Research Paper

Kate Jones
International Law Programme | November 2019

Online Disinformation and Political Discourse

Applying a Human Rights Framework

CHATHAM HOUSE
The Royal Institute of International Affairs

Contents

Summary

- Online political campaigning techniques are distorting our democratic political processes. These techniques include the creation of disinformation and divisive content; exploiting digital platforms' algorithms, and using bots, cyborgs and fake accounts to distribute this content; maximizing influence through harnessing emotional responses such as anger and disgust; and micro-targeting on the basis of collated personal data and sophisticated psychological profiling techniques. Some state authorities distort political debate by restricting, filtering, shutting down or censoring online networks.

- Such techniques have outpaced regulatory initiatives and, save in egregious cases such as shutdown of networks, there is no international consensus on how they should be tackled. Digital platforms, driven by their commercial impetus to encourage users to spend as long as possible on them and to attract advertisers, may provide an environment conducive to manipulative techniques.

- International human rights law, with its careful calibrations designed to protect individuals from abuse of power by authority, provides a normative framework that should underpin responses to online disinformation and distortion of political debate. Contrary to popular view, it does not entail that there should be no control of the online environment; rather, controls should balance the interests at stake appropriately.

- The rights to freedom of thought and opinion are critical to delimiting the appropriate boundary between legitimate influence and illegitimate manipulation. When digital platforms exploit decision-making biases in prioritizing bad news and divisive, emotion-arousing information, they may be breaching these rights. States and digital platforms should consider structural changes to digital platforms to ensure that methods of online political discourse respect personal agency and prevent the use of sophisticated manipulative techniques.

- The right to privacy includes a right to choose not to divulge your personal information, and a right to opt out of trading in and profiling on the basis of your personal data. Current practices in collecting, trading and using extensive personal data to 'micro-target' voters without their knowledge are not consistent with this right. Significant changes are needed.

- Data protection laws should be implemented robustly, and should not legitimate extensive harvesting of personal data on the basis of either notional 'consent' or the data handler's commercial interests. The right to privacy should be embedded in technological design (such as by allowing the user to access all information held on them at the click of a button); and political parties should be transparent in their collection and use of personal data, and in their targeting of messages. Arguably, the value of personal data should be shared with the individuals from whom it derives.

- The rules on the boundaries of permissible content online should be set by states, and should be consistent with the right to freedom of expression. Digital platforms have had to rapidly develop policies on retention or removal of content, but those policies do not necessarily reflect the right to freedom of expression, and platforms are currently not well placed to take account of the public interest. Platforms should be far more transparent in their content regulation policies and decision-making, and should develop frameworks enabling efficient, fair, consistent internal complaints and content monitoring processes. Expertise on international human rights law should be integral to their systems.

- The right to participate in public affairs and to vote includes the right to engage in public debate. States and digital platforms should ensure an environment in which all can participate in debate online and are not discouraged from standing for election, from participating or from voting by online threats or abuse.

1. Introduction

The framers of the Universal Declaration of Human Rights (UDHR) saw human rights as a fundamental safeguard for all individuals against the power of authority. Although some digital platforms now have an impact on more people's lives than does any one state authority,[1] the international community has been slow to measure and hold to account these platforms' activities by reference to human rights law. And although international human rights law does not impose binding obligations on digital platforms, it offers a normative structure of appropriate standards by which digital platforms should be held to account. Because of the impact that social media can have, a failure to hold digital platforms to human rights standards is a failure to provide individuals with the safeguards against the power of authority that human rights law was created to provide.

While the emergence of internet technology has brought human rights benefits, allowing a plurality of voices, a new freedom of association and more widespread access to information than ever before, it has also brought distortions to electoral and political processes that threaten to undermine democracy. The rapid pace of technological change has facilitated non-compliance with existing human rights law and related regulation, because the activities are new and because the infrastructure has not been in place to explain, monitor or enforce compliance with existing laws.[2] Urgent action is needed, as the challenges we are currently seeing to our democracies are challenges of the scale being tackled when the UDHR was drafted in the late 1940s.

There is a widespread desire to tackle online interference with elections and political discourse. To date, much of the debate has focused on what processes should be established[3] without adequate consideration of what norms should underpin those processes. Human rights law should be at the heart of any discussion of regulation, guidance, corporate or societal responses.[4] The UN Secretary-General's High-level Panel on Digital Cooperation has recently reached a similar conclusion, stating 'there is an urgent need to examine how time-honoured human rights frameworks and conventions should guide digital cooperation and digital technology'.[5] This paper attempts to contribute to this examination.

[1] For example, as of 31 March 2019 Facebook had over 2.38 billion monthly active users and 1.56 billion daily active users. Facebook Investor Relations (2019), 'Facebook Reports First Quarter 2019 Results', https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-First-Quarter-2019-Results/default.aspx (accessed 14 Oct. 2019).
[2] For example, bodies charged with monitoring and enforcement, such as the Information Commissioner's Office and the Electoral Commission in the UK, have not had the resources, powers or sanctioning capacity fully to absorb cyber challenges into their work.
[3] For example, whether and how Facebook should establish an Oversight Board; whether and how the UK government should establish a regulator.
[4] Other international laws also apply, e.g. the non-intervention principle in respect of interference in elections by a foreign state.
[5] UN Secretary-General's High-level Panel on Digital Cooperation (2019), *The Age of Digital Interdependence*, UN Secretary-General's High-level Panel on Digital Cooperation, https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf (accessed 24 Oct. 2019) p. 4.

Chapter 2 of this paper clarifies terms and concepts discussed. Chapter 3 provides an overview of cyber activities that may influence voters. Chapter 4 summarizes a range of responses by states, the EU and digital platforms themselves. Chapter 5 discusses relevant human rights law, with specific reference to: the right to freedom of thought, and the right to hold opinions without interference; the right to privacy; the right to freedom of expression; and the right to participate in public affairs and vote. Chapter 6 offers some conclusions, and sets out recommendations on how human rights ought to guide state and corporate responses.

2. Clarifying Core Concepts

2.1 Digital platforms

This paper focuses on digital platforms that host content generated by other users, the content then being accessible to all, or a subset of, users. While the paradigmatic platforms are websites designed specifically to host others' content, other websites, such as media and trading sites, will also fall within this definition if they allow their audiences to post comments. Digital platforms of most relevance in elections currently include website platforms such as Google and Yahoo!; relatively open social media and microblogging sites such as Facebook and Twitter; shared news websites such as Reddit; photo sharing sites such as Instagram and Snapchat; video sharing sites such as YouTube; and closed messaging applications such as WhatsApp and Facebook Messenger.

It is arguable that the most dominant platforms globally – Facebook, Twitter, and Google/YouTube – with perhaps those of dominance in specific countries – such as WeChat in China, and VK and OK in Russia – deserve different treatment from others as, although private entities, they are essentially providing a public service. This paper does not consider this point, instead focusing primarily on those largest platforms.

While not the primary originators of content, digital platforms can have extensive control over who sees what content on their sites. The distribution of content on digital platforms is dependent on the algorithms used by the platforms to prioritize and deprioritize material on them. The platforms only publish limited information about their algorithms.[6] For maximum advertising revenue, the platforms are actively designed, including through their algorithms, to be addictive, and to encourage users to spend as long as possible on them.[7]

Platforms also control who sees what adverts, for example through real-time bidding.[8] Advertising is usually their principal source of revenue, and can be extremely lucrative. Four of the five largest publicly traded companies by market capitalization are technology companies.[9]

In general, digital platforms currently have little legal responsibility ('intermediary liability') for content hosted on their services. In the US, Section 230 of the Communications Decency Act (CDA)[10] provides that digital platforms shall not be treated as the 'publisher or speaker' of information provided by others on their sites; consequently they have immunity from legal liability for content that may be (for example) defamatory, false, or threatening. This immunity is subject to exceptions in respect of federal criminal liability and some intellectual property claims.[11] The

[6] For example, Search Engine Journal maintains a list of developments in Google algorithms. Search Engine Journal (2019), 'History of Google Algorithm Updates' https://www.searchenginejournal.com/google-algorithm-history/ (accessed 5 Oct. 2019).
[7] For example, Harris, T. (2016), 'How Technology is Hijacking Your Mind — from a Magician and Google Design Ethicist', *Medium* blog, 18 May 2016, https://medium.com/thrive-global/how-technology-hijacks-peoples-minds-from-a-magician-and-google-s-design-ethicist-56d62ef5edf3 (accessed 5 Oct. 2019).
[8] Information Commissioner's Office (2019), *Update report into adtech and real time bidding*, https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf (accessed 5 Oct. 2019).
[9] PwC (2019), *Global Top 100 companies by market capitalisation*, London: PwC, https://www.pwc.com/gx/en/audit-services/publications/assets/global-top-100-companies-2019.pdf (accessed 29 Oct. 2019).
[10] 47 USC § 230 (2011).
[11] Ibid s230(e)(1) and s230(e)(2) respectively.

immunity extends to platforms' decisions to take content down (so not only may they host unlawful content, but they may also freely take down lawful content). CDA 230 was amended in 2018 to limit immunities in sex-trafficking cases.[12]

In the EU, the eCommerce Directive[13] exempts digital platforms from liability for illegal content hosted on their services, provided that, if aware of such content, they will remove it or disable access to it expeditiously. The Directive prohibits EU member states from imposing general obligations on digital platforms to monitor user-generated content. On 1 March 2018 the Commission issued a Recommendation concerning processes that online platforms should adopt in order to expedite the detection and removal of illegal content.[14] Some European states have now imposed, or are considering imposing, stricter obligations for platforms to consider taking down allegedly illegal content once made aware of it; examples are discussed in Chapter 4 below.

2.2 Disinformation

This paper avoids the term 'fake news', as that phrase has been used to encompass a range of information including information that the user of the term may wish to dispute as opposed to provable falsehoods.

This paper uses the term 'disinformation' to mean false or manipulated information that is knowingly shared to cause harm.[15] 'Disinformation' does not encompass information that is false but not created with the intention of causing harm (sometimes labelled 'misinformation').[16] Although standard definitions of 'disinformation' do not include true information, information that is true but knowingly shared to cause harm (sometimes termed 'mal-information')[17] can be as pernicious as false information: for example, when private material is made public, or when information is taken out of context or labelled so as to arouse emotion.

The paper contrasts the current debate over disinformation with past discussions of propaganda. The term 'propaganda' can cause confusion. As used in Article 20(1) of the 1966 International Covenant on Civil and Political Rights (ICCPR), it 'refers to the conscious effort to mould the minds of men [sic] so as to produce a given effect'.[18] Disinformation is therefore a subset of propaganda: whereas propaganda includes both true and false persuasive material, disinformation is only false or manipulated information that is knowingly shared to cause harm. Propaganda is not only material generated by governments, although this is the paradigm case.

[12] Allow States and Victims to Fight Online Sex Trafficking Act, H.R. 1865, 115th Cong. (2018).
[13] Parliament and Council Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce') [2000] OJ L 178/1.
[14] European Commission, 'Recommendation of 1.3.2018 on measures to effectively tackle illegal content online' (C(2018) 1177 final).
[15] The UK government defines disinformation as 'the deliberate creation and sharing of false and/or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain'. Digital, Culture, Media and Sport Committee (2018), *Disinformation and 'fake news': Interim Report: Government's Response to the Committee's Fifth Report of Session 2017-2019,* London: House of Commons, p. 2, https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/1630.pdf (accessed 5 Oct. 2019).
[16] Wardle, C. and Derakhshan, H. (2017), *Information Disorder: Toward an interdisciplinary framework for research and policymaking,* Strasbourg: Council of Europe, p, 20, https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c (accessed 14 Oct. 2019).
[17] Ibid.
[18] Whitton, J. B. (1948), 'Propaganda and International Law', in Hague Academy of International Law (ed) (1948), *Collected Courses of the Hague Academy of International Law*, Leiden: Brill, and Boston: Nijhoff, p. 547.

Some disinformation, but by no means all, is 'hate speech', a loose term that includes both speech that states are obliged or entitled to restrict or discourage as a matter of human rights law, and other hateful speech that may not be so restricted. The parameters of these restrictions are discussed in section 5.4.1.

2.3 Personal data

Personal data, the oil of the 21st century economy, is the commodity funding free-to-use digital platforms. Digital platforms are currently mining and using personal data at an unprecedented rate. There are many positive uses of personal data: analysis of big data has the potential to bring major benefits to society, from increasing safety, to diagnosing and treating illness and epidemics, to facilitating access to services. The collection and use of personal data is key to the effective operation of websites in many domains, for example marketing (e.g. universities attracting potential applicants) and legacy media (e.g. newspaper websites attracting and retaining readers).

However, the collection of data by digital platforms comes with major concerns about surveillance, the targeting of individuals to receive curated information or advertising based on the profile of them developed by the platform, and consequent discrimination as a result of this profiling. Browsing a web page, far from being as anonymous as entering a shop or reading a physical newspaper, in practice often requires consenting to cookies that collect data and potentially share those data with third parties. Most people are not aware of digital platforms and political campaigners' knowledge (and assumptions) about them, nor of the rapidly increasing scale on which data is shared, traded and used to develop personal profiles. Nor are they easily able to find out.

2.4 Elections and political discourse

Disinformation and similar challenges have potential to affect all political discourse, threatening political engagement not only in healthy democracies but in all societies. While their impact may be starkest in election and referendum campaigns, there is scope for disinformation, misuse of personal data and all the other phenomena discussed here in all political discourse, conducted on an ongoing basis – for example, as regards attitudes to President Trump in the US, or Brexit in the UK.[19] Disinformation in elections is part of a broader problem arising from the spread of disinformation in day-to-day online discourse, which has encouraged tribalism and a polarization of views on a wide range of societal issues ranging from vaccination of young children to gender and transgender issues. This polarization feeds into voters' preferences in elections and into the tenor and content of political debate.

[19] High Level Group on fake news and online disinformation (2018), *A multi-dimensional approach to disinformation*, Luxembourg: European Commission, p. 12, https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation (accessed 5 Oct. 2019): 'Equally important [with communication around elections] is the threat of more insidious and low profile disinformation strategies which are not linked to any political event. By creating repeated distortions impacting citizens' perceptions of events, these can give rise to deep-seated misinformed beliefs and cause significant harm. The fact that these streams of disinformation are potentially less recognisable and harder to track compounds their potential damage.'

3. Cyber Activities That May Influence Voters

It is important to dispel the misconception that the core challenge posed by disinformation and other election manipulation techniques is the transmission of incorrect information. The veracity of information is only the tip of the challenge. Social media uses techniques not just to inform but also to manipulate audience attention. Evidence shows that determining whether a message is appealing and therefore likely to be read and shared widely depends not on its veracity, but rather on four characteristics: provocation of an emotional response; presence of a powerful visual component; a strong narrative; and repetition. The most successful problematic content engages moral outrage and 'high-arousal' emotions of superiority, anger, fear and mistrust. Capturing emotion is key to influencing behaviour. As for repetition, the reiteration of a message that initially seems shocking and wrong can come to seem an acceptable part of normal discourse when repeated sufficiently: repetition normalizes.

Moreover, communication through digital platforms is not just about the sharing of information. Facebook's experiments on so-called social contagion demonstrated that emotions can be manipulated without people's awareness, through sight of others' expressions of emotion on social media.[20] In addition, communication plays a 'ritualistic function' in representing shared beliefs.[21] It drives connections within online communities and 'tribes', and reinforces the sense of 'us against them'. There is a 'emotional allure' to 'having our worldviews supported and reinforced by "confirmatory news"'.[22] This can lead to a 'proliferation of outrage' that supports 'mob rule' online and encourages populism.[23]

There is a clear picture emerging that while a free, uninterrupted flow of information is key for effective political discourse and participation through elections, the channels permitting this currently lend themselves to disinformation and potential manipulation of political participation. Researchers at the Computational Propaganda Research Project (COMPROP), based at the Oxford Internet Institute, University of Oxford, found evidence of 'formally organized social media manipulation campaigns' on the part of political parties or government agencies in 48 countries in 2018, and that 'political parties and governments have spent more than half a billion dollars on the research, development, and implementation of psychological operations and public opinion manipulation over social media' since 2010.[24] They conclude: 'The manipulation of public opinion over social media platforms has emerged as a critical threat to public life.'[25] Yet there is a widespread lack of awareness of this manipulation: 62 per cent of people don't realize that social

[20] Kramer, A. et al (2014), 'Experimental evidence of massive-scale emotional contagion through social networks', Proceedings of the National Academy of Sciences of the United States of America 111 (24): pp. 8788–8790, doi: https://doi.org/10.1073/pnas.1320040111 (accessed 13 Oct. 2019).
[21] Wardle and Derakhshan (2017), *Information Disorder: Toward an interdisciplinary framework for research and policymaking*, p. 7, drawing on the work of communications theorist James Carey.
[22] Ibid., p. 42.
[23] Williams (2018), *Stand out of our Light*, pp. 75-77.
[24] Bradshaw, S. and Howard, P. (2018), *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*, Oxford: Oxford Internet Institute and University of Oxford, http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf (accessed 5 Oct. 2019), p. 3.
[25] Ibid.

media can affect the news they see, and 83 per cent of people are unaware that companies can collect data that other people have shared about them.[26]

At present the boundaries between acceptable and unacceptable online political activity are difficult to discern. The emergence of cyber campaigning in elections has been a phenomenon of the last decade. For example, the proportion of campaign advertising spend declared as spent on digital advertising during UK election campaigns rose from 0.3 per cent in 2011 to 42.8 per cent in 2017.[27] In addition to paid advertising, campaigners may make use of social media platforms 'for free' (e.g. posting, sharing and liking material),[28] known as 'organic reach'.[29] Political consultancies now focus on digital capabilities as these are regarded as fundamental to effective campaigning.[30] Technology has developed more quickly than the norms that guide it, such that digital platforms are currently operating in largely unregulated fields.

Anyone with a computer or mobile device now has the power to create content, to disseminate it to others, and to encourage its dissemination at great speed and with endorsements from third parties along the way. They may do so at minimal cost, without the need for office facilities. They may do so from anywhere in the world, making overseas interference in elections straightforward. For audiences struggling to assess the credibility of any political post online, there is increasing reliance on friend or family endorsement, or the endorsement of a (private or public) group to which they belong.[31] Political parties and campaigning organizations, and their advisers, are increasingly well equipped to harness this potential.

Cyber operations that may influence voters include the *creation* of disinformation and divisive content; the use of specific methods to maximize the *distribution* of that material; and the use of personal data in order to maximize the *influence* of the material over individuals. As above, in some countries politically-motivated state *disruption* to networks presents a further challenge. Each of these is discussed in more detail below.

[26] Doteveryone (2018), *People, Power and Technology: the 2018 Digital Understanding Report*, London: Doteveryone, https://doteveryone.org.uk/wp-content/uploads/2019/07/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf (accessed 5 Oct. 2019), p. 6.
[27] Electoral Commission (2018), *Digital Campaigning: increasing transparency for voters*, https://www.electoralcommission.org.uk/sites/default/files/pdf_file/Digital-campaigning-improving-transparency-for-voters.pdf, p. 4 (accessed 14 Oct. 2019).
[28] Ibid., p. 5.
[29] Ibid., p. 13.
[30] Moore, M. (2018), *Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age*, London: Oneworld Publications, p. xiii.
[31] Wardle and Derakhshan (2017), *Information Disorder: Toward an interdisciplinary framework for research and policymaking*, p. 12.

