1 Mis- and disinformation: What governments can do to reinforce democracy

The spread of mis- and disinformation poses a fundamental threat to the free and fact-based exchange of information that underpins democracy. This chapter discusses how governments can respond to mis- and disinformation through a whole-of-government and whole-of-society approach. This includes preparing for and responding to the publication and spread of mis- and dis disinformation; preventing the publication and spread of mis- and dis-disinformation through increasing transparency; and reducing the economic and structural drivers of mis- and disinformation.

1.1. Introduction

The spread of mis- and disinformation poses a fundamental threat to the free and fact-based exchange of information that underpins democracy. The most evident way in which misleading and false information distorts democratic engagement is by convincing people to believe things that are not true, which can be especially harmful if focused on demonising political opponents, distorting policy debates or undermining democratic institutions. By making it more difficult to access timely, relevant and accurate information and data, the amplification of mis- and disinformation content can undermine the public's willingness and ability to engage constructively in democratic life, and down the line the ability of society to forge consensus. Furthermore, by blurring the line between authentic political speech and purposefully deceptive content, disinformation can also fuel polarisation, spread confusion and prop up and support authoritarian leaders.

As such, the spread of mis- and disinformation can weaken countries' abilities to protect their national interests and preserve national security and democracy. The COVID-19 pandemic, the 6 January 2021 attack on the US Capitol and the large-scale Russian aggression against Ukraine have all underscored the threats posed by the spread of false and misleading information and the dangers caused by widespread erosion of trust in institutions and information.

76% of respondents to the Edelman Trust Survey in 27 countries indicated that they worry about false information or fake news being used as weapon. In addition, 67% of respondents were worried that journalists and reporters are purposely trying to mislead people by saying things they know are incorrect or exaggerated, and 66% say the same about government and political leaders (Edelman, 2022[1]). This lack of institutional trust severely limits the scope for constructive democratic engagement.

Such concerns are at an all-time high and flag the urgency and importance of ensuring governments have the capacity to respond and create systems that can counteract the threats posed by the spread of mis- and disinformation. More widely, reinforcing democracy also requires strengthening media and information ecosystems – understood as the space where citizens, journalists and institutions (governmental, civic and private) create, spread and engage with information, governance frameworks and each other – to build trust and facilitate engagement since information quality is the most powerful trust builder across democratic institutions (Edelman, 2022[1]).

Mis- and disinformation are not new phenomena and will continue to exist in all societies, regardless of the strength of the democracy or of the media institutions within them. Previously, however, technological limitations to how information was spread and a combination of limited government regulation and checks and balances offered by the press and news media institutions and their governance helped limit its spread. This helped create democracies with media and information ecosystems that, while less dynamic and diverse than today's, were relatively stable and able to prevent a certain degree of polarisation and susceptibility to disinformation spread by foreign or domestic actors.

The emergence of online communication spaces and social media platforms that allow for virtually anyone to instantaneously be a source of information (or misinformation) and to amplify such content globally has been a systemic and fundamental shift. The Internet has changed and facilitated the ability for content to be created and shared in ways that are only beginning to be understood (Lesher, Pawelec and Desai, 2022[2]; OECD, forthcoming[3]).

These technological changes have allowed for the growth in the diversity of sources and the opportunities to access global information, offering an essential counterweight to proscribed, anticompetitive or otherwise restricted media (particularly notable in the context of Russia's aggression in Ukraine). Digital technologies have enabled more participatory, innovative and agile ways for institutions (government, media, other private sector and civil society) to communicate with citizens and for citizens to communicate with each other. Recent OECD analysis suggests how governments can also use social media platforms more effectively to promote interactive communication in ways that help counteract mis- and disinformation (OECD, 2021[4]).

Nevertheless, these same technologies are used by malign actors to threaten basic elements of democratic life. Both online and offline engagement can be shaped by information flows on social media platforms targeted by domestic and foreign actors to undermine the functioning of democracies. Indeed, the impact of social media goes beyond its use as a direct source of information, given that feedback loops between the platforms and traditional media can also serve to amplify mis- and disinformation. These threats can reduce trust and risk shrinking the space for democratic engagement and weaken the strength of free speech. Disinformation, in particular, can be used to deliberately alter understanding of public figures' integrity and competency, as well as to confuse and discourage the public in ways that reduce willingness to engage in debate or seek political office. Such campaigns affect segments of the population in different ways. For example, women in politics are disproportionately targeted by gendered disinformation campaigns, a pattern that is even more pronounced for female political leaders from racial, ethnic, religious, or other minority groups (Meco and Wilfore, 2021[5]). Understanding the nature and impact of mis- and disinformation, including the intersection of narratives targeting specific segments of the population, will be an important consideration moving forward.

Efforts to curb mis- and disinformation must also be considered hand in hand with the full preservation of free speech. Laws that define mis- and disinformation broadly can be used to restrict legitimate speech. Governments may also require or exert pressure on platforms to restrict otherwise legal content, forcing or coercing private owners of *de facto* public engagement spaces to be more restrictive than laws may require. Ultimately, maintaining freedom of expression will mean that false and misleading content will always exist; the aim is to mitigate the harm to democratic engagement that such content can cause and to reinforce information spaces that are conducive to democratic engagement.

As a result, a new governance model is needed by which governments, together with traditional media organisations, social media, academics and civil society organisations, jointly help redesign the shape of information ecosystems. Some governments have moved in that direction and have flagged the scale of the threat and need for an internationally co-ordinated and whole-of-society approach. Lithuania's constructive relationships with independent fact-checkers, Finland's engagement with civil society to support media literacy efforts and the European Union's efforts to develop a co-regulatory instrument via its Code of Practice on Disinformation are a few of the many, and increasing, examples of such a necessarily collaborative approach.

This chapter is focused on governance responses and provides an overview of the measures that governments are taking or could take on their own or in partnership with media and civil society organisations (as such, it does not explore self-regulatory measures taken by the private sector). It suggests a comprehensive strategy to prevent and combat mis-and disinformation and promote a governance of information ecosystems that strengthens democracies. With the Internet facilitating largely borderless information sharing, like-minded countries must work together and with a wide range of non-government partners to tackle challenges posed by mis- and disinformation.

1.2. Identifying government responses to mis- and disinformation

"Misinformation" can be defined as false or inaccurate information that is shared unknowingly and is not disseminated with the intention of deceiving the public, whereas "disinformation" is usually defined as false, inaccurate, or misleading information deliberately created, presented and disseminated (Wardle and Derakshan, 2017[6]; Lesher, Pawelec and Desai, 2022[2]). Misinformation is sometimes used as a catchall term for many similar but ultimately different practices, for example disinformation, information influence operation, and foreign interference in the information space,[1] each of which may require a different approach. Mis-and disinformation are furthermore not to be confused with the dissemination of terrorist, violent or illegal content online, which often require a set of specific measures not covered below.[2]

The measures discussed here represent the array of actions that Governments need to consider, to a greater or lesser extent and depending on their local context. Focusing on a single threat, such as foreign influence or the use of bot farms, or on a single response, would only render a government vulnerable to other sources of mis- and disinformation, and not address the full problem. In this sense, governments should consider a comprehensive strategy that considers a wide range of measures; deploys them together with a continuous effort to assess, address, and avoid the threats and harm caused by mis- and disinformation; and evaluates initiatives in light of potential impacts on freedom of speech and expression.

