GRIFFITH LAW REVIEW
2022, VOL. 31, NO. 4, 513–536
https://doi.org/10.1080/10383441.2022.2138140

Regulating disinformation on Twitter and Facebook

Corinne Tan

Division of Business Law, Nanyang Technological University, Singapore

ABSTRACT

The spread of disinformation in recent years has caused the international community concerns, particularly around its impact on electoral and public health outcomes. When one considers how disinformation can be contained, one often looks to new laws imposing more accountability on prominent social media platforms. While this narrative may be consistent with the fact that the problem of disinformation is exacerbated on social media platforms, it obscures the fact that individual users hold more power than is acknowledged and that shaping user norms should be accorded high priority in the fight against disinformation. In this article, I examine selected legislation implemented to regulate the spread of disinformation online. I also scrutinise two selected social media platforms – Twitter and Facebook – to anchor my discussion. In doing so, I consider what these platforms have done to self and co-regulate. Thereafter, I consider the limitations on regulation posed by certain behavioural norms of users. I argue that shaping user norms lie at the heart of the regulatory approaches discussed and is pivotal to regulating disinformation effectively.

KEYWORDS

Regulation; fake news; disinformation; social media; norms; Twitter; Facebook

1. Introduction

Structural changes in news content and delivery have caused almost every government across the world to be concerned with 'fake news'.[1] A survey conducted in 2017 found that 28.6% of the survey's respondents received news primarily online, either from social media platforms or websites.[2] In 2020, this percentage increased significantly, in part due to the COVID-19 pandemic increasing news consumption across both traditional sources of news as well as online news sources.[3] Social media is also one of the main sources of information for audiences worldwide on the Russian invasion of Ukraine in 2022.[4]

CONTACT Corinne Tan corinne.tan@ntu.edu.sg Nanyang Business School (Division of Business Law), Nanyang Technological University, 50 Nanyang Avenue, 639798 Singapore

[1]Frank Fagan, 'Optimal social media content moderation and platform immunities' (2020) 50 *European Journal of Law and Economics* 437.

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

[3]See, eg, Nic Newman et al, *Reuters Institute Digital News Report 2020* (Reuters Institute and the University of Oxford, 2020).

[4]See, eg, Collette Snowden, 'Guns, tanks and Twitter: how Russia and Ukraine are using social media as the war drags on' (The Conversation, 5 April 2022) <https://theconversation.com/guns-tanks-and-twitter-how-russia-and-ukraine-are-using-social-media-as-the-war-drags-on-180131>.

Media scholars have often described 'fake news' as containing two types of information: misinformation and disinformation.[5] While *misinformation* has been defined by some as simply comprising any information that is incorrect, regardless of intent, *disinformation* is characterised as the purposeful dissemination of false reports intended to mislead the public.[6] In general, 'fake news' which warrants concern needs to be differentiated from critical speech constituting opinions, even if the opinions are biased or exaggerated.[7] In this article, I am less concerned with intent and consider *both misinformation* and *disinformation* that can cause public harm, by way of threatening democratic political process, health, environment or security.[8] I will thus refer to false information as *disinformation*, *false information* or *false news* generally, given the negative connotations of the term 'fake news' due to its overuse by politically interested parties to dismiss narratives inconsistent with theirs.[9] In doing so, I acknowledge that there are challenges in defining disinformation – whether due to the sharing of characteristics with other categories of harmful content such as hate speech and terrorist incitement, *or* in relation to establishing intent and proving harm.

Disinformation is not a novel problem.[10] In recent times, as profits can readily be made from authoring and circulating false news, there is incentive to create and circulate such news. The 2016 United States (US) presidential election and its outcome is often discussed together with the impact of disinformation and commentators commonly deplore the effect of disinformation across the internet on American politics and the public.[11] There have also been dangerous consequences resulting from the proliferation of false news, such as a gunman showing up at pizzeria to liberate children he believed Hilary Clinton was holding hostage based on a false news article which was widely circulated.[12] To exacerbate these problems, disinformation is readily proliferated on social media platforms due to the speed of dissemination and wide reach of these platforms, particularly since the platforms adopt business models which rely on scraping as much data as possible from the online activities of their users to earn advertising revenue from the attention economy.[13] It has been suggested that social media platform Facebook influenced outcomes in national elections worldwide simply because users did not read

[5]See, eg, Amy Kristin Sanders and Rachael L. Jones, 'Clicks at Any Cost: Why Regulation Won't Upend the Economics of Fake News (2018) 2 *Business Entrepreneurship and Tax Law Review* 343.

[6]See, eg, Government of the United Kingdom (UK), *Online Harms White Paper* (2019) <https://www.gov.uk/government/consultations/online-harms-white-paper>; Digital, Culture, Media and Sport Committee, *Disinformation and 'fake news': Final Report* (2019) <https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf>; Sanders and Jones, above n 4; Dean Jackson, *Issue Brief: Distinguishing Disinformation from Propaganda, Misinformation and Fake News* (National Endowment for Democracy, 17 October 2017) < https://www.ned.org/issue-brief-distinguishing-disinformation-from-propaganda-misinformation-and-fake-news/>.

[7]See, eg, Annie C Hundley, 'Fake News and the First Amendment: How False Political Speech Kills the Marketplace of Ideas' (2017) 92 *Tulane Law Review* 502.

[8]This is how 'disinformation' is defined in the European Commission, *Code of Practice on Disinformation* (European Commission, September 2018).

[9]See, eg, Joshua Habgood-Coote, *The term "fake news" is doing great harm* (The Conversation, 27 July 2018) <https://theconversation.com/the-term-fake-news-is-doing-great-harm-100406>.

[10]See, eg, Arianna Huffington and Ari Emanuel, *Fake News: A New Name for an Old Problem* (Huffington Post, 21 December 2016) <https://www.huffpost.com/entry/fake-news-a-new-name-for-an-old-problem_b_585acd94e4b0eb586484eab2>.

[11]See, eg, Andrea Butler, 'Protecting the Democratic Role of the Press: A Legal Solution to Fake News' (2018) 96 *Washington University Law Review* 419; Kevin Wagner and Jason Gainous, 'Trending Politics: How the internet has changed political news coverage' in David Taras and Richard Davis (eds), *Electoral Campaigns, Media, and the New World of Digital Politics* (University of Michigan Press, 2022) 44.

[12]See, eg, ibid.

[13]See, eg, Lili Levi, 'Real Fake News and Fake Fake News' (2017) 16 *First Amendment Law Review* 290.

past headlines and try to confirm information before passing false news on.[14] In addition, social networks potentially allow users to be split into echo chambers of like-minded people with similar views that reinforce their own biases.[15] Moreover, the structural design of the economy comprising social media platforms leads such platforms to design algorithms to keep users scrolling, posting and commenting for as long as possible, by way of displaying content curated to entertain each user so as to encourage their engagement.[16] Specifically, the display of disinformation and extremist content can worsen the polarisation of users.[17]

Research on the regulation of disinformation has mainly centred its discussions around how regulating disinformation could conflict with constitutionally protected free speech rights encapsulated within the US First Amendment[18] and how different countries and regions such as the European Union (EU) implement laws and institute measures to regulate disinformation.[19] There is also research exploring patterns of news and media consumption[20] and how these impact on electoral outcomes.[21] There is room for further research which evaluates the effectiveness of laws, platform initiatives and user norms on disinformation in a summative manner.

This article seeks to understand the broader regulatory perspective first. I start off with providing an overview of the global approach towards disinformation, through looking at the legislation on intermediary liability in selected jurisdictions, such as in the US, Singapore, Australia and Germany. The purpose of this is not to provide an exhaustive or comprehensive analysis with respect to each jurisdiction. Instead, it is to illustrate with examples direct regulation (or regulation via laws) and to outline the key features of the legislation implemented in some jurisdictions under which online platforms would be liable if they do not act to curb the spread of disinformation in a timely manner. Additionally, I will discuss the key features of the codes of practice adopted by online platforms (including Twitter and Facebook) as a collective to regulate disinformation.

[14]See, eg, Jacob Finkel et al., *Fake News & Misinformation Policy Practicum* (Stanford Law School – Law and Policy Lab, October 2017).

[15]See, eg, Murdoch Watney, *The Legal Position of Social Media Intermediaries in Addressing Fake News* (European Conference on Cyber Warfare and Security, Reading, June 2018). It is noted in recent studies, however, that the threat posed by echo chambers (or filter bubbles) has been overblown, and personalisation by algorithms of news accessible by users is smaller than often assumed, see, eg, Efrat Nechushtai and Seth C. Lewis, 'What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations' (2019) 90 *Computers in Human Behaviour* 298; Mario Haim, Andreas Graefe and Hans-Bernd Brosius, 'Burst of the Filter Bubble?' (2018) 6(3) *Digital Journalism* 330; and Axel Bruns, 'Filter bubble' (2019) 8(4) *Internet Policy Review* 1.

[16]See, eg, Clara Hendrickson and William A. Galston, *Big tech threats: Making sense of the backlash against online platforms* (Brookings, 28 May 2019) <https://www.brookings.edu/research/big-tech-threats-making-sense-of-the-backlash-against-online-platforms/>; Finkel et al., above n 14.

[17]See, eg, ibid.

