11

Responding to Disinformation

Ten Recommendations for Regulatory Action and Forbearance

CHRIS MARSDEN, IAN BROWN, AND MICHAEL VEALE ■

INTRODUCTION

This chapter elaborates on challenges and emerging best practices for state regulation of electoral disinformation throughout the electoral cycle. It is based on research for three studies during 2018–2020: into election cybersecurity for the Commonwealth (Brown et al. 2020); on the use of artificial intelligence (AI) to regulate disinformation for the European Parliament (Marsden and Meyer 2019a; Meyer et al. 2020); and for UNESCO, the United Nations body responsible for education (Kalina et al. 2020). The research covers more than half the world's nations, and substantially more than half that population, and in 2019 the two largest democratic elections in history: India's general election and the European Parliamentary elections.

We found the claim of Tambini (2018, 270) still holds true, that there is 'surprisingly little analysis of the messages themselves, or of the validity of some of the more worrying claims about new forms of propaganda'. We do not know, and cannot measure, the individual or combined effect of the digitally dominant platforms on elections. We here refer to dominance in both the competition law definition (including Facebook/WhatsApp/Instagram, Apple, Google/Alphabet, Amazon, Microsoft—the so-called GAFAM platforms) (Competitions and Markets Authority 2020, particularly Appendix W), and also more broadly the more vernacular notion of large platforms (including Twitter, TikTok, and others) (Barwise and Watkins 2018).

We use the EU High Level Group's definition of disinformation as 'false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit' (High level Group on Fake News and

Online Disinformation 2018, 10), as distinguished from misinformation, which refers to unintentionally false or inaccurate information (Wardle and Derakhshan 2017).[1] We include accurate information presented with deceptive provenance or authorship—for example, emails stolen from Hillary Clinton's 2016 presidential campaign and the Democratic National Committee leaked via WikiLeaks and a fake hacker persona, 'Guccifer 2.0', which turned out to be a front for the Russian military intelligence agency GRU (Rid 2020, 377–86). The topic of disinformation became even more high-profile during the World Health Organization-diagnosed 'infodemic' relating to the COVID19 pandemic of 2020 (Silverman 2014), but we do not analyse that specific phenomenon here, since '[t]actics that work against dangerous health misinformation are likely to be less effective and more harmful when applied to political speech' (Kreps and Nyhan 2020).

Regulating digital dominance in electoral disinformation presents specific challenges in three very distinctive fields: election law, media law, and mass communications regulation, and targeted online advertising (including data protection law). International human rights law places strict limits on state actions restricting freedom of opinion and expression, summarized recently by UN Special Rapporteur David Kaye:

> In accordance with Article 19 (1) [of the International Covenant on Civil and Political Rights (ICCPR)], freedom of opinion may not be subject to any interference. Article 19 (2) robustly defines freedom of expression as one that is multidirectional ('seek, receive and impart'), unlimited by viewpoint ('information and ideas of all kinds'), without boundaries ('regardless of frontiers'), and open-ended in form ('or through any other media'). Article 19 (3) provides narrow grounds on which Governments may restrict the freedom of expression, requiring that any limitation be provided by law and be necessary for respect of the rights or reputations of others, or for the protection of national security or public order, or of public health or morals. That is, such limitations must meet the tests of necessity and proportionality and be aimed only towards a legitimate objective. (Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression 2020 para. 11)

The ICCPR also states, 'all peoples have the right of self-determination' (Article 1) and the opportunity to 'take part in the conduct of public affairs, directly or through freely chosen representatives', and to 'vote and to be elected at genuine periodic elections . . . guaranteeing the free expression of the will of the electors' (Article 25 ICCPR).

Russian state interference in the 2016 US presidential elections, via false information shared on social media (Corera 2020 chap. 26; Special Counsel Robert S. Mueller III 2019), has been a high-profile global media story ever since it was first revealed. But these tactics go back at least to 2011, when Russia was accused of deliberately faking news of political corruption in Ukraine (Sanovich 2018), and since when Iran's national broadcaster has used fake accounts to post on Facebook about 'a wide range of themes, from perennial Iranian concerns, such as the country's enmity with Israel and Saudi Arabia, to more surprising and momentary topics, such as the Occupy Movement of 2012 and the Scottish independence referendum of 2014' (Nimmo et al. 2020, 1). The interference by foreign state actors in digital platforms continued through the 2020 US election season. Organized disinformation tactics were used pre-Internet throughout the Cold War by the Soviet Union, United States, and their allies, and date back to the Bolshevik Revolution itself (Rid 2020). There have been (fiercely disputed) allegations of similar tactics by other states, such as China in Taiwan's 2020 elections (Aspinwall 2020). They are increasingly used by domestic political actors during election campaigns—which can be difficult to differentiate from foreign state influence operations (Aspinwall 2020). Group messaging tools owned by digitally dominant platforms, such as WhatsApp and Instagram, have also been used to spread electoral disinformation in countries including Brazil and India (see, e.g., Machado et al. 2018).

Since 2015, targeting of voters with disinformation worldwide has significantly increased (Communications Security Establishment of Canada 2019)—although the impact on electoral outcomes is still unclear. As Karpf noted: 'Generating social media interactions is easy; mobilizing activists and persuading voters is hard' (Karpf 2019). For instance, much of the disinformation shared by Russia before and during the 2016 elections focussed on stoking general partisan tensions, including building group identities. Rid (2020) found that

> [t]he [St Petersburg Internet Research Agency's] most engaged content, perhaps counterintuitively, was not designed to polarize but to build communities. The IRA's overall outreach on Facebook achieved approximately 12 million shares in the United States before Election Day in 2016, just under 15 million 'likes,' and just over 1 million comments. The majority of these interactions, however, happened with benign crowd-pleasing posts, not with the most polarizing and vile content.

Political parties and campaigning organizations make heavy use of voter data and social media to reach voters via direct marketing and targeted adverts (Select Committee on Democracy and Digital Technologies 2020). There has been a corresponding drive in parties and campaigns for detailed information about voters beyond that in electoral rolls. In some cases, such data been obtained or processed illegally (see, e.g., Information Commissioner's Office 2020a, 2020b), or used in the context of high levels of microtargeting online, resulting in illegal electoral overspend (BBC 2019).

Claims of disinformation must be expected from partisans at all stages of the electoral lifecycle, from voter registration to voting processes, to training of poll workers and observers, the location and timing of polling station opening, to vote tallying, to announcement of winners, to accounting for party and third-party finances, and even to post-election evaluation and proposals for reform. The losing party may call the integrity of the electoral management body (EMB) into doubt, and in sophisticated information campaigns, is met by similar calls from the winning party, to 'even the score'. The EMB is thus traduced on all sides. Without a robust independent media, public broadcast ethos, and political support, EMBs struggle under this onslaught.

Implementing best practices against electoral disinformation will require action by EMBs, data protection agencies, communications and media regulators, parliamentary authorities, and ministries of justice and equivalent (Brown et al. 2020). However, neither effective implementation nor a disinterested assessment of best practice can be guaranteed. Electoral laws are—like much history—written by the winners, often immediately after their victory. The UN Special Rapporteur also noted: 'disinformation is an extraordinarily elusive concept to define in law, susceptible to providing executive authorities with excessive discretion to determine what is disinformation, what is a mistake, what is truth' (Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression 2020 para. 42). Legal frameworks need to be updated as a response to disinformation challenges discovered during electoral processes, as well as encompassing international best practice (European Commission 2020).

ELECTORAL CAMPAIGNS, INTERFERENCE, AND DISINFORMATION

Digital political campaigning began in the 1990s as the World Wide Web popularized the Internet outside universities, with Canada and Singapore two of the first countries to deploy broadband on a large scale to the general public. High- and middle-income countries have since seen a huge growth in broadband Internet coverage, with the deployment of high-speed mobile networks and smartphone ownership in the past decade further impacting political and electoral information. Low-income countries have also seen significant increases in national communications infrastructure, international connectivity, and mobile phone penetration (ITU Telecommunication Development Bureau 2019).