While all measures presented can play an important role, the geopolitical and social context of each country requires a tailored analysis of which areas need more attention, as well as where to best allocate resources. In addition, individual countries have differing legal systems, precedents and approaches to the protection of freedom of speech that will inform their approaches. Thus, while there are common objectives and lessons regarding the key measures needed to tackle the threats, responses must also take into account the specific country context. By acknowledging the context-dependent nature of the threats and responses, governments will be better able to prepare a customised strategy informed by common principles and lessons.

That said, there are a number of common principles to guide governance responses. Promoting freedom of speech and reinforcing the space for democratic debate and engagement are at the root of this work. To that end, regulatory responses discussed in this chapter do not explore content-specific regulations, which risk impeding information distribution and restrict freedom of speech and expression. Facilitating the independence of actors – in civil society, the media, as well as regulators – will help encourage checks and balances. As noted by the OECD, a regulatory agency's independence from the government and from those it regulates is a useful element in providing confidence that decisions are fair and impartial, which may be particularly important where decisions can have significant financial and market consequences (OECD, 2012[7]).

Furthermore, social media platforms are inherently international, and there is much to be gained from a cross-border, comparative and analytical approach. Indeed, the spread of mis- and disinformation cannot be considered outside of the wider context of global challenges facing democracy and institutional trust. Understanding the extent to which the spread of mis- and disinformation is driven by groups and individuals that feel disenfranchised and alienated from democratic processes will need to be considered as part of the overall governance and societal response to the spread of false and misleading content.

Current and proposed measures to prevent and combat mis- and disinformation can be grouped under:

1. Governance policies and initiatives that help prepare for and respond to the publication and spread of mis- and disinformation.
2. Regulatory and policy measures to increase transparency and prevention.
3. Policy and regulatory responses that reduce economic and structural drivers of mis- and disinformation.

1.2.1. Preparing for and responding to the publication and spread of mis- and disinformation

Recent events have highlighted the need to develop capacity to response to the spread of false and misleading information, while simultaneously building more resilient societies better prepared to handle crises. A range of communication and domestic and international engagement efforts can help governments respond to mis- and disinformation content directly, and building a more effective public communication function and promoting media literacy can support more resilient information ecosystems. These initiatives are largely non-regulatory responses, many of which OECD Members have started to put in place.

A common thread through many of these responses is the critical and mutually reinforcing role played alongside government by the media, civil society organisations, and the private sector. Information does not spread in a vacuum – traditional media and fact-checkers, technology companies, civil society, and citizens themselves are essential to generate and amplify trustworthy content. Such interventions include:

Collaborating with media, civil society organisations, fact-checkers, and social media platforms

Measures to address mis- and disinformation must be pursued in the context of promoting the fundamental importance of ensuring freedom of expression, preserving the role of objective, independent and fact-based journalism, and securing the space for civil society organisations and innovative, non-traditional, local or community media to grow and help enable information ecosystems to thrive, free from undue government interference. More narrowly regarding specific responses, governments must also engage with media and civil society organisations to legitimately and transparently address urgent threats posed by mis-and disinformation. The lack of clarity on problems and solutions, combined with the complex, global and rapidly evolving nature of the challenges faced, calls for a more conscious effort to facilitate collaboration between various actors (OECD, forthcoming[8]). As noted by the European Commission, "the best responses are likely to be those driven by multi-stakeholder collaborations" (European Commission, 2018[9]).

There are numerous private organisations, fact-checkers, media and NGOs that seek to debunk mis- and disinformation (Credibility Coalition, 2021[10]; Khan, 2021[11]), and governments may support or benefit from the work of fact-checkers to serve as independent and trusted voices. Notably, from the onset of the COVID-19 crisis, youth organisations have launched information campaigns to combat false information, including the international campaign #youthagainstcovid19 and the national campaign #QuédateEnCasa in Mexico, to map and share myth-busting, fact-checking websites and resources targeted at young people (OECD, 2020[12]). In Italy, the government convened a group of experts in fact checking, debunking and disinformation and launched a joint action campaign called "***#bastabufale***", meaning, "stop hoaxes".

Governments may also seek to build transparent and constructive relationships with online platforms to monitor, flag and respond to mis- and disinformation. In preparation for the 2021 elections, the German Federal Office for Information Security (BSI) co-ordinated with social networks to facilitate rapid reactions to potential threats and established a unit to detect automated bots and synchronised inauthentic behaviour (Miguel, 2021[13]). These efforts should be undertaken carefully, however, so as not to inhibit freedom of speech, particularly regarding content takedowns (OECD, forthcoming[8]).

Pre-bunking – or attempting to "inoculate" the public to misleading messages – is another approach that requires anticipating potential misunderstandings or disinformation attacks and that has benefited from partnerships and engagement (Blastland et al., 2020[14]). For example, the Go Viral! game was developed by the University of Cambridge in partnership with the UK Cabinet Office. It builds on research that found that by exposing people to the techniques used to spread misinformation online, they can better identify and disregard false and misleading content. The game exposes players to examples of false news stories and memes to help them detect such content (Roozenbeek and van der Linden, 2019[15]).

Over the longer term, co-ordinating and engaging with a wide range of actors can help raise awareness, share knowledge and collect data on effective responses. For example, the Government of Latvia has engaged with the Baltic Centre for Media Excellence, which has established an informal network bringing together journalists, election officials, security services and government officials to enhance communication and co-operation in case of threats to election processes.

The value of engaging with media and civil society organisations is likewise relevant to moving the research agenda forward. Independent support for and collaboration between governments and researchers on topics such as understanding how disinformation is created and spread, why and by whom; which actions

are most effective; and what lessons can be drawn from previous technological changes, will be valuable in designing appropriate responses. In addition to the value of increased direct funding for research, ensuring academia, regulatory bodies and other relevant agencies are engaged in conversations about research needs is a useful step in promoting coherent and effective responses (Matasick, Alfonsi and Bellantoni, 2020[16]). For example, Canada's Digital Citizen Initiative funds research and digital/civic literacy activities with an aim to better understand the sources, spread, and impact of disinformation in Canada, and how might literacy activities best inoculate citizens and build resilience.

Collaborating at the international level through exchange of information

The ability for myths and false information to spread as widely and rapidly as they do – clearly seen throughout the COVID-19 pandemic – has added urgency to efforts to work across national boundaries to counter such narratives. Facilitating dialogue between all relevant actors is therefore an important element to tackle these threats. To that end, governments can build on efforts to collaborate and exchange information, threat analysis and good practices. For example, the EU Rapid Alert System (RAS) facilitates the sharing of insights related to disinformation campaigns and co-ordinate responses between EU member states. The RAS is based on open-source information and draws upon insights from academia, fact-checkers, online platforms and international partners. Similarly, in 2014, NATO founded an independent StratCom Centre of Excellence, which aims to contribute to the strategic communications capabilities of NATO allies and partners, including via research and preparations regarding threats posed by disinformation.[3] The G7 Rapid Response Mechanism (RRM) was launched in 2018 to strengthen G7 members' co-ordination and identification of threats to democracy, including those posed by mis- and disinformation.