3.1 The *creation* of disinformation and divisive content

Content may be created by:

- The use of words, pictures and videos with the intention of influencing political opinions in a divisive manner.
- This includes the creation of memes, i.e. pictures with a few words, which are particularly effective in creating an impression on the human consciousness.[32] In the words of New Knowledge's 2018 report *The Tactics & Tropes of the Internet Research Agency* (generally referred to as the 'Disinformation Report'):

  > Memes turn big ideas into emotionally-resonant snippets, particularly because they fit our information consumption infrastructure: big image, not much text, capable of being understood thoroughly with minimal effort. Memes are the propaganda of the digital age.[33]

- The curation of material that may be untrue, deliberately misleading, exaggerated, or true but intentionally divisive. It may also include hate speech and divisive speech.
- The 'trolling' of people or issues: creation of posts with the aim of annoying, angering, dividing or harassing. This may be done in an organized fashion: there are, for instance, people who are paid to act as 'trolls'.
- The use of 'deep fakes', such as audio and video whose fabrication is increasingly difficult to detect (by humans or algorithms). This is a growing risk as technology becomes more sophisticated.
- The impersonation of news websites.
- The use of fake websites and fake identities to impersonate authentic actors.
- Presentation of a political campaign as a 'war', with the intention of polarizing views and suppressing the reach of rational debate.[34]

3.1.1 Examples

In the UK, in evidence to the House of Commons Digital, Culture, Media and Sport Committee, Arron Banks described how, during the 2016 referendum campaign, Leave.EU deliberately focused on emotive issues, the 'pressure points' such as 'anger' about immigration, as these were issues that

[32] Wardle and Derakhshan (2017), *Information Disorder: Toward an interdisciplinary framework for research and policymaking*, pp. 39–40.
[33] Diresta et al (2018), *The Tactics and Tropes of the Internet Research Agency*, Austin: New Knowledge, https://cdn2.hubspot.net/hubfs/4326998/ira-report-rebrand_FinalJ14.pdf, p. 50 (accessed 5 Oct. 2019).
[34] Moore (2018), *Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age*, pp. 25–27.

would 'set the wild fires burning' on social media: 'Our skill was creating bush fires and then putting a big fan on and making the fan blow.'[35]

New Knowledge's 'Disinformation Report',[36] an extensive report on the activities of the Russia-based Internet Research Agency in US politics 2015–17 prepared for the US Senate Select Committee on Intelligence, identifies that the agency adopted a number of themes, taken from across the political spectrum, 'to create and reinforce tribalism within each targeted community',[37] including by demonizing groups outside that community. Its themes were primarily social issues – from racial and religious, to party political, to 'issue political' such as feminist culture, gun rights and trust in the media. For example, one of its many memes featured in its top half a picture of Donald Trump and Mike Pence, backed by the Stars and Stripes, with the slogan 'Like for Jesus Team', and in its lower half a much darker image of Hillary Clinton and Tim Kaine with devil's horns, with the slogan 'Ignore for Satan Team'.[38]

New Knowledge also reports how the Internet Research Agency impersonated state and local news, with approximately 109 Twitter accounts falsely presenting as genuine news organizations.[39] In addition, the agency was found to have deliberately amplified conspiracy theories.[40] It aimed to foster division, including through involvement in the Brexit debate and the Catalonian independence movement in Europe, and in support for Texan and Californian independence in the US.[41]

As regards the UK, the House of Commons Digital, Culture, Media and Sport Committee's 2019 report on disinformation and fake news[42] concluded that there is 'strong evidence that points to hostile state actors influencing democratic processes', citing research undertaken by Cardiff University and the Digital Forensics Lab of the Atlantic Council. The research agency 89up found that 'RT and Sputnik published no fewer than 261 media articles on the EU referendum in the UK, or with a strong anti-EU sentiment which mentioned Brexit from 1 January 2016 until 23 June 2016.'[43] There is evidence that the Internet Research Agency attempted to undermine UK government communications in the aftermath of the March 2018 Skripal poisonings, as well as to stoke social divisions and religious tensions in the aftermath of terrorist attacks in the UK – including the Westminster, Manchester, London Bridge and Finsbury Park attacks.[44]

By way of example of a 'deep fake', a YouTube video was widely disseminated in Mexico prior to the 2018 elections that appeared to be authentic Russia Today (RT) coverage of Vladimir Putin's strong

[35] Digital, Culture, Media and Sport Committee (2018*), Oral Evidence on Fake News of Arron Banks and Andy Wigmore*, London: House of Commons, pp. 30–31, http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/85344.pdf (accessed 14 Oct. 2019).
[36] Diresta et al (2018), *The Tactics and Tropes of the Internet Research Agency*.
[37] Ibid., p. 12.
[38] Ibid., p. 12.
[39] Ibid., p. 66.
[40] Ibid., p. 69.
[41] Ibid., p. 71.
[42] Digital, Culture, Media and Sport Committee (2019), *Disinformation and 'fake news': Final Report*, London: House of Commons, https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf (accessed 5 Oct. 2019).
43 89up (2018), 'Putin's Brexit? The Influence of Kremlin media & bots during the 2016 UK EU referendum', https://www.slideshare.net/89up/putins-brexit-the-influence-of-kremlin-media-bots-during-the-2016-uk-eu-referendum (accessed 5 Oct. 2019).
[44] Cardiff University Crime and Security Research Institute (2017), *Russian influence and interference measures following the 2017 UK terrorist attacks*, Cardiff: University Crime and Security Research Institute, https://crestresearch.ac.uk/resources/russian-influence-uk-terrorist-attacks (accessed 5 Oct. 2019).

endorsement of the leftist candidate for the presidency (and eventual victor), Andrés Manuel Lopez Obrador, as 'the next protégé of the regime'. This had been edited from a genuine video discussing Putin's support for the Russian bodybuilder Kirill Tereshin.[45]

To illustrate the framing of a political campaign as a conflict, Martin Moore discusses the tactics of Breitbart and others in the 2016 US election campaign in presenting the campaign as a war in order to recruit extreme voices, such as individuals and groups who post on the imageboard sites 4chan and 8chan, with their extreme methods such as creation of hyper-partisan memes, hacking of opinion polls and harassment of opponents. Moore describes such tactics as a 'deliberate transgression and destruction of democratic norms in the digital sphere'.[46]

Allegedly, disinformation has been rife in other countries in election periods. The University of Oxford's Computational Propaganda Research Project found evidence of disinformation or manipulated media in 52 of 70 countries surveyed in 2018, ranging from Italy and Germany to Brazil, China, South Africa and Nigeria.[47]

3.2 The *distribution* of disinformation and divisive content

The methods used by political campaigners to influence populations are not restricted to advertising. They also include a range of other methods which may be subtle and hard to detect. Disinformation for the influence of elections or other political campaigns may be distributed online:

- Through adverts.
- Through posts of targeted political messages, as well as likes, shares, retweets, etc.
- Through an understanding of how best to exploit the digital platforms' algorithms for promotion of content.
- Through encouraging others to innocently like, share, retweet, etc. ('content laundering'), thereby harnessing the motivation of peer pressure.
- Through development of an appearance of grassroots support ('astroturfing') by means of multiple posts, using the following tools:
    - Through the use of bots – i.e. software programs that mimic real people by posting, sharing and liking posts, usually at scale. Campaigners can purchase bots to help their campaigns.[48]

[45] Fregoso, J. (2018), 'Mexico's Election and the Fight against Disinformation', European Journalism Observatory, 27 September 2018, https://en.ejo.ch/specialist-journalism/mexicos-election-and-the-fight-against-disinformation (accessed 5 Oct. 2019).
[46] Moore (2018), *Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age*, p. 35.
[47] Bradshaw and Howard (2019), 'The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation' University of Oxford Computational Propaganda Research Project, https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf (accessed 31 Oct. 2019).
[48] Electoral Commission (2018), *Digital Campaigning: increasing transparency for voters*, p. 7. The OII Computational Propaganda Research Project defines high-frequency accounts as those which tweet more than 50 times per day on average, while recognising that not all bots tweet that frequently and some humans do so. See Wardle and Derakhshan (2017), *Information Disorder: Toward an interdisciplinary framework for research and policymaking*, p. 38.

Related to bots are cyborg accounts, operated in part by software and in part (so as to appear legitimate and avoid detection) by humans.

- Through 'fake' accounts (operated by humans under fictitious names) that post, share, like and otherwise build electoral support (astroturfing). [49]

- Not only through 'open' social media sites such as Facebook and Instagram, but increasingly through 'closed' peer-to-peer distribution networks such as WhatsApp and Facebook Messenger. Such private channels are increasingly the means for transmission of political material including disinformation (the 'pivot to private').[50]

3.2.1 Examples

Twitter reported that in May 2018, its systems 'identified and challenged more than 9.9 million potentially spammy or automated accounts per week'.[51] This gives some sense of the scale of the problem. In 2016, it was estimated that more than half of all web traffic was attributable to bots.[52]

The use of Facebook has influenced political sentiment all over the world, from its contributory role in the alleged genocide in Myanmar,[53] to a resurgence of intercommunal violence in Sri Lanka,[54] to rampant mistrust of information in India.[55] There is little information available as to the extent to which this influence results from disinformation and the extent to which it results from the non-strategic circulation of rumour and inflammation of emotions, without intention to cause harm.

As regards disinformation, New Knowledge's 'Disinformation Report'[56] states that, between 2014 and 2017:

- On Facebook, the Internet Research Agency attracted 3.3 million page followers, who generated 76.5 million engagements. These included 30.4 million shares, 37.6 million likes, 3.3 million comments, and 5.2 million reactions across the content. The agency held 81 Facebook pages, of which 33 had over 1,000 followers.[57] Facebook estimated that the agency's content was seen by 126 million Facebook users.[58]

[49] Electoral Commission (2018), *Digital Campaigning: increasing transparency for voters*, p. 8.
[50] Newman, N. et al (2019), *Reuters Institute Digital News Report 2019*, Oxford: Reuters Institute for the Study of Journalism, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_1.pdf (accessed 5 Oct. 2019), p. 9.
[51] Roth, Y. and Harvey, D. (2018), 'How Twitter is fighting spam and malicious automation', *Twitter* blog, 26 June 2018, https://blog.twitter.com/official/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html (accessed 5 Oct. 2019).
[52] Zeifman, I. (2017), 'Bot Traffic Report 2016', *Imperva* blog, 24 January 2017, https://www.imperva.com/blog/bot-traffic-report-2016/ (accessed 5 Oct. 2019).
[53] UN Human Rights Council (2018), *Report of the independent international fact-finding mission on Myanmar*, UN Doc A/HRC/39/64 (12 September 2018), para. 74.
[54] Easterday, J. and Ivanhoe, H. (2018), 'Tech companies' inability to control fake news exacerbates violent acts', https://www.openglobalrights.org/tech-companies-inability-to-control-fake-news-exacerbates-violent-acts/ (accessed 5 Oct. 2019).
[55] Kleis Nielsen, R. (2019), 'Disinformation is everywhere in India', *The Hindu*, 25 March 2019, https://www.thehindu.com/opinion/op-ed/disinformation-is-everywhere-in-india/article26626745.ece (accessed 5 Oct. 2019).
[56] Diresta et al (2018), *The Tactics and Tropes of the Internet Research Agency*. See also Mueller R. (2019), *Report on the Investigation into Russian Interference in the 2016 Presidential Election (The Mueller Report)*, Washington: U.S. Department of Justice, https://www.justice.gov/storage/report.pdf (accessed 13 Oct. 2019). pp. 14–34.
[57] Diresta et al (2018), *The Tactics and Tropes of the Internet Research Agency*, p. 21.
[58] Ibid., p. 33. As of 31 March 2019, Facebook had over 2.38 billion monthly active users: see footnote 1.

- The agency held 3,841 accounts on Twitter, generating nearly 73 million engagements with their content from approximately 1.4 million people.[59]
- On Instagram, the agency held 12 accounts with over 100,000 followers each, and its top accounts each had 'millions to tens of millions of interactions'.[60] Facebook estimates that content was seen by 20 million Instagram users; the report authors consider the true figure to be higher.[61]
- The agency produced 1,107 videos across 17 YouTube channels.[62]
- The agency used merchandising to increase engagement, particularly on Instagram.[63]
- The agency's advertising operation consisted of 3,519 ads (video and images), used to encourage users to 'like' pages, follow Instagram accounts, join events and visit websites.[64]

In evidence to the Mueller inquiry, Facebook testified that the Internet Research Agency reached an estimated 126 million people on Facebook. The Mueller report further states that Twitter informed 1.4 million users that they may have had contact with an account controlled by the agency.[65]

89up's research on Russian influence in the Brexit referendum found that the Russian articles that went most viral were those with the heaviest anti-EU bias. It found that 'the social reach of these anti-EU articles published by the Kremlin-owned channels was 134 million potential impressions, in comparison with a total reach of just 33 million and 11 million potential impressions for all content shared from the Vote Leave website and Leave.EU website respectively'.[66] 89up estimated that the comparable cost for a paid social media campaign would have been between £1.4 million and £4.14 million.[67]

As regards the growing use of private networks: in Brazil and Malaysia, 53 per cent and 50 per cent respectively of those using social media use WhatsApp as a source of news.[68] In India there are between 200 million and 300 million WhatsApp users, more than half of whom are estimated to obtain news on WhatsApp.[69] WhatsApp was a significant channel for political campaigning in recent elections in India[70] and in Brazil.[71] Although the maximum number of contacts in a WhatsApp group is 256, there are ways of avoiding this limit. In July 2018 WhatsApp reduced the number of contacts/groups to whom a message could be forwarded in one action, from 100 to 20

[59] Ibid., p. 18.
[60] Diresta et al (2018), *The Tactics and Tropes of the Internet Research Agency*, pp. 26–27.
[61] Ibid., p. 21.
[62] Ibid., p. 16.
[63] Ibid., pp. 30–31.
[64] Ibid., p. 34.
[65] Mueller, R. (2019), *The Mueller Report*, p. 15.
[66] Digital, Culture, Media and Sport Committee (2019), *Disinformation and 'fake news': Final Report*, paras 242–243.
[67] 89up (2018), 'Putin's Brexit? The influence of Kremlin media & bots during the 2016 UK EU referendum', https://www.slideshare.net/89up/putins-brexit-the-influence-of-kremlin-media-bots-during-the-2016-uk-eu-referendum (accessed 1 Nov. 2019).
[68] Newman et al (2019), *Reuters Institute Digital News Report 2019*, p. 38.
[69] Iqbal, M. (2019), 'WhatsApp Revenue and Usage Statistics (2019)', https://www.businessofapps.com/data/whatsapp-statistics/ (accessed 5 Oct. 2019).
[70] Ponniah, K. (2019), 'WhatsApp: The 'black hole' of fake news in India's election', BBC News, 6 April 2019, https://www.bbc.co.uk/news/world-asia-india-47797151 (accessed 5 Oct. 2019).
[71] Belli, L. (2018), 'WhatsApp skewed Brazilian election, proving social media's danger to democracy', https://theconversation.com/whatsapp-skewed-brazilian-election-proving-social-medias-danger-to-democracy-106476 (accessed 5 Oct. 2019).

globally, and to five specifically in India.[72] It extended the restriction to five worldwide in January 2019.[73] Nonetheless, there have been ways of circumventing this restriction,[74] and there are volunteers who spend their days forwarding WhatsApp political messages to others, a few groups at a time.[75]

3.3 Maximization of the *influence* of disinformation and divisive content

As already discussed, the most influential content is content that is divisive and that provokes an 'arousal' emotional response such as anger or disgust. In addition to tailoring and promoting content so as to maximize influence, political campaigners are using personal data to target their advertising and campaign materials at specific voters so as to have maximum impact. The digital availability of personal data, and the capacity to deduce personality traits and decision-making drivers from it,[76] mean that political messages can be much more precisely targeted to specific audiences than in the days when campaign material was simply posted through letterboxes.

In the election context, political campaigners and advertisers harness personal data for the purposes of micro-targeting posts and advertisements using sophisticated psychological profiling techniques – i.e. using 'targeting techniques that use data analytics to identify the specific interests of individuals, create more relevant or personalised messaging targeting those individuals, predict the impact of that messaging, and then deliver that messaging directly to them' such as to maximise their impact upon their audience.[77] While there is currently a move towards transparency in political advertising, there is no move away from micro-targeting, an accepted and growing means of election campaigning.

3.3.1 Examples

The UK Information Commissioner's Office has found that all the major UK political parties are using a wide range of personal data to create a personal profile on each voter, from which to target them individually; that they are failing to explain what data they are gathering and how they would use it;[78] and that they are failing to apply due diligence when obtaining information from data brokers[79] to satisfy themselves that data has been obtained lawfully.[80] The House of Commons Digital, Culture, Media and Sport Committee and the Information Commissioner's Office have exposed significant data harvesting, trade and use for targeted political advertising in advance of

[72] The lower limit in India was introduced in July 2018 in response to violence, including five deaths, caused by false WhatsApp rumours of an active child-abducting gang.
[73] WhatsApp blog, 'More changes to forwarding', 19 July 2018, updated 21 January 2019, https://blog.whatsapp.com/ (accessed 27 Oct. 2019).
[74] Banaji, S. and Bhat, R. (2019), 'WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India', London School of Economics and Political Science, http://www.lse.ac.uk/media-and-communications/assets/documents/research/projects/WhatsApp-Misinformation-Report.pdf (accessed 27 Oct. 2019).
[75] Perrigo, B. (2019), 'How volunteers for India's ruling party are using WhatsApp to fuel fake news ahead of elections', *Time*, 25 Jan, 2019, https://time.com/5512032/whatsapp-india-election-2019/ (accessed 19 Jan. 2016).
[76] Moore (2018), *Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age,* pp. 65–66.
[77] UK Information Commissioner's Office (2018), *Democracy Disrupted*, 11 July 2018, https://ico.org.uk/media/2259369/democracy-disrupted-110718.pdf (accessed 5 Oct. 2019) p. 27.
[78] Ibid., pp. 28–29.
[79] Businesses that aggregate and sell data on individuals, businesses, etc such as (in the UK) Oracle and Acxiom.
[80] Ibid., p. 31.

the UK's 2016 referendum on EU membership.[81] In the EU, 67 per cent of 28,000 people surveyed by the European Commission in November 2018 were concerned that their personal data online could be used to target them with political messages.[82]

3.4 State *disruption* to networks

Digital platforms have reported that they have been the subject of an 'exponential increase in politically motivated demands for network disruptions, including in election periods'.[83] These can include restrictions, filtering or shutdown as well as censorship. The UN Secretary-General's High-level Panel on Digital Cooperation reports that governments directed 188 internet shutdowns in 2018 (more than 100 of which in India), up from 108 in 2017.[84]

3.4.1 Examples

Sudan[85] and the Democratic Republic of the Congo[86] were both subject to internet shutdowns in 2018, as was Gabon[87] in January 2019. Also in January 2019, at a time of civil unrest due to rising fuel prices, the Zimbabwean government ordered service providers to block the internet. After a 30-hour shutdown, the internet was turned back on, only for another shutdown to be ordered. Internet service providers had no choice but to comply with the government's decrees. A Zimbabwean court later ruled that the government's orders to shut down the internet were illegal, and required them to resume full services.[88] The transitional government of Sudan again ordered the shutdown of mobile access to the internet during political turbulence in June 2019.[89] And in August 2019 the Indian government, following its revocation of the special status of Kashmir, shut down internet, mobile phone networks and television channels in Jammu and Kashmir.[90]

[81] Ibid.; Digital, Culture, Media and Sport Committee (2018), *Disinformation and 'fake news': Interim Report,* London: House of Commons, https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf (accessed 14 Oct. 2019); Digital, Culture, Media and Sport Committee (2019), *Disinformation and 'fake news': Final Report.* For details of data harvesting and targeting practices, see Digital, Culture, Media and Sport Committee (2018), *Oral Evidence on Fake News of Brittany Kaiser,* London: House of Commons, http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/81592.pdf (accessed 14 Oct. 2019).
[82] European Commission (2018), Special Eurobarometer Reports, 'Special Eurobarometer 477 Report: Democracy and Elections', https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/SPECIAL/surveyKy/2198 (accessed 29 Oct. 2019), p.56.
[83] UNESCO and Global Network Initiative (2018), *Improving the communications and information system to protect the integrity of elections: conclusions*, Paris: UNESCO and Global Network Initiative, p. 4, https://en.unesco.org/sites/default/files/633_18_gni_intergrity_of_elections_final_report_web.pdf (accessed 5 Oct. 2019).
[84] UN Secretary-General's High-level Panel on Digital Cooperation (2019), *The Age of Digital Interdependence*', p. 16.
[85] Taye, B. (2018), 'Amid countrywide protest, Sudan shuts down social media on mobile networks', Access Now, 21 December 2018, https://www.accessnow.org/amid-countrywide-protest-sudan-shuts-down-social-media-on-mobile-networks/ (accessed 5 Oct. 2019).
[86] Burke, J. (2019), 'DRC electoral fraud fears rise as Internet shutdown continues', *The Guardian*, 1 January 2019, https://www.theguardian.com/world/2019/jan/01/drc-electoral-fears-rise-as-internet-shutdown-continues (accessed 5 Oct. 2019).
[87] Internet Sans Frontières (2019), 'Coup attempt and Internet shutdown in Gabon', 8 January 2019, https://internetwithoutborders.org/coup-attempt-and-internet-shutdown-in-gabon/ (accessed 5 Oct. 2019).
[88] Dzirutwe, M. (2019), 'Zimbabwe court says Internet shutdown illegal as more civilians detained', Reuters, 21 January 2019, https://www.reuters.com/article/us-zimbabwe-politics/zimbabwe-court-says-internet-shutdown-during-protests-was-illegal-idUSKCN1PF11M (accessed 5 Oct. 2019).
[89] Internet Society, 'Turn the Internet back on in Sudan, and keep it on', https://www.internetsociety.org/news/statements/2019/turn-the-internet-back-on-in-sudan-and-keep-it-on/ (accessed 5 Oct. 2019).
[90] Holroyd, M. and Davis, S., 'Kashmir: The misinformation spreading online #TheCube', Euronews, 3 September 2019, https://www.euronews.com/2019/09/03/kashmir-the-misinformation-spreading-online-thecube (accessed 5 Oct. 2019).