For states, electoral disinformation is a complex multi-agency issue to regulate. Existing political coverage rules (for instance, requirements of impartiality and declarations of spending and origin of advertising) often only apply to political parties and the use of broadcast media, not print (newspaper), online or outdoor posters. Broadcast rules can apply to all broadcasting political coverage, with a 'fairness rule' and hate speech laws, with specific regulation of electoral periods for public service broadcasters. The sheer volume of social media posts has led many governments to adopt rules (such as Codes of Conduct; see Brown et al. 2009; Tambini et al. 2007) for the social media platforms on which posts, videos, and other content is shared, rather than quixotically pursuing the numerous and often anonymous posters of disinformation themselves.

The increase in advertising and content production both inside and outside electoral periods in online media is capable of causing disruption to existing electoral campaign rules, with so-called troll factories producing large volumes of often distorted or untrue posts which cannot be easily traced to any single source in domestic politics, and state actors even using unwitting local activists to promote them (Corera 2020; Special Counsel Robert S. Mueller III 2019). However, the impact of the notorious St. Petersburg 'Internet Research Agency' (IRA) may have been overstated, according to Rid (2020):

> It is unlikely that the trolls convinced many, if any, American voters to change their minds. . . . On Twitter, the IRA's impact practically vanished in the staggering number of election-related tweets. . . . The St. Petersburg troll den generated less than 0.05 percent of all election-related posts. The IRA, according to the data released by Twitter, boosted candidate Donald Trump's retweet count with only 860 direct retweets over the entire campaign.

Disinformation consumption in the United States (where most large-scale studies have so far taken place) appears to vary by political viewpoint, with one study finding '63% of all page-level traffic to untrustworthy websites observed in our data during the study period came from the 20% of news consumers with the most conservative information diets' (Guess et al. 2020). A second study examining American Twitter users in 2016 concluded: 'Only 1% of individuals accounted for 80% of fake news source exposures, and 0.1% accounted for nearly 80% of fake news sources shared. Individuals most likely to engage with fake news sources were conservative leaning, older, and highly engaged with political news' (Grinberg et al. 2019). Benkler (2019) summarized his research findings: 'Even where we did find traces of Russian origins in campaigns that did make it into the mainstream, the propaganda pipeline ran through Infowars, Drudge, and Fox News. That is, the primary vectors of influence were willing amplification of the information operations by the mainstays of the domestic American outrage industry'. He concluded: 'The crisis of democratic societies may be helped along by disinformation and propaganda, and certainly is fanned and harnessed by political opportunists, but it is fuelled by a decades-long extractive and distorted political economy'.

Given this ideological distribution, regulatory responses to disinformation have unsurprisingly became an intensely partisan affair in the United States, with the head of Facebook's Washington, DC, office reportedly telling colleagues: 'We can't remove all [false news] because it will disproportionately affect conservatives' (Timberg 2020). Another former Facebook employee (and current member of the UK legislature) has written that 'suppression of speech that is predominantly being made by one party in a contested space is an unavoidably partisan action', which social media companies are understandably eager to avoid (Allan 2020). But much of the short-term impact of

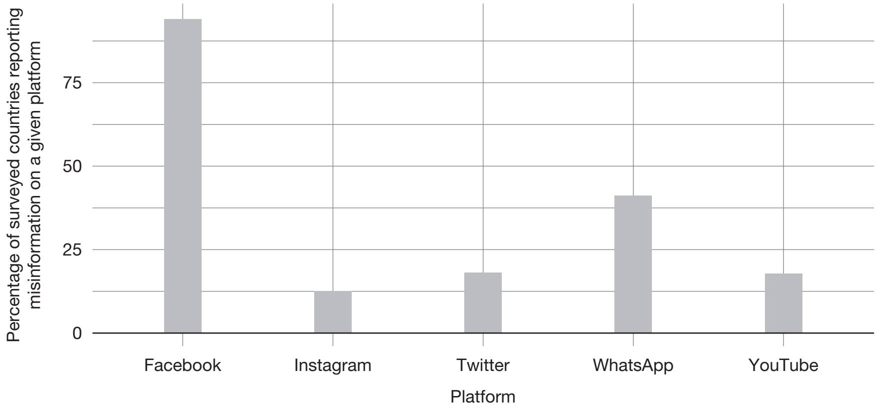


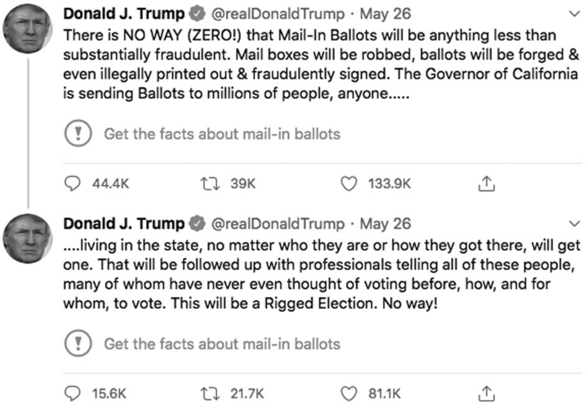
Figure 11.1 Reported cases of misinformation on social media platforms in surveyed Commonwealth countries.

disinformation comes when it is repeated by traditional media outlets and politicians, 'who are the primary sources of political news and commentary' (Weintraub 2020).

One broad challenge to regulating the intersection of modern campaigning, electoral integrity, and disinformation is the varying abilities of EMBs to monitor the online and platform environment, as 'there are many unfounded claims and opportunities for misinformation about the process of our elections' (Department of Homeland Security 2020). All the Commonwealth countries responding to our survey in mid-2019 have experienced the dissemination of misinformation in relation to their elections processes. Figure 11.1 shows a breakdown of the amount of reported cases per social media platform, showing the dominance of the Facebook properties in permissive misinformation.

The straightforward EMB response for high-profile false information relating to elections is to rapidly publicize corrections, using EMB websites and social media channels, as well as interviews on broadcast media and briefings for journalists. EMBs can also report clear instances of disinformation relating to elections—such as false information about the polling date or location of polling stations—to social media platforms, which will remove it where it breaches their terms and conditions.

Less clear-cut examples of electoral disinformation can be much more controversial to deal with. US President Donald Trump reacted furiously (Figure 11.2) when Twitter added a fact-checking link to his tweet claiming (with no evidence) that greater use of postal votes during the Covid-19 pandemic would lead to higher levels of electoral fraud: '@Twitter is now interfering in the 2020 Presidential Election'.[2] The tweets also contained a specific falsehood: that California was sending postal ballots to all residents 'no matter who they are or how they got there' (in reality, only to registered voters):



Donald J. Trump @realDonaldTrump · May 26

There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone.....

(!) Get the facts about mail-in ballots

44.4K 39K 133.9K

Donald J. Trump @realDonaldTrump · May 26

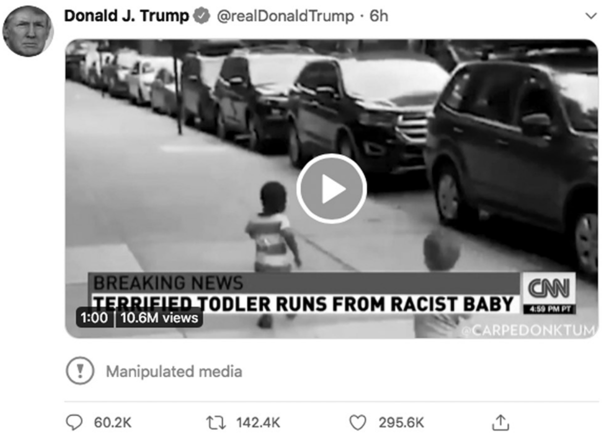
....living in the state, no matter who they are or how they got there, will get one. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way!