Moving forward, international collaboration can further promote more effective regulatory responses and engage like-minded countries in identifying regulatory priorities and understanding options and impact. Building an international knowledge base and applying lessons from other industry experiences is particularly useful to respond to the rapidly changing and complex trade-offs concerning the response to mis- and disinformation, and points to the utility of continuing to explore new avenues for international co-operation.

Building capacity for more responsive and effective public communication in counteracting mis- and disinformation

Building the capacity of the public communication function[4] to promote a more informed citizenry and support a healthy information ecosystem can be an essential tool to counteract the threats posed by mis- and disinformation. By providing proactive, timely, and transparent communication, governments can both react to and prevent the spread of such content. Specific examples of how governments can strengthen this function include:

- **Governance and institutionalisation of public communication responses.** Governments should formalise definitions,[5] policies and approaches to help shift from ad hoc and fragmented approaches to counteracting mis- and disinformation, to more structured and strategic approaches. Supporting the governance and institutionalisation of the public communication function can provide clarity of purpose, help set concrete metrics for measuring impact of public communication activities and justify allocations for resources to this government function (OECD, forthcoming[8]). Ultimately, strengthening the public communication function by assigning clear mandates, allocating appropriate institutional resources and establishing effective co-ordination mechanisms can enhance governments' ability to disseminate accurate content and foster citizen participation (Matasick, Alfonsi and Bellantoni, 2020[16]).
- **Identifying, tracking, monitoring, analysing and assessing problematic content and its sources**. Monitoring public and open channels and platforms to identify problematic content and

emerging narratives is a key feature of government efforts to understand emerging mis- and disinformation narratives and to develop effective communication responses. Given that disinformation campaigns often seek to elicit emotional reactions to content and to undermine trust in target audiences, governments should develop clearer understandings of how to track and react to such content. For example, in Lithuania, potential disinformation is assessed using three criteria: its source; its content and context; and the timing of when the content was spread. When taken together, this information helps provide a clearer picture of the potential disinformation threat.[6] Such activities should be conducted consistently to ensure the timeliness of possible responses, as well as transparently and within the limits of data privacy so as to help maintain democratic legitimacy (OECD, forthcoming[8]). To that end, governments should put in place structures, staff and resources, such as the UK's Rapid Response Unit (RRU), to ensure departments at all levels are sufficiently equipped and that counter-misinformation efforts are mainstreamed in a transparent and accountable way.

Nevertheless, the role and impact of closed groups and messages shared on encrypted services such as WhatsApp will need to be better understood. These platforms provide users with valuable privacy and safety functions, but can also be important channels to spread mis- and disinformation, while their private and encrypted nature make understanding content spread on these channels impossible to analyse. Monitoring the evolution of public discourse on social media and online searches in real-time can help governments understand emerging narratives and respond quickly and effectively to emerging threats. At the same time, governments must ensure mechanisms are in place to prevent tracking from being misused or for users to be identified and monitored in ways that can restrict speech, infringe on privacy or limit democratic participation.

- **Understanding audience needs, using behavioural insights to prevent the spread of falsehoods.** Gathering and analysing data on the public's needs and expectations is essential for devising effective communication against mis- and disinformation (OECD, forthcoming[17]). Using audience research and evaluation of communication efforts, governments can ensure messages are tailored, relevant and responsive. Furthermore, evidence from behavioural insights (BI) shows that behavioural failures such as information overload and confirmation bias can undermine government response to mis- and disinformation. A sophisticated understanding of human behaviour is fundamental to developing effective responses to mis- and disinformation and declining trust in institutions. BI enables governments to better understand who are the most vulnerable populations, design innovative policy solutions to mitigate the spread of misinformation and its effects, and build evidence on what approaches work best (see Box 1.1) (OECD, 2021[18]).

Box 1.1. An International Collaboration to tackle Misinformation with Behavioural Insights

The Government of Canada, in partnership with the OECD and the French Government, conducted an experiment to investigate and influence Canadians' intentions to share false and true news on social media. Implementing a randomised controlled trial (RCT) design in one wave of the longitudinal Impact Canada COVID-19 Snapshot Monitoring Study (COSMO Canada), two behaviourally informed interventions were tested. Both interventions were drawn from a rapidly growing research literature, and both aimed at improving the quality of news shared online (that is, the preference for sharing verifiably true over verifiably false news links) while prioritising individuals' autonomy. The first intervention was a simple accuracy evaluation prompt, attuning respondents' attention to accuracy by asking them to rate the accuracy of a single random headline prior to engaging with Facebook-style headlines online. The second intervention was a list of media literacy tips.

This collaboration found:

1. First, that early results indicate a disconnect between participants' (*N* = 1872 participants) beliefs and sharing intentions. People rate verifiably true headlines as significantly more accurate than verifiably false headlines (as determined by third-party fact-checkers), but are much less discerning in their sharing intentions – in other words, people may share news headlines they believe to be false, or questionable.
2. Second, that early results of the experimental intervention suggest that exposure to both the simple attention-to-accuracy prompt and the digital media literacy tips significantly increased participants' intentions to share true over false headlines. In contrast to previous published work, the effectiveness of the media literacy intervention far exceeded the effectiveness of the accuracy prompt, with the tips reducing false news sharing intentions by over 20%.

These results provide compelling support for how simple and scalable online interventions presented to individuals before they engage with news stories may improve the quality of information circulating online. For some, it may be surprising to hear that individuals are (sometimes) willing to share news that they believe to be false or questionable. This study provides evidence that this does indeed happen, likely due to a failure to pay attention to the accuracy of news content confronted in the social media context. Although additional research and analysis is required to determine why individuals may choose to share false or misleading headlines online, studies like these remain vital for challenging assumptions about human behaviour, creating more effective and scalable solutions based on those they aim to serve, and indicating areas of future exploration that can enhance the robustness of knowledge on global behavioural challenges like mis- and disinformation.

Source: Government of Canada

More broadly, using evidence, analytics and BI can help governments promote evidence-based and innovative policy outcomes, support their ability to manage information in the digital age, and increase their capacity to address global challenges to reinforce democracy.

- **Inclusive-minded content design and delivery.** Using appropriate channels and delivering clear and tailored messages can help ensure communications reach all segments of society, including groups that are less likely to be exposed to or trust official information. To that end, preparing and implementing strategic communication campaigns and ensuring accurate information reaches target audiences proactively is essential in counteracting the spread of mis- and disinformation. Throughout the COVID-19 response, many countries developed processes that utilise credible messengers, such as members of a particular community, scientists and doctors, or influencers to present relevant information in a timely, authoritative and non-politicised way to help ensure it reached as wide a segment of the population as possible (OECD, forthcoming[8]). Governments can also support trusted messengers to counteract mis- and disinformation by providing information and guidance. Along these lines, the US "Community Toolkit for Addressing Health Misinformation" provides "trusted messengers" with practical, step-by-step recommendations and actions for trusted community messengers.