4. State, EU and Platform Responses

While some governments have responded bluntly to the pluralism of political expression online, disrupting, shutting down, restricting or filtering internet and telecom network services, others are considering carefully how to respond to disinformation and hate speech while continuing to enable that pluralism. However, governments adopting less scrupulous approaches sometimes do so on the pretext that they are following the more nuanced approaches of responsible states; a factor that those responsible states ought to take into account in deciding on and presenting their responses.[91]

The ultimate aim of responses to interference in political discourse should be to retain the advantages that social media bring – in terms of far greater access to information and alternative sources of information than ever before – while tackling disinformation and manipulation that undermine that discourse. Currently, there is a tendency for responses to focus on the latter without due regard to the former.[92]

The activities of digital platforms have in some areas run ahead of regulation, and in others are subject to general regulation (for example, data protection laws) whose implementation is currently difficult to monitor in the digital sector. Two overarching unresolved issues are the extent to which platforms' activities ought to be specifically regulated, and how transparent their activities ought to be for the purposes of oversight of implementation of regulation. As regards specific regulation, one open issue is the extent to which digital platforms, as private commercial entities, should be obliged to take account of public interests – for example in being accessible to all, and in providing a platform for free expression. To the extent that digital platforms do take account of public interests, there is then a question as to whether the public sector ought to have a role in their decision-making or its oversight.[93]

This section provides an overview of the current legal position and some of the regulatory debates or initiatives under way. It discusses the approaches adopted or under discussion in the US and in some parts of Europe, as well as one example of a state – Singapore – banning online falsehoods. It also touches on cross-jurisdictional initiatives. In total, more than 30 states have introduced legislation designed to combat disinformation online since 2016.[94] The Poynter Institute website features a comprehensive mapping of state actions to combat disinformation globally.[95]

In the absence of international or comprehensive national regulation, many digital platforms have developed, or are developing, their own standards on impermissible speech and procedures for reporting it, and on transparency of and imprints[96] in political advertising. These rules and procedures are liable to criticism from all sides, as might be expected given both the novel, cross-jurisdictional challenges that platforms are faced with and the lack of institutional legitimacy that

[91] Kaye, D. (2019), *Speech Police: The Global Struggle to Govern the Internet*, New York: Columbia Global Reports, pp. 101, 113.
[92] Ibid., p. 125.
[93] Ibid., p. 112.
[94] Bradshaw and Howard (2018), *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation,* p. 6.
[95] Funke, D. and Flamini, D. (2019), 'A guide to anti-misinformation actions around the world', https://www.poynter.org/ifcn/anti-misinformation-actions/ (accessed 5 Oct. 2019).
[96] The labelling of advertisements with their source and payment origin.

private entities have to develop systems of regulation. Human rights law has played little role in the calibration of these standards, save that many decisions not to regulate have been taken in the name of the principle of free expression. Some platforms have called for greater state regulation or guidance as regards both impermissible content and data protection.[97] However, others argue that state regulation of freedom of speech carries significant risks as it may incentivize removal over retention of disputed content and may give undue weight to vested interests in restricting comment.[98]

Private and public fact-checking and digital literacy initiatives have also been established in order to improve the veracity of information and audience responses to it. Both of these have an important contribution to make in improving user engagement with a plurality of expression. However, they are not sufficient to address the manipulative techniques discussed above.

4.1 The US

4.1.1 Expression

In line with the US Constitution's First Amendment, the US has probably the strongest protections of freedom of expression in the world, and a culture of vigorous respect for those protections. As the US hosts most of the major digital platforms and other tech companies, this culture, with its reliance on the 'marketplace of ideas' to counter false information, has to date played a sculpting role in a global approach that disfavours interference with content on digital platforms.[99] As already discussed, digital platforms generally have immunity from legal liability in the US for third party content hosted on their sites.[100]

There are no moves towards comprehensive regulation of all digital platforms, or of posters of material in the election context. In October 2017 Senators Mark Warner, Amy Klobuchar and John McCain proposed the Honest Ads Act (S.1989).[101] This legislation would require digital platforms with at least 50 million monthly views to increase the transparency of election adverts,[102] as some digital platforms have already done,[103] and to make all reasonable efforts to ensure that adverts are not purchased 'directly or indirectly' by foreign individuals or entities. An identical bill was proposed in the House of Representatives. The bills did not make progress, but were reintroduced in May 2019.

[97] For example, see Zuckerberg, M. (2019), 'Four Ideas to Regulate the Internet', https://newsroom.fb.com/news/2019/03/four-ideas-regulate-internet/ (accessed 14 Oct. 2019).
[98] For example, Walker, K. (2019), 'How we're supporting smart regulation and policy innovation in 2019', Google *The Keyword* blog, 8 January 2019, https://www.blog.google/perspectives/kent-walker/principles-evolving-technology-policy-2019/ (accessed 4 Nov. 2019).
[99] Klonick, K. (2018), 'The New Governors: The People, Rules and Processes Governing Online Speech' *Harvard Law Review,* 131(6): pp. 1598–1670, https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf (accessed 14 Oct. 2019).
[100] 47 USC § 230 (2011).
[101] Honest Ads Act, S. 1989, 115th Congress (2017-2018).
[102] The legislation would require these digital platforms maintain a public file of all election adverts bought by anyone who spends more than $500 on their platform, with a copy of each advert, a description of the target audience, the number of views generated, the dates and times of publications, the rates charged.
[103] For example, see Facebook Business (2019), 'Bringing More Transparency to Political Ads in 2019', https://www.facebook.com/business/news/bringing-more-transparency-to-political-ads-in-2019 (accessed 5 Oct. 2019); Twitter (2019), 'Expanding Transparency Around Political Ads on Twitter', *Twitter* blog, 19 February 2019, https://blog.twitter.com/en_us/topics/company/2019/transparency-political-ads.html (accessed 5 Oct. 2019).

4.1.2 Privacy

The US has no comprehensive data protection legislation. There are calls for legislation similar to the EU's General Data Protection Regulation (GDPR) (see below). Some social media companies would support more regulation in this field. For example, Facebook has called for a 'globally harmonized framework' for effective privacy and data protection, in line with the standards of the GDPR.[104] California has passed the 2018 California Consumer Privacy Act (CCPA),[105] scheduled to come into force on 1 January 2020, which, in default of national legislation, will impose standards similar to GDPR on the California-based tech industry. Talks on a draft national bill in the Senate Commerce Committee (regarded by many as a way of reducing the impact of the CCPA through the imposition of weaker standards at federal level) have recently stalled.[106]

4.1.3 Media literacy

California passed a law in September 2018 to promote media literacy education in schools.[107]

4.2 The EU

4.2.1 Expression

As already noted, the eCommerce Directive[108] prohibits EU member states from imposing general obligations on digital platforms to monitor user-generated content. It exempts digital platforms from liability for the content provided on their services, on condition that, if aware of illegal content, they will remove it or disable access to it expeditiously. The European Court of Justice has held that the Directive does not preclude a court of a member state from ordering a host provider to remove information that is identical or equivalent to information previously found to be unlawful, provided that any differences between the original and equivalent content are not so great as to require the platform to carry out an independent assessment of the content.[109] On 1 March 2018, the European Commission issued a recommendation concerning processes that online platforms should adopt in order to expedite the detection and removal of illegal content.[110] Some EU member states, notably Germany, France and the UK, have adopted or are considering further regulation (see below).

As regards hate speech, in May 2016 the European Commission agreed, with Facebook, Microsoft, Twitter and YouTube, a Code of Conduct on Countering Illegal Hate Speech Online.[111] Under the Code of Conduct, the platforms agreed to have clear processes in place to review notifications

[104] Zuckerberg (2019), 'Four Ideas to Regulate the Internet'.
[105] California Department of Justice, Office of the Attorney General (2019), *California Consumer Privacy Act (CCPA): Fact Sheet*, https://oag.ca.gov/system/files/attachments/press_releases/CCPA%20Fact%20Sheet%20%2800000002%29.pdf (accessed 28 Oct. 2019).
[106] Stacey, K. (2019), 'Senate talks on US data privacy law grind to a halt', *Financial Times*, 12 June 2019, https://www.ft.com/content/ecbc11d0-8bad-11e9-a24d-b42f641eca37 (accessed 13 Oct. 2019).
[107] *An act to add Section 51206.4 to the Education Code, relating to pupil instruction*, ch 448, § 51206.4, 2018 Stat 830.
[108] Directive on electronic commerce [2000] OJ L 178/1.
[109] Case C-18/18, *Eva Glawischnig-Piesczek v Facebook Ireland Limited* (CJEU, 3 October 2019).
[110] European Commission, 'Recommendation of 1.3.2018 on measures to effectively tackle illegal content online'.
[111] European Commission (2016), 'Code of Conduct on Countering Illegal Hate Speech Online' Luxembourg: European Commission, https://ec.europa.eu/newsroom/just/document.cfm?doc_id=42985 (accessed 5 Oct. 2019).

regarding illegal hate speech; to review the majority of notifications of such content within 24 hours and take down the content if necessary; and to improve user awareness, feedback and transparency. Instagram, Google+, Snapchat, Dailymotion and jeuxvideo.com have also signed up since the beginning of 2018.[112] The Commission monitors compliance with the Code of Conduct. In its most recent evaluation, dated February 2019, it found that the companies had assessed 89 per cent of notifications within 24 hours, up from just 40 per cent in 24 hours in 2016. Of the notifications submitted in the six-week review period, 72 per cent resulted in removal of material as illegal hate speech. The proportion of content removed varied significantly, from 85 per cent by YouTube and 82 per cent by Facebook to just 44 per cent by Twitter.[113] Some commentators have been concerned that the Code of Conduct essentially legitimates the practices of the platforms while not allowing for any public-sector involvement in decision-making.[114] The quality of platforms' decision-making is not reviewed. As with all review initiatives, there is a risk that the Code of Conduct incentivizes digital platforms to take down, rather than retain, contentious content that would otherwise be protected under international human rights norms.

The European Court of Human Rights[115] has upheld a finding of defamation against a news portal which failed to take down clearly unlawful hate speech posted in comments to one of its articles *before* receiving a complaint about the speech, as being within the state's margin of appreciation (i.e. within the discretion that the state has in deciding how to meet its human rights obligations);[116] but the case is distinguished by its specific circumstances.[117] The Court has overturned a domestic finding of liability on the part of a social media company and a news portal in a case in which the language complained of, while offensive, did not constitute unlawful speech.[118]

In January 2018, the European Commission mandated a High Level Expert Group on Fake News and Online Disinformation, which reported in March 2018.[119] The Group took as its starting points both freedom of expression and the right it includes to receive and impart information. Its recommendations focused on increasing the transparency of online news, by means of sharing data about the systems that enable its circulation online; promoting media and information literacy to help users navigate the digital environment; developing tools to empower users and journalists to tackle disinformation; and safeguarding the diversity of European news.

In December 2018, the European Council endorsed the European Commission and the High Representative's joint Action Plan against Disinformation.[120] Initiatives to be taken under the

[112] European Commission (2019), 'The EU Code of conduct on countering illegal hate speech online', https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en (accessed 5 Oct. 2019).
[113] Jourová, V. (2019), 'How the Code of Conduct helped countering illegal hate speech online', *European Commission* Factsheet, February 2019, https://ec.europa.eu/info/sites/info/files/hatespeech_infographic3_web.pdf (accessed 5 Oct. 2019).
[114] Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet,* p. 72.
[115] The European Court of Human Rights monitors the 47 Council of Europe Member States' compliance with the European Convention on Human Rights, rather than the law of the European Union.
[116] *Delfi AS v Estonia* (2015) ECtHR 64669/09.
[117] The case concerned a professional and commercial news portal with an economic interest in the posting of comments. Its website professed to prohibit inappropriate and unlawful comments and stated that only it could modify or delete comments posted. Therefore, the Court found it to have had 'a substantial degree of control' over the comments on its portal, its involvement in making public the comments being 'beyond that of a passive, purely technical service provider' (paras 145–146). The offending comments were on its website for six weeks before they were complained of and taken down. The Court distinguished the news portal from other Internet fora, such as discussion fora and social media platforms (para. 116).
[118] *MTE and Index.hu ZRT v Hungary* (2016) ECtHR 22947/13.
[119] High Level Group on fake news and online disinformation(2018), *A multi-dimensional approach to disinformation.*
[120] European Commission (2018), *Action Plan Against Disinformation*, Brussels: European Commission, https://ec.europa.eu/commission/sites/beta-political/files/eu-communication-disinformation-euco-05122018_en.pdf (accessed 5 Oct. 2019).

Action Plan include the identification of disinformation, and response to it through a rapid alert system intended to disseminate accurate and relevant facts. In addition to online platforms' compliance with the Code of Conduct, scrutiny of transparency of political advertising and closing down of fake accounts, the Action Plan aims to raise awareness of disinformation and build resilience to it, including through the use of civil society, independent fact-checkers and researchers.

Notwithstanding these efforts, the May 2019 elections to the European Parliament were found to have been the target of disinformation campaigns, including 'coordinated inauthentic behaviour aimed at spreading divisive material … including through the use of bots and fake accounts'.[121] However, there was variation in the scale of problems reported.[122] The European Commission has urged digital platforms to provide more information to enable identification of malign actors, to work more closely with fact-checkers and to empower users to detect disinformation, and has not ruled out further action.[123] At its June 2019 Summit, the European Council welcomed an initiative by the Commission to evaluate in depth digital platforms' implementation of their Code of Conduct commitments, noting: 'The evolving nature of the threats and the growing risk of malicious interference and online manipulation associated with the development of Artificial Intelligence and data-gathering techniques require continuous assessment and an appropriate response.'[124]

4.2.2 Privacy

The General Data Protection Regulation (GDPR) is founded on the right to protection of personal data in EU law.[125] Applicable to all companies processing personal data of individuals within the EU, regardless of the company's location, the GDPR imposes controls on the processing of personal data, requiring that data be processed lawfully, fairly and transparently.[126] It provides a right of access not only to data held, but also to information about profiling.[127] Nonetheless, such information, particularly on profiling, is not currently quickly or easily available. Even when rules are in place for the effective management of personal data, there is not currently a culture that embeds them in the design of technology, allows for easy access to data, or sees compliance as a human rights rather than merely a technical issue.[128] In the political context, European data

[121] European Commission (2019), 'A Europe that protects: EU reports on progress in fighting disinformation ahead of European Council', Press release, 14 June 2019, https://ec.europa.eu/digital-single-market/en/news/europe-protects-eu-reports-progress-fighting-disinformation-ahead-european-council (accessed 1 Nov. 2019).
[122] Meyer-Resende, M., 'Six takeaways on digital disinformation at EU elections', EU Observer, 4 June 2019, https://euobserver.com/opinion/145062?fbclid=IwAR0wz4x6PqqFBh6Y3R5J4pg8N837lpcllrv_4ANiGQFcwhger6yTUe8kzvo (accessed 5 Oct. 2019).
[123] European Commission (2019), 'A Europe that protects: EU reports on progress in fighting disinformation ahead of European Council'.
[124] European Council (2019), 'European Council conclusions on the MFF, climate change, disinformation and hybrid threats, external relations, enlargement and the European Semester', Press release, para. 6, 20 June 2019, https://www.consilium.europa.eu/en/press/press-releases/2019/06/20/european-council-conclusions-20-june-2019/ (accessed 29 Oct. 2019).
[125] Charter of Fundamental Rights of the European Union, Article 8(1); Treaty on the Functioning of the European Union, Article 16(1).
[126] Parliament and Council Regulation EU 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (GDPR), OJ L 119/1, Articles 5 and 6.
[127] GDPR Article 15. Some aspects of its operation are not yet resolved. See for example the ongoing *Reference for a Preliminary Ruling from the High Court (Ireland) made on 9 May 2018, Data Protection Commissioner v Facebook Ireland Limited, Maximilian Schrems* (CJEU,C-311/18, 16 July 2018), concerning the standards applicable to transfer of personal data from a private company in the EU to one in a third state, particularly if the data may be further processed in the third state for reasons of national security.
[128] Denham, E., 'Elizabeth Denham's speech at the Data Protection Practitioners' Conference on 8 April 2019' (Speech, Information Commissioner's Office, 8 April 2019).

protection standards include limited exemptions for political parties, but not in respect of processing data of prospective members or voters.[129]

In March 2019, in advance of the European Parliament elections, the EU adopted legislation allowing financial sanctions on European political parties that have breached EU data protection regulations with the aim of influencing the outcome of the elections.[130]

4.3 Germany

Germany has taken the lead in legislating to require digital platforms to act to remove hate speech published via their sites. After finding that a request to digital platforms to remove hate speech within 24 hours did not work, Germany adopted specific legislation in the form of the Network Enforcement Act (Netzwerkdurchsetzungsgesetz), known as NetzDG.[131] With effect from 1 January 2018, NetzDG requires digital platforms with more than 2 million members to have an effective and transparent procedure in place to handle content removal requests by reference to the German criminal code. Platforms must take down the most egregious ('obviously illegal') content within 24 hours of notification, and must decide on less obvious cases within seven days (or longer in exceptional cases). Fines can be imposed – not in respect of a platform's judgment in a particular case, but for 'systematic' failures on the part of a platform.

During its passage through the Bundestag and on introduction, the law was subject to significant criticism as interfering unduly with freedom of expression, particularly on the grounds that content removal decisions would be made entirely by private companies without input from public authorities, and that platforms would err on the side of taking down disputed content. In the first six months of operation, however, removal rates were relatively low: Facebook removed 21.2 per cent of reported content YouTube 27.1 per cent, Google+ 46.1 per cent, and Twitter 10.8 per cent.[132] In the second half of 2018, Facebook received 500 reports under the legislation, of which 159 resulted in content deletion or blocking.[133] More time and data are needed to establish whether NetzDG is provoking a spike in unmerited requests for content removal, as well as to establish whether the law is actually preventing hate speech.

4.4 France

Two new laws,[134] together constituting the Law Against the Manipulation of Information, came into force in France in December 2018, despite considerable opposition and a challenge in the

[129] Political parties may process sensitive personal data without compliance with the legitimate interest test, but only in respect of members, former members, or persons in regular contact with them. For a full explanation, see European Commission (2018), *Guidance on the application of Union data protection law in the electoral context*, Brussels: European Commission, https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-data-protection-law-electoral-guidance-638_en.pdf (accessed 5 Oct. 2019).
[130] Parliament and Council Regulation (EU, Euratom) amending Regulation (EU, Euratom) No 1141/2014 as regards a verification procedure related to infringements of rules on the protection of personal data in the context of elections to the European Parliament, [2019] OJ L 851/7.
[131] *The Network Enforcement Act* (Germany) 1 September 2017 (BGBI. I S. 3352).
[132] Gollatz, K., Riedl, M. and Pohlmann, J. (2018), 'Removals of online hate speech in numbers', *The Alexander von Humboldt Institute for Internet and Society (HIIG)* blog, 9 August 2018, https://www.hiig.de/en/removals-of-online-hate-speech-numbers/ (accessed 5 Oct. 2019).
[133] Facebook (2019), *NetzDG Transparency Report*, https://fbnewsroomde.files.wordpress.com/2019/01/facebook_netzdg_january_2019_english71.pdf (accessed 5 Oct. 2019).
[134] LOI n° 2018-1201 du 22 décembre 2018 [Law No 2018-1201 of 22 December 2018] (France) JO, 23 December 2018, 1 ; LOI n° 2018-1202 du 22 décembre 2018 [Law No 2018-1202 of 22 December 2018] (France) JO, 23 December 2018, 2.

Constitutional Court. The Court concluded that the legislation complies with French constitutional principles, including freedom of expression, in light of its embedded safeguards. These safeguards include that the inaccuracy or misleading nature of the impugned material must be obvious; that it must not comprise opinions, parodies, partial inaccuracies or exaggerations; and that the removal obligation imposed on online platform operators is limited to the duration of the election campaign to which it applies.[135] In summary, the impact of the law is as follows:[136]

- It establishes a duty of cooperation for digital platforms, requiring them to introduce measures to combat disinformation and to make these measures public. The Conseil Supérieur de l'Audiovisuel (the French broadcasting authority) will check for compliance with this duty.
- During election campaigns, digital platforms must comply with a transparency obligation, reporting any sponsored content by publishing the author's name and the amount paid. Platforms that receive a prescribed number of hits a day must publish their algorithms and have a legal representative in France.
- During election campaigns, a judge may order an injunction to halt the circulation of disinformation if it is manifestly false, being disseminated deliberately on a massive scale, and may lead to a disturbance of the peace or compromise the outcome of an election.
- The French government has mandated a taskforce to produce a proposal for a new press ethics body, bringing together journalists, publishers and representatives of civil society.
- Users must be provided with 'fair, clear and transparent' information on how their personal data are being used, and sites must disclose money they have been given to promote information.

