening, seriously intimidating, seriously harassing or seriously humiliating'.[56] In March 2021, the House Standing Committee on Social Policy and Legal Affairs made a number of recommendations directed at further increasing the regulation of social media companies, including that:

> There should be greater acknowledgement that appropriate technology use is a shared community responsibility. It is not simply a responsibility of platforms to host and police content[;]
>
> There should be greater clarity around a platform's obligation to remove content, including through the Online Safety Act ... [; and]
>
> There should be a substantial increase in criminal and civil penalties for technology-facilitated abuse to act as a greater deterrent for errant behaviour.[57]

More stringent legislation was subsequently passed by the Australian Parliament in the form of the *Online Safety Act 2021* (Cth), increasing what is required of online service providers and enhancing the powers of the eSafety Commissioner. These include content removal notices, content blocking notices and link deletion notices, backed up with more significant penalties if service providers fail to comply. The eSafety Commissioner is now able to enforce civil penalties of up to A$555,000 for body corporates and issue infringement notices, enforceable undertakings and injunctions.[58]

The Social Media (Basic Expectations and Defamation) Bill 2021 was introduced into the Australian Parliament in October 2021 by Anne Webster MP. As discussed above, she has been a victim of serious online defamation. The Bill would give the responsible minister the power to set minimum expectations for social media providers and a role for the eSafety Commissioner to

[55] See Australian Government, eSafety Commissioner (initially established as the Children's eSafety Commissioner) (Web Page) <https://www.esafety.gov.au/>.
[56] Enhancing Online Safety Act 2015 (Cth) pt 1-5.
[57] Inquiry into Domestic Violence Report (n 6) 164–5.
[58] *Online Safety Act 2021* (Cth) pt 10. In April 2022, the European Union also upgraded its legislative response in this area, agreeing to introduce the Digital Services Act. It imposes a range of more stringent standards on social media and other online companies, including in relation to taking down illegal content, regulating advertising aimed at children and obligations for third-party sellers.

respond to complaints and report on compliance with the basic expectations. In relation to allegedly defamatory material, the Bill's innovative approach would impose an obligation on social media providers hosting the material to remove it within 48 hours of being notified, or face the consequence that the provider is statutorily declared to be co-liable for defamation.[59]

Finally, then–Australian Prime Minister Scott Morrison announced an aggressive new approach to respond to anonymous online trolling, whereby complainants would have new means to discover the identity of alleged offenders:

> The reforms will give victims of defamatory online comments two ways to unmask trolls and resolve disputes. First, global social media platforms will be required to establish a quick, simple and standardised complaints system that ensures defamatory remarks can be removed and trolls identified with their consent. This recognises that Australians often just want harmful comments removed. Second, a new Federal Court order will be established that requires social media giants to disclose identifying details of trolls to victims, without consent, which will then enable a defamation case to be lodged. Importantly, the reforms will also ensure everyday Australians and Australian organisations with a social media page are not legally considered publishers and cannot be held liable for any defamatory comments posted on their page, providing them with certainty.[60]

Then–Attorney-General Michaelia Cash stated that this was in response to the decision in the *Voller* case, which found that Australians who maintain social media pages can be 'publishers' of defamatory comments made by others on social media, stating: 'This is not fair and it is not right. Australians expect to be held accountable for their own actions but shouldn't be made to pay for the actions of others that they cannot control'.[61]

In December 2021, a Draft Exposure Social Media (Anti-Trolling) Bill 2021 (Cth) was released, which, as foreshadowed, statutorily declares that a social media service is to be taken to be a publisher of an online comment posted to that service and thus capable of being sued in defamation, but with a defence where end-user information is disclosed by the service to the person

[59] Social Media (Basic Expectations and Defamation) Bill 2021 (Cth) cl 28.
[60] Hon Scott Morrison MP, Prime Minister and Senator the Hon Michaelia Cash, Attorney-General, 'Combatting Online Trolls and Strengthening Defamation Laws' (Media Release, 28 November 2021) [11]–[14] <https://www.pm.gov.au/media/combatting-online-trolls-and-strengthening-defamation-laws>.
[61] Ibid [17].