(!) Get the facts about mail-in ballots

15.6K 21.7K 81.1K

Figure 11.2 President's election Twitter: mail-in ballots

By producing its own 'fact-check' (which heavily quoted journalists' tweets), rather than relying on independent third-party fact-checkers citing more in-depth analysis (as Facebook policy held, while exempting posts from politicians), Twitter and its staff became targets for highly critical responses from President Trump and his supporters. This continued when a Trump tweet was labelled by Twitter as containing 'manipulated media' for containing video falsely edited to look as if it had been broadcast by CNN (Figure 11.3):



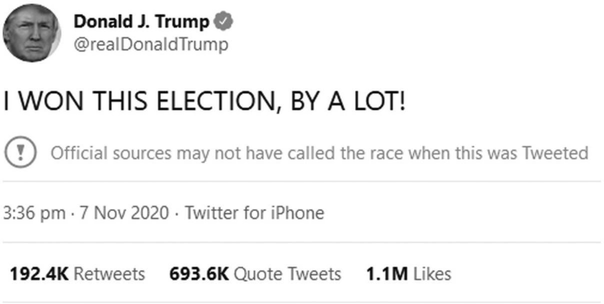
Donald J. Trump @realDonaldTrump · 6h

(!) Manipulated media

60.2K 142.4K 295.6K

Figure 11.3 President's election Twitter: false news
Source: Conger, Kate. 2020. 'Twitter Labels Trump Tweet about "Racist Baby" as Manipulated Media'. *New York Times*.

Trump also had his tweets doubting the integrity of the vote count in the 2020 presidential election flagged by Twitter (Figure 11.4), despite which it attracted over a million 'Likes':



Donald J. Trump
@realDonaldTrump

I WON THIS ELECTION, BY A LOT!

Official sources may not have called the race when this was Tweeted

3:36 pm · 7 Nov 2020 · Twitter for iPhone

192.4K Retweets **693.6K** Quote Tweets **1.1M** Likes

Figure 11.4 President's election Twitter: election integrity

More broadly than disputed tweets or social media in a hotly disputed election, Edelman explained that: 'political truth is a difficult thing to pin down, because politics is fundamentally about convincing other people to accept your interpretation of reality' (Edelman 2020). Candidates are likely to always challenge the media's framing of news as the first rough draft of history.

An emerging legal battleground in this area is whether there should be obligations on social media firms to carry certain speech. In Poland, a recent bill published by the Ministry of Justice would lead to moderation decisions by platforms being overseen by a 'Freedom of Speech Council'. However the appointment and oversight processes of this council do not guarantee independence and lend themselves to being populated by the ruling party of the day, and while its decisions can be appealed, they can only be appealed to administrative courts who do not typically deal with, or understand, the complex area of governing freedom of expression (Panoptkyon Foundation 2021). Over the pond, in a recent opinion, US Supreme Court Justice Clarence Thomas has suggested that social media firms should have common carry obligations and restricted First Amendment rights, indicating the potential for future flashpoints of this type there too.[3]

INTERNET 'SWITCH-OFF' AND DISINFORMATION LAWS

Internet shutdowns (general removal of transit so that all services are restricted by telecoms companies on order of the government) have been resorted to by governments in the immediate election and vote counting period (see, e.g., RSF 2019). There were 213 documented incidents of full and partial closure in 2019 alone (Access Now 2020). The African Commission on Human and Peoples' Rights in 2016 noted 'the emerging practice of State Parties of interrupting or limiting access to telecommunications services such as the Internet, social media and messaging services, increasingly during elections' (African Commission on Human and Peoples' Rights 2016). Fuller reported: '[i]n Zimbabwe, following sporadic internet blackouts in the midst of civilian protests in January 2019, the country's High Court issued a ruling in which it declared the shutdown illegal and ordered telecom operators to restore access' (Fuller 2019). Purdon et al. (2015) found 'Pakistan has often instructed telecommunication operators to suspend mobile and/or Internet networks where intelligence indicates a threat to national security'.

General Internet shutdowns are contrary to international standards. The UN Human Rights Council (2016) stated that it

> condemns unequivocally measures to intentionally prevent or disrupt access to or dissemination of information online in violation of international human rights law and calls on all States to refrain from and cease such measures.

A 2015 Joint Declaration by global and regional human rights bodies stated Internet 'kill switches' can never be justified under international human rights law, even in times of conflict (United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al. 2015 para. 4(c)).

Colonial-era criminal defamation laws were used in the period before the Commonwealth in response to claimed hate speech and to prevent opposition to colonial government. These have continued in many Commonwealth jurisdictions. However, in 2019, several countries passed antidisinformation and hate speech laws for online media, extending controls into the Internet environment. Singapore in May 2019 passed the Protection from Online Falsehoods and Manipulation Act 2019 (POFMA). It gives ministers powers to require online actors to remove disinformation, and regulators to stop access to Internet providers in Singapore that continue to carry such messages. Part 2 of POFMA criminalizes the communication of false statements of fact in Singapore in certain circumstances, and acts which enable or facilitate the communication. Section 7 provides that a person must not do any act in or outside Singapore in order to communicate in Singapore a statement knowing or having reason to believe that it is a false statement of fact that may affect political stability. Individuals who contravene section 7(1) face a fine of up to $50,000 and/or imprisonment for up to 5 years. Organizations face a fine of up to $500,000. The punishment is enhanced if an unauthentic online account or a bot is used to communicate the statement and for the purpose of accelerating the communication. Under Part 3, 'the Minister may direct the Infocomm Media Development Authority of Singapore (IMDA) to order an internet access service provider (ISP) to take reasonable steps to disable local access to the online location where the false statement of fact is communicated' (Chng 2019).

In the context of recent laws, it is important to consider their impacts in a framework of freedom of expression and human rights more generally. A joint declaration from the freedom of expression rapporteurs of several intergovernmental organizations, in collaboration with international civil society groups, called for the abolition of criminal defamation laws and the wholesale avoidance of general prohibitions on disinformation.[4] The UN Human Rights Committee, established by the International Covenant on Civil and Political Rights, emphasizes in General Comment No. 34 that restrictions on speech online must be strictly necessary and proportionate to achieve a legitimate purpose. The 2017 Joint Declaration by global and regional human rights bodies notes the existence of

> attempts by some governments to suppress dissent and to control public communications through such measures as:
>
> - repressive rules regarding the establishment and operation of media outlets and/or websites;
> - interference in the operations of public and private media outlets, including by denying accreditation to their journalists and politically motivated prosecutions of journalists;
> - unduly restrictive laws on what content may not be disseminated;
> - the arbitrary imposition of states of emergency;
> - technical controls over digital technologies such as blocking, filtering, jamming and closing down digital spaces; and
> - efforts to 'privatise' control measures by pressuring intermediaries to take action to restrict content. (UN Special Rapporteur on Freedom of Opinion and Expression et al. 2017)

Responses that seek to generally censor the Internet or even shut it down during elections may be disproportionate as well as illegal under international law. The Commonwealth, for example, has already concluded that direct government regulation is seen as censorship, and is not the best practice to respond to potential social media disinformation (Commonwealth Electoral Network 2016, 2). Suspending social media platforms during elections can potentially impact large numbers of voters, whose wider communication could be jeopardized by such a restriction (for instance, suspending WhatsApp or Skype, which are vital communications tools for users).

More proportional practices in democratic elections are ensuring that EMBs can liaise with social media platforms to remove and counter deliberate disinformation regarding electoral registration and voting, and ensuring that claims about disinformation that form hate speech, defamation, or fraud are promptly dealt with by the independent judiciary. Political name-calling can be classified by political opponents as disinformation or hate speech, which is one reason for the continued role of the independent judiciary as the arbiter of such decisions.

REGULATING THE USE OF SOCIAL MEDIA FOR ELECTORAL DISINFORMATION

We now detail state responses to those—typically GAFAM—self-regulatory processes for disseminating and/or controlling disinformation that we explained in the previous section. Many proposed approaches to tackling disinformation issues would extend broadcast rules to non-broadcast content, whether text-based or in any case at the user's individual choice. Yet care must be taken here, as this would, in all likelihood, increase the concentration of online communication in the hands of the largest platforms that can employ economies of scale in deploying proprietary filters to remove harmful content.

Digitally dominant platforms are essential actors in regulation of disinformation (GAFAM is the acronym used to denote Google/Apple/Facebook/Amazon/Microsoft-owned platforms and other entities). Google, Facebook ,and Twitter have deployed AI tools at large scale to combat disinformation, claiming this is the only cost-effective response to the billions of messages passed across their platforms daily (Marsden and Meyer 2019a)—particularly when their content moderation staff cannot easily work remotely during the 2020 Covid-19 pandemic, due to the general insecurity of home offices. Opinions are divided over whether regulating such platforms is a legitimate point of intervention, in particular because this could lead to two types of content moderation 'arms race':

- in reporting disinformation, where trolls are as likely to overwhelm well-meaning citizens when each reports against the other;
- in coding debates, so that fact checkers and other self-regulatory enforcers cannot control the amount of disinformation as it emerges in images and videos as well as text.

