Regulating Misinformation
Policy Brief

Introduction

The 2016 U.S. elections were clearly a watershed moment for misinformation. Following Donald Trump’s victory, it soon became clear that some social media platforms were unknowingly circulating inaccurate information. Worryingly, some of these were produced by external state actors who aimed to (unsuccessfully) directly influence the result (Parlapiano & Lee, 2018).

Misinformation is now a top-level policy issue. In addition to concerns about external nations interfering in democratic processes, governments are also concerned about information accuracy (Alba, 2020). This is especially relevant during the COVID-19 pandemic, where inaccurate information has circulated through social media, countering public health messaging and engaging in conspiracy theories about the origins and severity of the disease (Bruns et al., 2020; Meese et al., 2020).

Other popular conspiracies, such as the prominence of QAnon in the lead up to the 2020 U.S. election, offer another example of this growing misinformation problem. It also shows that citizens can circulate information-based threats to democracy.

Governments across the world have introduced new laws and regulatory frameworks in response. In this brief, we:

- Provide an overview of this reform moment;
- Identify key regulatory trends;
- Discuss recent platform decisions;
- Assess the benefits and risks associated with the identification and regulation of news and information.

Dr James Meese
Dr Edward Hurcombe
Defining misinformation

Distinctions have usually been made between disinformation and misinformation. The U.K. Government defines disinformation as false or manipulated information that has been produced or distributed for political ends. They also note that disinformation can cause personal and financial harm (DMCS, 2018). In contrast, misinformation has mainly been used to describe people unknowingly sharing false information (DMCS, 2018).

Recent work has questioned the usefulness of separating the issue into these two categories. Research has shown that ‘domestic political candidates, journalists, and citizens have played major roles in promoting disinformation’ (Tenove, 2020). The Australian Media and Communications Authority (ACMA) has also noted that ‘it is difficult to distinguish between these types of information disorder’ (ACMA, 2020: 10), and points out that inaccurate information can also cause significant harm. The fact that small groups of citizens are circulating conspiracy narratives that also have the potential to harm democratic processes similarly aligns these activities with external disinformation efforts.

While state-based information warfare and transnational conspiracy narratives operate independently of one another, Molly McKew has noted that they often converge around certain discussion points and can cause similar democratic harms (McKew & Farnham, 2020). We recognise this growing convergence and adopt the ACMA’s definition. This involves using misinformation as an ‘umbrella term’ and positioning disinformation as a subset of this wider concept (ACMA, 2020: 11).

The reform moment

Germany was one of the first jurisdictions to attempt to regulate misinformation (along with hate speech) and their model has influenced later reform efforts (Mchangama & Fiss, 2019). The Network Enforcement Act (or NetzDG) was introduced in 2018 and places a range of new obligations on social media platforms with two million users (Heldt, 2020). These platforms have to:

- Allow users to report content easily;
- Remove ‘manifestly unlawful content’ within 24 hours;
- Remove ‘unlawful content’ within 7 days; and
- Produce transparency reports.

While the law focuses predominantly on hate speech and content moderation, the prevalence of misinformation was an additional motivation for passing this Act. Since then a number of countries have implemented their own laws in an effort to tackle misinformation. Justitia, a Danish think tank, has identified fourteen countries that have followed Germany in attempting to restrict the spread of misinformation (Mchangama & Fiss, 2019). Since then Ethiopia, Turkey and Azerbaijan have passed similar laws, taking the total to seventeen countries. Brazil is also likely to adopt similar legislation in the near future (see Appendix One).

Justitia provides a good overview of the laws in each jurisdiction by November 2019. Our brief does not focus on specific jurisdictions, but instead identifies similarities and differences between them, compares the effectiveness of different interventions and discusses reform options that could be taken up in the future.
Key trends

Regulatory attempts at tackling misinformation vary, but they can be broadly categorised in terms of regulatory approaches and policy themes.

Three distinct regulatory approaches can be identified: voluntary co-regulation, legislation, and what we are calling ‘supporting activities’. Our identified key themes are information quality, criminalising speech, and national security. These approaches and themes are shaped by regional and political contexts.

Approaches

Voluntary co-regulation refers to policy interventions that do not involve mandated regulation or state oversight. Instead government bodies encourage and/or incentivise platforms to develop measures to address misinformation. These approaches allow platforms to consult with government and other relevant stakeholders to achieve a broad set of aims.