claiming to have been defamed.[62] The mechanism for such disclosure is a court order called an 'end-user information disclosure order', which is to be granted to an applicant who is unable to ascertain the identity of the maker of a defamatory post, whether the post is made in Australia or overseas. The court is to make an order unless doing so is likely to present a risk to the safety of the post's maker.[63]

> The Draft Bill provides a simplified outline as follows:
>
> For the purposes of the general law of the tort of defamation:
>
> (a) an Australian person who maintains or administers a page of a social media service is taken to not be a publisher of a third party comment posted on the page; and
>
> (b) if a comment is posted on a page of a social media service (and the comment is made in Australia), the provider of the social media service is taken to be a publisher of the comment.
>
> If the provider of a social media service is a publisher of a comment posted on a page of the social media service, the provider of the social media service has a defence in a defamation proceeding relating to the comment if certain conditions are satisfied.
>
> If a person has posted a comment on a page of a social media service, an application may be made to a court for an end-user information disclosure order that requires the provider of the social media service (or the provider's nominated entity) to disclose the commenter's relevant contact details or country location data to the applicant for the order.
>
> If the provider of a social media service is a foreign body corporate and the service has at least 250,000 Australian account-holders (or the service is specified in the legislative rules), the provider of the social media service must have a nominated entity in Australia.[64]

The intent of paragraphs (a) and (b) above is to clarify who is a publisher for the purposes of a defamation action and which defences apply. This is addressed in the Explanatory Notes:

> This subclause [ie paragraphs (a) and (b)] addresses the implications of the *Voller* decision. It overrides part of that decision, by providing that end-users who maintain or administer pages on

[62] Consideration of the Bill carried over to the following calendar year, and it subsequently became the Social Media (Anti-Trolling) Bill 2022.
[63] Exposure Draft: Social Media (Anti-Trolling) Bill 2021 (Cth) cl 18 <https://www.ag.gov.au/system/files/2021-11/social-media-anti-trolling-bill-2021-exposure-draft.PDF>.
[64] Explanatory Notes, Exposure Draft: Social Media (Anti-Trolling) Bill 2021 (Cth) 6.

> social media (as page owners) are not taken to be publishers of third-party comments on their pages, and clarifies that decision by providing that the provider of the social media service is a publisher of such a comment if it is made in Australia. The provider may, in any case, be a publisher as a consequence of the High Court's reasoning in *Voller*. The subclause clarifies that this is definitely the case, and avoids the need for a complainant to have to prove this point by applying the test espoused in *Voller*. The status of the comment originator under general law as a 'publisher' is not affected.[65]

Having provided that social media services are publishers for the purpose of Australian defamation proceedings, and having displaced defences available both under section 235 of the *Online Safety Act 2021* (Cth) and the defence of innocent dissemination (whether under general law or the law of a State or Territory), the Bill then provides a statutory defence through disclosure of end-user information. This puts social media services in an interesting position:

> At its core, and except where an applicant chooses to not progress past relevant points in the process, this mechanism ensures an applicant will either have the relevant contact details in order to bring defamation proceedings against the originator, or will be able to bring defamation proceedings against the social media provider. A consequence of this mechanism is that a social media provider may be unable to disclose the relevant contact details, which could occur where, for example, the originator has not consented to disclosure or where the social media provider simply does not have those details at its disposal. In these circumstances, the provider will not have access to the defence. This ensures an applicant will have an available respondent to bring defamation proceedings against.[66]

To avoid potential liability in defamation, a provider will have to ensure both that it collects identifying information of its users (eg, through a robust subscription process) and will hand over this information when served with an 'End-user information disclosure order'. This may give rise to difficulties with any privacy undertakings contained in the terms of service offered to new or existing users, but it is to be noted that these usually stipulate that the service will disclose information

[65] Ibid.
[66] Ibid 7–8.

required under a court order or other legal process requiring disclosure (eg, warrants, subpoenas and production orders).[67]