Examples of these content moderation arms races from the Internet's regulatory history include the attempts to prevent child abuse image and terrorist video distribution, as well as unauthorized sharing of copyrighted files. In each case, the use of technologies (such as comparing hash values) in theory permitted removal before publication by the platforms deploying the technology, specifically YouTube and Facebook. In practice, the proliferation of content was restricted but by no means prevented by such technological intervention.

The use of AI and machine learning to detect content has seen success in some areas, but struggles heavily in areas which are as value-laden, subjective, and complex as disinformation (Marsden and Meyer 2019b). Social media platforms have claimed that AI will be able to spot disinformation. It is broadly the case that disinformation cannot be effectively automatically detected by new techniques such as machine learning, as it is highly context-specific, and there is no clear canonical reality against which to judge.

Automated filtering is a heavy-handed tool, and will result in a large number of 'false positives', where bona fide statements are confused with 'fake news'. These open up new cybersecurity threats, as machine learning systems are capable of being fooled and 'poisoned', for example by political actors wishing to suppress the speech of specific opponents' voices (Biggio and Roli 2018).

Canada has focussed on traceability of political advertising, to ensure transparency in the advertising spend by major parties and to prevent violations of campaign finance laws by 'shadow' advertising by groups closely associated with political causes or parties. In Canada in 2019, 'online platforms that accept political advertising in Canada will be required to show more transparency than they have in the past' (Thompson 2019). While Facebook and Twitter complied with these rules, Google chose instead to prohibit Canadian political advertising (Boutilier 2019).

The most detailed state regulation of political advertising on social media platforms found in our survey was in India, where political parties must get approval for ads from a committee organized by the EMB. The committee provides a QR code that must be included in an approved ad, and which is checked by online platforms. The committee checks the ad content, and that the publisher is certified, and also logs the price paid for the ad, which is made publicly available.

Politicians must follow a code of conduct, and the EMB maintains a public website listing actions taken against violators. A 150-person Electronic Media Monitoring Committee monitors media articles during elections, and transmit specific articles to districts to check. An EMB app also allows citizens to report code of conduct violations, with a photo and location; districts must deal with these reports within 100 minutes. Platforms are required to take down illegal content notified to them within 15 minutes.

The Election Commission of India (ECI) convened a meeting with representatives of social media platforms and the Internet and Mobile Association of India (IAMAI) preceding the May 2019 general elections. Social media platforms submitted a 'Voluntary Code of Ethics for the 2019 General Election' (Press Information Bureau 2019). Platforms voluntarily undertook to create a dedicated reporting mechanism for the ECI, create fast response teams to take action on reported violations, and facilitate political advertisement transparency. The mechanism allows ECI to notify platforms of violations under s. 126, Representation of the People Act 1951. In the event of conflict between the Voluntary Code of Ethics and the legal framework, the latter prevails. Platforms must take down reported content within three hours, during the two-day non-campaigning 'silence period' before polling. Platforms provide reports to IAMAI and ECI on their actions.

Disinformation during elections in South Africa is regulated by Section 89(2)(c) of the Electoral Act and Item 9(1)(b) of the Electoral Code of Conduct, which prohibits a false statement of fact, and not the expression of comments and ideas. These issues were tested in the Constitutional Court in a case concerning a text message sent by a political party to 1.5 million citizens in 2014, concerning allegations of corruption about then-President Zuma.[5] The text was found to be permitted electoral communication, and not prohibited by Section 89(2). There is also a defamation offence, which has led to recent jurisprudence requiring removal of false online content.[6]

South Africa's EMB noted in 2016 the growth of online disinformation. The Directorate of Electoral Offences was established ahead of the 2016 Municipal Elections to investigate alleged breaches of the Code of Conduct and prohibited conduct. To help distinguish between official and fake adverts, political parties contesting the May 2019 elections were asked to upload all official advertising material used by the party to an online political advert repository (http://www.padre.org.za). Complaints relating to alleged breaches of the Code of Conduct must be submitted to the Electoral Court or the Directorate for Electoral Offences. In August 2019, the number of complaints and the success rate in examination were not evaluated. In addition, the Electoral Commission launched an innovative online reporting platform for citizens to report instances of alleged digital disinformation, the 411 Campaign (Electoral Commission of South Africa n.d.). '411' is Internet slang in southern Africa for disinformation. Developed in conjunction with Media Monitoring South Africa, the platform provided for the online submission and tracking of complaints relating to disinformation encountered on social media platforms, hosted on www.real411.org. The digital platform was intended for complaints related only to social media and not to replace existing channels and processes for investigating alleged breaches of the Code of Conduct. 156 complaints were logged by election day, to be considered by a panel of relevant experts including those with expertise in media law and social and digital media.[7] They will make recommendations for possible further action, which could include:

- Referring the matter for criminal or civil legal action;
- Requesting social media platforms to remove the offensive material;
- Issuing media statements to alert the public and correct the disinformation.

The UN Special Rapporteur has specifically argued that:

> a. General prohibitions on the dissemination of information based on vague and ambiguous ideas, including 'false news' or 'non-objective information', are incompatible with international standards for restrictions on freedom of expression, and should be abolished.
> b. Criminal defamation laws are unduly restrictive and should be abolished. Civil law rules on liability for false and defamatory statements are legitimate only if defendants are given a full opportunity and fail to prove the truth of those statements and also benefit from other defences, such as fair comment.
> c. State actors should not make, sponsor, encourage or further disseminate statements which they know or reasonably should know to be false (disinformation) or which demonstrate a reckless disregard for verifiable information (propaganda).
> d. State actors should, in accordance with their domestic and international legal obligations and their public duties, take care to ensure that they disseminate reliable and trustworthy information, including about matters of public interest, such as the economy, public health, security and the environment. (UN Special Rapporteur on Freedom of Opinion and Expression et al. 2017)

There is scope for standardizing (the basics of) notice and appeal procedures and reporting, and creating a self-regulatory multi-stakeholder body, such as the UN Special Rapporteur's suggested 'social media council' (Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression 2018 paras 58, 59, 63, 72). Such a multi-stakeholder body could, on the one hand, have competence to deal with industry-wide appeals and, on the other hand, work towards a better understanding and minimization of the effects of AI on freedom of expression and media pluralism.

ONLINE BEHAVIOURAL ADVERTISING AND DISINFORMATION

Targeted online behavioural advertising (OBA)—shown to Internet users based on profiles of their previous online and increasingly offline behaviour and characteristics—may use both legitimate and illegally obtained data sources. Disinformation threats may seek to suppress or increase voter motivation in specific targeted segments of the population by geography or expressed political motivation—so-called microtargeting to 'fire up the base' (motivate) or to suppress voter turnout via demotivational messages. The algorithms used to select which adverts are shown to social media users often promote adverts that users are more likely to click on—favouring emotional and partisan appeals even at a lower bidding price by advertisers.[8]

The European Union and the major global social media platforms have co-operated to create a new Code of Conduct to combat electoral disinformation via this route. The European Union was the first multilateral organization to develop a response to disinformation, investing very substantially in research and then regulation in the period since 2014 (Meyer et al. 2020). The EU-orchestrated multi-stakeholder forum industry self-regulatory 'Code of Practice on Online Disinformation' was intended to demonstrate the voluntary commitments of the major social media platforms to achieve greater transparency in political advertising, prior to the European Parliament elections of May 2019 (European Commission 2018). This was the world's second-largest democratic election after India's 2019 parliamentary election.