It is notable that while Germany's NetzDG is designed to promote expeditious removal of content that is illegal by reference to other provisions of the German criminal code, the French legislation entails the combating of disinformation and the removal of material that would not otherwise be unlawful, albeit only by order of a judge and during an election campaign.

President Emmanuel Macron has proposed the creation of a European agency for the protection of democracies, to safeguard the political process from cyberattacks and manipulation, and the development of EU rules prohibiting online incitement to hatred and violence.[137]

4.5 The UK

In April 2019 the UK government released its Online Harms White Paper,[138] in which it proposes to establish a new statutory duty of care for digital platforms that will require them to tackle harm caused by content on their services. An independent regulator will develop codes of practice on the content of the new duty, and will oversee and enforce its implementation. The regulator will be able

[135] Conseil constitutionnel [French Constitutional Court], décision n° 2018-773, 20 décembre 2018, reported in JO, 23 December 2018.
[136] This summary is derived from Gouvernement.fr (2018), 'Against information manipulation', https://www.gouvernement.fr/en/against-information-manipulation (accessed 5 Oct. 2019).
[137] Macron, E. (2019), 'For European Renewal', https://www.elysee.fr/emmanuel-macron/2019/03/04/for-european-renewal.en (accessed 5 Oct. 2019).
[138] Department for Digital, Culture, Media & Sport; Home Office (2019), *Online Harms White Paper* (White Paper, CP 57, 2019).

to require annual transparency reports and further information from digital platforms, explaining the prevalence of harmful content on their sites and what they are doing to combat it. The White Paper recognizes the threats posed by disinformation and online manipulation.[139] The White Paper does not propose to make digital platforms liable as 'publishers' of content they host;[140] nor does it require platforms proactively to identify and remove harmful content.[141] Instead, it requires companies 'to ensure that they have effective and proportionate processes and governance in place to reduce the risk of illegal and harmful activity on their platforms, as well as to take appropriate and proportionate action when issues arise'.[142] And specifically as regards disinformation:

> Companies will need to take proportionate and proactive measures to help users understand the nature and reliability of the information they are receiving, to minimise the spread of misleading and harmful disinformation and to increase the accessibility of trustworthy and varied news content.[143]

The government expects the regulator to require companies to take steps to prevent misrepresentation of identity, to spread and strengthen disinformation, to make disputed content less visible, to promote authoritative news sources and diverse news content, to identify automated accounts, and to ensure that algorithms 'do not skew towards extreme and unreliable material in the pursuit of sustained user engagement'.[144] The focus will be on protecting users from harm, not on judging truth.[145]

The UK government announced its plans to legislate in the October 2019 Queen's Speech,[146] and is currently reviewing responses to its consultation on the White Paper. While the government is attempting to create a flexible approach to regulation, various human rights concerns have been raised. These include concerns that there is an undue focus on content regulation; that the scope of 'online harms' is too broad and unspecific, as it includes open-ended categories of 'legal but harmful' content such as disinformation; that the potential liability for failure to remove harmful content could have a chilling effect on freedom of expression; and that very small companies have been included in the potential scope of the measures.[147]

Separately, in response to the Cairncross Review into a sustainable future for high-quality journalism,[148] the UK government is considering a proposal that social media companies be subject to a 'news quality obligation' that would require companies to help audiences understand the

[139] Ibid., Box 12-13, pp. 23–24.
[140] Ibid., para. 6.15.
[141] Ibid., para. 6.16.
[142] Ibid., para. 6.16.
[143] Ibid., para. 7.27.
[144] Ibid., para. 7.28–7.30.
[145] Ibid., para. 7.31.
[146] Cabinet Office and Prime Minister's Office, 10 Downing Street (2019), 'Queen's Speech 2019', https://www.gov.uk/government/speeches/queens-speech-2019 (accessed 1 November 2019).
[147] See for example Index on Censorship (2019), 'The UK Government's Online Harms White Paper: implications for freedom of expression', https://www.indexoncensorship.org/2019/06/the-uk-governments-online-harms-white-paper-implications-for-freedom-of-expression/ (accessed 5 Oct. 2019); and Article 19 (2019), 'Response to the Consultations on the White Paper on Online Harms', https://www.article19.org/wp-content/uploads/2019/07/White-Paper-Online-Harms-A19-response-1-July-19-FINAL.pdf (accessed 5 Oct. 2019).
[148] Cairncross, F. (2019), *The Cairncross Review: A sustainable future for journalism*, London: Department for Digital, Culture, Media & Sport, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf (accessed 5 Oct. 2019).

origins of a news article and trustworthiness of its source.[149] The government is not proposing to ban disinformation; nor is it proposing to impose increased removal obligations on digital platforms.

The UK government is also planning to extend the existing electoral law requirement for an imprint on campaign materials – indicating who created the advert and who paid for it – to include electronic communications.[150] The requirement will apply all year round (not just during election periods); it is not yet clear whether it will apply to what Facebook calls 'issue ads', i.e. advertisements supporting a political issue (e.g. taking a side on Brexit, immigration, etc.) rather than a political party specifically. The government also plans to introduce a new offence in electoral law of intimidating candidates and campaigners during election periods, online and offline, including through hate speech.[151]

4.6 Singapore

Singapore is an example of a state that has responded to online disinformation with a blanket prohibition on falsehoods. This approach is widely criticized as violating the right to freedom of expression.[152] In fact, Singapore legislated against disinformation long before it became an online phenomenon. By section 45 of the 1999 Telecommunications Act: 'Any person who transmits or causes to be transmitted a message which he knows to be false or fabricated shall be guilty of an offence …'

In January 2018 the Singaporean government issued a Green Paper on proposed legislation to combat disinformation online, and the Singaporean Parliament appointed a Select Committee to examine the issues. Among its recommendations, the Select Committee recommended that online falsehoods be disrupted through government intervention to correct and limit exposure to them. On 1 April 2019 the government tabled the Protection from Online Falsehoods and Manipulation Bill, which requires the correction or, in serious cases, the take-down of falsehoods and the disabling of inauthentic online accounts or bots that are spreading falsehoods. The law came into force on 2 October 2019.[153]

4.7 International initiatives

Except at EU level, there are as yet no major international initiatives, particularly with a focus on human rights, to build consensus on the tackling of disinformation and other cyber interference in elections.[154] The UN Human Rights Council has expressed concern about the spread of

[149] Department for Digital, Culture, Media & Sport; Home Office (2019), *Online Harms White Paper*, para. 7.29.
[150] Cabinet Office (2018), *Protecting the Debate: Intimidation, Influence and Information*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730209/CSPL.pdf (accessed 5 Oct. 2019); Cabinet Office (2019), *Protecting the Debate: Intimidation, Influence and Information: Government Response*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/799873/Protecting-the-Debate-Government-Response-2019.05.01.pdf (accessed 5 Oct. 2019).
[151] Ibid.
[152] See for example, Human Rights Watch (2019), 'Singapore: Reject Sweeping "Fake News" Bill', https://www.hrw.org/news/2019/04/03/singapore-reject-sweeping-fake-news-bill (accessed 5 Oct. 2019); Article 19 (2019), 'Singapore: New law on "online falsehoods" a grave threat to freedom of expression', https://www.article19.org/resources/singapore-new-law-on-online-falsehoods-a-grave-threat-to-freedom-of-expression/ (accessed 5 Oct. 2019).
[153] *Protection from Online Falsehoods and Manipulation Act 2019 (Singapore)*.
[154] See UN Secretary-General's High-level Panel on Digital Cooperation (2019),*The Age of Digital Interdependence.*

disinformation and propaganda, noting that it can violate human rights including the rights to privacy and freedom of expression, and can incite violence, hatred, discrimination or hostility.[155] The UN Secretary-General's High-level Panel on Digital Cooperation's 2019 report,[156] which includes discussion of human rights in the digital realm and makes proposals for improvement of 'global digital cooperation architecture', has provided some impetus for developments at UN level. The UN Human Rights Council's Resolution 41/11 of 11 July 2019 mandates the Human Rights Council's Advisory Committee to prepare a report 'on the impacts, opportunities and challenges of new and emerging digital technologies with regard to the promotion and protection of human rights' by June 2021, informed by a panel discussion to be held at the Human Rights Council in June 2020.[157] This initiative, which focuses primarily on the impact of technology on human rights rather than on human rights law as a normative framework in addressing challenges, was led by a cross-regional core group of Austria, Brazil, Denmark, Morocco, Singapore and South Korea.[158]

The Special Rapporteurs on freedom of expression from the UN, the Organization for Security and Co-operation in Europe (OSCE), the Organization of American States (OAS) and the African Commission on Human and Peoples' Rights (ACHPR) adopted a Joint Declaration on 'Freedom of Expression and 'Fake News', Disinformation and Propaganda' in March 2017.[159] The Declaration calls for restrictions on freedom of expression to be imposed only in accordance with Articles 19 and 20 of the ICCPR, and states that digital platforms should not be liable for content posted on them unless they 'specifically intervene' in that content or fail to comply with an oversight body's order to take it down. The Special Rapporteurs are clear that general prohibitions on false news are incompatible with the right to freedom of expression, but that state actors should not promote statements that they know, or reasonably should know, to be false. Where digital platforms restrict third-party content beyond legal requirements, their policies should be clear, predetermined and based on objectively justifiable criteria. The UN Special Rapporteur on promotion and protection of the right to freedom of opinion and expression has repeatedly called for alignment of digital platforms' policies on content moderation with freedom of expression standards.[160]

Participants in the Global Network Initiative (GNI), a global group of around 56 internet and telecommunications companies, human rights and press freedom groups, investors and academic institutions, have all committed to implement the GNI's Principles on Freedom of Expression and Privacy.[161] The Principles include commitments to freedom of expression and privacy, responsible decision-making, collaboration and good governance, accountability and transparency.

[155] UN Human Rights Council Resolution (2018), *The safety of journalists*, UN Doc A/HRC/RES/39/6 (27 September 2018), para. 7.
[156] UN Secretary-General's High-level Panel on Digital Cooperation (2019), *The Age of Digital Interdependence*.
[157] UN Human Rights Council Resolution (2019), *New and emerging digital technologies and human rights*, UN Doc A/HRC/RES/41/11 (11 July 2019).
[158] UN Web TV (2019), 'A/HRC/41/L.14 Vote Item:3 - 39th Meeting, 41st Regular Session Human Rights Council', http://webtv.un.org/meetings-events/human-rights-council/forum-on-business-and-human-/watch/ahrc41l.14-vote-item3-39th-meeting-41st-regular-session-human-rights-council-/6057970116001/?term=&page=57?lanrussian (accessed 5 Oct. 2019).
[159] UN Office of the High Commissioner for Human Rights (2017), 'Freedom of Expression Monitors Issue Joint Declaration on 'Fake News', Disinformation and Propaganda', https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21287&LangID=E (accessed 14 Oct. 2019).
[160] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), *Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms,* UN Doc A/73/348 (29 August 2018); Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet*.
[161] Global Network Initiative (2017), *GNI Principles on Freedom of Expression and Privacy*, https://globalnetworkinitiative.org/wp-content/uploads/2018/04/GNI-Principles-on-Freedom-of-Expression-and-Privacy.pdf (accessed 14 Oct. 2019)..

Participating companies are expected to put the Principles into practice, and their progress to do so in good faith is evaluated every two years by independent assessors.[162]

The Freedom Online Coalition[163] is a group of 30 governments that is working to 'advance Internet freedom' and to protect fundamental human rights of free expression, association, assembly and privacy online. While in the past it has focused on preventing and limiting state-sponsored restrictions to human rights online and supporting civil society voices, its 2019–20 programme of action will also begin to consider both artificial intelligence and disinformation.[164]

4.8 Initiatives by digital platforms

All the large digital platforms have standards by which they monitor content with a view to removal, suppression and/or deprioritization of certain material. Community standards differ from human rights standards, and also differ between platforms. For example:

- Facebook's Community Standards[165] delimit the scope of content allowed and prohibited on the platform, and are designed to be comprehensive. They include standards on violence and criminal behaviour, on safety, on objectionable content, and on 'integrity and authenticity'. Human rights are not referenced as a basis for these standards, albeit that there is some degree of overlap between Facebook's standards and human rights law. Content is monitored both by algorithm (e.g. to detect terrorist content) and as a result of reports of posts. After widespread consultation, Facebook has recently announced the creation and structure of an independent Oversight Board for content decisions.[166]

- The Twitter Rules[167] set out the parameters of acceptable content on Twitter, including by reference to safety, privacy and authenticity. Twitter cites human rights laws as the basis for its commitment to freedom of expression and privacy,[168] championed by its Trust and Safety team. Twitter asserts its commitment to being 'fair, informative, responsive and accountable' in enforcing its rules.[169] Separately, in October 2019 Twitter announced a ban on political advertising.[170]

- Verizon Media (parent of Yahoo! and AOL) has Community Guidelines,[171] Reddit has a Content Policy,[172] and YouTube has Community Guidelines,[173] all of which are enforced by automated

[162] Participating ICT companies include Microsoft, Google, Facebook and Verizon Media (corporate group of Yahoo!); but not Twitter or Reddit.
[163] Freedom Online Coalition (2019), 'The Freedom Online Coalition' https://freedomonlinecoalition.com/ (accessed 5 Oct. 2019).
[164] Freedom Online Coalition (2019), 'Program of Action 2019-2020', https://freedomonlinecoalition.com/wp-content/uploads/2019/03/FOC-Program-of-Action-2019-2020.pdf (accessed 27 Oct. 2019).
[165] Facebook (2019), 'Community Standards', https://en-gb.facebook.com/communitystandards/introduction (accessed 5 Oct. 2019).
[166] Harris, B. (2019), 'Establishing structure and governance for an Independent Oversight Board', https://newsroom.fb.com/news/2019/09/oversight-board-structure/ (accessed 5 Oct. 2019).
[167] Twitter (2019), 'The Twitter Rules', https://help.twitter.com/en/rules-and-policies/twitter-rules (accessed 5 Oct. 2019).
[168] Twitter (2019), 'Defending and Respecting the Rights of People Using our Service', https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice (accessed 12 July 2019).
[169] Twitter (2019), 'Our Approach to Policy Development and Enforcement Philosophy', https://help.twitter.com/en/rules-and-policies/enforcement-philosophy (accessed 12 July 2019).
[170] Dorsey, J. (@jack) (2019), 'We've made the decision to stop all political advertising on Twitter globally. We believe political message reach should be earned, not bought. Why? A few reasons…', tweets, 30 Oct. 2019, https://twitter.com/jack/status/1189634360472829952 (accessed 31 Oct. 2019).
[171] Verizon Media (2019), 'Verizon Media Community Guidelines', https://www.verizonmedia.com/policies/us/en/verizonmedia/guidelines/index.html (accessed 5 Oct. 2019).
[172] Reddit (2019), 'Reddit Content Policy', https://www.redditinc.com/policies/content-policy (accessed 5 Oct. 2019).

and human decision-makers. Google has content policies for various aspects of its operation, such as for publishers in Google News.[174]

These standards originally entailed minimal restriction of content, in line with the US culture of free speech,[175] but have evolved over time as platforms have appreciated the need for effective standards in order not to deter users and advertisers.[176] The full extent of the standards being applied is not clear; nor is their implementation, as there is very little transparency in content assessment or removal.[177] Platforms are slowly beginning to reveal their policies and practices, but at present the overall picture remains opaque. It is not possible to see, for example, how much content is removed, and on what grounds.

While, as discussed, digital platforms have a responsibility to respect international human rights law, domestic law in most states allows them, as private entities, to adopt standards that restrict more expression than does human rights law. One unresolved issue is the extent to which it ought to be permissible for digital platforms to adopt their own content standards – for instance because they wish to market themselves by reference to standards of behaviour of users (e.g. 'child-friendly', 'mutually respectful debate'), or by reference to content they support (e.g. 'Christian'). By analogy with freedom of association and private clubs, it can be argued that it should be permissible for smaller platforms to adopt their own rules within domestic law. But as regards the largest platforms such as Facebook, Google and Twitter, the operation and accessibility of which have a major impact on public conversation, the responsibility to respect human rights means that the standards by which the platforms assess content should be no more restrictive than human rights law entails.

[173] YouTube (2019), 'Policies and Safety', https://www.youtube.com/intl/en-GB/about/policies/#community-guidelines (accessed 5 Oct. 2019).
[174] Google (2019), 'Content Policies', https://support.google.com/news/publisher-center/answer/6204050?hl=en-GB (accessed 5 Oct. 2019).
[175] Klonick (2018), 'The New Governors: The People, Rules and Processes Governing Online Speech'.
[176] Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet,* pp. 48–51.
[177] Ibid., p. 51.

5. Relevant Human Rights Law

5.1 Application of human rights law

International human rights law obliges states to secure enumerated rights to individuals within their territories, and to some extent to individuals outside their territories.[178] The UN Human Rights Council has adopted resolutions affirming that 'the same rights that people have offline must also be protected online'.[179] Human rights law does not impose obligations directly on private companies. In 2011, the UN adopted by consensus the Guiding Principles on Business and Human Rights,[180] also known as the Ruggie Principles, which elucidate how states have a legally binding 'duty to protect' individuals against human rights abuse by business enterprises within their territory, as well as how businesses have a non-binding 'responsibility to respect' human rights (which means they should avoid interference with human rights law and provide remedies where necessary).

Consequently, digital platforms have no international legal obligation to comply with international human rights law, but have a responsibility to respect it. International human rights law provides the only available international legal framework to guide the activities of companies whose impacts on people's lives may be as significant as those of a national government, and whose reach (in terms of numbers affected) may be far greater than a national government.

The extent to which governments bear human rights obligations to individuals outside their territories in international law is disputed. However, it is clear from the Ruggie Principles that the corporate responsibility to respect human rights 'is a global standard of expected conduct for all business enterprises wherever they operate',[181] regardless of where they are headquartered or of national laws and regulations that may vary between states.

This paper therefore proceeds on the basis that digital platforms have a responsibility to respect the human rights in the ICCPR wherever they operate, in parallel with states' 'duty to protect' obligations to ensure that digital platforms operating in their countries respect the human rights of individuals present there, and to some extent overseas.[182]

The application of international human rights law does not mean that there are simple solutions to all the challenges presented in this paper. Establishing how existing norms apply in new contexts is likely to be contested, and reaching settled views takes time whether it is done through expert opinion, through the drafting of normative guidance, through state negotiation or litigation. Within the parameters of the international norms, precise standards may vary from country to country, and case by case assessment will still be necessary. Nonetheless, international human rights law can

[178] The parameters of extraterritorial obligations are disputed.
[179] UN Human Rights Council Resolutions (2012-2018), *The promotion, protection and enjoyment of human rights on the Internet*, UN Doc A/HRC/RES/38/7 (5 July 2018), A/HRC/RES/32/13 (1 July 2016), A/HRC/RES/26/13 (26 June 2014), A/HRC/RES/20/8 (5 July 2012).
[180] UN OHCHR (2011), *Guiding Principles on Business and Human Rights*, New York and Geneva: United Nations Office of the United Nations High Commissioner for Human Rights, www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf (accessed 5 Oct. 2019).
[181] Ibid., p. 14 (Commentary to Foundational Principle 11).
[182] Ibid.

provide a valuable, overarching normative framework for making carefully calibrated decisions in this space, and the assistance of decades of jurisprudence and analysis on how to balance rights that are as valid in the online as in the offline context.

While the activities of digital platforms may engage a broad range of human rights,[183] this chapter will focus on five key rights, considering two of them together:

- The right to freedom of thought, and the right to hold opinions without interference;
- The right to privacy;
- The right to freedom of expression;
- The right to vote in elections.

Each section sets out the text of the relevant right in the 1948 UDHR, which is not of itself legally binding but largely reflects customary international law and to that extent is legally binding on all states; and in the 1966 ICCPR, which is legally binding on its 172 states parties and in parts also reflects customary international law and therefore is binding on all states. As relevant, each section discusses the views of the Human Rights Committee, the body of independent experts established to monitor the implementation of the ICCPR. The Human Rights Committee's views are not legally binding but are to a large degree authoritative.