Improving media literacy through awareness campaigns and civic education

Maintaining freedom of expression and an open internet means that mis- and disinformation will never disappear. A focus on reducing systematic risks to its spread, therefore, suggests that governments should also build long-term resilience at the level of individual citizens, who should be better equipped to

differentiate between accurate and false or misleading information and be more aware of their role in preventing its spread. Media and information literacy plays an important role in helping to protect society from the relevant threats, while building capacity for the public to take advantage of the benefits to online and social platforms.

Media and information literacy efforts aim to build capacity of individuals to recognise and dismiss false and misleading information. Efforts can be campaign-based or achieved through civic education. Examples of campaign-based initiatives include Australia's "Stop and Consider" campaign (Buckmaster and Wils, 2019[19]), which sought to encourage the electorate to check sources of elections information carefully (Australian Electoral Commission, 2019[20]). Belgium also used media literacy measures through a website informing people about mis- and disinformation (Funke and Flamini, 2020[21]; Mon Opinion, 2021[22]). In Latvia, the Ministry of Culture used social media advertisements to promote false news titles; if clicked, the user was redirected to media literacy resources. In one month, the campaign reached more than 895 000 people, with 129 000 people seeing the media literacy information (Ministry of Culture of Latvia, 2021[23]).

Civic education can also be carried out by incorporating media literacy into existing school and university curricula, as well as providing training for teachers to deliver the content (Burns and Gottschalk, 2020[24]).[7] Existing efforts include those by France's Ministère de la Culture and Belgium's High Council of Media Literacy (Conseil Supérieur d'Éducation aux Médias) which provide tools, training courses and engagement opportunities between students and journalists to increase resilience to disinformation (Matasick, Alfonsi and Bellantoni, 2020[16]; Suarez-Alvarez, 2021[25]).[8] Additionally, in 2008 the Ministry of Education, Culture and Science of the Netherlands established the Dutch Media Literacy Network, which brings together a wide range of partners from across society to promote awareness and share knowledge, expertise and relevant media literacy resources (Dutch Media Literacy Network, 2022[26]). In Finland, the National Media Education Policy lays out the national effort to provide high-quality, systematic and comprehensive media education, using a variety of actors. Finland's efforts are structured as part of the country's broader effort to strengthen democracy and education, and are built on media literacy activities that began in the 1950s.[9]

1.2.2. Preventing the publication and spread of mis- and disinformation through increasing transparency

In addition to actions that address immediate threats or that strengthen the resilience of societies to mis- and disinformation, governments can also adopt regulations and other policy measures aimed at increasing online platforms' transparency. Given the asymmetry in knowledge between online platforms and governments about how content is spread and what interventions work, transparency is an essential component in helping government and non-government actors develop better understanding to inform policy making.

Along these lines – though focused instead on combatting terrorist and violent extremist content (TVEC) online – the OECD Science, Technology and Innovation Directorate has developed the Voluntary Transparency Reporting Framework (VTRF). This tool offers a common standard for TVEC transparency reporting for online content-sharing services to provide information about their TVEC-related policies and actions.[10] Its application and the analysis derived from the VTRF reports can help inform regulatory responses to promote transparency reporting around mis- and disinformation, as well as future international voluntary efforts to collect relevant information from platforms.

Regarding regulatory responses, the design and application of transparency regulations depend on governments partnering with media and civil society organisations to ensure the utility of the measures and public benefit, as well as to provide assurances that there is no government interference in the free flow of information. These measures can include:

Data sharing requirements for online platforms

Data collection and targeted advertising fuel the 'attention economy'. By encouraging companies to highlight the kind of content that keeps viewers' attention, regardless of whether that content is true or not, this attention economy also helps fuel the spread of mis- and disinformation (Balkin, 2020[27]; 2016[28]). Conducted with appropriate privacy safeguards and oversight mechanisms to prevent infringements on freedom of speech and expression, the increased ability to identify and trace mis- and disinformation and to collect data on which interventions are effective will help build understanding of the challenges and the design of effective policy responses. Specifically, for example, building the understanding of the sources and content of disinformation campaigns from foreign actors can support law enforcement, security and intelligence agencies to better assess and understand the threat of foreign influence in domestic matters (McCallum, 2021[29]).

To facilitate public-private access to and sharing of information and data on mis- and disinformation, governments can consider promoting partnerships with external researchers and platforms to share and analyse data from online platforms (OECD, forthcoming[8]). For example, co-ordinating the sharing of data between platforms and government could be done through an information sharing and analysis organisation (ISAO) or information sharing and analysis centre (ISAC). As long as there are sufficient protections and structures in place to comply with privacy laws, these platforms can gather relevant information and can enable voluntary sharing of information between the private and public sector (DiResta, 2021[30]). Partnerships, to the extent that companies are willing the share information, may include open data initiatives related to mis- and disinformation, in addition to closed or secure data access and sharing arrangements between governments, technology companies and independent researchers. In 2015, the US Government encouraged the creation of ISAOs for private companies, non-profits, and government departments and agencies to share cyber threat information and best practices. It also established limited liability protections for organisations that voluntarily share threat intelligence with each other and the government via these venues (US Government Office of the President, 2015[31]).

Moreover, new legislation could establish a legal framework to require the sharing of metadata with external researchers, including information related to disinformation and removed content. The Australian Government announced that they would propose legislation based on the Australian Communications and Media Authority's (ACMA) report on the adequacy of digital platforms' disinformation and news quality measures.[11] This legislation will aim to provide ACMA with the ability to collect information on Australia-specific content, as well as data on the steps taken to address mis- and disinformation from social media platforms. Australia will also establish a Misinformation and Disinformation Action Group to support collaboration and information-sharing between government, the private sector, researchers and civil society (Minister for Communications, Urban Infrastructure, Cities and the Arts, 2022[32]). Additionally, the draft Platform Accountability and Transparency Act in the US Senate would create a process through which academic researchers could gain access to information about the operation of social media platforms. The companies would be required to disclose certain internal data and respond to independent research requests. The proposal would also protect researchers from legal liability and would require that platforms proactively make certain information available to researchers or the public.[12] All efforts to increase data-sharing, however, should be conducted under reasonable privacy protections and in ways that protect individuals' civil liberties (Stamos et al., 2019[33]).

While enabling the collection of this data from social media platforms will likely require regulation mandating increased data sharing, governments should also build their own capacity to monitor, understand and make sense of data collected. This will necessitate developing public servants' skills and establishing constructive partnerships with media, academics and civil society partners to facilitate greater understanding of – and more effective responses to – the challenges faced.