The Senate referred the Bill for review, and it was not subsequently passed prior to the dissolution of the Parliament; however, the Senate Legal and Constitutional Affairs Committee reported on the Bill in March 2022.[68] Amendments to the Bill were recommended by government senators, in relation to the circumstances in which a social media service provider or page owner could be found liable for defamation, and to protect a poster's safety where it may be at risk as a result of an end-user information disclosure order. The Opposition senators were highly critical, asserting that the Bill:

> appears to reflect an assumption that a remedy for this problem can be achieved by ad hoc adjustments to defamation law. This attitude ignores the fact that trolling on social media is a pernicious problem created by profound cultural and technological changes. Labor members of this committee believe that an effective remedy to the problem must be part of broader reform of the existing regulatory regime for the media, including social media.[69]

This is a complex issue to address, and many potential limitations of the Bill were aired in the inquiry report, including that there would be significant privacy impacts in requiring social media companies to collect and maintain personal details of their users to a greater extent than they presently do. Further, the Bill may also impose an unrealistic regulatory burden on relevant companies to monitor content on their pages, given the increasingly widespread use of social media in the community. The inquiry heard from witnesses that had been significantly impacted by online trolling and had subsequently initiated defamation proceedings. They submitted that defamation proceedings were time consuming, expensive and emotionally draining, and that it was unlikely most Australians would be able to afford them.[70]

---

[67] See, eg, *Facebook Data Policy* (Web Page, 2021) <https://m.facebook.com/about/privacy/update?_rdr#legal-requests-prevent-harm>.

[68] Senate, Legal and Constitutional Affairs Legislation Committee, Parliament of Australia, *Social Media (Anti-Trolling) Bill 2022 [Provisions]* (Report, March 2022) ('*Social Media (Anti-Trolling) Bill 2022 Report*'). A House Select Committee also reported in March 2022, examining a broader range of issues relating to online safety, including education, privacy, system design, legislative frameworks and related international developments: House Select Committee on Social Media and Online Safety, Parliament of Australia, *Select Committee on Social Media and Online Safety* (Report, March 2022).

[69] *Social Media (Anti-Trolling) Bill 2022 Report* (n 68) 50.

[70] Ibid 57–8.

While the proposed legislation provides an innovative approach in the field of defamation law, there are questions regarding whether it would effectively address the problem of online trolling and the serious impact this can have on people's lives. Not every social media user who is aggrieved by a negative post will have the resources or motivation to pursue a defamation action, when in most instances all that is wanted is removal of the offending comment. Further, even very harmful posts, such as those urging others to harm or kill themselves, are not in themselves defamatory in the absence of specific claims that damage reputation. Therefore, existing alternatives, such as complaints to the eSafety Commissioner and criminal liability for use of a carriage service in a manner that offends against the *Criminal Code Act 1995* (Cth), are necessary but remain insufficient. Similarly, the defamation reforms in the Bill may assist in some specific instances, though there may be unintended effects. A further interesting but unresolved question, not considered by the House Standing Committee on Social Policy and Legal Affairs, is whether the newfound status of a social media service as a publisher of material posted by others might give rise to novel arguments of accessorial liability for the criminal transgressions of some users.[71]

V SYSTEM ARCHITECTURE

Thus far, three regulatory approaches have been examined, all responding to issues that have arisen in relation to social media: self-regulation, litigation and legislation. There is a further measure, not yet implemented, that is more proactive and could potentially address the anonymity of social media, which is at the heart of many of the issues that have been canvassed: system architecture to implement an identity verification requirement. As it is likely that there will be technical limitations to the effectiveness of the above-discussed legislation, further reforms of this type may eventually be implemented.