The EU Commission has implemented a voluntary code and the Australian Media and Communication Authority (ACMA) is currently developing one. In both cases, independent regulatory authorities ask platforms to devise policies to combat harmful misinformation, empower users to identify quality news sources and increase transparency around paid political advertising campaigns. Regulatory oversight for both codes comes in the form of a reporting regime, which allows signatories to provide regular updates to governments or the general public against agreed upon targets. In Australia’s case, a more direct regulatory approach may be considered if the voluntary code is not effective.

Legislation, on the other hand, refers to more direct attempts at combating misinformation, in the form of state intervention. Typically, legislation bypasses the consultative and voluntary approach discussed above and instead creates new laws or, in the case of France, new judicial processes to regulate platforms. This approach has been more popular than voluntary co-regulation. Germany, France, Russia, Singapore, the UK, Malaysia, Taiwan, and Brazil have all created or indicated support for new anti-misinformation laws (often dubbed ‘fake news’ laws). In Myanmar’s case, the government has used this approach to simply block websites outright (IFJ, 2020). Some of these legislative approaches have been criticised as being politically motivated, which we discuss in more detail below.

Finally, we have identified ‘supporting activities’, a broad category of government initiatives to tackle misinformation that go beyond regulation. This category includes funding for digital literacy programmes (Indonesia and the E.U.) and efforts at improving ‘critical thinking’ (the E.U.) (EU Code of Practice, 2018). Indonesia’s fact-checking Telegram chatbot, developed in partnership with a local digital start-up, is another example (Putera, 2019). These supporting activities tackle misinformation from the bottom-up, largely through educating people to better identify or be less susceptible to falsehoods and other viral ‘hoaxes’.

Themes

The first major regulatory theme we identified was information quality. By this, we mean approaches of various kinds that seek to monitor and/or improve the quality of information on platforms. Efforts to ensure ‘news quality’ are also included in this thematic category. This theme is
present across the three approaches discussed above. For example, ACMA’s position paper emphasises information quality. It wants the voluntary code to make misinformation more clearly identifiable on platforms, easily reportable by users and improve the ‘discoverability’ of ‘high quality’ news and information (ACMA, 2020: 28-29). The EU’s Code of Practice, which commits signatories to more effective fact-checking, is another attempt at voluntary co-regulation aimed at information quality.

Information quality is also the focus of legislation. The UK’s Department of Digital, Culture, Media and Sport Committee has recommended that platforms introduce ‘online tools to help [users] distinguish between quality journalism, and stories coming from organisations that have been linked to disinformation or are regarded as being unreliable sources’ (DCMS, 2018: 350). They suggest that platforms could be required to do so through the implementation of a statutory code overseen by an independent regulator. The Cairncross Review also recommended that platforms work with an independent regulator to assess and improve news quality (DCMS, 2019).

Supporting activities could also be included in this thematic category insofar as these programmes seek to empower people with the skills and the tools to identify information quality. Often, ‘quality’ is not explicitly defined in these approaches, although as evident from the proposals just described, it is usually associated with ‘accuracy’ (US) and contrasted against ‘harmful’ information (UK) or ‘misinformation’ (Australia).

The second regulatory theme is criminalising speech. Here, misinformation is addressed by outlawing certain kinds of speech or actors. Approaches in this thematic category are predominately legislative. A notable example here is Germany’s 2017 Network Enforcement Act, discussed above. Content banned under the Act includes speech already outlawed in Germany’s criminal code (such as incitement to crime and violence, unconstitutional symbols, ‘treasonous messaging’, and ‘propaganda’). However, the Act has been criticised for overreach. For instance, there have been concerns that platforms have been overly cautious in removing content as they attempt to comply with the law. The use of automated moderation tools have also contributed to the widespread removal of problematic content. The Act is currently under review (Earp, 2020) and proposed amendments may require platforms to ‘proactively report hate speech to law enforcement’, although it would also provide more transparency to users who would also be able to contest takedown decisions (Delcker, 2020).

France tried to pass a similar law, drawing on the German model. It would require platforms to remove any illegal content that was flagged by users within 24 hours or face fines of up to $1.25 million euros (Breeden, 2020). However, the bulk of the legislation was struck down by the French Constitutional Court, who claimed that the threat of overly cautious moderation raised free speech concerns (Conseil Constitutionnel, 2020).