The Code of Practice includes commitments relating to scrutiny of ad placements, political and 'issue-based' advertising, integrity of services, empowering users, and empowering the research community. Part of the industry response to the Code of Practice concerned rectifying the limited access platforms provide to political advertisements using their systems. In April 2021, the European Union committed to a revision of the Code, demonstrating the extent to which it is heavily guided and sponsored self-regulation: '[our] assessment found: inconsistences in application; gaps in the scope; lack of common definitions; lack of key indicators and independent oversight. This initiative issues guidance on how to better apply the code' (European Commission 2021). Explicitly paid for political advertisements are increasingly placed in online 'ad archives', such as those provided by Facebook and by Google (Facebook n.d.; Google n.d.). The main intention of these systems is to allow civil society actors and regulators to identify and audit the political advertising spend by actors deemed political by the platform. Users themselves can access such an archive, but the information in the archive is not currently presented to them when, for example, they browse a site and view an ad.

The Mozilla Foundation has proposed, along with over seventy researchers, standards for effective political advertising archives that should be enforced upon platforms (Mozilla 2017). Their recommendations include that the ad archive should be comprehensive, including 'direct electioneering content, candidates or holders of political office, matters of legislation or decisions of a court, functions of government'. They also argue that the ad archive should 'provide information about targeting criteria and information about impressions, content, payment, and microtargeting features' and 'support research, by allowing bulk access and download and persistent, well-documented meta-data . . . both up-to-date and historical data'. The advertising archive should be accessible to the public. A consistent challenge is ensuring that companies deliver workable advertising archives, such as those in line with the above guidelines. In the European Union, Facebook's attempt to create such a system has been described as 'inadequate', pointing to challenges in enforcement more broadly (Mozilla 2019).

NATO has reported the continued need for EMB and wider government readiness against disinformation threats (NATO StratCom COE 2018). In a cybersecurity context, it points to the large and changing 'scale of the black-market infrastructure for developing and maintaining social metric manipulation software, generating fictitious accounts, and providing mobile proxies and solutions for SMS activation' (NATO StratCom COE 2019, 16). These systems rely on security loopholes, data breaches, and the use of bots at scale in order to influence disinformation on a large scale.

The European Commission, announcing its European Democracy Action Plan in December 2020, stated it: "will steer efforts to overhaul the Code of Practice on Disinformation into a co-regulatory framework of obligations and accountability of online platforms, in line with the upcoming Digital Services Act. To that end, the Commission will issue guidance to enhance the Code of Practice in spring 2021 and set up a more robust framework for monitoring its implementation." (European Commission, 2020, point 3)

DISINFORMATION AND DATA PROTECTION POLITICAL EXEMPTIONS

Privacy and data protection legislation is a key component of electoral integrity in the complex ecosystem of behavioural advertising data sharing and brokerage that has developed over the last two decades. Significant damage to perceived electoral integrity can be done if a party, campaign, or candidate misuses data to manipulate voters via targeted disinformation. The European-style data protection laws implemented in at least 142 countries usually require a regulator which is truly independent from government, and which has powers and resources effective for and commensurate with its role (Greenleaf and Cottier 2020).

Data protection laws commonly provide higher protection for 'sensitive' or 'special category' data commonly used in electoral processes, such as *data revealing political opinions or affiliations*.[9] Such restrictions tend to have exemptions for electoral processes, but these exemptions must be balanced against the need for high protection, risk assessment, and scrutiny. A report commissioned by the UK data protection regulator concluded:

> To the extent that contemporary elections are 'data-driven', their worst effects have been apparent in countries whose data protection laws do not cover political parties. In most democratic countries where parties are covered by data protection law, and have been for decades, there is little evidence that these restrictions have impeded their ability to perform their basic

> democratic roles of political mobilization, elite recruitment and policy development. (Bennett and Oduro Marfo 2019)

Depending on the local data protection or privacy regime, some organizations involved in elections may be allowed to process these specific categories of data with lower restrictions than other actors, or even be out of scope of the law entirely. In the United Kingdom, for example, political parties are bound by the requirements of the Data Protection Act 2018, although registered parties benefit from being able to process data on political opinions without relying on consent (with voters allowed to opt out in writing from processing by specific parties and campaigns).[10] A similar framework can be seen in South Africa's Protection of Personal Information Act 2013.[11] South Africa's Electoral Commission is in discussions with the national Information Commissioner about protection of electoral data, with public concerns expressed about commercial marketers and debt collectors accessing voter records. Political parties want access to the full voter list to verify voters.

In Malta, the Office of the Data Protection Commissioner has stated that political parties must get consent before processing political opinions (Pace 2018). In Australia, political parties are not considered as organisations for the purposes of privacy law,[12] and other organizations undertaking political activities, such as parties' (sub-)contractors and volunteers, are also exempted.[13] In Canada, the political parties 'fall between the cracks' of the national privacy regime (Bennett and Bayley 2012), as they are not governmental institutions for the purposes of public sector privacy law,[14] and are exempted from federal private sector privacy legislation by virtue of not meeting the definition of 'federal work, undertaking or business'.[15] Some issues around the use of the electoral roll are regulated by electoral law, but the application and scope of this is inconsistent and patchy (Bennett and Bayley 2012).

In addition to data protection law, many countries have related provisions that focus on unsolicited and direct marketing. These fall in different types of law depending on the jurisdiction. In the Commonwealth's EU members (Cyprus and Malta and, until/unless it modifies its EU-based law, the United Kingdom), these follow the e-Privacy Directive,[16] which in large part focusses on implementing the fundamental right to confidentiality of communication (Zuiderveen Borgesius and Steenbruggen 2019). In Canada, this issue became controversial with the illegal impersonation of the EMB, in the 2011 elections, by a campaigner for a major party using automated calling (Brown et al. 2020, 31–32).

While much of the EU policy debate relating to platform regulation takes place in narrow competition economics terms, it could have a much broader impact on European societies, given the increasing use of major Internet platforms as essential 'social infrastructure'—used by families to share news and photos; schools to communicate with parents; sports teams to arrange games; politicians to communicate with constituents; campaign groups to organize protests; and many other aspects of modern-day life (Brown and Marsden 2013). It is no longer easy for many individuals to 'opt out' of using such platforms.

Many issues around elections, such as the use of data on social media platforms or in the advertising technology domain, span borders and jurisdictions. They cannot be tackled on the domestic level alone. Regulators must therefore be part of global and regional groupings to share information and build a coherent strategy for international challenges.

TEN RECOMMENDATIONS FOR FURTHER REGULATORY ACTION AND FORBEARANCE

The starting point for democratic governments in dealing with electoral disinformation must always be the importance of freedom of expression, especially for political speech, and compliance with international human rights law. Special rapporteurs of the United Nations (UN), Organization for Security and Co-operation in Europe (OSCE), and Organization of American States(OAS) highlighted in a joint statement in 2020 'the essential role that freedom of expression and information, free, independent and diverse media and a free and accessible Internet play in ensuring free and fair elections, including referenda, in particular by informing the public about parties and candidates and their platforms'. They also expressed their alarm at 'the misuse of social media by both state and private actors to subvert election processes, including through various forms of inauthentic behaviour and the use of "computational propaganda" ' (UN Special Rapporteur on Freedom of Opinion and Expression et al. 2020).

At this relatively early stage of large-scale use of online tools of political persuasion, there is significant uncertainty about the overall impact of disinformation on election outcomes. Karpf notes: 'There is no evidence that psychographic targeting actually works at the scale of the American electorate, and there is also no evidence that Cambridge Analytica in fact deployed psychographic models while working for the Trump campaign. The company clearly broke Facebook's terms of service in acquiring its massive Facebook dataset. But it is not clear that the massive dataset made much of a difference' (Karpf 2019). Rid (2020) concluded:

> On Twitter, the IRA's impact practically vanished in the staggering number of election-related tweets. Approximately 1 billion tweets related to the campaigns were posted during the fifteen months leading up to the election. The St. Petersburg troll den generated less than 0.05% of all election-related posts. The IRA, according to the data released by Twitter, boosted candidate Donald Trump's retweet count with only 860 direct retweets over the entire campaign.

Researchers will need much greater access to data held by social media platforms to assess (with any level of confidence) the impact of disinformation on users' political opinions, activities, and voting processes, including failure to vote.