[18]See, eg, Jill I Goldenziel and Manal Cheema, 'The New Fighting Words: How U.S. Law Hampers the Fight against Information Warfare' (2019) 22 *U Pa J Const L* 107; Jessica Stone-Erdman, 'Just the (Alternative) Facts, Ma'am: The Status of Fake News under the First Amendment' (2017) 16 *First Amendment Law Review* 43; Fernando Nunez, 'Disinformation Legislation and Freedom of Expression' (2020) 10 *UC Irvine L. Rev.* 783.

[19]See, eg, Oreste Pollicino and Elettra Bietti, 'Truth and Deception across the Atlantic: A Roadmap of Disinformation in the US and Europe' (2019) 11 *Italian J. Pub. L.* 43; Kenny Chng, 'Reflections on thinking about the POFMA' (Symposium on POFMA, 2019) <https://ink.library.smu.edu.sg/sol_research/2986/>; Andrea Carson and Liam Fallon, Fighting Fake News: A Study Of Online Misinformation Regulation in the Asia Pacific (La Trobe University, 2021); Antonios Kouroutakis, 'EU Action Plan Against Disinformation: Public Authorities, Platforms and the People' (2020) 53(2) *International Lawyer* 1.

[20]See, eg, Nic Newman et al., above n 3.

[21]See, eg, Edda Humprecht, Frank Esser and Peter Van Aelst, 'Resilience to Online Disinformation: A Framework for Cross-National Comparative Research' (2020) 25(3) *The International Journal of Press/Politics* 493.

Next, I highlight and discuss: the main purposes and inherent features on the selected platforms; the policies against disinformation and the tools available to users on each platform; as well as how health information relating to the COVID-19 pandemic has been regulated. For the purpose of the article, I have chosen to look at the two social media platforms Twitter and Facebook, to delineate how social networking platforms deal with the challenges posed by disinformation. In this respect, I acknowledge that there are many other online platforms including Reddit, Google (i.e. in particular Google News, the Google search engine and YouTube), Instagram, et cetera, that should play their part in containing disinformation. I have, however, chosen not to look at these other platforms in order to confine the scope of the article.

Thereafter, I consider certain behavioural norms of users on online platforms that pose challenges to the forms of regulation discussed. Among other forms of regulation, I argue that shaping user norms lies at the heart of and is most crucial to regulating disinformation effectively.

2. Overview of laws against disinformation

Here, I consider the extent of intermediary liability where there is a spread of disinformation through looking at legislation in selected jurisdictions such as the US, Singapore, Australia and Germany. I will start off with the US, given that this is where the companies operating the selected platforms are registered. The laws implemented in other jurisdictions also have an impact on the way the platforms operate, as users of online platforms – including the selected social networking sites studied here – are located worldwide. In this respect, I have chosen to discuss laws enacted in recent years in jurisdictions such as Singapore, Australia and Germany, as they address online disinformation to varying extents more directly.

In the US, the spreading of politically divisive content or even blatant disinformation by Americans is constitutionally protected free speech under the First Amendment.[22] This protection is supported by the theory on the marketplace of ideas – all ideas, including false ones, should be available to the community, moreover, false information would somewhat be weeded out through exposure to the truth over time.[23] As such, there is no specific legislation targeting disinformation. On the contrary, the broad immunity conferred under the Communications Decency Act (CDA)[24] for defamation has been said to contribute to creating an environment where disinformation, alternative facts or plain lies are ubiquitous on online platforms and largely unchecked.[25] These platforms are immune from any liability so long as they did not author the defamatory material in the first place. The platforms have frequently disclaimed editorial control over content shared by their users and benefitted from the complete immunity offered under the

[22]United States Constitution amend I. See also Nicole Perlroth, *A Former Fox News Executive Divides Americans Using Russian Tactics* (The New York Times, 21 November 2019) <https://www.nytimes.com/2019/11/21/technology/LaCorte-edition-news.html>.

[23]Stanley Ingber, 'The Marketplace of Ideas: A Legitimizing Myth' (1984) *Duke Law Journal* 1, 2–3. It is noted that this theory is imperfect as its assumptions are that people in the marketplace must be able to distinguish between truth and falsehood *and* must actually be searching for the truth, see, eg, Butler, above n 11; Goldenziel and Cheema, above n 18.

[24]Communications Decency Act (CDA) of 1996 (United States), § 230.

[25]See, eg, Butler, above n 11, 435.

CDA to internet distributors.[26] The extent to which this immunity shields online platforms is illustrated by the recent case of *Nunes v Twitter Inc,*[27] where it was held by the court that lawsuits seeking to hold platforms like Twitter liable for exercising a publisher's conventional editorial functions (i.e. whether to withdraw, publish or alter content) were barred. In earlier decisions, the courts in the US did not attempt novel interpretations of § 230 of the CDA to ascertain if it applies to new services offered by online platforms (for example, one providing online dating services[28] and another helping people find roommates through their preferences[29]), but instead chose to retrofit the original provision under the CDA to their rulings.[30] These decisions therefore affirmed the broad applicability of a wide immunity conferred on internet service providers.[31] The First Amendment, together with the CDA, thus create strong barriers to the statutory and judicial regulation of false news and largely allow online platforms to avoid legal responsibility.[32] In spite of this, it is generally recognised that there is a need to regulate online platforms including social media platforms more so than in the past, given their political, cultural and social influence,[33] and how they can 'nudge' users through their technological features to create and disseminate content online.[34]

When jurisdictions grant online platforms immunity for content generated by their users, there is wide discretion on the part of these platforms to make private decisions on content – the platforms adopt a private lawmaking function and can decide on the types of speech to suppress.[35] On the other hand, other jurisdictions, including Singapore, Australia and Germany, explicitly hold online platforms accountable to varying extents for the speech of their users, via legislation such as Singapore's Protection from Online Falsehoods and Manipulation Act (POFMA),[36] Australia's Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act (Australian Criminal Code Amendment),[37] and Germany's Network Enforcement Act (NetzDG).[38]

In addition, I will discuss the key features of the codes of practices adopted by online platforms as a collective, including Twitter and Facebook, for the purpose of regulating disinformation.

[26]See, eg, ibid 436. Cf Eric Goldman, 'Of Course the First Amendment Protects Google and Facebook (and It's Not a Close Question)' in David E. Pozen et al., *Straining (Analogies) to Make Sense of the First Amendment in Cyberspace* (Columbia University Press, 2020) 146, 148. The platforms (for example, Facebook) have also said the opposite, subject to who the audience is. Regardless of what is claimed, platforms such as Facebook do make editorial decisions in choosing content to publish and withdraw. See also Ashutosh Bhagwat, 'Do platforms have editorial rights?' (2021) 1 *Journal of Free Speech Law* 97.

[27]*Devin Nunes v Twitter Inc,* Case CL19-1715-00 (United States Circuit Court, Virginia, 24 June 2020).

[28]*Anthony v Yahoo! Inc*., 421 F. Supp. 2d 1257 (N.D. Cal. 2006); 376 Fed. Appx. 775 (9th Cir. 2010).

[29]Fair Hous. Council of San Fernando Valley v Roommates.com, LLC, 521 F. 3d 1157 (9th Cir. 2008).

[30]Dallas Flick, 'Combatting Fake News: Alternatives to Limiting Social Media Misinformation and Rehabilitating Quality Journalism' (2017) 20 *SMU Science and Technology Law Review* 385.

[31]Ibid.

[32]See, eg, ibid 375.

[33]See, eg, ibid 393.

[34]See the discussion in Corinne Tan, *Regulating Content on Social Media: Copyright, Terms of Service and Technological Features* (UCL Press, 2018), where the concept of 'nudge' discussed in Richard H. Thaler and Cass R. Sunstein, *Nudge* (Penguin, 2009) is applied to social media platforms.

[35]See Fagan, above n 1, 439.

[36]Protection from Online Falsehoods and Manipulation Act (No. 18 of 2019) (Singapore) (POFMA).

[37]Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (Cth) (Australian Criminal Code Amendment).

[38]Network Enforcement Act (valid as from 1 October 2017) (Germany) (NetzDG).