Establishing an effective transparency framework around content moderation

One of the fundamental shifts in how information is spread is the role played by online platforms in curating, amplifying and moderating user-generated content. Many of the regulatory frameworks that apply to traditional media (with the exception of otherwise illegal content) do not apply to social media or other online platforms – nor would such frameworks make sense given the fundamentally different models of content creation and distribution. To date, platforms have conducted content moderation largely in response to social and government pressure. This predominantly self-regulatory approach gives private companies de facto control over what information is shared on these important spaces for news dissemination and engagement. At the same time, content-specific government regulations that expand restrictions beyond otherwise illegal speech present clear risks to freedom of speech and expression.

It is therefore important to explore process questions and encourage platforms to establish a framework around which their content moderation activities can be structured. This could include requiring platforms to put in place safeguards for users, such as allowing for the possibility of challenging platforms' content moderation decisions, as well as mandating transparency measures for online platforms that clarify their approach and decisions. Such measures have been proposed, for example, in the European Commission's Digital Services Act (DSA). In addition, governments should focus on ensuring capacity to monitor self-regulatory practices, understand the incentives that underpin participation in self-regulatory regimes, and consider the costs and benefits of regulatory flexibility.

The underlying principles of increasing transparency of content moderation aim to protect users from false or misleading content, while also providing clarity and protection for users around decisions made concerning their content. These goals can be carried out through:

- Requiring online platforms to clarify in plain language through their content moderation policy or Terms of Service how they moderate content, including on algorithmic decision making and human review (OECD findings show, however, that only a small share of consumers read platform terms and conditions in full, suggesting that important and relevant information should be communicated in ways that users can more easily access and understand (OECD, 2017[34]).
- Requiring clear policies regarding users who break terms of service repeatedly.
- Increasing transparency for users regarding content removal or downgrading actions, the reasons for the decision, and the tools used to reach the decision.
- Requiring online platforms to report regularly on actions taken against mis- and disinformation, potentially including an overview of how content was removed or de-prioritised, the number of accounts suspended, how content was flagged, etc.[13]
- Requiring that online platforms be subject to regular audits of how their content management policy is implemented to identify areas for potential misuse or abuse.

A risk-based approach, which takes into account size and scale of platforms and services, could also help meet the dual demands of encouraging innovation while protecting human rights and democratic discourse. As social media platforms connect more people and play an increasingly important role in public discourse, for example, their potential impacts and the risks they pose to individuals and society also increase. Larger platforms may therefore face more stringent obligations compared to smaller platforms. In practice, this trade-off is particularly complicated when it comes to social media platforms, which blur the lines between consumers and producers of information, making regulations all the more difficult to design (OECD, 2018[35]).

Notably, the European Commission's proposed DSA places additional requirements on very large online platforms to undertake annual risk assessments to identify systemic risks, including "intentional and (…) co-ordinated manipulation of the platform's service, with a foreseeable impact on health, civic discourse, electoral processes and public security" as well as mitigation measures to address such risks (European Commission, 2020[36]). The DSA also requires large platforms to take into account how their content

moderation, recommender and advertisement selection and display systems influence the spread of illegal or manipulative content.

Other approaches are informed by efforts to provide broad guidance on good practices, pre-established rules and standards for transparent content moderation.[14] In the European Union, the Code of Practice on Disinformation was the first self-regulatory instrument that leading industry actors, including Facebook, Google, Microsoft, Mozilla, TikTok and Twitter, voluntarily agreed to. The Code sets out a "wide range of commitments, from transparency in political advertising to the closure of fake accounts and demonetisation of purveyors of disinformation."[15] It also includes an annex identifying best practices that signatories will apply to implement its commitments. The European Commission recently made proposals to strengthen its implementation,[16] with the aim that the Code can become a co-regulatory instrument, as outlined in the European Commission's proposed DSA. It sets out the transparency standards that platforms need to establish regarding how they detect, identify and address content that is incompatible with their terms and conditions (European Commission, 2020[36]).

Content moderation touches on normative questions about freedom of expression, access to information, and rights to hold different opinions, as well as what type of content could be considered 'factual or truthful' and what type of content is 'false or misleading'. Governments must take care that any regulation is not used to limit freedom of speech or expression or does not unfairly burden smaller platforms, thereby skewing the market further toward the largest and most powerful platforms. Governments therefore must engage with industry and civil society groups to ensure that freedom of speech and other rights of users are protected in ways conducive to democratic engagement.

Increasing transparency and understanding of algorithms

The personalisation of user experiences that online communication platforms can provide also represent a novel and fundamental change to how people engage with information. The algorithms used by some social media platforms may attempt to predict what each user wants to view based on evidence collected from metadata such as location, time spent on specific content, or app usage (Jarboe, 2020[37]). The content that users receive responds to its relevance, rather than to a first-come first-served basis. Also, some algorithms do not distinguish between advertisements, propaganda, disinformation or fact-checked data (DiResta, 2018[38]).

Moreover, algorithms can feed users information that tends to agree with their views and beliefs, thereby risking the creation of "echo chambers" or "filter bubbles" that reinforce and confirm already held beliefs. The European Commission acknowledges how the use of opaque algorithms by widely used platforms has enabled the spread of false information and polarising messages, including through disinformation campaigns (European Commission, 2020[39]). To combat mis- and disinformation, governments could require online platforms to make transparent the parameters of their algorithms. Promoting transparency is a main priority of some government initiatives, such as the UK Draft Online Safety Bill (Minister of State for Digital and Culture, 2021[40]).

In addition to transparency requirements, governments, media and civil society organisations could encourage greater accountability and offer guidance or help build safeguards in the way algorithms are designed to feed users with more diverse content and reduce the spread of mis- and disinformation. As noted by New Zealand Prime Minister Jacinda Ardern, while social media platform algorithms may personalise user experiences, they can also make user experiences more extreme and radicalised. As such, in addition to the "pressing and urgent need" for greater transparency of how algorithms work and the outcomes they deliver, the Prime Minister also called for creating a shared approach for responsible algorithm development and deployment (Prime Minister Jacinda Ardern, 2022[41]). Through the DSA, for example, the European Commission has proposed rules requiring platforms to develop risk assessments and submit to independent audits focused on how their algorithms prioritise and target information as a means of promoting platform accountability (European Commission, 2020[39]).

Increasing authentic online activity and countering foreign interference

Social media platforms have pursued efforts to limit the effect of co-ordinated inauthentic and manipulated behaviour and content, and efforts to increase the authenticity of online engagement is an important avenue. For example, validating that social media platform accounts can be linked to real persons can limit the spread of disinformation, since by eliminating the risk of bots purposefully triggering tipping points of disinformation, platforms can decrease the risks of harm (Gladwell, 2000[42]; The Economist, 2009[43]). Concerning preventing or removing false accounts, governments could require bots to be labelled or provide more guidance on social media platform requirements to boost authentic activity. Governments could also establish guidance on how to identify false ID documents of their own nationality. Of course, any measures taken toward promoting authentic activity and limiting inauthentic engagement should be informed by human rights regarding freedom of expression and the right to privacy and ensure, to the greatest possible extent, individual awareness, participation and control over the use and sharing of personal information.