Governments across South East Asia have also taken a similar approach to misinformation. Singapore passed a law in 2019 that required platforms to label content which the government deemed ‘false’, and those accused of spreading ‘falsehoods’ faced fines and jail time (BBC, 2019). In 2018, the Malaysian government also passed a law that would lead to fines and jail time for those found spreading ‘fake news’. The law was scrapped
following the election of a new government (Aljazeera, 2019). As noted earlier, the Myanmar government have also suppressed what they described as ‘misinformation’ via the blocking of over 200 websites in 2020. Outside of Asia, Russia passed several laws in 2019 to fine those ‘knowingly’ spreading falsehoods (Russian Federation, 2019), and Ethiopia in 2020 passed a law criminalising hate speech and disinformation online (Aljazeera, 2020).

All of these five cases have been accused of using concerns about misinformation to achieve political ends. For example, the Singapore law has been used to target news organisations critical of the government (BBC, 2019). The Brazilian government has also proposed a similar law targeting misinformation. The law would require users to register social media accounts with identity information (such as phone numbers and official identity documents) although it’s not yet clear how this would happen. It also threatens fines and jail time for those found to be ‘abusing’ or ‘manipulating’ others online, alongside other measures (Senado Federal, 2020). These reforms all approach the misinformation problem in a blunt fashion: outlawing certain speech and behaviour, with the latter countries often doing so for politically expedient reasons.

Finally, the third regulatory theme focuses on national security and election integrity. This theme captures reforms that attempt to protect the integrity of domestic elections and public debate from hostile foreign actors. These approaches tend to address disinformation – that is, the strategic and deliberate spreading of false information. In 2018, France passed legislation specifically targeting these issues (Conseil Constitutionnel, 2020). It allows the French broadcasting authority to suspend foreign-owned television channels if these channels are found to be interfering with domestic elections. The law also requires platforms to report sponsored content to the government during election periods. During the three months preceding election campaigns, individuals or groups can also bring a case of suspected misinformation (including disinformation) before a judge, who must provide a ruling within 48 hours.

Taiwan and Germany have also passed similar laws, although with a broader application beyond election periods. In the case of Taiwan, the government established powers in 2019, which allowed them to fine those found to be funded or aided by foreign actors to ‘infiltrate’ and threaten national stability (ROC, 2019). National security concerns for Germany, on the other hand, are included in Network Enforcement Act discussed above: specifically, the ability for the government to draw upon existing bans on ‘treasonous’ and ‘propaganda’ activity in the criminal code to target suspected disinformation on platforms (NetzDG, 2019). There is some overlap in this theme with the ‘criminalising speech’ category, in so far as both categories include attempts at outlawing certain kinds of activity. However, the emphasis here is on targeting activity coordinated externally by foreign actors.

Platforms

In recent years, platforms have taken steps to stop misinformation. These have emerged in the form of sporadic updates. In Facebook’s case, notable updates include the platform’s tightening of app permissions and privacy controls in March 2018 (to prevent third-party apps from misusing user data (Lomas, 2018)) and its
January 2018 decision to make the newsfeed algorithm favour posts from friends over other posts (Chaykowski, 2018). Facebook also expanded fact-checking and paid ad transparency in March 2018 (Levin, 2018), and throughout 2020 has limited or banned Groups and Pages linked to ‘militarised’ social movements, as well as those espousing unfounded conspiracy theories such as QAnon (Facebook, 2020).

Other platforms have also acted in a similar fashion. In August 2018 Instagram made a significant transparency update, implementing an information tab for accounts with large followings (Purtill, 2017). This allowed users to evaluate the ‘authenticity’ of these accounts. In October 2020, the platform also began labelling accounts that were linked to or affiliated with state-controlled media, a transparency measure that followed a similar move by Twitter in August 2020 (Bonifacic, 2020). In 2016, YouTube demonetised videos and provided brands with more control over ad placement after brands threatened to pull advertising following concerns about violent and extremist videos on the platform. Reddit addressed hate speech and extreme content in a unique way, when in 2015 it began to ‘quarantine’ subreddits (Reddit’s term for discussion forums) based around such content. ‘Quarantined’ subreddits are not banned outright, but have their reach limited by being restricted from appearing in search, recommendations, and non-subscription newsfeeds (Stephen, 2018).