2.1. Singapore's POFMA

Prior to the POFMA being enacted in Singapore, there were criminal, judicial and executive levers that could be used to counter online falsehoods.[39] Existing legislation[40] can be applied to online falsehoods, albeit not being drafted specifically for this purpose. Notwithstanding this applicability, in reality, there will be limitations of scope, speed and adaptability,[41] particularly in respect of the removal of the relevant falsehoods from access online. Indeed, prior to the enactment of the POFMA, it was not clear if an online social media platform such as Twitter or Facebook could face executive action such as an imposed prohibition against broadcasting or the cancellation of a license from the Infocomm Media Development Authority (IMDA) under existing legislation – this is as the platforms may not fall under the definition of an 'internet content provider'.[42] With the POFMA, platforms such as Twitter and Facebook would fall clearly within the definitions of 'internet intermediaries' providing 'internet intermediary services'.[43] The doing of any act for the purpose of, or incidental to, the provision of an internet intermediary service is, however, exempt from the prohibitions against the communication of false statements of fact as well as the provision of services for such communication.[44] Furthermore, if directed by the Minister, the IMDA may issue access blocking orders to disable the access of users in Singapore to relevant online locations with subject matter contravening the provisions under the POFMA, in the event that internet intermediaries fail to comply with the directions and orders issued. These orders can be issued to either the internet access service providers or internet intermediaries, as the case may be.[45] Arguably, the POFMA is mainly directed at the actual persons who communicate false statements (directly or via bots) and news platforms.[46] Internet intermediaries such as the social media platforms this article is concerned with are given a wide berth to operate *as is*, so long as they comply with orders issued to disable access to online locations identified under the POFMA. In the event of non-compliance by internet intermediaries, not only could such intermediaries face having the relevant internet access service providers cut off access to online locations via their platforms,[47] they could also face potential fines.[48] The POFMA has been criticised for its wide coverage of many types of falsehoods[49] which could deter legitimate speech and public debate around government policy, as well as expressions of controversial opinions.[50]

[39]Goh Yihan, 'Written Representations to the Select Committee on Deliberate Online Falsehoods – Effectiveness of Current Legislative Tools' (Parliament of Singapore, 7 March 2018) <https://www.parliament.gov.sg/docs/default-source/sconlinefalsehoods/written-representation-129.pdf>.
[40]See, eg, Telecommunications Act (Cap 323, 2000 rev ed) (Singapore), s 45; Internal Security Act (Cap 143, 1985 rev ed) (Singapore), s 26; Sedition Act (Cap 290, 1948) (Singapore), ss 4 and 10; Protection from Harassment Act (Cap 256A, 2014) (Singapore), s 15; Penal Code, (Cap (Cap 224, 2008 rev ed) (Singapore), ss 298 and 298A.
[41]See Goh, above n 39.
[42]Broadcasting Act (Cap 28, 2012 rev ed), ss 12–16, together with Condition 16 of the Schedule in the Notification. See also Goh, above n 39.
[43]POFMA, above n 36, s 2.
[44]Ibid ss 7(4) and 9(4) respectively.
[45]Ibid ss 28, 33, 34 and 43.
[46]Ibid ss 7–9.
[47]Ibid ss 28(2), 33(3), and 43(2).
[48]Ibid s 34(5).
[49]Ibid ss 7, 8 and 9.
[50]See, eg, Natalie Leal, *Controversial fake news law passed in Singapore* (Global Government Forum, 20 May 2019).

The POFMA Office, in exercise of the powers conferred under the POFMA discussed above, has issued three Codes of Practices for three targeted objectives, namely: firstly, the Code of Practice for Transparency of Online Political Advertisements which provides that prescribed intermediaries must implement measures to disclose information on online political advertisements targeted at end-users in Singapore as well as provide an annual report on such measures;[51] second, the Code of Practice for Giving Prominence to Credible Online Sources of Information which stipulates that prescribed intermediaries must put in place measures to prioritise and increase the visibility of relevant information, as well as furnish an annual report on the same;[52] and third, the Code of Practice for Preventing and Countering Abuse of Online Accounts which provides that prescribed intermediaries must implement measures to reduce the likelihood of inauthentic online accounts being used to engage in malicious activities as well as submit an annual report on the measures taken[53] (collectively, the Singapore Codes). The companies operating the selected platforms examined in this article – Twitter and Facebook – are included in the list of prescribed intermediaries that have to comply with the three Singapore Codes.[54] It is noted, however, that the measures that are required are '*reasonable* due diligence measures' in all three codes,[55] hence giving some breadth of flexibility and subjectivity to the relevant platforms to ascertain what is reasonable in line with their differences in operations.

2.2. Australian Criminal Code Amendment

In Australia, the Australian Criminal Code Amendment requires hosting and content services to remove violent content in an expeditious manner.[56] This legislation is narrow in scope, as the definition of 'abhorrent violent material' includes mainly the recording of abhorrent violent conduct, which is further defined as engagement in terrorism, murder, torture, kidnapping, et cetera.[57] Platforms like Twitter and Facebook are only obliged to remove content, including disinformation, which meets the very narrow definition for 'abhorrent violent material'. If they fail to do so, they can be subject to heavy fines of up to 10% of their annual turnovers.[58] There is arguably room to expand its coverage to other types of harmful content, although legislation existed to tackle foreign influence in elections before the Australian Criminal Code Amendment.[59] It is recognised that the effectiveness of the Australian Code of Practice is hampered by the excessively restrictive definition of harm, such that the signatory platforms need only act against content that will result in serious and imminent harm. In this respect, it has been suggested that

[51]POFMA Office, 'Code of Practice for Transparency of Online Political Advertisements' (Political Advertisements Code) (POFMA Office, 2019), paras 5–7.
[52]POFMA Office, 'Code of Practice for Giving Prominence to Credible Online Sources of Information' (Prominence Code) (POFMA Office, 2019), paras 5 and 6.
[53]POFMA Office, 'Code of Practice for Preventing and Countering Abuse of Online Accounts (Online Accounts Code) (POFMA Office, 2019), paras 5–7.
[54]POFMA Office, 'Prescribed Intermediaries subject to Code' (POFMA Office, 31 January 2020).
[55]See Political Advertisements Code, Prominence Code and Online Accounts Code, para 5.
[56]Australian Criminal Code Amendment, Schedule 1.
[57]Ibid ss 474.31 - 474.34.
[58]Ibid s 474.34.
[59]See, eg, Electoral and Other Legislation Amendment Act 2017 (Cth); Foreign Influence Transparency Scheme Act 2018 (Cth).

chronic harm resulting from the cumulative effect of disinformation over a sustained period of time such as the reduction of trust in public institutions and of community cohesion would inevitably be excluded from this narrow definition.[60]

In early 2021, the Digital Industry Group Inc. (DiGi) drafted the Australian Code of Practice on Disinformation and Misinformation (Australian Code of Practice),[61] in response to the Australian government asking the major digital platforms to develop a voluntary code of conduct to address concerns around disinformation and credibility for content.[62] This code is similarly focussed narrowly on false content which is likely to cause serious and imminent harm, and is mainly applicable to services and products delivered to end users in Australia.[63] Twitter, Facebook, Google, Microsoft and Apple, together with other technology companies, have adopted the code. All signatory platforms have to comply with the core objective to provide safeguards against harms that will arise from disinformation. Other objectives include but are not limited to disrupting advertising and monetisation incentives for disinformation, ensuring the integrity of services delivered (for example, through reducing the risks of inauthentic behaviours such as the use of fake accounts and automated bots to spread disinformation), as well as empowering consumers to make better informed choices with digital content.[64] In addition, the platforms can opt into particular commitments they deem appropriate, to accommodate variations in their business models and offerings.[65] Each signatory platform can further commit to providing, on an opt-in basis, an annual report to DiGi setting out its progress in achieving the outcomes aimed for under the Australian Code of Practice.[66] There is also an obligation to establish a facility to address non-compliance by the signatory platforms with their general commitments under the Australian Code of Practice, including for appeals from complaints of breaches of the code.[67]

Notably, in 2022, the Australian government has announced plans for the introduction of new legislation against disinformation which would, among other things, empower the Australian Communications and Media Authority (ACMA) to establish industry standards and to start a disinformation action group, so as to improve access to information on the effectiveness of measures to address disinformation.[68]

[60]See, eg, Australian Government, 'New disinformation laws' (Australian Government, 21 March 2022) <https://www.infrastructure.gov.au/department/media/news/new-disinformation-laws>; Emma Croft and Sophie Dawson, 'Australian Government to introduce new disinformation and misinformation laws' (Bird & Bird, 28 March 2022) <https://www.twobirds.com/en/insights/2022/australia/australian-government-to-introduce-new-disinformation-and-misinformation-laws>; Australian Communications and Media Authority, 'Report to government on the adequacy of digital platforms' disinformation and news quality measures' (Australian Government, June 2021) <https://www.acma.gov.au/report-government-adequacy-digital-platforms-disinformation-and-news-quality-measures>.

[61]Digital Industry Group Inc. (DiGi), *Australian Code of Practice on Disinformation and Misinformation* (DiGi, 22 February 2021).

[62]Australian Government, 'Governmental Response and Implementation Roadmap for the Digital Platforms Inquiry' (Australian Government, 12 December 2019) <https://treasury.gov.au/publication/p2019-41708>.

[63]DiGi, above n 61.

[64]Ibid.

[65]Ibid.

[66]Ibid.

[67]Ibid.

[68]See, eg, Australian Government, 'New disinformation laws' (Australian Government, 21 March 2022) <https://www.infrastructure.gov.au/department/media/news/new-disinformation-laws>; Australian Communications and Media Authority, 'Report to government on the adequacy of digital platforms' disinformation and news quality measures' (Australian Government, June 2021) <https://www.acma.gov.au/report-government-adequacy-digital-platforms-disinformation-and-news-quality-measures>; Emma Croft and Sophie Dawson, 'Australian Government to introduce

2.3. Germany's NetzDG

In the EU, under Article 14 of the E-Commerce Directive,[69] social networks that store information can amount to hosting providers which are generally not liable for content, provided that they have no actual knowledge of the unlawful content or, upon such knowledge, acts promptly to remove or disable access to such content. There is also no obligation on the part of social networking platforms to monitor content or to verify facts.[70] In Germany, however, the NetzDG specifically obliges social network providers like Twitter and Facebook to report their processes to counteract unlawful content online and to establish systems to handle complaints on content to ensure that unlawful content can be deleted or access thereto can be blocked within 24 h. Should they fail to do so, the providers face hefty fines in millions of dollars.[71] The NetzDG mainly applies to social networks with more than two million registered users in Germany[72] – this will include platforms such as Twitter and Facebook, but will exclude platforms which are new and trying to grow their presence. In February 2022, amendments to the NetzDG came into effect, introducing obligations on large online platforms such as to share information with authorities for criminal prosecutions and to report unlawful content to the Federal Criminal Police Office.[73] Notably, lawsuits have been filed by Twitter and Facebook in relation to these newly imposed obligations.[74] There are criticisms of the NetzDG, particularly around its incompatibilities with the right of freedom of expression in Article 10 of the European Convention on Human Rights[75] and in Article 11 of the EU's Charter of Fundamental Rights,[76] as well as with regard to the principle of territoriality (where the legislation can be enforced against an anonymous individual in another country).[77] Indeed, in respect of the amended NetzDG, a German administrative court has held that it violates the country of origin principle under EU law (in particular, the E-Commerce Directive, where the legal requirements for a provider of electronic services must be based on the law of the member state in which they are located).[78]

new disinformation and misinformation laws' (Bird & Bird, 28 March 2022) <https://www.twobirds.com/en/insights/2022/australia/australian-government-to-introduce-new-disinformation-and-misinformation-laws>.