Other measures aim to address mis- and disinformation campaigns spread by lobbyists or influencers who knowingly or accidentally promote misleading content on certain topics or products (Alderman, 2021[44]; Fisher, 2021[45]; OECD, 2021[46]). For example, during the COVID-19 pandemic, hoax companies posing as public relations firms approached influencers and content producers in France and Germany to lure them into spreading disinformation (Alderman, 2021[44]). This technique has been used by foreign actors seeking to affect domestic affairs, by undemocratic governments against their own population, and by domestic groups pursuing more power (Fisher, 2021[45]).

Addressing this threat requires more transparency in beneficial ownership registries and disclosure of the companies or individuals sponsoring certain content (Khan, 2021[11]). To this end, for example, the Australian Government's Foreign Influence Transparency Scheme seeks to reduce the risk posed by foreign interference, which includes covert, deceptive and coercive efforts to affect political or governmental processes driven or undertaken by or on behalf of foreign actors (Australian Government, 2019[47]). Governments could also improve incentives and enforce duties of care for companies to enhance their levels of transparency regarding the metadata related to groups and actors pushing specific content (Balkin, 2020[27]).

In addition, there is scope for governments to require the disclosure of certain types of digitally manipulated and misleading content. Notably, deepfakes are audio or visual media content that seem authentic, but are in fact synthetic or manipulated. They present a disinformation risk by featuring people saying or doing things they have never said or done (van Huijstee et al., 2021[48]). Deepfakes have been used in fraud schemes to mimic the voice of the CEO of a company and trick an employee to transfer money from the company (Stupp, 2019[49]). They have also been used in disinformation campaigns to target civil rights activists and trigger hate speech (Mezzofiore, 2018[50]; MIT Open Documentary Lab, 2020[51]). The use of deepfakes for disinformation campaigns has been a matter of special concern since it has become easier to make and harder to detect (Sen, 2021[52]). Efforts to tackle deepfakes have entered a technological race to develop AI tools to detect such content (Andrews, 2020[53]; Diaz, 2021[54]).

To counteract this risk, the European Union is proposing to enforce an obligation to disclose that certain content is generated using automated means, with the exception of those cases that have a legitimate purpose (law enforcement and freedom of expression) (European Commission, 2021, p. 5.2.4[55]). Other measures proposed by the European Union include bans on certain applications, legal obligations for deepfake technology providers, and institutionalised support for victims of deepfakes (van Huijstee et al., 2021[48]). In the United States, the proposed Deepfakes Accountability Act aims, inter alia, to require producers of deepfakes to comply with certain digital watermarks and disclose information on content (US Congress, 2019[56]), and the State of California criminalised the use of deepfakes in political advertising (though the law does not apply to news media, parody or satire) (Statt, 2019[57]).

1.2.3. Reducing the economic and structural drivers of mis- and disinformation

Governments can also implement measures that are indirectly connected to mis-and disinformation but nevertheless have significant implications on the underlying structural and economic drivers that affect its spread. The European Commission (EC) has taken this approach through the DSA and the Digital Markets Act (DMA). The EC has focused on creating and maintaining a level playing field for digital services; ensuring responsible behaviour of online platforms; fostering trust, transparency and ensuring fairness on online platforms; and keeping markets open by promoting a fairer business environment and encouraging new services to enter the market.[17]

Analysing and applying lessons from policy responses and approaches undertaken in similar and other rapidly evolving markets can help governments better understand new technologies and implications and may help develop more flexible approaches. Some of the relevant responses to economic and structural drivers to explore further are:

Leveraging competition measures

The innovations brought by digitalisation have introduced substantial consumer benefits, including lower prices, greater accessibility and convenience, more variety, and new products. At the same time, several concerns have been identified with respect to competition in many digital markets, for example in terms of market structure and anticompetitive conduct and merger activity. Digital-intensive sectors have also demonstrated a tendency toward greater market concentration and falling entry rates of new firms (OECD, 2019[58]; OECD, 2022[59]). These trends are a concern because evidence shows that healthy market competition helps spur innovation, as well as promote long-term growth and well-being (OECD, forthcoming[17]).

Competition measures may also play a role in addressing the behaviour (and the incentives) of large online platforms that can breed mis- and disinformation. Indeed, the most influential media companies benefit from large resources, global and networked user bases, and access to vast amount of data that can be used to strengthen network effects, target products and steer consumer decision making. Together, these factors can make it harder for consumers to easily switch services and potentially lead to anti-competitive conduct that can stifle innovation (OECD, forthcoming[17]). Changing market dynamics have also affected the news and information industry – and how people get and share information. These factors can lead to market distortions, as well as dominance of algorithms that may facilitate the amplification of mis- and disinformation. To that end, encouraging new entrants and innovation may spur competition between online platforms with regards to privacy issues, data portability, platform content moderation policies, among others.

To address these risks, market regulation and competition tools could be applied, including:

- **Governments could require large online platforms to ensure a "fair" remuneration to news media companies for the use of their content.** Large online platforms have disrupted advertising markets, drastically altering the incentives of traditional and ad-funded news publishers. Australia adopted a news media bargaining code in February 2021 to address the bargaining power imbalances between large online platforms and news publishers (Australian Competition and Consumer Commission, 2020[60]). The code requires designated digital platforms and news businesses that have indicated an intention to bargain to do so in good faith. If an agreement about remuneration cannot be reached within three months, there is an arbitration mechanism within the framework to resolve disputes over remuneration (Australian Competition and Consumer Commission, 2020[60]).[18] Similarly, in July 2019, France enacted a law transposing the EU directive on copyright and related rights, and providing remuneration criteria for the use of news abstracts on online platforms (Autorité de la concurrence, 2020[61]). In April 2020, the French competition

authority imposed interim measures requiring Google to negotiate in good faith with publishers and news agencies the remuneration due to them under the law (Autorité de la concurrence, 2020[61]).

- **Governments could increase scrutiny of mergers and 'killer acquisitions'**. A 'killer acquisition' is the practice of large tech companies acquiring smaller firms or start-up companies that may represent competition. Large online platforms have been very active in acquiring other businesses. For example, between 2001 and 2021, Google bought 258 companies, meaning they closed more than one deal per month; Facebook employed a similar practice, buying 90 companies in a period of 16 years (2005 to 2021), meaning they closed one deal every two months (Nadler and Cicilline, 2020[62]; American Economic Liberties Project, 2021[63]). Some of these transactions may have chilled innovation, and concentration may have reduced competition for and availability of trustworthy sources of news (Nadler and Cicilline, 2020[62]). Moreover, with few options available for consumers, concentration may also reduce incentives for large online platforms to compete on quality aspects.
- **Governments could consider more structural reforms to address digital platforms' market power.** For instance, competition policies to address mis- and disinformation might lead to preventing social networks from engaging in various functions. Social media platforms also play an important role in advertising and as news distributors (e.g. when a user streams an event live). Separating their social network functions from their ad business could potentially be justified based on promoting media diversity and protecting journalism, though more research is needed to understand the implications and efficacy of such an approach.