In 2021, the House Standing Committee on Social Policy and Legal Affairs of the Australian Parliament proposed (as has been discussed in other countries around the world) that to use a social media account, it be compulsory for users to provide evidence of their identity, such as a copy of a passport or drivers' licence.[72] The objective of such an approach is to address the anonymity problem that contributes to many of the issues described above. People using anonymous accounts to harass

[71] Noting especially ss 473.5 and 474.17 of the *Criminal Code Act* (n 41), discussed earlier.
[72] Inquiry into Domestic Violence Report (n 6) xxxi.

and abuse online and undertake technology facilitated abuse would be deterred if such actions could be more directly and more easily ascribed to an individual.[73] In combination with legislative developments, identity verification could play an important role in preventing these issues from continuing to create significant problems for contemporary society and the legal system. As has been discussed, the spectrum ranges from defamation to vitriolic comments, hate speech and the potential spread of these views and incitement of serious violent actions, of which the above-discussed New Zealand right-wing terrorist act is a poignant example.[74]

There is an argument that 'social media is too powerful now to be anonymous' and that just as identification and registration are required to drive a car or own a firearm, so should they be required to operate a social media account.[75] While there have been increasing powers granted to government agencies such as the Office of the eSafety Commissioner to remove content and block accounts, there are growing calls for more to be done. The recommendation that a registration system be implemented, requiring individuals to verify their identity to obtain a social media account, was made with a view to both deterring and, where necessary, identifying individuals to facilitate investigation and prosecution of offences. It has been proposed that the federal government introduce laws requiring that 100 points of identification[76] be required to obtain, or maintain an existing, social media account:

> In order to open or maintain an existing social media account, customers should be required by law to identify themselves to a platform using 100 points of identification, in the same way as a person must provide identification for a mobile phone account, or to buy a mobile SIM card. Social media platforms must provide those identifying details when requested by the eSafety Commissioner, law enforcement or as directed by a court.[77]

The introduction of laws requiring compulsory identity verification for social media account holders would be complex and controversial and, to date, have not been widely discussed. However, they should be examined further as they are likely to provide effective deterrence for online

[73] Ibid.
[74] See generally Barton (n 40).
[75] See generally Will Burns, 'Is it Time to Require Identity Verification for Everyone Using Social Media?' *Forbes* (online, 2 February 2018) <https://www.forbes.com/sites/willburns/2018/02/22/is-it-time-to-require-identity-verification-for-everyone-using-social-media/?sh=5193bc198683>.
[76] Under the current Australian system, a passport is valued at 70 points and a driver's licence is valued at 40 points.
[77] Inquiry into Domestic Violence Report (n 6) xxxi.

harassment and abuse, hate speech and misinformation; facilitate improved investigation and prosecution; and complement legislative reform and litigation through the courts. However, there are potential issues with the approach that require further examination. Regarding data security, if identity documents, such as copies of passports and drivers licences, were provided to multinational technology companies such as Google and Facebook, which already have a great deal of personal data about users' online and real-world behaviour,[78] this would be a security concern and would increase the level of risk associated with the detailed information that social media companies already hold about individuals. There would need to be confidence that data security risks could be adequately mitigated before implementation.

However, using system architecture as a form of regulation, particularly in combination with legislation, is likely to be more effective than legislation alone, due to the complexity of information and communication technologies, the pace at which they evolve (especially in comparison with the time taken to develop and implement new laws), the internationalisation of the technology sector and the Internet, and the lack of understanding of technology and cybersecurity in the community. Regulatory theorists refer to 'law' imposed by technological capabilities and system designs, in place of, or in combination with, legally proscribing activities with legislation.[79] There are already examples of such approaches being developed in Australia and other countries. System architecture to regulate and facilitate smart contracts and digital currencies is being implemented by the Australian Government to provide the foundation for blockchain to become a mainstream part of the future private sector, providing authentication, security and auditability for digital currency transactions and throughout the lifecycle of contracts.[80] A consortium of the federal government and private sector is establishing the Australian National Blockchain to enable businesses to digitally manage contracts, exchange information and conduct authentication.[81]

[78] See Alex Druce, '"It's a Long Bow": Social Media ID Push Dubbed Ineffective, a Privacy Risk', *Sydney Morning Herald* (online, 2 April 2021) <https://www.smh.com.au/politics/federal/it-s-a-long-bow-social-media-id-push-dubbed-a-privacy-risk-20210402-p57g7d.html>.