Occasionally, updates come in the form of cross-platform bans on specific high-profile accounts: for example, in August 2018 the far-right conspiracy theorist Alex Jones and the fake news outlet that he hosts, InfoWars, were simultaneously banned from Facebook, Apple, YouTube and Spotify (Coaston, 2018). Platforms have also worked to promote correct information at specific times. The most notable example saw Facebook and Google ensure that official information from the World Health Organisation and national governments was prominent across their platforms during the pandemic (Farr & Rodriguez, 2020).

We can identify two major themes emerging from platform self-regulation. Firstly, self-regulation has largely been reactive: platforms generally respond to significant political and media pressure. For instance, Facebook’s tightening of third-party app controls followed the Cambridge Analytica scandal, and Instagram’s labelling of state-controlled media was in belated response to concerns about alleged Russian-backed disinformation campaigns on the platform. Secondly, there has been a steady retreat from a US-centric first amendment view of free speech. Platforms increasingly recognise that they need to at least be seen as competent moderators. Taken together, these broad themes suggest that platform self-regulation might be viewed as an attempt to evade state regulation by demonstrating that platforms are capable of governing themselves. These trends continued during the recent U.S. election and platforms (in particular, Twitter) took an increasingly proactive moderation role.
Concluding analysis

Countries are approaching the misinformation problem in a number of different ways. Here we provide some brief concluding thoughts on this busy period of policy activity.

Limiting foreign interference and securing election integrity are the most critical reform areas for democracies. We argue that with respect to misinformation (as opposed to broader foreign interference efforts) targeted reforms are the best way of addressing these problems. The French approach, which focuses specifically on media activity and foreign interference around elections stands as a salutary example of this kind of intervention.

Attempts to fight misinformation on a wider scale presents challenges. It is clear from the reforms discussed above that laws addressing these issues can only be successfully implemented in strong democracies. Without appropriate checks and balances, the regulation of information can become easily politicised.

Democracies have their own difficulties. Regulating hate speech and other illegal content associated with misinformation raises important questions around free speech. The different experiences of Germany and France in this regard point to the importance of historical context when considering reforms in this area. Germany has a long legacy of criminalising certain types of speech following the Second World War, allowing them to introduce relatively strong legislation. This was not the case in France and as a result, similar reforms were derailed. These outcomes suggest that regulations will struggle to be translated across different democracies. Interventions will need to be sensitive to national approaches to speech.

Similar democratic tensions also emerge around the issue of news quality. Regulators are seeking to address a wide range of harms associated with misinformation. These include serious harms around electoral activity and personal safety but also include lesser harms like false information and rumours. As a result, news quality has become an important regulatory category. Actions from platforms suggest that they are keen to limit their regulatory exposure and their reactive approach that they have a narrow view about where they should intervene. They focus on serious harms and generally avoid making value judgments about information.

We are sympathetic to the focus on quality but note that it is difficult to make hard and fast distinctions between news, information and misinformation. Sometimes news is not accurate, and news organisations may occasionally fall outside established definitions of quality. We make no recommendations but simply observe that foundational questions about the freedom of the press, free speech and press ethics emerge here. A number of common law countries are able to pursue concerns about accuracy on radio and television under existing broadcasting regulations. One unanswered question is whether regulators and platforms should play a similar role with respect to all news content published online.

We end by arguing that formal regulatory instruments will not solve the problem. Communication researchers have noted that misinformation emerges from a wider political and cultural context, which sees people choose to believe certain information because it supports their worldview (Marwick and Lewis, 2017). Others in this field also note that
misinformation regulations can discourage political participation (particularly in repressive regimes) and that political parties can easily conduct ‘dirty campaigning’ in response to what they see as misinformation from the opposing side (Ong, 2020). As a result, bottom up approaches like supporting media literacy initiatives, addressing political disengagement and countering partisanship and populism are also likely to be just as important.

References


Facebook (2020). An update to how we address movements and organizations tied to violence.


and-indonesian-ministry-fight-fake-news-with-anti-hoax-chatbot


Countries that have introduced or are currently developing misinformation reforms (Brazil is planning to introduce legislation).