Nevertheless, a fragmented legislative landscape for platforms carries costs for firms and consumers, increases uncertainty, and may preclude welfare-enhancing innovation. Promoting a more coherent global approach to identifying and implementing relevant regulation would enhance the effectiveness of government efforts (OECD, forthcoming[17]). Due to how recently many of the policies have been enacted, the complexity and scale of the businesses affected and the rapid pace of technological and market change, furthermore, additional analysis and ongoing engagement with relevant stakeholders will be required to understand the impact of such measures on the spread of false and misleading content and on the economic trade-offs.

Promoting quality and safety in platform design

Similar to other areas of engineering and design, where a common set of technical standards include safety and quality requirements, online platforms could also respond to new and higher-level safety and quality requirements that can mitigate mis- and disinformation risks. Governments' – and societies' – interests in ensuring quality and safety of platforms is related to the scale, utility and impact of the platforms on society. Specific considerations on platform design are related to technical considerations of the choices in design, architecture and engineering that affect what information is shared and how it is spread (Forum on Information and Democracy, 2020[64]). For example, governments could focus efforts on developing specific and quantifiable tests, standards and processes to support responsible business conduct and promote safety of online services, as well as engaging with technical experts and stakeholders to encourage the design of guidelines or codes for social platforms and other digital commons (Forum on Information and Democracy, 2020[64]).

To that end, the OECD Guidelines on Multinational Enterprises, as well as other international standards,[19] can help direct government efforts to create an enabling policy environment for online platforms that supports responsible business conduct (RBC) to facilitate companies' efforts to identify and address negative impacts they may cause or to which they may contribute (OECD, 2014[65]; 2011[66]). Examples of government efforts to apply an RBC approach include the proposed UK Draft Online Safety Bill, which introduces duties to protect "content of democratic importance", as well as measures that require risk assessments and the implementation of due diligence procedures to reduce the risk of harm. Similarly, the

EU's DSA includes due diligence obligations to promote a transparent and safe online environment (European Commission, 2020[36]).

A related issue arises when considering how social media companies' access to significant amounts of personal data can be used as a vehicle for spreading mis- and disinformation. Dis-information campaigns use private data to categorise individuals and seek users that can be targeted through customised messaging (Privacy International, 2021[67]; Khan, 2021[11]). Thus, efforts to regulate how third parties and online platforms can use, or not use, private data will affect the ability of disinformation campaigns to inflict harm. For example, the draft UK Online Safety Bill places special duties regarding both freedom of expression and privacy, in particular by requiring that all service providers protect their users, within the law, against censorship and unwarranted infringements of privacy (Minister of State for Digital and Culture, 2021[40]). Efforts to protect privacy can also be seen in the EU through the implementation of the General Data Protection Regulation, or the Data Protection Law Enforcement Directive (European Commission, 2021[68]).

These areas will require feedback and engagement from a wide range of partners to ensure any regulatory responses meet broader democratic needs and do not unduly stifle innovation. Exploring how to engage effectively with non-governmental partners is also important to consider in the design of regulations in this rapidly evolving and complex space. Indeed, reliance on traditional policy tools and actors is difficult in situations where the direction of technological innovation and impact cannot be determined.

Tools that encourage flexibility and innovation are necessary and require that governments improve their ability to engage with technology developers and users (OECD, 2018[69]). For example, regulatory sandboxes and testbeds are co-creation processes designed to help governments better understand new technologies and regulatory implications, while at the same time giving external partners an opportunity to test new technology and business models (OECD, 2018[69]). For online products and businesses, regulatory models often differ significantly to those in traditional markets and may not fit well with existing frameworks. This challenge is exacerbated by the fast pace of digital transformation, which makes market developments and future policy concerns difficult to predict. In these cases, sandboxes can help increase flexibility and decrease regulatory uncertainty, while enabling closer relationships between regulators and firms (Attrey, Lesher and Lomax, 2020[70]). Codes of conduct and real-time technology assessments are other examples of more flexible solutions.

Broadly, collaborative approaches to developing regulation should focus on drawing on a range of stakeholders that include civil society, fact-checkers, media and academic organisations. Bringing a diverse range of stakeholders together may prove particularly useful given the rapid change, complexity and critical role the sector plays in affecting democratic engagement (Koulolias et al., 2018[71]).

Promoting and maintaining a diverse and independent media sector

Governments can also address mis- and disinformation by strengthening the press and news media sector through encouraging diversity, editorial independence, and ensuring high-quality news provision. Government efforts in this regard will be rooted in part in policies that facilitate an enabling environment for civil society organisations, such as media watchdog groups, and that counteract media capture by special interests (Nelson, 2017[72]). The objective is to discourage market concentration and encouraging innovation and the development of new online platforms owned by different companies to diversify options for users (Balkin, 2020[27]). The creation of new platforms can develop their own social media environment, norms, and communities, which would expand the options for users.

Additionally, governments can apply anti-monopoly measures and foster fair competition to address media capture. Measures could address situations in which large technology or media companies buy independent news agencies, and then use their control to unduly influence media content (Stiglitz, 2017[73]). In addition to private models of news provision, not-for-profit foundations (such as Pro Publica) and public service media (such as the BBC) can play an important role in the information and media space. Notably,

public broadcasters tend to have the highest trust scores, at least in countries where their independence is not in doubt (Newman et al., 2018[74]).

Transparent and independent financial support for high-quality journalism can also combat media capture by large technology companies, who may threaten to withdraw financial incentives from traditional media, for example by blocking subscribers and advertisers. Some countries, such as Austria, Sweden, Norway and the Netherlands, are mitigating this risk by giving subsidies to newspapers that provide political, cultural and economic content (Greenwell, 2017[75]; Schiffrin, 2017[76]). Governments could also support initiatives, both domestically and via international development mechanisms, that provide training to citizen journalists and to traditional outlets on how to manage public engagement to foster participation in news production through citizen and community journalism. Clear and independent oversight can maintain impartiality and help ensure any government support provided to news providers is done in a way that promotes democratic engagement and the free exchange of information.

More broadly, governments can think strategically about how the media can facilitate effective information exchange and the implications of changing technologies on how people get and share news, and to identify ways to build on civil society initiatives to support effective and independent media. For example, the Government of Ireland set up an independent Future of Media Commission, which examined the challenges faced by public service broadcasters, commercial broadcasters, print and online media platforms. The commission also held public dialogues focused on issues related to funding sources, changes in audience behaviour and changes in technology.[20] Furthermore, the Journalism Trust Initiative (JTI), operated by Reporters Without Borders, promotes a healthier information space via indicators for trustworthy, professional and ethical journalism. The JTI can be used by regulators and state actors as an independent, self-regulatory mechanism to allocate subsidies and benefits to media outlets.