Regional human rights instruments, such as the European Convention on Human Rights[184] and the American Convention on Human Rights,[185] include rights framed in similar terms to those in UDHR and ICCPR. The jurisprudence of the regional courts charged with interpreting those instruments, which can be legally binding on the states parties to cases,[186] in some cases gives useful specific guidance on the meaning and content of the rights within regional contexts and is referred to as appropriate in this chapter. These courts, too, are in the early stages of applying rights to online activities and the regulation of digital platforms.

The concluding chapter of this paper offers recommendations on the application of human rights law in general, and each right discussed in this chapter.

[183] For example, Ohlin discusses election interference as a breach of the right of self-determination, viewed as the 'right of all peoples to determine for themselves their political destiny'. The lack of clarity internationally in the parameters of this right, and its group nature, do not lend it to being the source of specific recommendations and commitments for digital platforms. Ohlin, J. (2018), *Election Interference: The Real Harm and the Only Solution*, Cornell Law School Legal Studies Research Paper Series, Ithaca, NY: Cornell University, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3276940 (accessed 5 Oct. 2019).
[184] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (European Convention on Human Rights).
[185] American Convention on Human Rights, opened for signature 22 November 1969, 1144 UNTS 123 (entered into force July 18, 1978) (Inter-American Convention on Human Rights).
[186] For example, European Convention on Human Rights, Article 46(1): 'The High Contracting Parties undertake to abide by the final judgment of the Court in any case to which they are parties.'

5.2 Rights to freedom of thought and to hold opinions without interference

UDHR Article 18

'Everyone has the right to freedom of thought, conscience and religion …'

UDHR Article 19

'Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.'

ICCPR Article 18

'1. Everyone shall have the right to freedom of thought, conscience and religion …'

5.2.1 The content of the rights

> [H]uman liberty … comprises, first, the inward domain of consciousness; demanding liberty of … thought and feeling; absolute freedom of opinion and sentiment on all subjects, practical or speculative, scientific, moral or theological.[187]

While the core human rights treaties clearly reflect this absolute freedom of the *forum internum* of the mind, it is a relatively unexplored area on which there is little jurisprudence. The boundaries between freedom of thought, associated with freedom of conscience and religion, and the freedom to hold opinions, associated with freedom of expression, are not yet clear.[188] The American Convention on Human Rights changes the schematization by providing for a 'right to freedom of thought and expression'.[189]

In its 2011 General Comment on Article 19 ICCPR,[190] in which it discusses both freedom of opinion and freedom of expression, the UN Human Rights Committee states that 'Freedom of opinion and freedom of expression are indispensable conditions for the full development of the person. They are essential for any society. They constitute the foundation stone for every free and democratic society'[191] and 'The freedoms of opinion and expression form a basis for the full enjoyment of a wide range of other human rights'.[192] There can be no derogation from freedom of opinion,[193] which is 'a right to which the Covenant permits no exception or restriction'.[194] The Committee observes that the obligation to respect freedom of opinion and expression 'also requires States parties to ensure that persons are protected from any acts by private persons or entities that would impair the

[187] Mill, J.S. (1859), *On Liberty,* London: Longman, Roberts & Green, Chapter 1, para. 12.
[188] This paper uses 'freedom of thought' as a convenient shorthand for the right to freedom of thought and the right to hold opinions without interference.
[189] *American Convention on Human Rights,* Article 13.1.
[190] UN Human Rights Committee, *General Comment No. 34: Article 19 (Freedoms of Opinion and Expression)*, Human Rights Committee 102nd session, UN Doc CCPR/C/GC/34 (12 September 2011).
[191] Ibid., para. 2.
[192] Ibid., para. 4.
[193] Ibid., para. 5.
[194] Ibid., para. 9.

enjoyment of the freedoms of opinion and expression to the extent that these Covenant rights are amenable to application between private persons or entities';[195] and that 'Any form of effort to coerce the holding or not holding of any opinion is prohibited,' including 'inducements of preferential treatment' in prison.[196] The 'right to form an opinion and to develop this by way of reasoning' is an essential element of the right to freedom of opinion.[197]

In its 1993 General Comment on Article 18 ICCPR,[198] the Committee states: 'The right to freedom of thought, conscience and religion … in article 18.1 is far-reaching and profound; it encompasses freedom of thoughts on all matters … this provision cannot be derogated from, even in time of public emergency …'[199]

While the Committee's views are clear, there have been few individual complaints brought to it alleging violation of the rights to freedom of thought and freedom of opinion.[200] Similarly, these rights have not yet been well explored in regional systems of human rights protection. For example, while Article 9 of the European Convention on Human Rights provides that 'Everyone has the right to freedom of thought, conscience and religion …', the European Court of Human Rights' Guidance on that article,[201] which reflects the Court's case law, focuses almost entirely on freedom of conscience and religion as there is very little jurisprudence on the right to freedom of thought. It has been little contemplated in the context of online activity; even the Council of Europe's 'Guide to Human Rights for Internet Users' does not discuss the right to freedom of thought.[202]

In reality, despite the avowedly absolute nature of the right, we are constantly the recipients of influences on our thoughts, and frequently subject to deliberate attempts to influence our thoughts and opinions, for example from the media and from advertising. Indeed, the freedom to be subject to a wide range of influences is itself a dimension of our autonomy. We generally consider ourselves to have control over our thoughts despite these influences. Moreover, there has been relatively little challenge to deradicalization programmes on grounds of compatibility with freedom of thought. Nowak has suggested that as it may be difficult to distinguish between permissible and impermissible influences on thought, infringements may be limited to occasions on which one's opinion is involuntarily influenced.[203] In a slightly different context – that of education and

[195] Ibid., para. 7.
[196] Ibid., para. 10; UN Human Rights Committee Communication No. 878/1999, *Yong-Joo Kang v Republic of Korea*, UN Doc CCPR/C/78/D/878/1999 (15 July 2003) (Kang v Republic of Korea).
[197] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), *Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms*, para. 23.
[198] Human Rights Committee, *'General Comment No. 22'*, Human Rights Committee 48th session, UN Doc CCPR/C/21/Rev.1/Add.4 (30 July 1993).
[199] Ibid., para. 1.
[200] *Yong Joo Kang v. Republic of Korea,* is the only individual complaint in which the Committee has found a violation. Mr Kang challenged the 'ideology conversion system', a system by which political prisoners were offered incentives to change their political beliefs. The Committee opined that this system 'restricts freedom of expression and of manifestation of belief on the discriminatory basis of political opinion and thereby violates articles 18, para. 1, and 19, para. 1, both in conjunction with article 26'. But this formulation suggests that the Committee's focus may have been on *manifestation* of opinion and belief, rather than on the 'forum internum'. In two other cases (Human Rights Committee, Communication No. 1119/2002, *Jeong-Eun Lee v Republic of Korea,* 84th session, UN Doc CCPR/C/84/D/1119/2002 (20 July 2005) and Human Rights Committee, Communication No. 628/1995 *Tae Hoon Park v. Republic of Korea,* 64th session, UN Doc CCPR/C/64/D/628/1995 (3 November 1998)), the Committee has avoided addressing claims that conviction for membership in an illegal political party violates the rights to freedom of thought and freedom of opinion.
[201] European Court of Human Rights, *Guide on Article 9 of the European Convention on Human Rights: Freedom of thought, conscience and religion*, Strasbourg: European Court of Human Rights, https://www.echr.coe.int/Documents/Guide_Art_9_ENG.pdf (accessed 5 Oct. 2019). There is not yet an equivalent Guide to Article 10 (freedom of expression, including 'freedom to hold opinions').
[202] Council of Europe (2014), *Guide to Human Rights for Internet Users*, Council of Europe, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680d5b31 (accessed 5 Oct. 2019).
[203] Nowak, M. (2005), *UN Covenant on Civil and Political Rights: CCPR Commentary*, 2nd edition, Germany: NP Engel, p. 442.

freedom of religion – the European Court of Human Rights has emphasized the importance of not indoctrinating children against the will of parents: 'The State is forbidden to pursue an aim of indoctrination that might be considered as not respecting parents' religious and philosophical convictions. That is the limit that the States must not exceed'.[204]

As a minimum, freedom of thought entails a right not to have one's opinion unknowingly manipulated or involuntarily influenced, fundamentally linked with the concept of human agency.[205] Clearly the parameters of this right are difficult to establish, and proving a breach is not straightforward. It also entails a right not to reveal one's thoughts or opinions, and not to be penalized for one's thoughts.[206] There is now a pressing need to explore whether cyber interference in elections and other online political discourse may be breaching this right.

5.2.2 Freedom of thought and opinion: potential breaches

5.2.2.1 Structure and activities of digital platforms

The last few years have seen the emergence of an online 'industry of influence' on an unprecedented scale, in two senses. Firstly, most digital platforms and services are driven by profit, and for many of them, profit is driven by the ability to influence their users through advertising. As James Williams discusses in his book *Stand out of our Light*,[207] advertising, or 'the industrialisation of persuasion', has become the 'default business model' for digital platforms and services. Digital platform designers exploit the 'catalogue of decision-making biases' compiled in the advertising industry, and magnify it through the use of digital techniques, in order to encourage users to spend maximal time on their sites and to micro-target advertising to individual users on whom it is most likely to be effective. Their use of personal data in this endeavour is discussed in the context of the right to privacy, below.

The quest for attention favours (through algorithms) content that goes viral.[208] Virality is biased towards bad news; news that inspires emotion; and emotions that produce an 'arousal' response such as anger.[209] At present, the quest for virality and maximum attention means that digital platforms tend to use algorithms that prioritise bad news and news that provokes emotional responses. In the context of political discourse, this structure favours content that shocks and scandalizes over reasoned democratic debate. Williams quotes the Egyptian activist Wael Ghonim: 'We who use the Internet now 'like' or we flame – but there's [very little] now happening [algorithmically] to drive people into the more consensus-based, productive discussions we need to have, to help us make civic progress.'[210]

[204] *Lautsi v Italy* (2011) 30814/06, para. 62.
[205] UN Secretary-General's High-level Panel on Digital Cooperation (2019), *The Age of Digital Interdependence*, p. 17.
[206] Alegre, S. (2017), 'Rethinking freedom of thought for the 21st century', *European Human Rights Law Review* 3: pp. 221–233 (accessed 14 Oct. 2019).
[207] Williams, J. (2018), *Stand out of our Light,* Cambridge: Cambridge University Press.
[208] Ibid., pp. 33–35.
[209] See also Wardle and Derakhshan (2017), 'Information Disorder: Toward an interdisciplinary framework for research and policymaking', Council of Europe, DGI (2017)09
[210] Williams, J. (2018), *Stand out of our Light*, pp. 79–80.

The prioritization of news, through algorithms, can have a significant effect on voting in elections. Even in 2015, one study found that internet search rankings can change the voting preferences of undecided voters by 20 per cent or more.[211] The lack of transparency in algorithms means that the individual may be under the impression that they are receiving the most relevant or objective information, when in fact the prioritization of information may be based on entirely different, unseen factors.[212] This may have a significant yet unseen impact on individuals' capacity to form and develop their opinions.[213]

In the offline world, illegitimate manipulation has been regulated. For example, the UK's Broadcasting Act 1990 banned subliminal advertising.[214] As the development of digital platforms has run ahead of regulation, they have been subject to few of the carefully formulated checks and balances developed over many years to avoid undue influence in the industries of advertising, broadcasting and political campaigning. The perception that digital platforms are 'neutral' hosts of content has meant that there has been little thought given to the application of those conventional checks and balances. The result is that social media platforms have had unbridled opportunity to influence thought around their goals, in particular their fundamental, profit-driven goal of retaining attention.

Looking ahead to future technology raises further issues, as thought and digital platforms are likely to become more closely linked. In 2015 Facebook filed a patent for detecting emotions from computer and smartphone cameras.[215] And in 2017 Facebook discussed development of a brain-computer interface, stating that it would not 'invade your thoughts' but, rather, decode 'words you've already decided to share by sending them to the speech center of your brain'.[216] Similarly, the technology entrepreneur Elon Musk's Neuralink[217] is developing 'brain-machine interfaces' via which users may interact with computers without the need for keyboards, mice or trackpads. This raises fresh concerns about the impact of technology on the *forum internum* – the unregulated space inside our heads; or as Samuel Warren and Louis Brandeis put it, the 'right to be let alone'.[218]

There is a fundamental tension between the legal treatment, and consumers' perception, of digital platforms as neutral hosts of others' content, and the actual profile and aims of many digital platforms. The overwhelming impetus of social media companies as advertising companies means that they curate both their treatment of posted content and consumers' personal data for their own profit, while having little responsibility to do so in a way that upholds human rights or democratic

[211] Epstein, R. and Robertson, R. (2015), 'The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections', Proceedings of the National Academy of Sciences of the United States of America, 112(33), National Academy of Sciences, doi https://doi.org/10.1073/pnas.1419828112 (accessed 5 Oct. 2019).
[212] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), *Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms*, para. 25.
[213] The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression has noted that AI-assisted content curation online, for example through the algorithmic ranking yet apparent objectivity of search facilities of large platforms, raises concerns about individuals' ability to form opinions that call for further research, as well as more transparency in algorithms on the part of the platforms. Ibid., paras 25–26.
[214] Broadcasting Act 1990, section 6(1)(e), succeeded by Communications Act 2003, section 319(2)(l).
[215] United States Patent Application 20150242679, 27 Aug. 2015, pdfaiw.uspto.gov/.aiw?PageNum= ... 9&IDKey=47BC4614A23D (accessed 30 Oct. 2019).
[216] Williams (2018), *Stand out of our Light*, p. 93. See also, Harvard Law School via YouTube (2019), 'Zittrain and Zuckerberg discuss encryption, 'information fiduciaries' and targeted advertisements', video, 20 February 2019, https://www.youtube.com/watch?v=WGchhsKhG-A (accessed 5 Oct. 2019).
[217] Neuralink (2019), 'Neuralink', http://www.neuralink.com (accessed 5 Oct. 2019).
[218] Warren, S. and Brandeis, L. (1890). 'The Right to Privacy' *Harvard Law Review*, 4(5): pp 193-220, http://links.jstor.org/sici?sici=0017-811X%2818901215%294%3A5%3C193%3ATRTP%3E2.0.CO%3B2-C (accessed 10 Oct. 2019).

principles. The steps now being taken by some digital platforms, for example to address harmful content, are minor obstacles to the rushing waters of the industry of influence, as they have no impact on the underlying models and structures of the platforms.

5.2.2.2 Disinformation campaigns

The second dimension of this industry of influence consists of abuse of digital platforms by other actors in order deliberately and cynically to manipulate audience opinions and emotions for their own political, financial or other purposes. Of these, the most egregious that have been comprehensively recorded are those of the Internet Research Agency in their attempts to influence the US and other elections and subsequent political discussions.[219] New Knowledge's 'Disinformation Report' found that Russian interference in the 2016 US presidential election included 'a sweeping and sustained social influence operation consisting of various coordinated disinformation tactics aimed directly at US citizens, designed to exert political influence and exacerbate social divisions in US culture'.[220] The report found that:

> Throughout its multi-year effort, the Internet Research Agency exploited … social unrest and human cognitive biases. The divisive propaganda Russia used to influence American thought and steer conversations for over three years wasn't always objectively false. The content designed to reinforce in-group dynamics would likely have offended outsiders who saw it, but the vast majority wasn't hate speech. Much of it wasn't even particularly objectionable. But it was absolutely intended to reinforce tribalism, to polarize and divide, and to normalize points of view strategically advantageous to the Russian government on everything from social issues to political candidates. It was designed to exploit societal fractures, blur the lines between reality and friction, erode our trust in media entities and the information environment, in government, in each other, and in democracy itself. This campaign pursued all of those objectives with innovative skill, scope, and precision.[221]

Moreover, the report concluded, the Internet Research Agency recruited unsuspecting Americans to further its own ends, and would likely continue to use them for 'human-exploitation tradecraft and narrative laundering'.[222] The authors comment that society is still treating this as a problem of false stories, requiring counter-messaging and counternarratives; rather than as an information war requiring more strategic responses. Similarly, the EU's responses of fact-checking and correction seem woefully inadequate in the face of such orchestrated campaigns.

In this context, it is worth noting that the Chief of General Staff of the British Army has discussed disinformation as a weapon of war,[223] and disinformation is under discussion by NATO as a tactic of hybrid warfare.[224] The academic Emma Briant has argued that some of the tactics of digital platforms grew from Western governments' defence and intelligence capabilities and developments.

[219] Diresta, R. et al (2018), *The Tactics and Tropes of the Internet Research Agency*; Howard, P. et al (2018), *The IRA, Social Media and Political Polarisation in the United States, 2012-2018*, Oxford: Oxford Internet Institute, https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf (accessed 5 Oct. 2019).
[220] Diresta et al (2018), *The Tactics and Tropes of the Internet Research Agency*, p. 4.
[221] Ibid., p. 99.
[222] Ibid., p. 100.
[223] General Sir Mark Carleton-Smith, 'Introductory Remarks' (Speech at RUSI's Land Warfare Conference 2019, 4 June 2019), https://rusi.org/annual-conference/rusi-land-warfare-conference/chief-general-staff-introductory-remarks-2019 (accessed 5 Oct. 2019).
[224] NATO (2018), 'Allied Intelligence Chiefs discuss countering cyber-attacks, disinformation', 29 November 2018, https://www.nato.int/cps/en/natohq/news_161119.htm?selectedLocale=en (accessed 5 Oct. 2019); NATO (2019), 'NATO and EU discuss defence against hybrid warfare'14 March 2019https://www.nato.int/cps/en/natohq/news_164603.htm (accessed 5 Oct. 2019).

She observes that tactics such as 'disinformation and deception techniques; methods used to demoralize an enemy; methods of harnessing psychological weaknesses or violent tendencies within a population or group; methods for influencing extremists, or increasing or decreasing inter- and intra-group tensions; techniques and specialist knowledge about surveillance and hacking' may all arise from training and/or knowledge acquired in a military or intelligence context.[225]

Here there is a critical role for freedom of thought, distinct from debates on freedom of expression. The challenge regarding the Internet Research Agency's tactics lies not so much in the content of its messaging but in the *techniques* used to influence. It is specifically the techniques, rather than the messages, that need to be combated in order to preserve freedom of thought.

5.3 Right to privacy

Article 12 UDHR

'No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.'

Article 17 ICCPR

'1. No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation.