[69] E-Commerce Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L178/1 (E-Commerce Directive).

[70] Ibid art 15.

[71] NetzDG, above n 38, ss 2, 3 and 4.

[72] Ibid s 1(2).

[73] See, eg, Oliver Noyan, 'Big tech opposes Germany's enhanced hate speech law' (Euractive, 1 February 2022) <https://www.euractiv.com/section/internet-governance/news/german-reinforcement-of-hate-speech-law-faces-opposition-from-big-online-platforms/>; Christoph Schmon and Haley Pedersen, 'Platform Liability Trends Around the Globe: Recent Noteworthy Developments' (EFF, 1 June 2022) <https://www.eff.org/deeplinks/2022/05/platform-liability-trends-around-globe-recent-noteworthy-developments>.

[74] Ibid.

[75] European Court of Human Rights, European Convention on Human Rights (Council of Europe, 1 August 2021), art 10.

[76] European Commission, Charter of Fundamental Rights of the European Union (Official Journal of the European Union, 2012), art 11.

[77] NetzDG, above n 38, s 4(3). See also Victor Claussen, 'Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation' (2018) *Fake news, pluralismo informative e responsabilita in rete* 131.

[78] See, eg, Ananaya Agrawal, 'Germany administrative court holds new online hate speech regulation violates EU law' (Jurist, 2 March 2022) <https://www.jurist.org/news/2022/03/germany-administrative-court-holds-new-online-hate-speech-regulation-violates-eu-law/>. This principle is important as it reduces the administrative burden for companies, allowing them to access the entire EU single market while complying with only the laws of the countries in which they are based, see, eg, EDiMA, 'The e-Commerce Directive' (September 2020) <https://euagenda.eu/upload/publications/edima-ecd-principles-sept-2020.pdf.pdf>.

In 2018, online platforms such as Twitter, Facebook, Google, Mozilla and other advertisers voluntarily committed to an EU Code of Practice on Disinformation (EU Code of Practice)[79] which obliges them to adhere to self-regulatory standards to fight disinformation, including the submission of monthly reports on their efforts to contain disinformation ahead of elections and adopting best practices relating to, among other things, transparency in political advertising, the closure of fake accounts and the demonetisation of purveyors of disinformation.[80] The battle against disinformation has become more relevant over the course of the COVID-19 pandemic, as false claims and conspiracy theories have become rife online. Indeed, online platforms are characterised as both the 'culprits' and 'antidotes' to the proliferation of disinformation over the pandemic.[81] Since it was first introduced, the EU Code of Practice on Disinformation has been revised further to become part of a broader regulatory framework in combination with newly proposed legislation such as the Digital Services Act which introduces strict requirements for online platforms.[82] The revised code provides that: practices are put in place to ensure more transparency of the recommender systems; fact-checking coverage be extended throughout the EU to ensure that platforms make consistent use of fact-checking on their services; a transparency centre and task-force is established to allow for an overview of the implementation of the code; and the monitoring framework is strengthened to include service level indicators to measure the impact of the code throughout the EU.[83]

As can be seen, laws such as the POFMA, the Australian Criminal Code Amendment and the NetzDG have all faced criticisms,[84] whether for over-inclusivity (in the case of the POFMA covering a wide scope of falsehoods and the NetzDG being enforceable outside Germany) or under-inclusivity (in the case of the Australian Criminal Code Amendment which only addresses 'abhorrent violent material'), as well as for contradiction with values such as the freedom of expression. These laws reflect the intention within the respective jurisdictions to hold online platforms, including social media platforms, accountable for their operations and design. There are other jurisdictions that are similarly concerned, and are looking at enacting relevant legislation.[85]

Additionally, across all the Codes of Practices discussed above, online platforms such as Twitter and Facebook commit to implement measures for the purposes of: achieving transparency in advertising (particularly political); disrupting the economic incentives for the creation and dissemination of disinformation; increasing the visibility for credible

[79]European Commission, above n 8.
[80]See, eg, ibid; Matthew Dando and Jack Kennedy, 'Combating fake news: The EU Code of Practice on Disinformation' (2019) 30(2) *Entertainment Law Review* 44.
[81]Roxana Radu, 'Fighting the "Infodemic": Legal Responses to Covid-19 Disinformation' (July-September 2020) *Social Media + Society* 1.
[82]For example, under the proposed Digital Services Act, independent audits must be conducted on very large platforms annually to assess for compliance with obligations, see European Commission, 'The Digital Services Act: ensuring a safe and accountable online environment' (2019) <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en>. See also Ilaria Buri and Joris van Hoboken, 'The Digital Services Act (DSA) proposal: a critical overview' (28 October 2021) <https://dsa-observatory.eu/wp-content/uploads/2021/11/Buri-Van-Hoboken-DSA-discussion-paper-Version-28_10_21.pdf> accessed 15 June 2022.
[83]European Commission, 'The 2022 Code of Practice of Disinformation' (June 2022) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>.
[84]See, eg, Leal, above n 50; and Claussen, above n 77.
[85]For example, in the UK, the Online Safety Bill 2022–2023 introduced to regulate how online platforms manage illegal or lawful but harmful content is currently being considered. See UK Parliament, *Parliamentary Bills: Online Safety Bill* <https://bills.parliament.uk/bills/3137>.

content and deprioritising false content; reducing inauthentic behaviours from fake accounts and automated bots; and giving users the context they need to make informed choices about the content they encounter online. These commitments are arguably to some extent aligned with the self-regulatory efforts the surveyed platforms have and continue to undertake, as the subsequent sections will show. The codes include the mechanisms that the platforms as a collective have developed to regulate disinformation. They further introduce a layer of accountability so that the platforms are not merely exercising self-regulation, but *also* co-regulation[86] with regulatory bodies to account to (i.e. the POFMA Office, the DiGi and the POFMA Office under the respective codes). Co-regulation occurs as platforms have to provide regular reports on the measures they implement against disinformation, thus providing a way for the regulatory bodies concerned to monitor and assess the platforms' effectiveness at regulating disinformation.

3. Platform regulation of disinformation

In the following sections, I look at the inherent characteristics of Twitter and Facebook as well as their efforts at regulating disinformation through: the policies and mechanisms available to users against disinformation; and the initiatives undertaken to regulate information on the COVID-19 pandemic. Together, these give the macro-level view of the systemic features of both social media platforms and further how each of them has chosen to regulate disinformation. Here, I take an objective approach in surveying what the platforms have done or are *professing* to do in the spirit of self-regulation.

3.1. Inherent purposes and features

I examine the purposes for which Twitter and Facebook are established, how users interact and exchange content on the platforms, as well as the accompanying vulnerabilities to disinformation owing to their inherent characteristics. This can lend some insights into how the spread of false news may be more effectively contained.

While Twitter is seen more as a platform for users to view and share opinions, Facebook has been characterised more as a platform to socialise with friends and family, as well as those with similar interests.[87] Thus, on Twitter, wrong information appears to abate more quickly than on platforms such as Facebook, where users interact less with those they disagree with and instead tend to consume information from those with similar views in defined groups based on interests.[88] More specifically, Twitter has been perceived as allowing for the cross-pollination of ideas and information which helps reduce the spread of false news, unlike Facebook which allows users to create silos of information through tailoring their preferences so only their friends can see their posts.[89] The latter platform (i.e. Facebook) arguably makes more room for individual cognitive biases to kick in.

[86]See the explanation of co-regulation in Chris Marsden, Trisha Meyer and Ian Brown, 'Platform values and democratic elections: How can the law regulate digital disinformation' (2020) 36 *Computer Law and Security Review* 1.
[87]See, eg, David John Hughes et al., 'A Tale of Two Sites: Twitter vs. Facebook and the Personality Predictors of Social Media Usage' (2012) 28(2) *Computers in Human Behaviour* 561; Facebook, 'Investor relations: FAQs' (Facebook, 2019) <https://investor.fb.com/resources/default.aspx>.
[88]See, eg, Finkel et al., above n 14.
[89]Ibid.