Given human rights imperatives, and uncertainties about the overall effect of electoral disinformation, policymakers must be extremely cautious in legislative and other responses, recognising also the unpredictable outcomes of regulation on other areas of freedom of expression, from broadcasting and newspaper publishing, to advertising regulation, and to disinformation in less overtly political arenas, including (topically in 2020) public health pandemics. Our ten recommendations for policymakers take account of these imperatives and uncertainties:

1. Electoral management boards (EMBs) should not request the operation of Internet shutdowns during election periods, or at any other point not objectively assessed a national emergency and sanctioned by a superior court. Such an injunction may be achieved with great speed, and the need for procedural legitimacy before such an extreme response is in our view essential.
2. Governments should avoid Internet shutdowns as a response to disinformation concerns, while ensuring false announcements are responded to where defamatory, fraudulent, or unjustifiably casting doubt on official EMB results and guidance. We acknowledge that heads of state may seek faster executive redress, but also that such action may not appear disinterested or legitimate prior to a court decision.
3. Disinformation is best tackled by governments through media pluralism and literacy initiatives, as these allow diversity of expression and choice. Governments should encourage social media platforms to consider the use of source transparency indicators, and highlighting and deprioritization of disinformation identified by independent fact-checkers. Users need to be given the opportunity to understand how their search results or social media feeds are built, and edit their search results/feeds where desirable.
4. Freedom of expression as a fundamental right should be protected from automated private censorship by dominant social media platforms, and thus disinformation should be regulated by legislation with independent appeal to courts of law. Options to ensure independent appeal and audit of platforms' regulation of their users should be introduced. When technical intermediaries need to moderate content and accounts, detailed and transparent policies, notice and appeal procedures, and regular reports are crucial (see, further, Meyer et al. 2020).
5. Governments should consider legislating to ensure that platforms and advertising networks are obliged to make political ads public, in line with best practices in the area which allow public research and scrutiny. They should also consider requiring major platforms to develop privacy-protective mechanisms by which independent researchers can investigate the societal impact of disinformation (and other phenomena).
6. Governments may be aided by a template agreement with social media companies for memoranda of understanding relating to disinformation, potentially based on the EU Code of Practice or India's example; and should make public such agreements.

7. Governments should strengthen reporting and publication of political spending online as well as offline, with independent electoral authorities monitoring all donations and uses of 'dark money' to try to influence campaigns.
8. Governments should ensure privacy and data protection laws are in place to protect voter data wherever it is held, including in the private sector. These laws should allow political parties and candidates to engage with voters; but any exemptions that affect voters' trust or data protection and security should be carefully limited. States without a data protection or privacy law should look to establish one in line with existing international standards and institutional practices.
9. Personal data and privacy issues around elections should be overseen by a regulator which is truly independent from government, and which has powers and resources effective for and commensurate with its role.
10. Governments should support accurate, independent identification and attribution of disinformation, working with neutral fact-checking organizations.

Our first three recommendations may be seen as an expression of the obvious, but media literacy and preventing Internet switch-offs are not yet accepted best practices in all parts of the world. The UN/OSCE/OAS special rapporteurs have expressed concern that 'many States are passing laws which . . . unduly limit freedom of expression, expand State control over the media, restrict Internet freedom and/or further the ability of various actors to collect personal data'. They deplore 'restrictions on the ability of the public to access the Internet, including complete or partial shutdowns, which seriously limit the ability of media, parties, candidates and others to communicate with the public, as well as the ability of members of the public to access information' (United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al. 2020, 2).

Note the reference to false announcements, far too common a claim on results day and thereafter in many nations. Both could be abused by governments and candidates lacking integrity, as seen in the Trump tweet. The UN/OSCE/OAS special rapporteurs recommend 'States should adopt appropriately clear and proportionate laws that prohibit the dissemination of statements which are specifically designed to obstruct individuals' right to vote, such as by intentionally spreading incorrect information about where or when to vote' (United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al. 2020, para. 1(c)(ii)).

Recommendation 5 has the regulatory aim of forcing social media platforms to display sources of all advertising, as third-party advertising may influence voters significantly even on single issues that later define their voting behaviour (as in the 2018 Irish abortion referendum), while recommendation 7 will ensure further transparency of the funders of political advertising. The UN/OSCE/OAS special rapporteurs have stressed 'the need for robust rules and systems requiring transparency of parties and candidates in relation to media spending on elections' (United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al. 2020, 2).

Recommendation 10 emphasizes the importance of independent assessment of the accuracy or otherwise of alleged disinformation, and identification of its origins. Claims and counterclaims about the accuracy of political statements can quickly become highly partisan in electoral contents, while the secrecy of (some) state information operations can make them difficult to identify. An inaccurate claim of disinformation by Taiwan's government in 2019 'gave the [opposition] KMT and its supporters the opportunity to question the [governing] DPP's credibility when it discusses Beijing's desire to meddle in Taiwanese politics'. The chair of Taiwan's National Communications Commission (NCC) resigned in April 2019, complaining: 'If there is any website, any TV station they report to be untrue or disinformation, [the governing party] want the NCC to punish them or block the website. For me, it is impossible to do that. That is not the responsibility of the NCC' (Aspinwall 2020).

Recommendation 5 also emphasizes the need for much better-quality data—currently held by the platforms—to be available to independent researchers, so that societies can gain a deeper understanding of how disinformation affects democracy, and what would therefore be proportionate responses. As Rid states: 'At-scale disinformation campaigns are attacks against a liberal epistemic order, or a political system that places its trust in essential custodians of factual authority'. Benkler concludes: 'No specific electoral outcome better serves Russia's standard propaganda agenda—disorientation and loss of trust in institutions generally—than the generalization of disorientation and de-legitimation of institutions represented by the present pitch of the debate over election interference' (Benkler 2019). This delegitimation effect has already been seen in Taiwan, where a 2019 'anti-infiltration law' was boycotted by the opposition in parliament but passed unanimously by the governing party.

Our recommendations are framed with a view to the different regulatory traditions inside national electoral frameworks—crudely Anglo-American (Westminster system) self-regulation, European co-regulation, and state regulation (Marsden et al. 2020). Recommendation 6 calls for explicit terms of engagement between social media platforms and governments, not necessarily for legislation such as that found in Singapore, Germany, and France.

It is in the enforcement of disinformation laws that we are most likely to see the regulatory outcomes diverge between different regulatory and electoral-political traditions, and research into the institutional models for regulating disinformation is urgently required. Researchers need to anchor disinformation research in national responses to digitally dominant platforms across electoral, media and communications and data protection laws. Moreover, disinformation regulation must be responsive to the international human rights law standards of legality, necessity, and proportionality.

The United Nations Special Rapporteur quotes Hannah Arendt on the totalitarian dangers of disinformation: 'If everybody always lies . . . nobody believes anything any longer' (Arendt 1978; Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression 2020, paras 58–59). The need to accurately analyse responses to disinformation will remain pressing in this increasingly GAFAM-dominated age.

NOTES

1. The EU's interinstitutional terminology database IATE (Inter-Active Terminology for Europe) specifically notes that disinformation should not be confused with misinformation, defined in IATE as 'information which is wrong or misleading but not deliberately so' (Bentzen 2015).
2. The president posted the Tweet at 1240 am at https://twitter.com/realdonaldtrump/status/1265427538140188676; it provoked 81,000 retweets and 223,000 "likes" (archived at https://web.archive.org/web/20200528040606/https://twitter.com/realdonaldtrump/status/1265427538140188676).
3. *Biden v. Knight First Amend. Inst. at Columbia Univ.*, No. 20-197, 2021 WL 1240931 (U.S. Apr. 5, 2021).
4. In very narrow specific circumstances pertaining to judicial reputation, criminal defamation with a financial penalty rather than imprisonment has been considered appropriate in the European Court of Human Rights—see *Peruzzi v. Italy* (App no 39294/09) judgment of June 30, 2015.
5. *Democratic Alliance v. African National Congress and Another* (CCT 76/14) [2015] ZACC 1.
6. *Trevor Manuel v Economic Freedom Fighters and Others* ([2019] ZAGPJHC 157) Johannesburg High Court, May 30.
7. Retrieved from data at https://www.real411.org/complaints.
8. Facebook disputed claims from the Trump 2016 presidential campaign their advertising costs were consequently much lower than Hillary Clinton's. See Breland (2018).
9. See, e.g., Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, 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), OJ 2016 L 119/1, art 9.
10. Data Protection Act 2018 (United Kingdom), sch 1 para 22.
11. Protection of Personal Information Act 2013 (South Africa) s 31.
12. Privacy Act 1988 (Australia), s 6C(1).
13. Privacy Act 1988 (Australia), s 7C.
14. Privacy Act 1982 (Canada), s 3.
15. Personal Information Protection and Electronic Documents Act S.C. 2000, c. 5 (Canada), s 4(1).
16. Directive 2009/136/EC of the European Parliament and of the Council of November 25, 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws, OJ L 337, 18.12.2009, 11–36.