2. Everyone has the right to the protection of the law against such interference or attacks.'

5.3.1 The content of the right

At its core, the right to privacy must entail a right to choose not to divulge your personal information, and a right to opt out of trading in and profiling on the basis of your personal data.[226] The impact of the right to privacy on the harvesting and use of personal data online, generally or in the election context, is yet to be fully clarified. The UN Human Rights Committee's General Comment on Article 17 ICCPR dates from 1988. But this situation is changing. In 2015 the Human Rights Council established the mandate of a new Special Rapporteur on the right to privacy. The Council has also commissioned research and dialogue led by the Office of the UN High Commissioner for Human Rights (OHCHR) on 'the right to privacy in the digital age'. In its report of August 2018,[227] which followed an expert workshop and the receipt of 63 written submissions, the OHCHR found: 'There is a growing global consensus on minimum standards that should govern the processing of personal data by States, business enterprises and other private actors', referring to

[225] Briant, E. (2018), 'Building a stronger and more secure democracy in a digital age', *Written Evidence FKN0071 on Fake News submitted to the Digital, Culture, Media and Sport* Committee,http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/fake-news/written/88559.pdf (accessed 5 Oct. 2019).
[226] While this paper does not analyse the meaning of 'privacy', it is worth noting Westin's definition: 'Privacy is the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others,' Cited in Roessler (2005), *The Value of Privacy*, Cambridge and Malden: Polity Press, p. 7.
[227] UN OHCHR (2018), *The right to privacy in the digital age*, UN Doc A/HRC/39/29 (3 August 2018).

various regional conventions including the Council of Europe's modernized Convention 108.[228] These minimum standards include that 'processing of personal data should be fair, lawful and transparent'; that individuals should be informed of the processing of their data; that processing should be 'based on the free, specific, informed and unambiguous consent of the individuals concerned, or another legitimate basis laid down in law'; that 'personal data processing should be necessary and proportionate to a legitimate purpose'; that changes of purpose without the data subject's consent should be avoided; that data should be held securely; that entities processing personal data should be accountable; and that sensitive data should enjoy a higher level of protection.[229] The OHCHR also stressed the corporate responsibility to respect these rights.[230]

5.3.2 Right to privacy: potential breaches

At present there is a widespread failure to meet these minimum standards in the online context: personal data is collected, traded and used widely in algorithmic processes and political campaigning on an unprecedented scale with very little choice, awareness or ability for individuals to see what is happening. Personal data is generating enormous revenues for digital platforms without individuals understanding or having consented to this, and without regard to human rights law.[231]

In the election context, personal data is harnessed by political campaigners and advertisers with little restraint. In the UK, for instance, the Information Commissioner's Office (ICO) has found that all the major UK political parties are using a wide range of personal data to create a personal profile on each voter; that they are failing to explain what data they are gathering and how they would use it;[232] and that they are failing to apply sufficient due diligence when obtaining information from data brokers.[233] The ICO found that personal data is used to micro-target[234] posts and advertisements so as to maximise their impact upon their audience. Micro-targeting in the political context can lead to people's viewing of selective content without their realizing that they are not seeing the full picture, and therefore to the polarization of views.[235] While some digital platforms, such as Facebook, are currently working to increase transparency in advertising, there is no move away from micro-targeting; this is despite widespread suspicion that micro-targeting is often discriminatory, whether on the basis of protected characteristics (e.g. age) or as a result of proxy characteristics (e.g. liking a particular band, or buying particular products). Moreover, regulatory guidance on cookies is not yet being fully implemented.[236] Many individuals believe they have some control over their data, for example because they can decide who can see their Facebook profile; but

[228] Ibid., paras 28–33.
[229] Ibid., para. 29.
[230] Ibid., paras 42–49.
[231] For example Information Commissioner's Office (2019), *Update report into adtech and real time bidding*, https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf (accessed 25 Oct. 2019).. While this report concerns commercial advertising, there would be no bar on political organisations or consultancies registering to participate in real-time bidding and so potentially accessing personal data profiles as described in the report.
[232] UK Information Commissioner's Office (2018), *Democracy Disrupted*, pp. 28–29.
[233] Ibid., p. 31.
[234] 'The term 'micro-targeting' describes targeting techniques that use data analytics to identify the specific interests of individuals, create more relevant or personalised messaging targeting those individuals, predict the impact of that messaging, and then deliver that messaging directly to them.' Ibid., p. 27.
[235] This is not to comment on micro-targeting in commercial or other contexts.
[236] Information Commissioner's Office, 'Guide to Privacy and Electronic Communications Regulations: Cookies and Similar Technologies', https://ico.org.uk/for-organisations/guide-to-pecr/cookies-and-similar-technologies/ (accessed 22 Jul. 2019). Equivalent guidance has been issued in some other jurisdictions, e.g. Ireland and France.

in reality they cannot opt out of much sharing on the part of platforms; nor do most individuals have any idea of the impact of clicking the 'consent to cookies' button online.

More generally, personal data, including metadata gleaned from large sets of personal data, is routinely used in algorithmic processes. Here there is an evident link between the right to freedom of thought and the right to privacy, as data gathered about individuals is being harnessed to impact their agency and autonomy in decision-making, including in the political context. Arguably, 'we value privacy for the sake of our autonomy'.[237] The recently adopted Council of Europe Declaration on the Manipulative Capabilities of Algorithmic Processes[238] declares that 'Public awareness … remains limited regarding the extent to which everyday devices collect and generate vast amounts of data. These data are used to train machine-learning technologies to prioritise search results, to predict and shape personal preferences, to alter information flows, and, sometimes, to subject individuals to behavioural experimentation.'[239] The Declaration observes that information inferred about individuals from readily available data can support segregation and discrimination, as well as facilitating micro-targeting.[240] Among its recommendations, the Declaration calls on states to consider 'the need for additional protective frameworks related to data that … address the significant impacts of the targeted use of data on societies' and to enhance 'public awareness of how many data are generated and processed by personal devices, networks, and platforms through algorithmic processes that are trained for data exploitation'.[241]

Similarly, the UN Human Rights Council Special Rapporteur on the right to freedom of expression has observed:

> The use by AI of such [personal data] datasets raises serious concerns, including regarding their origins, accuracy and individuals' rights over them; the ability of AI systems to de-anonymize anonymized data; and biases that may be ingrained within the datasets or instilled through human training or labelling of the data.[242]

All these technological developments have occurred with minimal regard to the right to privacy, and manifest widespread failure to respect it: first on the part of states in failing adequately to regulate; second on the part of the platforms themselves in their collection and use of personal data; and third on the part of data brokers, who act as large-scale traders in personal information. There is minimal transparency of companies' data harvesting, retention, or trading activities and policies.

Data protection is founded on human rights law, but vigilance is needed to ensure that it meets its standards rather than merely playing lip service to them. The EU GDPR, which 'protects fundamental rights and freedoms',[243] is the most advanced data protection regulation. Of the six

[237] Roessler (2005), *The Value of Privacy*, pp. 1, 71–76. The central argument of Roessler's important philosophical exposition of privacy is 'that the true realisation of freedom, that is a life led autonomously, is only possible in conditions where privacy is protected' (p. 72).
[238] Council of Europe (2019), 'Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies)', https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b (accessed 5 Oct. 2019).
[239] Ibid., para. 4.
[240] Ibid., para. 6.
[241] Ibid., para. 9.
[242] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), *Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms*, para. 7.
[243] GDPR Article 1(2).

bases for lawful processing of data, the two principal ones – consent, and the 'legitimate interests' of the data controller or a third party – both currently lend themselves to abuse.[244] The EU's ePrivacy Directive's requirement for 'consent to cookies' often manifests as notional 'consent' as a condition of using a website, rather than genuine choice on the part of consumers;[245] ongoing negotiations for replacement of the Directive with an ePrivacy Regulation offer an opportunity to review requirements for consent, but are currently stalled. As regards 'legitimate interests', currently many companies assume that all commercial purposes are legitimate, without adequate consideration; and the 'fundamental rights and freedoms' of the individual, which may override these,[246] are assumed not to be engaged. Nor is there adequate consideration of whether subsequent trading in personal data (even if anonymized) is consistent with requirements to avoid changes of purpose and to hold data securely. Full implementation and enforcement of the GDPR by powerful regulators is now needed for compliance with the right to privacy and identification of any human rights gaps, particularly as technology evolves.

5.4 Right to freedom of expression

Article 19 UDHR

'Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.'

Article 19 ICCPR

'2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.

3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:

(a) For respect of the rights or reputations of others;

(b) For the protection of national security or of public order (ordre public), or of public health or morals.'

Article 20 ICCPR

'1. Any propaganda for war shall be prohibited by law.

2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.'

[244] Information Commissioner's Office (2019), 'Update report into adtech and real time bidding'.
[245] If users do not consent to cookies they frequently do not have full access to the website.
[246] GDPR Article 6(1)(f): Processing of personal data shall be lawful if 'necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child'.

5.4.1 The content of the right

Both Article 19 UDHR and Article 19(2) ICCPR guarantee a broad right to freedom of expression; as per the UN Human Rights Committee, 'This right includes the expression and receipt of communications of every form of idea and opinion capable of transmission to others, subject to the provisions in article 19, paragraph 3, and article 20'.[247] This includes a 'free, uncensored and unhindered press or other media',[248] including new media resulting from developments in information and communication technologies.[249] The Committee has stressed that freedom of expression includes 'the free communication of information and ideas about public and political issues between citizens, candidates and elected representatives'.[250] The right to freedom of expression is a vital bulwark in combating attempts by governments to suppress dissent and to control the spread of information. Consequently, it entails a right to disseminate all information, subject to exceptions; it is not limited to true information.

The permissible restrictions on freedom of expression in Article 19(3) and Article 20 ICCPR are narrowly construed. As regards Article 19(3), the Human Rights Committee's view is that a restriction is permitted only if (i) it is 'provided by law', i.e. a law of sufficient precision, (ii) it is for one of the purposes set out in Article 19(3)(a) or (b), and (iii) it conforms to 'strict tests' of necessity and proportionality.[251] Consequently, in the view of the Human Rights Committee, atrocity denial laws and blasphemy bans untethered to advocacy of incitement of imminent harm are incompatible with the right to freedom of expression.[252] However, state practice has varied in relation to Article 19(3) ICCPR. For example, some EU member states have imposed bans on certain speech (such as blasphemy) that have been upheld by the European Court of Human Rights yet would be unconstitutional if imposed in the US.[253] Similarly, the European Court of Human Rights has upheld bans on hateful speech and holocaust denial, whereas under US law they are impermissible absent advocacy of imminent violence or a true threat of harm.[254] Thus, incitement to hatred and discrimination (without advocacy of violence) is illegal in many EU states, but constitutional protection of freedom of speech prevents its prohibition in the US.

The scope of Article 20 is also controversial, particularly as on its terms it proscribes advocacy of hatred that incites not only violence but also 'discrimination' or 'hostility'. Some states, including the US, have entered reservations to Article 20 ICCPR[255] and interpret it narrowly. The Human

[247] UN Human Rights Committee, *General Comment No. 34* (2011), para. 11.
[248] Ibid., para. 13.
[249] Ibid., para. 15.
[250] Ibid., para. 20.
[251] Ibid., para. 22.
[252] UN Human Rights Committee, *General Comment No. 34* (2011), paras 48, 49. For a detailed discussion of the UN standards, see Report of UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression(2012), *Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms* (7 September 2012), A/67/357.
[253] Aswad, E. (2018), 'The Future of Freedom of Expression Online', *Duke Law & Technology Review* 17(26) p. 43; *Otto-Preminger-Institut v Austria* (1994), 13470/87; *Joseph Burstyn, Inc v Wilson*, 343 US 495 (1952), para. 506. The Alex Jones defamation litigation in the US is currently testing whether the First Amendment's protection of free expression of opinion can extend to deliberate lies. Collins, D. (2019), 'Court hears Alex Jones' appeal in Sandy Hook case', *AP News*, 26 September 2019, https://apnews.com/b003eeb4ee9844a498c7013a204d8933 (accessed 1 Nov. 2019).
[254] Aswad (2018), 'The Future of Freedom of Expression Online'.
[255] United States' reservation to Article 20 ICCPR: 'That article 20 does not authorize or require legislation or other action by the United States that would restrict the right of free speech and association protected by the Constitution and laws of the United States.' See also, United Kingdom: 'The Government of the United Kingdom interpret article 20 consistently with the rights conferred by articles 19 and 21 of the Covenant and having legislated in matters of practical concern in the interests of public order *(ordre public)* reserve the right not to introduce any further legislation'.

Rights Committee's view is that all speech that falls to be prohibited as a result of Article 20 – i.e. 'propaganda for war', and 'advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence' – must also meet the tests in Article 19(3).[256] The OSCE Representative on Freedom of the Media has challenged that view in research on propaganda, reviewing the history, drafting and aims of the provisions before concluding that 'freedom of expression under the ICCPR should be interpreted as *not including* war propaganda and hate speech that constitutes incitement to discrimination, hostility or violence'.[257] She is clear that Article 20 not only tolerates but requires the prohibition of incitement to hatred and discrimination.

Article 20 demonstrates that fears over disinformation are not new, and were addressed during the drafting of ICCPR. The Second World War saw the first widespread dissemination of hostile propaganda by radio broadcast directly into people's homes, in the early days of mass availability of the wireless. Mass broadcasting gave rise to fears of disinformation not dissimilar to those of today, save that contemporary threats were perceived as more closely connected with incitement to war and violence. The Appendix to this paper outlines the 20-year, closely contested controversy over how to deal with hostile propaganda and demonstrates that liberal states were among those calling for its restriction, although the permissible restrictions in the final text of Article 20, in including incitement to discrimination or hostility, were too broad for them to support. The addressing of these issues during the drafting process provides reassurance both that freedom of expression does not entail that disinformation cannot be restricted at all, and that new international norms are not needed to grapple with disinformation today.

In the electoral context, Frank La Rue, when UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, explored in detail the realization of the right to freedom of expression,[258] elucidating human rights principles that apply equally today in the face of new challenges of widespread online disinformation. He recognized 'the responsibility of States to prohibit incitement of hatred, hostility, discrimination and violence'[259] while 'ensuring an open public debate where all the main stakeholders … can freely share information and opinions'.[260] He called for states to '[deploy] efforts … to promote the pluralism of the media and ensure a plural political debate, ensure transparency in the promotion and financing of political campaigns, and guarantee accountability and fair enforcement of political regulations to prevent those in power from taking advantage … to dominate and manipulate public debate'.[261] The themes of openness, pluralism, a fair playing field and avoidance of domination by those in power are equally relevant in considering how best to ensure freedom of speech in political discourse online. La Rue emphasized

United Kingdom: 'The Government of the United Kingdom interpret article 20 consistently with the rights conferred by articles 19 and 21 of the Covenant and having legislated in matters of practical concern in the interests of public order *(ordre public)* reserve the right not to introduce any further legislation'.

[256] UN Human Rights Committee, *General Comment No. 34*(2011), para. 50.

[257] OSCE Representative on Freedom of the Media (2015), *Propaganda and Freedom of the Media*, Non-paper, https://www.osce.org/fom/203926?download=true (accessed 5 Oct. 2019), p. 17.

[258] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Report of Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, 26th session, UN Doc A/HRC/26/30 (2 July 2014).

[259] Ibid., para. 5.

[260] Ibid., para. 12.

[261] Ibid., para. 13.

independence and diversity of the media as a 'conduit' between voters and politicians,[262] principles that should also apply to digital platforms as they act as an interface between politicians, campaigners and voters. Equally significant is La Rue's observation that openness in political debate does not entail avoidance of regulation, but instead that states must 'ensure that an equitable balance is struck in providing for a structural environment that will enhance freedom of expression while not hindering the independent role of the media or the content of political expression'.[263]

5.4.2 Freedom of expression online

The internet and social media provide wonderful new potential for free expression, including for minorities, dissenters and alternative voices. In the election context, they allow for wider-reaching political campaigning than ever before, and have potential to enable voters to be better informed in making their democratic choices. However, establishing the parameters of speech properly protected by freedom of expression is particularly challenging in the online environment, in light of the potential for huge quantities of speech with extensive transnational reach, at high speed and without editorial filter. Some states, such as the UK, have legislated specific offences in respect of the originators of indecent, grossly offensive or harassing online speech.[264]

For some years, the US's expansive approach to freedom of expression, coupled with commercial motivations, encouraged a hands-off approach to content regulation by largely Silicon Valley-headquartered digital platforms. There was a focus on fact-checking initiatives and improvements in digital literacy, both of which are important but insufficient to tackle disinformation, given its potentially manipulative effect. More recently, there has been an appreciation, precipitated in part by violence, of the potential harms of online speech and the potential role of digital platforms in tackling it. While care needs to be taken not to erode the right to freedom of expression, the right does not prohibit addressing such harms.

In practice, policing the parameters of acceptable speech online is now a huge task, and one fraught with controversy, not assisted by significant gaps in international and domestic guidance. Decisions to take down content lie largely with digital platforms at present. There are currently concerns both that platforms are taking down too little, and suspicions that they're taking down too much,[265] exacerbated by a lack of transparency. Larger platforms have set up algorithms to detect and remove some inappropriate content: for example, Facebook claims that its automated systems identify and take down 98 per cent of hate speech before it is viewed. There is a risk that algorithms, lacking the capacity to evaluate subtleties of culture or humour in human speech, err on the side of taking down too much speech.[266] As regards requests to take down speech, Facebook now receives some 2 million reports of inappropriate content per day, and has 15,000 reviewers working in 50

[262] Ibid., paras 41–42.
[263] Ibid., para. 78.
[264] Malicious Communications Act 1988; section 127 Communications Act 2003.
[265] Keller, D., *Internet Platforms: Observations on Speech, Danger and Money*, Stanford: Hoover Institution on War, Revolution, and Peace, https://cyberlaw.stanford.edu/files/publication/files/381732092-internet-platforms-observations-on-speech-danger-and-money.pdf (accessed 5 Oct. 2019).
[266] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), *Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms*, para. 29.

languages to consider these requests. These reports may not all be motivated by public interest concerns: for example, in the run-up to the 2018 abortion referendum in Ireland, each side of the debate made politically motivated complaints to Facebook about the other's posts. Facebook is in the process of establishing an Independent Oversight Board which will oversee its content policy and enforcement decisions.[267] Much of the feedback in the consultation process exhorted Facebook to incorporate international human rights norms and standards into its core decision-making functions.[268]

5.4.3 The right to receive information

The right to freedom of expression includes the 'freedom to seek, receive and impart information and ideas of all kinds' (Article 19(2) ICCPR).[269] For many years, this has primarily been interpreted as an obligation on the part of states not to impede the distribution of information and to promote an independent and diverse media, as well as to provide information about governmental activities.[270] The algorithms deployed by platforms have a direct impact on the information that is received, and therefore have potential to have serious impact on the implementation of this right.[271] Some commentators consider that the right entails that digital platforms, or at least those platforms that 'contribute to the structure of the information and communication space' must 'respect political, ideological and religious neutrality' in doing so, being 'neutral' in their distribution and curation of information.[272] If there is such a responsibility on the part of digital platforms, fact-checking would also contribute to meeting it.

5.4.4 Freedom of expression: challenges

5.4.4.1 Government interference with internet service

The right to freedom of expression is a right at perennial risk of abuse at the hands of governments that see it as constraining their capacity to prevent the silencing of dissent and challenges.

Blunt government action, such as restrictions of internet service,[273] as well as the taking down of websites, jamming of signals, and blocking of specified types of content, are all violations of the rights of freedom of expression, save in the rare circumstance that they are narrowly tailored so as to be compatible with Article 19(3) ICCPR.

5.4.4.2 Determining the boundaries of protected speech

Through their content moderation and prioritization, platforms can have very significant reach into political and other public discourse. As Professor David Kaye argues, social media are 'a new kind of

[267] Facebook (2019), 'Getting Input on an Oversight Board', 1 April 2019https://newsroom.fb.com/news/2019/04/input-on-an-oversight-board/ (accessed 5 Oct. 2019); Facebook (2019), *Global feedback & input on the Facebook Oversight Board for Content Decisions*,https://fbnewsroomus.files.wordpress.com/2019/06/oversight-board-consultation-report-2.pdf (accessed 5 Oct. 2019).
[268] Ibid., p. 34.
[269] Also European Convention on Human Rights Article 10(1), American Convention on Human Rights Article 13(1).
[270] For example, UN Human Rights Committee General Comment 34 (2011).
[271] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), *Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms*, para. 30.
[272] Information and Democracy Commission (2018), International Declaration on Information and Democracy, , https://rsf.org/en/news/international-declaration-information-and-democracy-principles-global-information-and-communication (accessed 5 Oct. 2019).
[273] See section 3.4 above.

speech police' with a 'hold' on public spaces.[274] Consequently, it is arguable that there should be a role for the public sector in setting and/or overseeing the standards that platforms apply. Just as the right to privacy is given shape in domestic and international law through data protection laws, arguably the right to freedom of expression should similarly be given shape through norms that limit the discretion of platforms to remove or deprioritize speech.[275] While such a public-sector role would need to be deployed carefully to avoid its misuse as state censorship, this is no reason to leave these important public interests entirely to the commercial imperatives of the private sector.

This requires a frank conversation over the extent of permissible restrictions on speech online. At present, states may be recognizing the harm of some online speech, but, for fear of censure, are reluctant to appear to be restricting free speech – instead turning a blind eye to platform practices of unilaterally restricting harmful content. This in turn creates a risk that platforms adopt overly restrictive policies.

Legislation that simply bans or censors fake news or disinformation, without nuance, is inconsistent with the right to freedom of expression as it does not meet the criteria in Articles 19(3) or 20 ICCPR. A number of human rights experts appointed by intergovernmental organizations have declared that 'prohibitions on disinformation may violate international human rights standards'.[276] Articles 19(3) and 20 allow for restrictions on freedom of expression that (as above) are not only provided by law, but meet one of the purposes in those provisions and are necessary and proportionate to the harm being addressed; restrictions must be carefully tailored, rather than sweeping and unfocused.

It is not necessarily the case that Articles 19(3) and 20 only permit restrictions on freedom of expression online that are identical to those in the offline environment. The scale, speed and reach of speech online may entail legitimate adjustment to these restrictions and/or restrictions specifically of online speech. For example, it has not yet been explored whether, or in what circumstances, large-scale online political disinformation campaigns similar to that of the Internet Research Agency discussed above, potentially being used as a 'weapon of war',[277] may be seen as advocating hatred that foments social unrest in the form of hostility, discrimination or violence, and so fall to be prohibited by Article 20 ICCPR. In addition, the right to freedom of expression does not entail that techniques for the manipulation of attention, such as use of bots and trolls, must be free of restriction.