Moreover, Facebook is seen to be more susceptible to disinformation than Twitter.[90] It has been suggested that features such as the character limit on tweets (adjusted from 140 to 280 in 2017) and the use of hashtags to share information present structural barriers, making it more difficult to link to sources with disinformation on Twitter via uniform resource locators (URLs) than on platforms such as Facebook.[91] Further, disinformation is said to require the careful cultivation of narratives that enhance the plausibility of the information and reduce the doubts of the relevant audiences – again, Twitter's character limit makes it more challenging to craft thorough narrative development which can disrupt opportunities for deliberately cultivated stories to spread. This stands in contrast to Facebook, where users can share more detailed narratives that are provocative, frequently accompanied by imagery and other multimedia, hence facilitating the spread of false stories.[92]

Overall, there are arguably less barriers to the creation of disinformation on Facebook than on Twitter. Based on the discussion above, Facebook thus appears to be the platform more conducive for the creation and dissemination of disinformation.

3.2. Policies and mechanisms to regulate disinformation.

3.2.1. Twitter

Users agree to Twitter's terms of service, privacy policy as well as its rules and policies in using Twitter.[93] Under Twitter's rules, Twitter prohibits users from engaging in inauthentic behaviours including: those of platform manipulation, such as artificially amplifying, suppressing information or engaging in behaviours which disrupts others' experiences on the platform;[94] manipulating or interfering in elections; impersonating individuals, groups or organisations in a manner that is intended to or could mislead others; and sharing manipulated media that could cause harm.[95] When Twitter's policies on content are violated, Twitter can choose to introduce warning notices or prevent the content from being shared altogether, after considering – among other factors and criteria – the context and apparent intent of the user sharing such content.[96] While content that is deceptively altered and shared is more likely to be labelled with warnings, content that can cause immediate harm is more likely to be removed.[97] Also, subtler forms of manipulated media, such as isolated editing, omission of context or presentation of false context may be labelled or removed upon assessment on a

[90]See, eg, Mark Travers, 'Facebook Spreads Fake News Faster Than Any Other Social Website, According to New Research' (Forbes, 21 March 2020) <https://www.forbes.com/sites/traversmark/2020/03/21/facebook-spreads-fake-news-faster-than-any-other-social-website-according-to-new-research/?sh=9d0ea046e1a9>.

[91]See, eg Sarah Perez, 'Twitter's doubling of character count from 140 to 280 had little impact on length of tweets' (TechCrunch, 30 October 2018) <https://techcrunch.com/2018/10/30/twitters-doubling-of-character-count-from-140-to-280-had-little-impact-on-length-of-tweets/>.

[92]See Finkel et al., above n 14.

[93]Twitter, 'Twitter Terms of Service' (Twitter, 18 June 2020) <https://twitter.com/en/tos>.

[94]Examples of platform manipulation are ingenuine engagements which can make accounts or content appear to be more popular than they are and coordinated activity that attempts to artificially influence conversations through the use of multiple fake accounts, see Twitter, 'Platform manipulation and spam policy' (Twitter, September 2020) <https://help.twitter.com/en/rules-and-policies/platform-manipulation>.

[95]Twitter Help Center, 'The Twitter Rules' (Twitter, 2021) <https://help.twitter.com/en/rules-and-policies/twitter-rules>.

[96]Twitter Help Center, 'Our approach to blocking links' (Twitter, July 2020) <https://help.twitter.com/en/safety-and-security/phishing-spam-and-malware-links>.

[97]Twitter Help Center, 'Synthetic and manipulated media policy' (Twitter, 2021) <https://help.twitter.com/en/rules-and-policies/manipulated-media>.

case-by-case basis.[98] Twitter could additionally exercise alternatives to labelling and removal such as: notifying users through warnings before they share or like content that has been manipulated; reducing the visibility of content on Twitter and preventing such content from being recommended; providing a link to additional explanations or clarifications (either to a Twitter page or an external trusted source), hence giving users more information on the claims within tweets and the context; as well as disabling users from liking, replying or retweeting manipulated content.[99] Finally, where there are impersonations of accounts that are meant to deceive or repeated violations of Twitter's policies, the relevant accounts could be permanently suspended (although appeals can be submitted in the case of errors).[100]

Twitter has also established a curated page containing Twitter 'Moments' which reflect the engaging conversations on Twitter. These 'Moments' are created algorithmically to cover sports events and television shows, et cetera, as well as manually by Twitter's curation team whose policy is to showcase content that is compelling, impartial and accurate.[101] This Twitter-curated page will be updated with corrections if it is found to contain inaccurate information.[102]

Further, Twitter relies on its community of users to report content which could be 'abusive or harmful' or 'suspicious or spam', following which Twitter may remove the content. This notification and removal mechanism is easy for users to use should they wish to report to Twitter the availability of disinformation on its platform.

In addition to the above, Twitter recently introduced 'Birdwatch', an enhanced community-based approach to address disinformation, where users can participate through identifying information in tweets they believe to be misleading, thereafter writing notes that provide valuable informative context for other users.[103] While 'Birdwatch' is currently only piloted in the US, Twitter's aim is to eventually make these notes from the community directly visible on tweets available to users worldwide.[104] This arguably appears to be a more robust approach than simply labelling information as true or false and could give users the context to understand and evaluate a tweet better through the facts provided by a broad and diverse community of Twitter users. Twitter has expressly committed to making 'Birdwatch' transparent, by making available data contributed to this initiative and publishing the code of algorithms powering 'Birdwatch', such as consensus and reputation systems.[105]

3.2.2. Facebook

In a similar vein, Facebook approaches fighting disinformation generally through: giving users context on the information they see so that they are better informed; removing

[98] Ibid.
[99] Ibid.
[100] Twitter Help Center, 'Impersonation policy' (Twitter, 2021) <https://help.twitter.com/en/rules-and-policies/twitter-impersonation-policy>.
[101] Twitter Help Center, 'Twitter Moments guidelines and principles' (Twitter, 2021) <https://help.twitter.com/en/rules-and-policies/twitter-moments-guidelines-and-principles>.
[102] Ibid.
[103] Keith Coleman, 'Introducing Birdwatch, a community-based approach to misinformation' (Twitter Blog, 25 January 2021) <https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation>.
[104] Ibid.
[105] Ibid.

content and accounts that are in violation of its policies; and reducing the distribution of false content and the economic incentives to create such content.[106]

Under Facebook's terms of service, Facebook states that while users can express themselves and share content that is important to them, this cannot be done at the expense of the safety and well-being of others.[107] Users have to agree not to use Facebook's services to share content that breaches Facebook's terms and community standards, as well as anything that is, among other things, 'unlawful, misleading, discriminatory or fraudulent'.[108] Should there be repeated breaches of Facebook's terms and policies by any user (including of community standards), Facebook would suspend or permanently disable his or her account.[109] In some circumstances, there could be opportunities to review decisions relating to the removal of a user's content or inactivation of the user's account.[110] Under Facebook's community standards, Facebook states its policies in, among other things: removing content which purposefully misrepresents, defrauds or otherwise exploits others for money;[111] prohibiting users from mispresenting themselves through using fake accounts or artificially boosting the popularity of their content;[112] reducing the spread of false news by lowering such content in the news feed (while recognising that there is a fine line between false news and satire or opinion, the latter of which Facebook does not want to prohibit);[113] removing media (including images, audio or video clips, et cetera) which have been edited to mislead (for instance, a technical deepfake that superimposes content on a video clip and that makes it appear authentic).[114]

Facebook has also established an Oversight Board in October 2020 comprising members from all over the world to review Facebook's decisions to remove content, including where there are appeals from individual users.[115] The Board's decisions are binding and issued in alignment with Facebook's stated policies, unless implementing them could violate the law.[116]

Similar to Twitter, Facebook relies on its community of users to report content which could be false, following which Facebook may remove the content. Users have to submit the report after considering if the content violates Facebook's community standards. This notification and removal mechanism is easy for users to adopt, should users wish to report disinformation accessed on Facebook's platform.

Finally, Facebook has a fact-checking programme since December 2016 where potentially false information is identified through other users' feedback, by way of machine

[106]Tessa Lyons, 'Hard Questions: How is Facebook's Fact-Checking Program Working?' (Facebook, 14 June 2018) <https://about.fb.com/news/2018/06/hard-questions-fact-checking/>.
[107]Facebook, 'Terms of Service' (Facebook, 22 October 2020) <https://www.facebook.com/terms.php>.
[108]Ibid.
[109]Ibid.
[110]Ibid.
[111]Facebook, 'Community Standards: 5. Fraud and deception' (Facebook, 2021) <https://transparency.fb.com/en-gb/policies/community-standards/fraud-deception/>.
[112]Facebook, 'Community Standards: 20. Inauthentic behaviour' (Facebook, 2021) <https://transparency.fb.com/en-gb/policies/community-standards/inauthentic-behavior/>.
[113]Facebook, 'Community Standards: 21. False news' (Facebook, 2021) <https://transparency.fb.com/en-gb/policies/community-standards/false-news/>.
[114]Facebook, 'Community Standards: 22. Manipulated media' (Facebook, 2021) <https://transparency.fb.com/en-gb/policies/community-standards/manipulated-media/>.
[115]Guy Rosen, 'Users can now appeal content left up on Facebook or Instagram to the Oversight Board' (Facebook, 2021) <https://about.fb.com/news/2021/04/users-can-now-appeal-content-left-up-on-facebook-or-instagram-to-the-oversight-board/>.
[116]Oversight Board, 'Oversight Board' (Oversight Board) <https://oversightboard.com/meet-the-board/>.

learning (in the US) or via the initiatives of the relevant independent fact-checkers from various countries who review the content, rate its accuracy and write articles explaining their ratings. Once any content is seen to be inaccurate, its distribution is reduced through de-prioritization on the news feeds of others.[117] Since 2021, Facebook users are informed by way of notifications if they are viewing content that has been rated by fact-checkers as false.[118] Users can hence make informed decisions with the context they are given before they choose to like or 'follow' pages that have repeatedly shared content rated to be false by fact-checkers as there will be pop-ups of notifications indicating this falsity.[119] Through notifications, Facebook discourages its community of users from sharing fact-checked content that has been rated as false. Should users choose to do so, Facebook informs them that their posts will be moved lower in Facebook's news feeds so that other users are less likely to see them.[120] On this note, Facebook has also shared that it uses its artificial intelligence (AI) tools to spot posts that may contain false information, to detect automatic iterations and versions of previous content and to prioritise new content for review, so that fact-checkers can benefit from a more efficient process where only genuinely new content is directed to them for checking.[121]

3.3. Regulating disinformation during the COVID-19 pandemic

In response to the overflow of information, including false information, occurring over the course of the COVID-19 pandemic (or 'infodemic'), Twitter and Facebook have undertaken some initiatives to regulate disinformation. These could serve as a good reference point to evaluate the effectiveness of their measures against disinformation. I will discuss some of the efforts taken.