[79] See, eg, Lawrence Lessig, *Code: Version 2.0* (Basic Books, 2006). For an overview of regulatory theory in technology law, see Smith and Urbas (n 23). See also Marcus Smith, *Technology Law: Cases, Commentary and Materials* (LexisNexis, 2022) ch 2.

[80] Marcus Smith, 'A Modern Approach to Regulation: Integrating Law, System Architecture and Blockchain Technology in Australia' (2020) 48 *Australian Business Law Review* 460.

[81] Department of Industry, Science, Energy and Resources (Cth), *National Blockchain Roadmap* (Report, February 2020). See also ibid.

Relevant work is also currently being undertaken by the Digital Transformation Agency in Australia to expand the digital identity system and create a trusted digital identity framework. Initial work has focused on digital identity in relation to government identity documents and service provision, and it will be some time until this approach becomes established and is expanded to private sector organisations such as social media companies. Integration within a broader digital identity system of this type, to store identity documents and manage identity security for social media users, would ensure a deterrent effect in relation to defamatory, misleading, abusive or criminal activity on social media. A statutory authority could be established that individuals register with, which verifies their identity to a social media company when the individual seeks to establish an account. The authority would keep the identity documents secure and only release details of the account holder's identity in specific circumstances, such as for legal proceedings or under warrant in a law enforcement investigation. The technical approach adopted—such as whether the system is centralised, federated or decentralised, and whether blockchain, a similar distributed ledger technology, or another form of escrow system arrangement is used for security—will be vital to the feasibility of this approach.[82] Further research and evaluation from both the technical and legal perspectives are required prior to such a system being implemented.

VI CONCLUSION

This article has discussed a range of responses to social media, including self-regulation, litigation, legislation and system architecture. Social media has been widely embraced around the world and provides great benefit by increasing human connectedness and as a vehicle for sharing ideas and opinions in real time. While social media will continue to have a positive influence on many aspects of society, it is also a vehicle for bullying, harassment, vilification and incitement of hate crimes, and

[82] Andrej Zwitter, Oskar Gstrein and Evan Yap, 'Digital Identity and the Blockchain: Universal Identity Management and the Concept of the Self-Sovereign' (2020) 3 *Individual Frontiers in Blockchain* 1, 7:

> *Centralized* identity systems are where a single organization establishes and manages identity and are typical for the direct relationship between a state and the individual; *Federated* identity systems, where different public and private institutions establish stand-alone systems; *Decentralised* identity systems, where the individual is at the center and institutions or private corporations just add (verified) credentials to a central 'identity hub', 'application', or 'vault' that is controlled by the individual.

See also Seumas Miller and Terry Bossomaier, *The Ethics of Cybersecurity* (Oxford University Press, 2022).

has created complexities for civil justice from the perspective of defamation law. Social media companies have implemented moderation policies to deal with fake news and extreme content. There has been significant case law in Australia in recent years relating to social media, as well as a steady stream of legislation, first in relation to criminal law, such as abhorrent and violent content and online safety, and, more recently, proposed civil law legislation addressing the issue of defamation.

It is apparent that piecemeal measures to regulate social media, even working in combination, will be insufficient and may in fact create further problems. There is no doubt that the challenge of responding to social media is exacerbated, like many areas of technology law, by the international operation of the companies involved. Moreover, the spread of users around the world, able to set up accounts in minutes and interact in real time, is a difficult environment to regulate. Black letter law alone will be insufficient; measures incorporating system architecture, such as identity verification, will be a necessary adjunct. Further work, involving consultation with users, and the collaboration of technology experts, lawyers, ethicists and regulators is needed to investigate the use of system architecture to facilitate an effective and secure identity verification system that can be integrated with complementary measures.

Australia is already a leading jurisdiction internationally in implementing social media law reform. As it did with its news media bargaining code, which successfully required technology companies to pay news organisations for content and led the way for laws that were subsequently proposed in the United States, United Kingdom and Canada, Australia can also lead the world in online safety and effective regulation of social media, drawing on litigation, legislation and system architecture.