As regards election material, some draw a contrast between 'false' and 'distorted' election information in determining legitimacy.[278] However, this distinction is hard to draw, and its legitimacy may be disputed. In the UK, no regulator monitors the truth of election material, nor does the Online Harms White Paper propose to establish a focus on truth.[279] As above, it is arguable

[274] Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet,* pp. 26, 112.
[275] Some of these norms may already exist, such as the UK offences in the Malicious Communications Act 1988 and Communications Act 2003, but they are currently addressed only to the originators of offensive speech, not to platforms.
[276] United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access to Information (2017), 'Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda', preambular para. 7.
[277] See section 5.2.2 above.
[278] Baade, B. (2018), 'Fake News and International Law', *European Journal of International Law*, 29(4): pp 1357-1376, doi: https://doi.org/10.1093/ejil/chy071 (accessed 15 Oct. 2019).
[279] Department for Digital, Culture, Media & Sport; Home Office (2019), *Online Harms White Paper,* para. 7.31.

that the freedom to seek, receive and impart information includes an obligation for those digital platforms that contribute to the 'architecture' of the digital space to be neutral in their distribution and curation of information, a neutrality that could include fact-checking or promotion of true over false information.

The right to freedom of expression does not prevent digital platforms, as private entities, from deciding not to host speech which is lawful. Indeed, like newspapers, different platforms may intentionally adopt different content policies so as to attract users of different profiles: for example, a platform may adopt a 'family-friendly' stance or act as a platform for certain political views. Some argue that the largest platforms should not adopt unduly narrow content policies, as they are acting akin to a public service in providing a forum for free expression.

5.4.4.3 Algorithms

The algorithms deployed by digital platforms, determining who sees what content, can impact both freedom of expression (the right may not be exercised if speech is suppressed) and the right to receive information (if some content is suppressed, and some promoted). As a first step, there is a need for more transparency in algorithms with a view to understanding better their impact on these rights.

5.4.4.4 Who should make content decisions

In order to meet the requirements of freedom of expression, there is a need to develop consistent, fair, effective and efficient internal complaints and content monitoring processes. Expertise on international human rights law should be integral to the system, whether at platform level or through independent or regulatory oversight. Decisions regarding the take-down of content should be made with adequate knowledge of both facts and context.[280] Decisions on content should not be weighted so as to incentivize take-downs, thereby having a chilling effect on legitimate speech.

At present, the onus of determining what speech to retain and what to take down is falling on digital platforms. However, digital platforms are not currently well-placed to make decisions that satisfy these conditions, particularly without independent input such as guidance, oversight and/or recourse to appeal. Digital platforms have their own vested interests, usually commercial (for example, to maximize user and advertiser figures), and potentially other (for example, they have no commitment to political neutrality). Even if they establish processes with the aim of making fair decisions, digital platforms, not being public-sector organizations, are not well-placed to make assessments as to the parameters of Articles 19(3) and 20 ICCPR, for example as to the requirements of 'public order' or 'public morals'. The differing interpretation of the right to freedom of expression in different jurisdictions complicates their task further. Consistency is hampered by each platform making its own decisions, with little communication between them and little institutionalized transparency[281] or accountability.[282] The challenges are amplified for smaller platforms.

[280] Consequently, it may not be compatible with freedom of expression, for example, for a US-national content moderator in the US to ascertain whether particular speech would have the effect of inciting violence in Myanmar.
[281] At present, there is little transparency in decision-making. For example, Facebook had 800 people working on content take-downs for the 2019 Indian elections, but very little is known about their decisions.
[282] Klonick (2018), 'The New Governors: The People, Rules and Processes Governing Online Speech', pp. 1666–1669.

The status quo, whereby digital platforms make content decisions with little external oversight or guidance, therefore poses significant risks of application of the wrong standards, unfairness and inconsistency. It also imposes a major burden on digital platforms.[283] In reality, there may be no pragmatic alternative to content decisions being made by platforms; but there is significant scope for process improvements that will help meet the requirements of human rights law. One essential element for the improvement of standards is more transparency in content decisions, fundamental to ensuring fairness and consistency. This transparency is currently lacking: for example, Facebook currently has a list of designated 'hate speakers' who are banned from using its platforms,[284] but this list is not published. Moreover, Facebook does not explain why certain content is kept up or taken down, nor its decision-making process in respect of any one piece of content.

A second essential element is the improvement of fairness and consistency through appropriate processes, such as regulation, impartial scrutiny, appeal and/or oversight bodies as well as international collaboration. Larger platforms may set up their own oversight bodies, as Facebook is planning to do. Independent bodies established by governments could also play a role. Some digital platforms, such as Facebook, are calling for a greater governmental role in regulating what speech digital platforms ought to take down. On the other hand, there is suspicion of governments given their political impetus to restrict speech. The UK government is taking a different approach in proposing that an independent regulator develop codes of practice to tackle online harms while respecting freedom of expression.[285] International collaboration on guidance would help to promote consistency of approach across jurisdictions.

The right to freedom of expression does not determine to what extent digital platforms should bear legal responsibility for illegal content ('intermediary liability') and/or responsibility for policing the parameters of lawful speech, provided that the burdens on platforms are not so great as to undermine their operations. Neither the general immunity from liability in the US, nor existing and growing requirements for digital platforms to take down illegal speech once aware of it in the EU, are clear violations of the right to freedom of expression.

5.4.4.5 Private speech

Urgent consideration, by reference to human rights law, is needed to consider what if any restrictions may be appropriate in respect of speech that human rights law permits or requires to be limited but is in private spaces, e.g. WhatsApp.[286]

[283] Keller, 'Internet Platforms: Observations on Speech, Danger and Money', p. 2.
[284] Ortutay, B. (2019), 'Facebook bans "dangerous individuals" cited for hate speech', *AP News*, 3 May 2019, https://www.apnews.com/7825d0df3fda4799a78da92b9e969cdc (accessed 5 Oct. 2019).
[285] Department for Digital, Culture, Media & Sport; Home Office (2019), *Online Harms White Paper*.
[286] See section 3.2 above.

5.5 Right to participate in public affairs and to vote

Article 21 UDHR

'(1) Everyone has the right to take part in the government of his country, directly or through freely chosen representatives …

(3) The will of the people shall be the basis of the authority of government; this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures.'

Article 25 ICCPR

'Every citizen shall have the right and the opportunity, without any of the distinctions mentioned in article 2 and without unreasonable restrictions: (a) To take part in the conduct of public affairs, directly or through freely chosen representatives; (b) To vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors …'

5.5.1 The content of the right

The right to take part in the conduct of public affairs includes not only the right to participate as members of an executive or legislative body, but also the right to engage in public debate and assembly.[287] The undermining of that debate, for example through interruptions to internet access or through speech contrary to the right to freedom of expression, is consequently a breach of this right.[288]

The right to vote guarantees the right to participate in free and fair elections, in an election system that is free from external interference and permits the 'free expression of the will of the electors'. The latter element entails the freedom of thought and opinion and freedom of expression discussed above, specifically in the context of elections and participation in public affairs.[289] As the UN Human Rights Committee has observed, states are obliged to ensure that 'Voters should be able to form opinions independently, free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind,'[290] and 'Freedom of expression, assembly and association are essential conditions for the effective exercise of the right to vote and must be fully protected.'[291] The Committee has further commented that 'the free communication of information and ideas about public and political issues between citizens, candidates and elected representatives is essential',[292] and that:

[287] UN Human Rights Committee, *General Comment No. 25: Article 25 (Participation in Public Affairs and the Right to Vote), The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service*, 57th session, UN Doc CCPR/C/21/Rev.1/Add.7 (1996), para. 8; UN OHCHR, *Promotion, protection and implementation of the right to participate in public affairs in the context of the existing human rights law: best practices, experiences, challenges and ways to overcome them* (23 July 2015), A/HRC/30/26, paras 9–11.
[288] Ibid. (OHCHR), para. 41 gives some states' recognition of a right of access to the Internet as an example of good practice.
[289] Ibid., para. 69.
[290] UN Human Rights Committee, *General Comment No. 25* (1996), para. 19.
[291] Ibid., para. 12.
[292] Ibid., para. 25.

> [The right to vote] requires the full enjoyment and respect for the rights guaranteed in articles 19, 21 and 22 of the Covenant, including freedom to engage in political activity individually or through political parties and other organizations, freedom to debate public affairs, to hold peaceful demonstrations and meetings, to criticize and oppose, to publish political material, to campaign for election and to advertise political ideas.[293]

It follows that states must guarantee an open flow of information, freedom of expression and an open and pluralistic media, also that voters must be able to make up their mind freely and without manipulation.

Further, the right to vote entails that everyone should have the right to stand for election, without being deterred by the risk of being a target of hate speech. It also raises issues as to the sources and appropriate limits of election campaign funding, which are outside the scope of this paper.

Facebook has called for legislation to protect elections,[294] envisaging that legislation could create common standards for digital platforms to verify political actors, as well as for the identification and archiving of political adverts.

5.5.2 Right to participate in public affairs and to vote: potential breaches

The potential interferences with freedoms of thought and opinion, privacy and expression discussed above are also potential interferences with the rights to participate in public affairs and to vote, as they impact citizens' ability to engage in democratic debate and voters' ability to glean information and make up their mind freely.

The use of algorithms that manipulate what voters see, and so affect their political activity, may also reduce capacity to influence through public debate, and/or be inconsistent with the right to vote. This would include, for example, algorithms that give a distorted impression of public debate; algorithms that prioritize disinformation; algorithms that give the impression that a preferred candidate is bound to win, or a preferred referendum result is bound to be secured, with the intention of dissuading voters from voting and/or actively campaigning for that candidate or result. The same is the case as regards collecting and trading in personal data with a view to manipulation of voting or participation intentions through micro-targeting.

An environment that dissuades potential candidates from standing, or encourages them to withdraw, through the use of unlawful speech such as incitement to hatred against them may be inconsistent with the right to stand for election.[295] The increase in threats to politicians over recent

[293] Ibid., para. 25. Similarly the European Court of Human Rights in *Bowman v UK*, 1998 26 EHRR 1, paras 42–43: 'Free elections and freedom of expression, particularly freedom of political debate, together form the bedrock of any democratic system…The two rights are interrelated and operate to reinforce each other…'
[294] Zuckerberg (2019), 'Four Ideas to Regulate the Internet'.
[295] The Human Rights Committee has stressed the 'importance of freedom of expression for the conduct of public affairs and the effective exercise of the right to vote': UN Human Rights Committee *General Comment No. 34* (2011) para. 20. This suggests a correlation between limitations on the rights, i.e. that speech not permitted by freedom of expression is also contrary to the right to vote and participate.

years may be linked with the rise in hate speech online. Some politicians have stood down for this reason.[296]

Online messaging that discourages voting may also breach the right to vote.[297] For example, the Internet Research Agency allegedly encouraged US minorities not to vote in the 2016 presidential election, or to vote for a non-mainstream candidate.[298]

[296] For example, UK Conservative Member of Parliament Caroline Spelman MP announced in September 2019 her intention to stand down as a result of abuse and death threats, saying 'Myself, my family and my staff, have borne an enormous brunt of abuse and I think quite frankly we've had enough. The anonymity the Internet affords allows people to say things which if they said it to your face or they wrote it down, would not be legal.' Brewis, H. (2019), 'Caroline Spelman quits', *Evening Standard*, 5 September 2019, https://www.standard.co.uk/news/politics/caroline-spelman-quits-tory-mp-to-stand-down-over-abuse-and-death-threats-which-left-her-wearing-a4230241.html (accessed 5 Oct. 2019).
[297] UN Human Rights Committee, *General Comment No. 25* (1996), para. 11: 'Any abusive interference with registration or voting as well as intimidation or coercion of voters should be prohibited by penal laws and those laws should be strictly enforced.'
[298] *United States of America v Internet Research Agency LLC and others*, 18 USC 2, 371, 1349, 1028A (2018), para. 46.

6. Conclusion and Recommendations in Respect of Human Rights Law

The regulatory environment has not kept pace with the rapid development of digital platforms and their harnessing for political use. There is a growing appreciation that the new digital campaigning environment, while increasing pluralism and engagement with politics, poses significant threats to democracy. It threatens to lend credence to untruths, to prioritize shocking content and emotion over rational debate, to polarize, and to distort narratives. Both digital platforms – in their quest for user attention – and deliberate disinformation campaigns adopt and amplify techniques used in the advertising industry to manipulate attention, emotions and reactions. There is growing discussion of what should be done, and some states are introducing legislation or other regulation. But to date there has been little recourse to established normative frameworks.

Human rights law should be at the heart of any discussion of international or domestic regulation, guidance, or societal responses to cyber interference in political thought; not because it imposes legal obligations on digital platforms, but because it is a framework established to safeguard individuals from the power of authority. This paper has explored the implications of the rights to freedom of thought and opinion, the right to privacy, the right to freedom of expression, and the right to participate in public affairs and to vote. Contrary to popular view, freedom of expression does not entail that there must be no restriction of online political content; rather, that any restriction must be properly tailored.

Current practice evidences failures to meet the standards of each of these rights: of freedom of thought in the impact of digital platform structures and online political campaigning on personal agency; of the right to privacy in the widespread harnessing, trade and use of personal data, including for close targeting of political material; of freedom of expression in how decisions on content removal and retention are made and overseen; and of the right to participate in public affairs and to vote in respect of all these matters in the political context. Urgent action is needed on the part of states and digital platforms to end these failures. The health of the world's democracies is at stake.

6.1 Application of human rights law: recommendations

6.1.1 States

In light of their duty to protect their inhabitants[299] from abuse of human rights by business enterprises, all states should put human rights at the heart of the debate on how to tackle cyber

[299] Regarding extraterritorial jurisdiction, please see discussion in Chapter 5, Section 5.1.

interference in elections. Any initiatives for legislation, regulation, codes of ethics or behaviour for digital platforms should incorporate consideration of human rights law.[300]

States should not rely on digital platforms to self-regulate human rights protections into their business models. Voluntary initiatives cannot compete adequately with commercial business imperatives, which currently favour divisive content and detailed personal profiling. Although market pressures encourage some action by companies that mitigates abuse – such as some removal of egregious content – they are not sufficient to respect the human rights of users. Often, voluntary commitments on the part of companies have not been implemented, and/or lack of transparency means that implementation has been impossible to measure.

As online activities are not naturally territorially bounded, international discussion, consensus and guidance are needed on the implications of existing human rights law for the activities of digital platforms in connection with elections and other political discourse, building on the Ruggie Principles and existing human rights law jurisprudence. UN human rights processes, including the forthcoming work of the Human Rights Council's Advisory Committee, are obvious loci for this. Discussion and guidance within the Council of Europe and other regional bodies would be valuable steps towards universal consensus.

6.1.2 Digital platforms

All companies have a responsibility to respect human rights, wherever they operate. This entails awareness of, and behaviour consistent with, internationally accepted standards, and not merely the standards of their home state. Implementation should take account of local circumstances that may affect the implementation of the rights at stake.

6.2 Rights to freedoms of thought and opinion: recommendations

Freedoms of thought and opinion raise complex issues because of the challenge, as yet barely explored, of differentiating between legitimate and illegitimate influence. Despite that complexity, freedom of thought and opinion offer an important lens through which to assess the structure and business models of our online environments, and their openness to abuse by disinformation campaigns.

Digital literacy campaigns, often advocated as a shield against inappropriate online influence, are important in helping individuals be aware of those influences, but alone such campaigns are unlikely to be adequate to combat them.

[300] For example, the UK House of Commons Digital, Culture, Media and Sport Committee recommended establishment of a compulsory Code of Ethics setting out what constitutes harmful content. Digital, Culture, Media and Sport Committee (2019), *Disinformation and 'fake news': Final Report*, para. 37.

6.2.1 States

At international level, multilateral discussion is needed to produce guidance on the rights to freedom of thought and freedom of opinion, their meaning, scope and parameters, and on the impact of digital platforms and disinformation campaigns on those rights.

The structure of many social media companies as advertising companies, and their openness to disinformation campaigns, pose significant threats to freedom of thought in political discourse. States should consider whether structural changes to digital platforms are needed to shield users against unknowing manipulation or involuntary influence of their thought and opinion. For example, there may be a need for far greater transparency on the use of persuasive techniques, and restriction of specific techniques that are proven to have manipulative effect (just as subliminal advertising has long been restricted offline). More radical possibilities would include the restructuring of platforms, for example by separation of political and personal content to different platforms.

As regards disinformation campaigns, there is an urgent need for states to consider in what ways the techniques deployed by, most notably, the Internet Research Agency breach freedom of thought, and therefore what measures should be used to constrain them. This consideration should be at a systematic level, rather than merely at the level of specific pieces of content. For example, states may wish to consider restricting techniques that do not add to content of expression but that serve to amplify its manipulative value, such as the use of bots, cyborgs and trolls.[301] Banning campaign funding from overseas would help to limit the reach of such techniques.[302]

6.2.2 Digital platforms

At a minimum, there should be much more transparency as to the aims and activities of all digital platforms and their use of algorithms. Platforms should respect freedom of thought and opinion in designing the operation of their services, and avoid unduly manipulative techniques. Platforms' uses of personal data and placement of advertising are linked to these issues, and will be considered in section 6.3.

Platforms could adjust their algorithms so as to reduce the amplification of disinformation and distortion, and to give individuals greater sight of, and control over, the algorithms that affect what content they see.[303]

6.2.3 Other private actors

All online campaign material, just like offline material in the UK, should carry an imprint stating from whom it originated and who paid for it.[304] Arguably, all political material should be imprinted in this way. All advertisements, including political adverts, should be visible to all audiences, either

[301] UK Electoral Commission proposes that posts by bots and paid trolls should carry imprints, like political advertisements. Electoral Commission (2018), *Digital Campaigning: increasing transparency for voters*, p. 9.
[302] Ibid., p. 18.
[303] Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet*, p. 92.
[304] Electoral Commission (2018), *Digital Campaigning: increasing transparency for voters*, pp. 7–9.

through open publication by digital platforms or through a central public register (as proposed in the UK by the Institute of Practitioners in Advertising[305]). Information would include not just the content of the advert, but the targeting, actual reach and amount spent. Facebook has launched such initiatives as regards political adverts, to a certain extent,[306] while Twitter has recently announced a ban on political advertising.[307] Other platforms have not yet done so.

6.3 Right to privacy: recommendations

6.3.1 States

States need to be aware of all the current practices of commercial organizations in harvesting, using, trading and storing personal data, including in algorithmic processes and advertising. In the UK, both the Parliamentary Joint Committee on Human Rights' inquiry on the Right to Privacy and the Digital Revolution[308] and the Information Commissioner's work on adtech and real-time bidding[309] are valuable initiatives in this regard.

States should place the right to privacy at the heart of regulation of the use of personal data in algorithms, political campaigning and advertising. Current practice, whereby vast amounts of data are harvested, curated and traded for commercial or political purposes, largely without the knowledge of the individuals whose data is collected, is likely inconsistent with the right to privacy: states should consider a significant tightening of requirements of consent and transparency. States should consider regulating the growing trade in personal data and the activities of data brokers, including for political purposes.

Where states give effect to the right to privacy through data protection laws, those laws should facilitate compliance with the right to privacy. Data protection laws should not permit technical compliance that has the effect of undermining the right to privacy, such as notional 'consent' to extensive personal data gathering, profiling and transacting through non-transparent, semi-optional 'consent to cookies'. The right to privacy may entail considerable change to current practices in harvesting and trading data on the basis of either notional consent or (in the EU) inappropriate reliance on a 'legitimate interests' basis for processing.[310]

States should foster a culture of respect for the right to privacy, for example by ensuring full implementation of data protection laws. Regulators should be able to see how commercial organizations are using data, and should have powers to require evidence and impose fines on a scale that encourages digital platforms and commercial companies to build data protection into

[305] Ibid., p. 13 para. 59.
[306] Facebook (2019), 'Facebook Ad Library', https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=GB (accessed 5 Oct. 2019); Twitter (2019), 'Twitter Ads Transparency Center', https://ads.twitter.com/transparency (accessed 5 Oct. 2019).
[307] Dorsey (@jack) (2019), 'We've made the decision to stop all political advertising on Twitter globally. We believe political message reach should be earned, not bought. Why? A few reasons…'.
[308] Joint Committee on Human Rights (2019), 'The Right to Privacy (Article 8) and the Digital Revolution Inquiry', https://www.parliament.uk/business/committees/committees-a-z/joint-select/human-rights-committee/inquiries/parliament-2017/right-to-privacy-digital-revolution-inquiry-17-19/ (accessed 5 Oct. 2019).
[309] Information Commissioner's Office (2019), *Update report into adtech and real time bidding*.
[310] Per the UK Information Commissioner's Office, 'the adtech industry appears immature in its understanding of data protection requirements'. Ibid., p. 23.

their operations. States should ensure that individuals can find out easily (at the click of a button, for example) what profiling data and other information about them is being held and shared, by whom, and for what value. States should consider treating data as an asset with a commercial benefit that should be shared with its subjects.