During the pandemic, both Twitter and Facebook have been active in monitoring information on their platforms – they have flagged, demoted and removed false information that could directly lead to harm, at the same time taking care to promote reliable information sources such as the Centers for Disease Control and the World Health Organization (WHO).[122]

Twitter has created a detailed COVID-19 misleading information policy, expressing its commitment to label or remove false information about, among other things, the nature of the virus, the efficacy of preventative measures, official restrictions and the risk of infection.[123] Instead of relying simply on reports from other users, Twitter is using, consulting and working with public health authorities like the WHO, non-governmental organisations and governments from across the world when it comes to reviewing

[117]Lyons, above n 106.
[118]Facebook, 'Taking Action Against People Who Repeatedly Share Misinformation' (Facebook, 26 May 2021) <https://about.fb.com/news/2021/05/taking-action-against-people-who-repeatedly-share-misinformation/>.
[119]Ibid.
[120]Ibid.
[121]Such artificial intelligence (AI) technologies include SimSearchNet++, an improved image matching model and ObjectDNA which helps to identify manipulated photographs even if objects are placed in front of a new background, see Facebook AI, 'Here's how we're using AI to help detect misinformation' (Facebook, 19 November 2020) <https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/>.
[122]See, eg, Tom Porter, 'Disinformation and the Freedom of Expression' (Bowdoin, 5 March 2021) <https://www.bowdoin.edu/news/2021/03/disinformation-and-the-freedom-of-expression.html>.
[123]Twitter Help Center, 'COVID-19 misleading information policy' (Twitter, 2021) <https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy>.

health information relating to the pandemic on Twitter.[124] Any violation of this policy can result in Twitter temporarily locking users out of their accounts, displaying warnings to users before they share or like erroneous tweets, reducing the visibility of such tweets and preventing them from being recommended, disabling likes, replies and retweets, as well as providing further context to the tweets (for instance, through informing users reading such tweets that the information in the tweets conflicts with public health experts' guidance before they actually view the tweets).[125] Twitter also applies a strike policy where repeated violations will result in Twitter locking the relevant accounts for longer durations, and further, in the case of five or more strikes, permanent suspension.[126] Additionally, since January 2020, Twitter halted auto-suggest results that are likely to direct users to non-credible content on Twitter.[127]

In August 2020, Facebook rolled a new campaign out across Europe, the Middle East and Africa to educate people about detecting false information.[128] In consultation with Facebook's fact-checking partners, three questions have been developed that will be displayed through a series of creative advertisements which will link to a dedicated website.[129] Users are reminded and encouraged to: question where the content is from and to search for a source if this is not clear; find out the whole story through filling gaps that are missing and not rely simply on the headlines; as well as question how the information makes them feel, as false information can manipulate feelings.[130] Users are further notified by way of notification screens when the information they are about to share is older than 90 days.[131] Such notifications can help them reconsider the recency and source of information before they share the information. Users are additionally directed to a COVID-19 information centre to ensure that they have access to credible information. On the other hand, when information shared is from recognised global or local health organisations like the WHO, this notification will not be present, so that the spread of information from credible sources is not slowed.[132] This way, Facebook users are given the resources they need to question and challenge the information they are exposed to, as well as to make informed decisions about what to share with others on the platform.[133]

The above initiatives arguably reflect that Twitter and Facebook are fairly committed to combatting disinformation on the COVID-19 pandemic.

4. User norms as limitations to regulation

As discussed above, legislation such as Singapore's POFMA, the Australian Criminal Code Amendment and Germany's NetzDG impose requirements to remove content

[124]Ibid.
[125]See ibid; Yoel Roth and Nick Pickles, 'Updating our approach to misleading information' (Twitter Blog, 11 May 2020) <https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information>.
[126]Ibid.
[127]Twitter Inc., 'Coronavirus: Staying safe and informed on Twitter' (Twitter Blog, 12 January 2021) <https://blog.twitter.com/en_us/topics/company/2020/covid-19>.
[128]Facebook for Media, 'A new campaign to help spot false news' (Facebook, 2021) <https://www.facebook.com/formedia/blog/a-new-campaign-to-help-spot-false-news>.
[129]The website is at www.stampoutfalsenews.com. See also ibid.
[130]Facebook for Media, above n 128.
[131]John Hegeman, 'Providing People With Additional Context About Content They Share' <https://about.fb.com/news/2020/06/more-context-for-news-articles-and-other-content/>.
[132]Ibid.
[133]Facebook for Media, above n 128.

and to disable access to specified locations on platforms such as Twitter and Facebook. These laws add a layer of accountability through the potential imposition of hefty fines.

Twitter and Facebook have also tried to reduce the spread of disinformation by way of their efforts at moderating content. Notwithstanding these self-regulatory efforts, given that the business models of many social media platforms are centred around user-generated content and engagement with such content,[134] the interests of these platforms are not always aligned to curbing the spread of disinformation.[135] There are financial conflicts of interests given that the short term economic incentives of the platforms are to gain revenue through increased engagement from users and advertising.[136] As such, co-regulation, demonstrated via the platforms' commitment to codes of practices, has evolved in response to the limitations of self-regulation.

The challenges posed by disinformation on online platforms arguably relate mainly to how users react and relate to false information. I posit that the behavioural norms of users *accessing, assessing, assimilating* and *disseminating* information (including disinformation) on the platforms play a significant role in causing harm to others who believe in the disseminated disinformation. Henceforth, in addition to regulation by laws, self-regulation by the platforms (i.e. through their policies and the infrastructure they set up) and co-regulation (i.e. through industry collaboration with regulatory bodies, et cetera),[137] regulating disinformation effectively on social media platforms will require a consideration of the norms around reading and sharing information on such platforms.

Indeed, as outlined in the section earlier, social media platforms such as Twitter and Facebook have, as part of their efforts to self-regulate, purported to give users the information they need to make decisions about the content they view. This commitment is also integrated within the codes of practices discussed earlier. There is further reliance on the community of users on each platform to identify and report on content which could be disinformation, among other things. Arguably, these measures reflect the platforms' understanding that tackling the broader challenge of disinformation requires improving users' abilities to assess the credibility of information seen before disseminating such information. In this respect, self and co-regulation by the platforms *involve trying to influence user norms* through 'nudging' users to exercise their discretion to share only information that is credible.

On a related note, human vulnerabilities and cognitive predispositions also need to be taken into consideration to ensure that the approaches to counter disinformation would be effective.[138] Generally speaking, notwithstanding the efforts expended in fact-checking and in notifying users of falsities undertaken by Twitter, Facebook and other online platforms, it appears that that early interventions to prevent exposure to disinformation in the first instance are more effective at countering their proliferation as compared to

[134]See Laura DeNardis, *The global war for internet governance* (Yale University Press, 2014). See also Tan, above n 34.
[135]See, eg, David Tan and Jessica Teng, 'Fake news, free speech and finding constitutional congruence' (2020) 32 *Singapore Academy of Law Journal* 207.
[136]Abby K Wood and Ann M Ravel, ' Fool Me Once: Regulating Fake News and Other Online Advertising' (2018) 91 *Southern California Law Review* 1245.
[137]This approach draws, to some extent, on Lessig, see Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books, 1999); Lawrence Lessig, *Code Version 2.0* (Basic Books, 2nd ed, 2006). The 'modalities' of regulation under Lessig's regulatory theory of cyberspace include not just law, but also architecture, markets and social norms.
[138]See, eg, Daniel T. Gilbert, 'How Mental Systems Believe' (1991) 46(2) *American Psychologist* 114; Norman Vasu et al., *Fake News: National Security in the Post-Truth Era* (S. Rajaratnam School of International Studies, January 2018).

measures taken after to correct users' misbeliefs.[139] In support of this view, early exposure to incorrect content is found to be positively associated with believing false information, regardless of the opportunities to fact-check and subsequent exposure to corrections.[140] Moreover, labelling content as false may have limited impact, as ironically, believers may choose to ignore the labels and remain steadfast in their misperceptions, although non-believers of the relevant content will simply rely on the labels to reinforce their views.[141] Tools which flag content to be disputed by fact-checkers may also have the unintended effect of exacerbating some users' negative engagement behaviours,[142] henceforth defeating such efforts against disinformation. Furthermore, even where accurate facts are accessible, users may ignore or fail to process information appropriately.[143]