REFERENCES

Access Now. 2020. 'Targeted, Cut Off, and Left in the Dark: The #KeepItOn Report on Internet Shutdowns in 2019'. *The #KeepItOn Coalition*. https://perma.cc/S7EL-37LE.

African Commission on Human and Peoples' Rights. 2016. 'Resolution on the Right to Freedom of Information and Expression on the Internet in Africa—ACHPR/Res.362(LIX)2016'. Retrieved October 18, 2020. https://www.achpr.org/sessions/resolutions?id=374.

Allan, Richard. 2020. 'Partisan—The Missing Word'. regulate.tech Blog26 April. Retrieved October 18, 2020. https://www.regulate.tech/partisan-the-missing-word-26th-april-2020/.

Arendt, Hannah. 1978. 'Hannah Arendt: From an Interview 26 October'. *New York Review of Books*.

Aspinwall, Nick. 2020. 'Taiwan's War on Fake News Is Hitting the Wrong Targets'. *Foreign Policy,* 10 January.

Barwise, T. P., and L. Watkins. 2018. 'The Evolution of Digital Dominance: How and Why We Got to GAFA' pp. 21–49 in Digital Dominance: The Power of Google, Amazon, Facebook, and Apple. Oxford University Press, New York, NY. Edited by Moore Martin and Tambini Damian. doi: 10.35065/PUB.00000914.

BBC. 2019. 'Brexit: Vote Leave Drops Appeal against Referendum Spending Fine'. *BBC News*. Retrieved October 18, 2020. https://www.bbc.co.uk/news/uk-politics-47755611.

Benkler, Yochai. 2019. 'Cautionary Notes on Disinformation and the Origins of Distrust'. *MediaWell (Social Science Research Council)*. https://mediawell.ssrc.org/expert-reflections/cautionary-notes-on-disinformation-benkler/.

Bennett, Colin, and Robin M. Bayley. 2012. 'Canadian Federal Political Parties and Personal Privacy Protection: A Comparative Analysis'. *Office of the Privacy Commissioner of Canada*. Retrieved May 12, 2019. https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2012/pp_201203/#toc5.

Bennett, Colin, and Smith Oduro Marfo. 2019. 'Privacy, Voter Surveillance and Democratic Engagement: Challenges for Data Protection Authorities'. Wilmslow: Information Commissioner's Office. doi: 10.2139/ssrn.3517889.

Biggio, Battista, and Fabio Roli. 2018. 'Wild Patterns: Ten Years after the Rise of Adversarial Machine Learning'. *Pattern Recognition* 84: 317–31. doi: 10.1016/j.patcog.2018.07.023.

Boutilier, Alex. 2019. 'Twitter Announces Rules for Canadian Political Advertising'. *The Star*. Retrieved October 18, 2020. https://www.thestar.com/politics/federal/2019/08/29/twitter-announces-rules-for-canadian-political-advertising.html.

Breland, Ali. 2018. 'Facebook Says Trump Paid More Than Clinton for Digital Advertising'. *The Hill*. Retrieved October 18, 2020. https://thehill.com/policy/technology/375915-facebook-says-trump-paid-more-than-clinton-for-digital-advertising.

Brown, Ian, Lilian Edwards, and Christopher T. Marsden. 2009. 'Information Security and Cybercrime'. In *Law and the Internet*, edited by Edwards Lilian and Waelde Charlotte. Hart: Oxford.

Brown, Ian, and Christopher T. Marsden. 2013. *Regulating Code: Good Governance and Better Regulation in the Information Age*. Cambridge, MA: MIT Press.

Brown, Ian, Christopher T. Marsden, James Lee, and Michael Veale. 2020. *Cybersecurity for Elections: A Commonwealth Guide on Best Practice*. London: Commonwealth Secretariat. doi: 10.14217/e56e4289-en.

Chng, Darren Grayson. 2019. 'POFMA: Singapore's Anti-Fake News Law'. *Society for Computers and Law*. Retrieved October 18, 2020. https://www.scl.org/articles/10541-pofma-singapore-s-anti-fake-news-law.

Commonwealth Electoral Network. 2016. *New Media and the Conduct of Elections*. London: Commonwealth Secretariat.

Communications Security Establishment of Canada. 2019. '2019 Update: Cyber Threats to Canada's Democratic Process'. Government of Canada. https://perma.cc/RU5S-9G77.

Competitions and Markets Authority. 2020. *Online Platforms and Digital Advertising: Market Study Final Report*. London: HM Government.

Conger, Kate. 2020. 'Twitter Labels Trump Tweet about "Racist Baby" as Manipulated Media'. *New York Times,* 16 September.

Corera, Gordon. 2020. *Russians among Us: Sleeper Cells, Ghost Stories and the Hunt for Putin's Agents*. London: Williams Collins.

Edelman, Gilad. 2020. 'Twitter Finally Fact-Checked Trump. It's a Bit of a Mess'. *Wired* 27 May, Retrieved October 18, 2020 https://www.wired.com/story/twitter-fact-checked-trump-tweets-mail-in-ballots/ .

Electoral Commission of South Africa. n.d. 'Report Digital Disinformation'. Retrieved October 18, 2020. https://www.elections.org.za/content/Elections/2019-National-and-provincial-elections/Report-digital-disinformation/.

European Commission. 2018. *Code of Practice on Disinformation*. Brussels: European Commission.

European Commisison (2020) European Democracy Action Plan: making EU democracies stronger, Press release,3 December 2020: Brussels, Retrieved June 28, 2021 https://ec.europa.eu/commission/presscorner/detail/en/IP_20_2250

Facebook. n.d. 'Ad Library'. Facebook. Retrieved October 18, 2020. https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=CN.

Fuller, Simon. 2019. 'Our Digital Future'. International Bar Association. Retrieved October 18, 2020. https://www.ibanet.org/Article/NewDetail.aspx?ArticleUid=60554B04-C95A-494B-845B-60BAFC7CA4C6.

Google. n.d. 'Political Advertising on Google'. Google Transparency Report. Retrieved October 18, 2020. https://transparencyreport.google.com/political-ads/home.

Greenleaf, Graham, and Bertil Cottier. 2020. '2020 Ends a Decade of 62 New Data Privacy Laws'. *Privacy Laws and Business International Report* 163: 24–26.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. 'Fake News on Twitter during the 2016 U.S. Presidential Election'. *Science* 363(6425): 374–78. doi: 10/gf3gmt.

Guess, Andrew M., Brendan Nyhan, and Jason Reifler. 2020. 'Exposure to Untrustworthy Websites in the 2016 US Election'. *Nature Human Behaviour* 4(5): 472–80. doi: 10/dpcn.

High Level Group on Fake News and Online Disinformation. 2018. *A Multi-Dimensional Approach to Disinformation*. Luxembourg: Publications Office of the European Union. doi: 10.2759/739290.

Information Commissioner's Office. 2020a. 'Bounty UK Fined £400,000 for Sharing Personal Data Unlawfully'. ICO. Retrieved October 18, 2020. https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/04/bounty-uk-fined-400-000-for-sharing-personal-data-unlawfully/.

Information Commissioner's Office. 2020b. 'ICO Issues Maximum £500,000 Fine to Facebook for Failing to Protect Users' Personal Information'. ICO. Retrieved October 18, 2020. https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/10/facebook-issued-with-maximum-500-000-fine/.

ITU Telecommunication Development Bureau. 2019. *Measuring Digital Development: Facts and Figures 2019*. Geneva: ITU.

Kalina, Bontcheva, Julie Posetti, Denis Teyssou, Trisha Meyer, Sam Gregory, Clara Hanot, and Diana Maynard. 2020. 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression: Broadband Commission Research Report on "Freedom of Expression and Addressing Disinformation on the Internet" '. Paris: UNESCO.