Because online data transfers are often unbounded by jurisdiction, international discussion is needed on standardizing norms and cultures of data protection internationally, building on the 2018 OHCHR Report.[311] Consistency of standards and implementation internationally would simplify compliance and monitoring.

6.3.2 Digital platforms

Digital platforms should be transparent about their collection, trading and aggregation of personal data. They should comply fully with data protection laws, where these are in place. Consistent with business' responsibility to respect human rights, digital platforms should embed the right to privacy into technological design. Arguably this should mean that they hold significantly less personal data than at present. Platforms should also normalize the right of access in corporate and social culture, so that users should be entitled and able to see, quickly and easily, all the data on them, and inferences and profiles drawn therefrom, held by digital platforms and political parties. Platforms should reduce or eliminate non-optional third-party cookies and similar tracking devices such as canvas fingerprinting,[312] placed only for commercial data-gathering purposes.

6.3.3 Other private actors

Political parties, and companies working for them, should be transparent about their collection, purchase and use of personal data; and about their generation and use of targeted messages. Like platforms, arguably their capacity to collect, draw inferences from and trade in personal data should be restricted significantly. Arguably, the right to privacy entails that political parties should not be entitled to use data for micro-targeted political prediction or persuasion.

6.4 Right to freedom of expression: recommendations

6.4.1 States

States must not impede individuals' unrestricted access to an open and uncensored internet, and must not disrupt networks, digital platforms or websites except in the most exceptional circumstances. Although states have already made this commitment,[313] the frequency of breach suggests that some states are not aware of it – or do not feel sufficient pressure to comply.[314]

[311] UN OHCHR (2018), *The right to privacy in the digital age*.
[312] A technique of tracking online users without using cookies (instead using HTML5 canvas element).
[313] Human Rights Council Resolution (2018), *The promotion, protection and enjoyment of human rights on the Internet*, UN Doc A/HRC/38/L.10/Rev.1 (5 July 2018) para. 13.
[314] UNESCO and Global Network Initiative (2018), *Improving the communications and information system to protect the integrity of elections: conclusions*, p. 5.

State guidance or jurisprudence is needed on the implications of Articles 19(3) and 20 ICCPR for speech online. For example, it should be clear in what circumstances a prohibition of disinformation in political discourse may be compatible with Article 19(3), and there should be a clear threshold for prohibitions required by Article 20.

States should not leave it to digital platforms to determine what freedom of expression requires of them, nor avoid difficult or unpalatable decisions by delegating them to businesses. As Professor David Kaye has written: '[T]he rules of speech for public space, in theory, should be made by relevant political communities, not private companies that lack democratic accountability and oversight.'[315] Governments should set clear boundaries for permissible and impermissible content by reference to human rights law. At least as regards the larger platforms, these boundaries should both prohibit removal of political discourse that is consistent with freedom of speech (such as speech critical of an incumbent administration) and require removal of discourse which is inconsistent with it (such as incitement to violence). Governments should take care that any regulatory efforts fully respect the right to freedom of expression: for example, they should not incentivize take-downs over retention of content in ambiguous cases.[316]

States should consider establishing impartial scrutiny mechanisms such as independent regulatory bodies or social media councils[317] to oversee digital platforms' decisions on removal or retention of contentious content by reference to the right to freedom of speech and other human rights law.[318] Such mechanisms would help ensure fairness, neutrality, transparency and consistency in decision-making.

6.4.2 Digital platforms

As they are required to make decisions about the take-down and retention of content, platforms should establish frameworks that enable efficient, fair, context-specific decision-making, reflecting the standards of human rights law. Larger platforms, or consortia of platforms, should establish impartial scrutiny mechanisms to oversee their decision-making. Digital platforms should provide clear, open information on the rules they are applying in making content decisions, and should provide data on their content management and take-down decisions.

Given the freedom to seek, receive and impart information, larger digital platforms should make transparent their algorithmic policies and take care to ensure that they facilitate rather than impede that freedom, so as to promote a more democratic and less polarized dialogue online. For example, they should consider algorithmic 'throttling' (i.e. down-ranking) of content that is designed to mislead.

[315] Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet*, p. 112.
[316] Ibid., p. 123. As regards the downsides of governance by platforms alone, see also Klonick (2018), 'The New Governors: The People, Rules and Processes Governing Online Speech', pp. 1664–1669.
[317] For example, Article 19 (2019), 'Social Media Councils: Consultation', https://www.article19.org/resources/social-media-councils-consultation/ (accessed 5 Oct. 2019).
[318] Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet*, p.116.

6.4.3 Other private actors

Freedom of expression entails entertaining a plurality of voices. The promotion of media literacy, not only among young people but also on the part of older audiences, will help them understand how to evaluate different sources of information and how to think critically about information that reaches them.[319] Audiences should be encouraged to check the truth of an online post before liking or sharing it. Similarly, journalists should be taught how to discern quality of sources. Fact-checking by journalists and civil society should be encouraged, and their sites promoted, in order to maximize the availability of reliable sources of verified information,[320] while not overlooking the role of 'strategic silence' [321] in combating disinformation. Public-service media and a plurality of journalism should be supported.

6.5 Right to participate in public affairs and to vote: recommendations

6.5.1 States and digital platforms

It is vital for democracy that states identify and combat activity that is inconsistent with the right to vote and participate in public affairs. To preserve capacity to engage in public affairs and the free expression of the will of electors, there should, as already discussed, be no undue manipulation of thought, use of personal data or discouragement of voting. Consideration should be given to measures to tackle hate speech, bots and trolls, algorithms that prioritize disinformation, and micro-targeting for the purpose of manipulating voter behaviour. Existing offline electoral safeguards should be applied to online campaigning, for example to require imprints on and transparency of political adverts, tackle overseas interference, enforce limits on campaign spending, enforce rules on political communications, and ensure equal treatment of candidates. These measures should be bolstered by robust safeguards such as media literacy and promotion of responsible, free journalism.

[319] For example, UK Government (2019), 'Use the S.H.A.R.E. Checklist', https://www.sharechecklist.gov.uk/?utm_source=twitter&utm_medium=cpx&utm_campaign=29032019-zn (accessed 5 Oct. 2019).
[320] For example, the CrossCheck project brought together 37 newsroom partners in France and UK to help report false, misleading and confusing claims that circulated online in the 10 weeks prior to the French presidential election in 2017. First Draft (2017), 'CrossCheck: Our Collaborative Verification Newsroom', https://firstdraftnews.org/about/crosscheck-newsroom/ (accessed 5 Oct. 2019).
[321] Wardle and Derakhshan (2017), *Information Disorder: Toward an interdisciplinary framework for research and policymaking*, p. 19.

Appendix: Historical background to the contemporary debate on propaganda and free expression

Propaganda has long been used as a tool of foreign policy and pursuit of power.[322] The challenge of how best, if at all, to regulate 'fake news' is by no means new. From the period of the French Revolution, 'subversive propaganda' employed by one state to stir up insurrection within another state was widely seen as unlawful. There was, for instance, widespread condemnation of the French National Assembly's decree of 19 November 1792, offering the aid of the French nation to all peoples desirous of recovering their liberty;[323] and there were extensive diplomatic protests against subversive propaganda attacks by Soviet Russia on foreign states from 1920.[324] Until the 1930s, it was not accepted that the state had a duty to prevent messaging intended to stir up insurrection in another state by private actors operating on its territory.[325]

In 1936 the League of Nations adopted the International Convention concerning the Use of Broadcasting in the Cause of Peace.[326] The Convention, which came into force on 2 April 1938, technically remains in force, albeit that Australia, France, the Netherlands and the UK denounced it after Russia became a party in 1983. The US is not a party.

Article 3 provides:

> The High Contracting Parties mutually undertake to prohibit and, if occasion arises, to stop without delay within their respective territories any transmission likely to harm good international understanding by statements the incorrectness of which is or ought to be known to the persons responsible for the broadcast …

The Convention was an early attempt to combat the perceived threat of disinformation spread by radio. However, despite attracting 22 states parties, it was not recognized as encapsulating a general rule of international law, and did not play a significant role in restricting the spread of propaganda during the Second World War.[327]

The Convention was novel not only in imposing treaty restrictions on states' behaviour in respect of false information, but also in requiring states to regulate the behaviour of private actors in accordance with its provisions.[328] While Article 3 prohibits 'incorrect' statements, Articles 1 and 2 prohibit transmission that would incite war or acts threatening to internal order or security, regardless of whether true.

[322] For a historical overview, see Whitton (1948), 'Propaganda and International Law', pp. 551–562.
[323] Ibid., p. 583.
[324] Ibid., p. 585.
[325] Ibid., p. 593.
[326] League of Nations (1936), 'International Convention concerning the Use of Broadcasting in the Cause of Peace (adopted 23 September 1936, entered into force 23 September 1936)', 186 LNTS 301.
[327] Baade (2018), 'Fake News and International Law', p. 12.
[328] League of Nations (1936), 'International Convention concerning the Use of Broadcasting in the Cause of Peace', Article 6.

After 1945, concerns regarding mass communication as a potential conduit for propaganda to some extent paralleled those of today. The advent of mass radio was seen as heralding opportunities to bring information and propaganda into people's homes internationally in an unprecedented way. Just as with the internet in the 21st century: 'The ominous potentialities of radio increase by leaps and bounds as the number of stations, the power and range of the signals, and the number of listeners, continue to grow with astounding rapidity.'[329] Radio propaganda was seen as one of the two 'most lethal weapons in the history of the world' – the other of course being the atomic bomb.[330]

Propaganda was distinguished from education and information not by its content (propaganda could be true or false) but by motive, in that the term 'refers to the conscious effort to mould the minds of men in a particular direction so as to produce a particular effect', the aim of a propaganda bureau being 'to persuade, not to inform or enlighten'.[331] Like disinformation, propaganda was distinguished by the motive or intention behind it; but whereas 'disinformation' refers only to false or distorted information that is knowingly shared to cause harm, 'propaganda' is a broader term that also covers true information shared with intention to persuade.[332] But propaganda shared with intention to cause harm was considered threatening, and the threat was perceived in not dissimilar manner to that posed by disinformation today: 'The remarkable cheapness of propaganda compared with the cost of other weapons of power, and its enormous potentialities, have greatly endeared this psychological arm to aggressive governments …'[333]

As a result of these post-war concerns, during the 1940s there were extensive debates about the future regulation of propaganda. The 1947 UN General Assembly adopted two resolutions on this topic: one condemning propaganda for war; and the other false and distorted information.[334] These were taken up at the UN Conference on Freedom of Information in 1948. The Conference formed in part a precursor to the adoption of the Universal Declaration of Human Rights (UDHR) and International Covenant on Civil and Political Rights (ICCPR), as it declared fundamental rights to freedom of information, freedom of thought and expression, freedom of opinion, and the right to seek, impart and receive information and ideas by any means and regardless of frontiers.[335] The Conference condemned peacetime censorship, and emphasized the rights to freedom of information and to listen, proposing practical measures such as the distribution of cheap radio sets to give effect to these rights.[336] While the Conference condemned 'all distortion and falsification of news through whatever channels, private or governmental, since such activities can only promote misunderstanding and mistrust between peoples of the world',[337] the US successfully argued that efforts to control such material should be voluntary, due to concerns over governmental suppression of speech.[338] France and Belgium, and even the UK, were more favourable than the US

[329] Whitton (1948), 'Propaganda and International Law', p. 550.
[330] Ibid., p. 549.
[331] Ibid., p. 547.
[332] Ibid., p. 567.
[333] Ibid., p. 548.
[334] UN General Assembly Res 110 (II), 'Measures to be taken against propaganda and the inciters of a new war' (3 November 1947), UN Doc A/RES/110(II) and UNGA Res 127 (II), 'False or distorted reports' (15 November 1947), UN Doc A/RES/127(II).
[335] UNGA, 'Draft Convention on Freedom of Information' (1948), UNYB 593; Whitton, J. B. (1949), 'The United Nations Conference on Freedom of Information and the Movement against International Propaganda', *The American Journal of International Law*, 43(1): p. 74.
[336] Ibid., p. 74.
[337] 'Final Act of the UN Conference on Freedom of Information' (21 April 1948), UN Doc. E/CONF.6/C.1/19, p. 22.
[338] Whitton (1949), 'The United Nations Conference on Freedom of Information and the Movement against International Propaganda', p. 76.

towards proposals for the international control of propaganda.[339] The Draft Convention on Freedom of Information, its drafting led by the British and approved by the majority (but not the US), expressly stipulated that freedom of information (similar to today's freedom of expression) may be subject to restrictions including the 'systematic diffusion of deliberately false or distorted reports which undermine friendly relations between peoples or states'.[340] Ultimately, however, scepticism with regard to the motivation behind the Eastern Bloc's insistence on regulation of propaganda did nothing to help the USSR's cause.[341] In the view of Whitton, the firm opposition of the US to any legal restriction on propaganda was shaped in part by the strong influence of the American press on the US delegation to the Conference, seven of the 10 US delegates to the Conference being representatives of the media. The Draft Convention on Freedom of Information was not agreed, and continued to be discussed for a number of years[342] before being absorbed into work on the draft ICCPR.

The debate on propaganda continued after the 1948 Conference, during negotiations on the draft UDHR. The primary, and consistently unsuccessful, proposer of restriction of subversive speech was the Soviet delegation,[343] but the USSR was not the only state to countenance limitations. As regards false news, a French proposal put forward by René Cassin would have limited freedom of expression 'by defamation of character or failure to present information and news in a true and impartial manner'.[344] The British proposal for a Bill of Rights included an exception to freedom of expression for 'publications intended or likely to incite persons to alter by violence the system of Government', albeit with an accompanying comment emphasizing that the exception 'is to be interpreted as strictly confined to such publications as advocate the use of violence, and does not apply to publications advocating a change of government or of the system of Government by constitutional means'.[345] While no such limitation was included in Article 19 UDHR, the general limitations on rights included in Articles 29–30 were intended to operate as limits on freedom of expression.[346]

Following the adoption of the UDHR, the debate over propaganda – both subversive propaganda and propaganda for war – continued during the ICCPR negotiations, with the text of what became Article 20 (Article 26 of the draft ICCPR) proving particularly controversial. This was a politicized debate, with the US championing unrestricted freedom of expression and the Soviets again taking a restrictive approach. But, distinct from the USSR view, several other states spoke against the use of propaganda to shape public opinion, including the Chilean and French delegates. For example,

[339] Ibid., p. 77. Whitton quotes the UK delegate, a young Vincent Evans, at footnote 19: 'all governments owed a duty not only to their own citizens but also to international law to suppress all activities which might prejudice international peace or law and order. Such activities did not always lend themselves readily to definition as incitement to violence or as criminal acts or offences inimical to peace.'
[340] UNGA, 'Draft Convention on Freedom of Information', Article 2(1)(j); United States Delegates, 'Report of the United States Delegates with Related Documents (Department of State Publication 3150, International Organization and Conference Series III, 5) p. 22.
[341] The US Delegation's Official Report on the Conference stated that 'stripped of propaganda phraseology designed to confuse, the Soviet offensive amounted to a drive for the institution in other countries of a state-controlled press system, with governments deciding what is true and what is false, what is friendly and what is unfriendly.' Whitton (1949), 'The United Nations Conference on Freedom of Information and the Movement against International Propaganda', p. 82.
[342] For example, UNGA Res 1840 (XVII) (December 1962), UN Doc A/RES/1840(XVII).
[343] For example, the USSR proposed a prohibition of 'war-mongering and fascist speech', UNGA Res 3rd Committee, General Assembly Official Records, 180th Plenary Meeting (9 December 1948), UN Doc A/PV.180, p. 855. The proposal was defeated by 41 votes to 6, with 9 abstentions (Ibid., pp. 930–31). The USSR made a similar proposal when Article 19 was discussed in the General Assembly.
[344] Drafting Committee on an International Bill of Human Rights, First Session, 'Report of the Drafting Committee to the Commission on Human Rights' (1 July 1947), UN Doc E/CN.4/21, p. 57.
[345] Drafting Committee of the Commission on Human Rights, 'Text of Letter from Lord Dukeston, the United Kingdom Representative on the Human Right Commission, to the Secretary-General of the United Nations (5 June 1947), UN Doc E/CN.4/AC.1/4, pp. 11–12.
[346] For a detailed discussion, see Farrior, S. (1996), 'Molding the Matrix: The historical and theoretical foundations of international law concerning hate speech', *Berkeley Journal of International Law* 14(1), pp. 17–21.

René Cassin, on behalf of France, proposed that the article prohibit 'any advocacy of national, racial or religious hostility that constitutes an incitement to violence *and hatred*' [*emphasis added*].[347] The US and the UK argued against such broad restrictions on grounds of the potential for abuse by unscrupulous governments. In substance, there appears to have been a measure of shared concern about the potential role of propaganda in forming public opinion, but disagreement over whether to respond to this by law or by other means. Eventually, by way of proposed compromise, 16 (non-Western) states introduced the text that was adopted as Article 20(2). The formulation in Article 20(2) was too sweeping to secure Western or Chilean support, and was adopted by vote.[348] The US and several Western states have entered reservations to Article 20.[349]

This brief account illustrates that the drafters of the ICCPR grappled with how to address disinformation, and that many of them considered that some restriction on it ought to be compatible with the freedom of expression. The point was extremely controversial, and a mutually acceptable compromise was never reached. But although Article 20 was not adopted by consensus, its existence and the debate leading to its adoption demonstrate that the threat of disinformation was not overlooked in the drafting of the ICCPR, and that the drafters did not intend that disinformation could never be restricted. Consequently, this history offers reassurance that the provisions of Articles 19 and 20 should be adequate to tackle the problems posed by disinformation today. The new challenge of our times is not one of substance but of means – i.e. not propaganda itself, but the potential that social media offers for its rapid and pervasive dissemination and inculcation into culture.

[347] UN Commission on Human Rights, Sixth Session 31 May 1950 (14 June 1949), UN Doc. E/CN.4/SR.123, p. 6.
[348] Farrior (1996), 'Molding the Matrix: The historical and theoretical foundations of international law concerning hate speech', pp. 36–42.
[349] As well as the US, Australia, Belgium, Denmark, Finland, Iceland, Luxembourg, the Netherlands, Switzerland and the UK have entered reservations to Article 20, while France, Ireland, Malta, New Zealand and Thailand have entered declarations concerning their understanding of its implications.

Acronyms and Abbreviations

ACHPR	African Commission on Human and Peoples' Rights
CJEU	Court of Justice of the European Union
COMPROP	Computational Propaganda Research Project, University of Oxford
DCMS	Department for Digital, Culture, Media & Sport
ECtHR	European Court of Human Rights
Euratom	European Atomic Energy Community
GDPR	General Data Protection Regulation
GNI	Global Network Initiative
HRC	United Nations Human Rights Council
ICCPR	International Covenant on Civil and Political Rights
ICO	Information Commissioner's Office
ICT	information and communications technology
IRA	Internet Research Agency
IT	information technology
JO	Journal Officiel
NATO	North Atlantic Treaty Organization
OAS	Organization of American States
OHCHR	Office of the High Commissioner for Human Rights
OJ	Official Journal of the European Union
OSCE	Organization for Security and Co-operation in Europe
RT	Russia Today
RUSI	Royal United Services Institute
UDHR	Universal Declaration of Human Rights

UNESCO	United Nations Educational, Scientific and Cultural Organization
UNGA	United Nations General Assembly
USC	United States Code

About the Author

Kate Jones is a member of the University of Oxford's Faculty of Law, and the director of its Diplomatic Studies Programme.

Acknowledgments

The author has benefited tremendously from the great and rapidly growing richness of literature and thought on these topics, only some of which it has been possible to reference in this paper. The author would like to thank Harriet Moynihan and Ruma Mandal of Chatham House for their ideas and support; colleagues and friends who provided food for thought on various topics discussed in this paper; all those who contributed to seminars on this topic in May 2019 at Chatham House and the Bonavero Institute, University of Oxford; and the peer reviewers for their insightful comments and suggestions.

Independent thinking since 1920

Cover image: A man votes in Manhattan, New York City, during the US elections on 8 November 2016.

Photo credit: Copyright © Mohammed Elshamy/Anadolu Agency/Getty

ISBN 978 1 78413 374 0

This publication is printed on FSC-certified paper.

FSC

Typeset by Soapbox, www.soapbox.co.uk

The Royal Institute of International Affairs
Chatham House
10 St James's Square, London SW1Y 4LE
T +44 (0)20 7957 5700 F +44 (0)20 7957 5710
contact@chathamhouse.org www.chathamhouse.org

Charity Registration Number: 208223