How people individually respond to information overload in the context of our current realities poses challenges to the goal of reducing disinformation. Individual user inclinations are relevant since they shape user reactions, perceptions and behaviours (constituting norms) towards disinformation. Among others, there are two specific aspects which can be looked at. *One,* users are suggested to have a baseline social-psychological tendency to seek out evidence that fits into their preconceptions (i.e. *confirmation biases*), to congregate with others who are like them and to avoid information that does not fit into what they like.[144] To elaborate further, a person is found to be typically less critical of information that is favourable to his or her views than otherwise, exhibiting such confirmation bias in his or her information seeking and processing behaviour.[145] Additionally, people are inclined to adopt the views of the peer groups most salient to them, even if other objective factual information contradicts those views.[146] *Two,* relating back to the point on fact-checks, how effective these checks are would be subject to the personal inclinations of users such as their *tolerance for negativity and political sophistication*.[147] User inclinations do matter – this means that some users would be more appropriately responsive to fact-checks than others.[148] For example, when users with a low tolerance for negativity see a negative fact-check, they are found to be least likely

[139]Gordon Pennycook, Tyrone Cannon and David G. Rand, 'Prior exposure increases perceived accuracy of fake news' (2018) 147(12) *Journal of Experimental Psychology: General* 1865.

[140]Ibid.

[141]See Danah Boyd, *Did media literacy backfire* (Data and Society, 6 January 2017).

[142]See, eg, Sam Levin, 'Facebook Promised to Tackle Fake News, But the Evidence Shows It's Not Working' (Guardian, 16 May 2017) <https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working>.

[143]Matthew T. Binford et al., 'Invisible transparency: Visual attention to disclosures and source recognition in Facebook political advertising' (2021) 18(1) Journal of Information Technology & Politics 70. It was found that disclosures on political advertising were not effective in increasing users' awareness of the parties who sponsored political advertisements, even if such users were visually attentive to the disclosures.

[144]Peter Wason coined the term 'confirmation bias', see Wikipedia, 'Peter Cathcart Wason' < https://en.wikipedia.org/wiki/Peter_Cathcart_Wason>. See also Uwe Peters, 'What is the Function of Confirmation Bias?' (2022) 87 *Erkenntnis* 1351; Masaki Suzuki and Yusuke Yamamoto, 'Characterizing the Influence of Confirmation Bias on Web Search Behaviour' (December 2021) 12 *Frontiers in Psychology* 771948; and Cass Sunstein, *Republic.com* (Princeton University Press, 2001).

[145]See, eg, Herbert Lin, 'On the Organization of the U.S. Government for Responding to Adversarial Information Warfare and Influence Operations' (2019) 15(1) *I/S A Journal of Law and Policy for the Information Society* 10.

[146]See, eg, William Hart et al., 'Feeling Validated Versus Being Correct: A Meta-Analysis of Selective Exposure to Information' (2009) 135 *Am. Psychol. Ass'n* 556–59; Solomon E. Asch, 'Effects of Group Pressure Upon the Modification and Distortion of Judgments' in Harold Guetzkow (ed), *Groups, Leadership, And Men* (1951) 226.

[147]See, eg, Waheeb Yaqub et al., 'Effects of Credibility Indicators on Social Media News Sharing Intent' (April 2020) *CHI* Paper 86.

[148]Kim Fridkin, Patrick J. Kennedy and Amanda Wintersieck, 'Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising (2015) *Political Communication* 127, 145.

to accept the claims in negative advertisements. Further, when users with more political sophistication access a fact-check challenging the truthfulness of a negative commercial, they will view the commercial more negatively than users with less sophistication.[149] On the other hand, tools which flag content to be disputed by fact-checkers have been suggested to have the unintended effect of exacerbating *some* users' negative engagement behaviours (again depending on their personal inclinations).[150] Henceforth, user inclinations could mean that the efforts made to provide more information through giving context to posts and fact-checking could have limited effect in alleviating the impact of disinformation with respect to some groups of users.

Finally, users 'don't know what they don't know'.[151] There is a concern in the reliance on users and fact-checkers to detect disinformation online. It has been suggested that most users, being susceptible to human errors and biases, would find it difficult to identify false information and inaccurate sources, resulting in fewer articles being reported as false and as requiring fact-checking.[152] It does follow that a huge volume of disinformation on social media platforms like Twitter and Facebook could go undetected and uncorrected.

On the whole, due to the cognitive predispositions of users, harm can occur as soon as disinformation spreads and is accessed, even if there are corrections thereafter on the accuracy of the relevant content. In addition, the inclinations of individual users including but not limited to confirmation biases amplify the harmful effects of disinformation. Last but not least, it is likely to be challenging for the average user to identify disinformation. These vulnerabilities and cognitive predispositions shape user norms on social media platforms, particularly in respect of how users respond to and interact with disinformation. As such, any sustainable approach against disinformation requires user norms to change for the better, such that users are less likely to be affected by and to share disinformation with others. The importance of shaping user norms towards disinformation is recognised by the platforms. To some extent, both Twitter and Facebook palpably attempt to influence user behaviours through the self and co-regulatory approaches undertaken, including but not limited to: displaying warning notices and corrections for content where there are inaccuracies; providing users with more information to give context to content so that users are better placed to assess such content; removing content that is inaccurate so that users cannot share such content in the first place; reminding users to question the sources and their credibility; and informing users that shared content with inaccuracies will be demoted in the news feeds so that users are less incentivized to share such content. Notwithstanding these efforts, providing more accurate information through giving context and fact-checking arguably have limited effectiveness in alleviating the spread of disinformation, in light of the vulnerabilities and cognitive predispositions of individual users online. Reminding users to question sources of information and their credibility could, however, plausibly be more effective if such questioning is ingrained within user behaviours and integrated as part of user norms. This will take time to cultivate.

[149] ibid.
[150] See, eg, Sam Levin, above n 142.
[151] See, eg, Arts Markman, 'Do you know what you don't know?' (Harvard Business Review, 3 May 2012) < https://hbr.org/2012/05/discover-what-you-need-to-know>.
[152] See, eg, Stone-Erdman, above n 18.

5. Evaluating regulation

In regulating disinformation on online platforms, an appropriate balance needs to be struck between the need to combat disinformation and values to be upheld such as freedom of speech as well as the continued innovation of online services.[153] There is a need to balance intermediary liability and accountability with intermediary immunity.[154] It is noted that multiple layers of regulation co-exist to address the challenge posed by disinformation. In the first part of this article, I outlined the extent of intermediary liability imposed in selected jurisdictions such as the US, Singapore, Australia and Germany. This form of regulation is direct via implemented laws and regulations, empowering state-directed authorities to issue orders to block websites, remove unlawful content, delete the relevant accounts and impose fines.

Online platforms, including social media platforms, also self-regulate.[155] This arguably reflects their intentions to be perceived as platforms for reliable information and to avoid eliciting further forms of direct regulation. Twitter and Facebook's self-regulation, for example, is reflected through their policies, tools and mechanisms against disinformation, as well as the specific COVID-19 initiatives they have undertaken. While the efforts taken in specific contexts such as the COVID-19 pandemic appear reasonable, online platforms such as Twitter and Facebook cannot be unilaterally responsible for the gargantuan task of regulating disinformation. This is due to the inherent conflicts of interests arising as a result of the platforms profiting from increased user engagement with content (regardless of its accuracy) and advertising.[156]

In response to this insufficiency of self-regulation, a model of co-regulation[157] has emerged, evidenced through the codes of practices discussed above. Cooperation is envisaged among the governments, 'big technology' companies operating the relevant online platforms, media organisations, researchers and other stakeholders.[158] Moreover, these companies have developed their practices and mechanisms – both individually and collectively as an industry – to regulate disinformation on their platforms, whilst being accountable to governments and other authorities which can monitor their practices for effectiveness. Beyond self-regulation, these codes allow for a layer of oversight by way of co-regulation.

I argued earlier that human vulnerabilities and cognitive predispositions shape user norms and limit the effectiveness of self and co-regulatory efforts against disinformation undertaken by Twitter and Facebook. Improving the digital literacy of individual users are therefore necessary as part of a holistic solution against disinformation.[159] In order to be better at distinguishing between false and credible information, users have to learn, in particular, to overcome developed heuristics (or mental shortcuts) resulting in over reliance on information from the internet and inherent biases. Further, in light

[153]See, eg, Finkel et al., above n 14.
[154]See Jack M. Balkin, *How to Regulate (and Not Regulate) Social Media* (Knight First Amendment Institute at Columbia University, 25 March 2020).
[155]See, eg, Flick, above n 30, 393; *Seran v Am. Online, Inc.*, 129 F.3d 327, 330–31 (4th Cir. 1997).
[156]See, eg, Nabiha Syed, 'Real Talk about Fake News: Towards a Better Theory for Platform Governance' (2017) 127 *Yale Law Journal Forum* 356.
[157]See Marsden, Meyer and Brown, above n 86.
[158]See, eg, Flavia Durach, Alina Bargaoanu and Catalina Nastasiu, 'Tackling Disinformation: EU Regulation of the Digital Space' (2020) 20(1) *Romanian Journal of European Affairs* 5.
[159]Ibid.

that the personal attributes of users are important in ascertaining how they interact with disinformation, digital literacy education can be centred around emotional management and digital self-care, as well as awareness of important lessons such as thinking before sharing information, avoiding filter bubbles and understanding the threats posed by mere exposure to wrong information.[160] More broadly speaking, providing civic education and allowing for the development of enhanced critical thinking skills would also improve digital literacy in users.[161] These 'user-centred' efforts could, in the longer term, result in an improvement of user norms with regard to disinformation, and ultimately in the establishment of a more educated digital community better placed to identify and to disregard disinformation.