Karpf, David. 2019. 'On Digital Disinformation and Democratic Myths'. *MediaWell (Social Science Research Council)*. https://mediawell.ssrc.org/expert-reflections/on-digital-disinformation-and-democratic-myths/.

Kreps, Sarah, and Brendan Nyhan. 2020. 'Coronavirus Fake News Isn't Like Other Fake News'. *Foreign Affairs,* 30 March.

Machado, Caio, Beatriz Kira, Gustavo Hirsch, Nahema Marchal, Bence Kollanyi, Philip N. Howard, Thomas Lederer, et al. 2018. 'News and Political Information Consumption in Brazil: Mapping the First Round of the 2018 Brazilian Presidential Election on Twitter'. Oxford Internet Institute Blogs. http://blogs.oii.ox.ac.uk/comprop/wp-content/uploads/sites/93/2018/10/machado_et_al.pdf.

Marsden, Chris, and Trisha Meyer. 2019a. *Regulating Disinformation with Artificial Intelligence: Effects of Disinformation Initiatives on Freedom of Expression and Media Pluralism*. Brussels: European Parliament. doi: 10.2861/003689.

Marsden, Chris, and Trisha Meyer. 2019b. 'How Can the Law Regulate Removal of Fake News?' *Society for Computers and Law*. Retrieved October 18, 2020. https://www.scl.org/articles/10425-how-can-the-law-regulate-removal-of-fake-news.

Marsden, Chris, Trisha Meyer, and Ian Brown. 2020. 'Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?' *Computer Law and Security Review* 36: 105373. doi: 10/ggjt4w.

Meyer, Trisha, Christopher T. Marsden, and Ian Brown. 2020. 'Regulating Internet Content with Technology: Analysis of Policy Initiatives Relevant to Illegal Content and Disinformation Online in the European Union'. Chapter 16, pp. 309–327. In *Disinformation and Digital Media as a Challenge for Democracy*, edited by E. Kużelewska, G. Terzis, D. Trottier, and D. Kloza. Cambridge: Intersentia.

Mozilla. 2017. 'Facebook and Google: This Is What an Effective Ad Archive API Looks Like'. *The Mozilla Blog*. Retrieved June 21, 2019/https://blog.mozilla.org/blog/2019/03/27/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like.

Mozilla. 2019. 'Facebook's Ad Archive API Is Inadequate'. *The Mozilla Blog*. Ret rieved July 9, 2019. https://blog.mozilla.org/blog/2019/04/29/facebooks-ad-archive-api-is-inadequate.

NATO StratCom COE. 2018. *The Black Market for Social Media Manipulation*. Riga: NATO StratCom COE.

NATO StratCom COE. 2019. *Protecting Elections: A Strategic Communications Approach*. Riga: NATO StratCom COE.

Nimmo, Ben, C. Shawn Eib, Léa Ronzaud, Rodrigo Ferreira, Thomas Lederer, and Melanie Smith. 2020. 'Iran's Broadcaster: Inauthentic Behavior: Facebook Takes Down Covert Assets Linked to State Broadcaster'. *Graphika*. https://perma.cc/8ACK-R7S2.

Pace, Yannick. 2018. 'Parties Face Hefty Fines over Electoral Profiling without Consent'. *Malta Today*. Retrieved October 18, 2020. http://www.maltatoday.com.mt/news/national/87128/gdpr__parties_face_hefty_fines_over_electoral_profiling_without_consent.

Panoptkyon Foundation. 2021. 'Ustawa o wolności słowa? Raczej ustawa inwigilacyjna 2.0!'. *Fundacja Panoptykon*. Retrieved April 12, 2021. https://panoptykon.org/inwigilacja-2-0.

Press Information Bureau. 2019. '"Voluntary Code of Ethics" by Social Media Platforms to Be Observed in the General Election to the Haryana and Maharashtra Legislative Assemblies and All Future Elections'. *Government of India*. Retrieved October 18, 2020. pib.gov.in/Pressreleaseshare.aspx?PRID=1586297.

Purdon, Lucy, Arsalan Ashraf, and Ben Wagner. 2015. 'Security v Access: The Impact of Mobile Network Shutdowns, Case Study Telenor Pakistan'. Internet Policy Observatory Retrieved October 18, 2020 at https://repository.upenn.edu/internet-policyobservatory/13/ .

Rid, Thomas. 2020. *Active Measures: The Secret History of Disinformation and Political Warfare*. London: Profile Books.

RSF. 2019. 'Benin's Citizens Deprived of Internet on Election Day | Reporters without Borders'. *Reporters without Borders*. Retrieved October 18, 2020. https://rsf.org/en/news/benins-citizens-deprived-internet-election-day.

Sanovich, Sergey. 2018. 'Russia: The Origins of Digital Misinformation'. Pp. 21–40 in *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, edited by Samuel C. Wooley and Philip N. Howard. Oxford: Oxford University Press. doi: 10.1093/oso/9780190931407.003.0002.

Select Committee on Democracy and Digital Technologies. 2020. *Digital Technology and the Resurrection of Trust (HL Paper 77)*. London: House of Lords.

Silverman, Craig, ed. 2014. *Verification Handbook: An Ultimate Guideline on Digital Age Sourcing for Emergency Coverage*. Maastricht: European Journalism Centre.

Special Counsel Robert S. Mueller III. 2019. 'Report on the Investigation into Russian Interference in the 2016 Presidential Election'. Washington DC. Retrieved from https://perma.cc/9R75-FSBR.

Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. 2018. 'Report of the Special Rapporteur to the Human Rights Council on online content regulation (A/HRC/38/35)'. *United Nations*. http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/38/35.

Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. 2020. 'Report on Disease Pandemics and the Freedom of Opinion and Expression (A/HRC/44/49)'. *United Nations*. Retrieved October 18, 2020. https://www.undocs.org/A/HRC/44/49.

Tambini, Damian, Danilo Leonardi, and Chris Marsden. 2007. *Codifying Cyberspace: Communications Self-Regulation in the Age of Internet Convergence*. London: Routledge.

Thompson, Elizabeth. 2019. 'Most of Canada's Top Websites Won't Post Federal Election Ads This Year | CBC News'. *CBC*. Retrieved October 18, 2020. https://www.cbc.ca/news/politics/online-election-advertising-canada-1.5116753.

Timberg, Craig. 2020. 'How Conservatives Learned to Wield Power inside Facebook'. 20 February *Washington Post*.

UN Human Rights Council. 2016. 'Resolution Adopted by the Human Rights Council on 1 July 2016—32/13. The Promotion, Protection and Enjoyment of Human Rights on the Internet (A/HRC/RES/32/13)'. Office of the High Commissioner for Human Rights. Retrieved October 18, 2020. https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/32/13.

UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression, and African Commission on the Human and People's Rights Special Rapporteur on Freedom of Expression and Access to Information. 2017. 'Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda'. Retrieved July 10, 2019. https://www.osce.org/fom/302796?download=true.

United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, and the Organization of American States (OAS) Special Rapporteur on Freedom of Expression. 2020. 'Joint Declaration on Freedom of Expression and Elections in the Digital Age'. *Office of the High Commissioner for Human Rights*. Retrieved October 18, 2020. https://www.osce.org/representative-on-freedom-of-media/451150.

United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, Organization for Security and Co-operation in Europe (OSCE) 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 (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. 2015. 'Joint Declaration on Freedom of Expression and Responses to Conflict Situations'. Office of the High Commissioner for Human Rights. Retrieved October 18, 2020. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=15921&LangID=E.

Wardle, Claire, and Hossein Derakhshan. 2017. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. DGI(2017)09. Strasbourg: Council of Europe Report DGI(2017)09.

Weintraub, Karen. 2020. '"Fake News" Web Sites May Not Have a Major Effect on Elections'. 2 March, *Scientific American*.

Zuiderveen Borgesius, Frederik J., and Wilfred Steenbruggen. 2019. 'The Right to Communications Confidentiality in Europe: Protecting Privacy, Freedom of Expression, and Trust'. *Theoretical Inquiries in Law* 20(1): 291–322. doi: 10/gf2xzk.