Evidently, the following approaches to regulate disinformation being: direct-regulation via laws; self-regulation through the voluntary efforts of Twitter and Facebook; co-regulation where there is a commitment made among the relevant authorities, the companies operating the platforms, media organisations and researchers, to collaborate; as well as user-centred solutions[162] – all have to exist in order for disinformation to abate. On this note, I argue that as direct regulation by way of laws is exclusive only to countries so regulated, self and co-regulation can play a *bigger* part than direct regulation in the shorter term as users across the world experience social media platforms like Twitter and Facebook uniformly. In light of the corporate motives of the companies operating the platforms, however, the individual susceptibilities of users does evoke concerns around relying mainly on the platforms to self and co-regulate disinformation, *even if* some of the their efforts are targeted towards shaping user norms against disseminating dissemination. While it may take time to nurture more digitally literate communities of users better placed to identify and disregard disinformation, shaping user norms around online disinformation would likely be enduring and hold more promise in combating disinformation in the longer term. This is particularly likely given that the problem of disinformation essentially arises from the acts of *accessing and disseminating* disinformation, so much so that this is recognised by the platforms and that their self and co-regulatory efforts aim at influencing these user norms. Therefore, I argue further that while there are multiple layers of regulation co-existing, influencing user behaviours around disinformation lie at the heart of most forms of regulation, whether referred to implicitly (i.e. self and co-regulatory approaches) or explicitly (i.e. user-centred solutions). Individual users hold more power over these platforms than is realised, as such, shaping user norms remains critical to the fight against disinformation.

6. Conclusion

In recent years, regulators and the public are increasingly distrustful of the 'big technology' companies operating platforms such as Twitter and Facebook. In March 2021, a lawsuit was filed by Reporters Without Borders (RWB) with the public prosecutor in France

[160]See, eg, Darren G. Lilleker, 'Evidence to the Culture, Media and Sport Committee "Fake news" inquiry presented by members of the Centre for Politics and Media Research' (Faculty for Media and Communication, Bournemouth University, 2017) <https://eprints.bournemouth.ac.uk/28610/3/Evidence%20Submission%20-%20Fake%20News%20FINAL.pdf>.

[161]See, eg, Vinton G. Cerf, 'Hazards of the Information Superhighway' (2019) 62(11) *Communications of the ACM* 5; Finkel et al, above n 14.

[162]See Durach, Bargaoanu and Nastasiu, above n 158.

against Facebook, accusing the latter of 'deceptive commercial practices' under the French consumer code – in particular, of allowing harmful content such as false information and hate speech (including hatred against journalists) to flourish on its platform, in spite of its contradictory promises in its terms of service and advertisements to provide a safe and error free online environment.[163] In the course of the COVID-19 pandemic, First Draft, a non-profit organisation set up to combat online disinformation, found that 84 per cent of all interactions relating to vaccine-related conspiracy content came from two platforms, Facebook and Instagram, both operated by Facebook.[164] Public scepticism of social media platforms will increase in future – this would arguably support a move *away* from reliance mainly on self-regulation by these platforms, towards direct regulation by laws, co-regulation, as well as regulation by way of the shaping of user norms.

Under the laws in Singapore, Australia and Germany, online platforms are held accountable through the potential imposition of fines and the blocking of access to their platforms under legislation such as Singapore's POFMA, the Australian Criminal Code Amendment and Germany's NetzDG. Upon examining the policies and mechanisms available on the selected social media platforms Twitter and Facebook, as well as their COVID-19 specific initiatives, it is observed that the current efforts of the social media platforms Twitter and Facebook at moderating content are, to some extent, aligned with the laws examined. Some of these efforts were made before the enactment for these laws, such as the notification and removal mechanisms for harmful content and even fact-checking on Facebook. In this sense, beyond adding accountability through the possibility of being fined heavily, there is no new obligation imposed on the platforms under the laws examined.

Holding platforms accountable via laws and self-regulation, is, however, inadequate, and as a result, the platforms co-regulate through collectively committing to codes of practices to set up room for further collaboration and accountability among themselves, regulators and other stakeholders. Given that the regulation of disinformation will likely be a moving target with novel technologies posing new challenges, multi-stakeholder collaboration is an important development that has to be retained and refined, so that conversations can continue among the online platforms, governments, other public authorities, academia, civil society and news organisations as key stakeholders. This allows for inclusive decision-making, strong transparency and monitoring mechanisms, as well as government interventions when there is ineffectiveness on the part of the platforms.[165]

In spite of layers of regulation by laws, self and co-regulation, much of the harm caused by disinformation arguably arises from the user norms of individual users accessing and disseminating disinformation on the social media platforms. A solution to this is to improve the literacy of all digital citizens using the online platforms so that they are better able to evaluate disinformation – this, however, is not an immediate fix to

[163]RSF, 'RSF files lawsuit in France accusing Facebook of "deceptive commercial practices"' (RSF, 22 March 2021) <https://rsf.org/en/news/rsf-files-lawsuit-france-accusing-facebook-deceptive-commercial-practices>.

[164]Rory Smith, Seb Cubbon and Claire Wardle, 'Under the surface: Covid-19 vaccine narratives, misinformation and data deficits on social media' (First Draft, 12 November 2020) <https://firstdraftnews.org/long-form-article/under-the-surface-covid-19-vaccine-narratives-misinformation-and-data-deficits-on-social-media/>.

[165]Private-public co-regulation aims for these outcomes in other sectors such as environmental governance, see Dennis Kolcava, Lukas Rudolph and Thomas Bernauer, 'Citizen preferences on private-public co-regulation in environmental governance: Evidence from Switzerland' (2021) 68 *Global Environmental Change* 1.

managing disinformation. The effectiveness of this strategy, if at all, may only be experienced in the longer term. Moreover, there are personal attributes of users (for example, in the form of confirmation biases, tolerance for negativity and political sophistication) on social media platforms such as Twitter and Facebook which account for the spread of disinformation. The importance of influencing user behaviours so that such behaviours are not conducive to the spread of disinformation consequently lies at the heart of self and co-regulatory approaches adopted by Twitter and Facebook, as well as within user-centred solutions.

As the COVID-19 pandemic becomes endemic, other pandemics and their 'infodemics' may emerge in future. Widespread disinformation will likely persist, along with social media platforms that make virality possible. Amid all complexities, investing in educating a well-functioning public sphere may well be the main bulwark against the amplified impact of disinformation on social media.[166] Governments and authorities will continue to regulate via laws and co-regulatory approaches. At the same time, social media platforms will self and co-regulate. The behavioural norms of individual users instrumental in spread of disinformation can, with the benefit of digital literacy education, be improved. These users, whose patronage is desired, could forge responsible digital practices unconducive to the spread of disinformation. They can constitute a formidable force holding social media platforms like Twitter and Facebook accountable for our digital future.

Acknowledgement

The author is most grateful to the two anonymous reviewers who generously provided constructive feedback to an earlier version. She is also grateful for comments received at the Digital Law Research Colloquium organised by the Digital Law Center at the University Geneva in 2021, as well as from Kalyan Takru and Kapilesh Taneja. Last but not least, the author will also like to thank Carmen Sin for her research assistance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research undertaken for this article is funded by a Ministry of Education Academic Research Fund Tier 1 grant [number RS07/20].

Notes on contributor

Corinne Tan PhD, LLM (Melbourne Law School, University of Melbourne), LLB (Hons, National University of Singapore) is an Assistant Professor at the Nanyang Business School (Division of Business Law), Nanyang Technological University. Email: corinne.tan@ntu.edu.sg

Corinne has taught and researched in Australia and Singapore for the past 10 years. Her research interests are on platform governance, regulation of social media, copyright on social

[166]Yochai Benkler, Robert Faris and Hal Roberts, *Network Propaganda: Manipulation, Disinformation and Radicalization in American Politics* (Oxford University Press, 2018) 386.

media, as well as copyright and accessibility for the visually impaired. Her monograph titled 'Regulating Content on Social Media: Copyright, Terms of Service and Technological Features' was published by the University of College London (UCL) Press in March 2018 and is available at https://www.uclpress.co.uk/products/95612. In addition, Corinne has published in international journals such as the Computer Law and Security Review, International Review of Intellectual Property and Competition Law, European Intellectual Property Review, the Journal of Banking and Finance Law and Practice, the Intellectual Property Quarterly, the Media and Arts Law Review, the Singapore Academy of Law Journal, the Competition and Consumer Law Journal and the Law Quarterly Review. She has given talks to present her research in Europe, Australia and Singapore.

Corinne holds a PhD and a LLM from the Melbourne Law School, University of Melbourne and a LLB from the National University of Singapore. She was called to the Singapore Bar in 2007.

ORCID

Corinne Tan http://orcid.org/0000-0002-1389-9356