1.3. Conclusions

The breadth and depth of the mis-and disinformation challenge call for a wide range of measures driven by a whole-of-government and whole-of-society perspective. Efforts cannot be limited to the national level, as mis- and disinformation transcend territorial boundaries, nor can they be limited to governance responses alone. Current and proposed initiatives must reflect the interdisciplinary and systemic challenges faced and be developed and implemented in partnership with media and civil society organisations focused on a wide range of issues, including legal and human rights, cyber security and privacy, competition, foreign interference, etc. Governments will need to explore policies and initiatives that respond to immediate threats as well as support more resilient societies, in addition to regulatory measures to increase transparency and prevention, and reduce economic and structural drivers of mis- and disinformation.

At present, there is a need to explore how governments, the media and civil society organisations can co-operate more effectively to build strong and meaningful relationships, as well as the mechanisms necessary to help ensure media and civil society organisations remain sufficiently independent to hold governments to account. The nature of the challenges means that only collective action will deliver the changes needed to strengthen information spaces, build trust and make democracy more resilient. Building on existing good practices, an OECD Action Plan has been developed, with concrete actions countries can take to address the issues outlined in this chapter with reforms that are ambitious and impactful: https://www.oecd.org/governance/reinforcing-democracy/.

References

Alderman, L. (2021), "Influencers Say They Were Urged to Criticize Pfizer Vaccine", *The New York Times*,, https://www.nytimes.com/2021/05/26/business/pfizer-vaccine-disinformation-influeners.html (accessed on 3 March 2022). [44]

American Economic Liberties Project (2021), "Big Tech Merger Tracker", American Economic Liberties Project, https://www.economicliberties.us/big-tech-merger-tracker/ (accessed on 3 March 2022). [63]

Andrews, E. (2020), "Using AI to Detect Seemingly Perfect Deep-Fake Videos", Stanford Institute of Human-Centered Artificial Intelligence, https://hai.stanford.edu/news/using-ai-detect-seemingly-perfect-deep-fake-videos (accessed on 5 October 2021). [53]

Attrey, A., M. Lesher and A. Lomax (2020), *The role of sandboxes in promoting Flexibility and innovation in the digital age*, https://goingdigital.oecd.org/data/notes/No2_ToolkitNote_Sandboxes.pdf. [70]

Australian Competition and Consumer Commission (2020), *Draft news media bargaining code*, Australian Competition and Consumer Commission, https://www.accc.gov.au/focus-areas/digital-platforms/draft-news-media-bargaining-code. [60]

Australian Electoral Commission (2019), "AEC encouraging voters to "stop and consider" this federal election", Australian Electoral Commission, https://www.aec.gov.au/media/media-releases/2019/04-15.htm. [20]

Australian Government (2019), *Foreign Influence Transparency Scheme - Factsheet 2*, https://www.ag.gov.au/sites/default/files/2020-03/influence-versus-interference.pdf. [47]

Autorité de la concurrence (2020), "Related rights: the Autorité has granted requests for urgent interim measures presented by press publishers and the news agency AFP (Agence France Presse)", Autorité de la concurrence, https://www.autoritedelaconcurrence.fr/en/press-release/related-rights-autorite-has-granted-requests-urgent-interim-measures-presented-press. [61]

Balkin, J. (2020), "How to Regulate (and Not Regulate) Social Media", https://knightcolumbia.org/content/how-to-regulate-and-not-regulate-social-media. [27]

Balkin, J. (2016), "Information Fiduciaries and the First Amendment", *UC David Law Review*, Vol. 49/4, pp. 1183–1234. [28]

Blastland, M. et al. (2020), "Five rules for evidence communication", *Nature*, Vol. 587/7834, pp. 362-364, https://doi.org/10.1038/d41586-020-03189-1. [14]

Buckmaster, L. and T. Wils (2019), "Responding to fake news", *Parliamentary Library Briefing Book*, https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook46p/FakeNews. [19]

Burns, T. and F. Gottschalk (eds.) (2020), *Education in the Digital Age: Healthy and Happy Children*, Educational Research and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/1209166a-en. [24]

Credibility Coalition (2021), *Our goal: to understand the veracity, quality and credibility of online information*, Credibility Coalition, https://credibilitycoalition.org/ (accessed on 3 March 2022). [10]

Diaz, J. (2021), "Facebook Researchers Say They Can Detect Deepfakes And Where They Came From", NPR, https://www.npr.org/2021/06/17/1007472092/facebook-researchers-say-they-can-detect-deepfakes-and-where-they-came-from?t=1633428935377. [54]

DiResta, R. (2021), *Dancing in the Dark: Disinformation Researchers Need More Robust Data and Partnerships*, National Endowment for Democracy, Global Insights Series, https://www.ned.org/wp-content/uploads/2021/01/Disinformation-Researchers-Robust-Data-Partnerships-DiResta-1.pdf. [30]

DiResta, R. (2018), "Free Speech Is Not the Same As Free Reach", *Wired*, https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/. [38]

Dutch Media Literacy Network (2022), *Dutch Media Literacy Network*, https://netwerkmediawijsheid.nl/over-ons/about-dutch-media-literacy-network/ (accessed on 2022). [26]

Edelman (2022), *Edelman Trust Barometer 2022*, https://www.edelman.com/sites/g/files/aatuss191/files/2022-01/2022%20Edelman%20Trust%20Barometer%20FINAL_Jan25.pdf. [1]

European Commission (2021), "Data protection in the EU", https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en. [68]

European Commission (2021), *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*, COM(2021) 206 final, European Commission, Brussels, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF. [55]

European Commission (2020), *Digital Services Act*, https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital. [36]

European Commission (2020), *European Democracy Action Plan: making EU democracies stronger*, COM(2020) 790 final, https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2250. [39]

European Commission (2018), *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation*, Directorate-General for Communication Networks, Content and Technology, European Commission. [9]

Fisher, M. (2021), "Disinformation for Hire, a Shadow Industry, Is Quietly Booming", *The New York Times*, https://www.nytimes.com/2021/07/25/world/europe/disinformation-social-media.html (accessed on 3 March 2022). [45]

Forum on Information and Democracy (2020), *Working Group on Infodemics: Policy Framework*, https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf (accessed on 7 September 2021). [64]

Funke, D. and D. Flamini (2020), "A guide to anti-misinformation actions around the world", Poynter, https://www.poynter.org/ifcn/anti-misinformation-actions/ (accessed on 3 March 2022). [21]

Gladwell, M. (2000), *The tipping point: how little things can make a big difference*, Little, Brown and Company, Boston. [42]

Greenwell, T. (2017), "Journalism is in peril. Can government help?", *Inside Story*, https://insidestory.org.au/journalism-is-in-peril-can-government-help/. [75]

Jarboe, G. (2020), "How Do Social Media Algorithms Work?", *Search Engine Journal*, https://www.searchenginejournal.com/how-social-media-algorithms-work/380642/. [37]

Khan, I. (2021), *Disinformation and freedom of opinion and expression*, United Nations, General Assembly, Human Rights Council, https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Report-on-disinformation.aspx (accessed on 3 March 2022). [11]

Koulolias, V. et al. (2018), *Combating Misinformation: An Ecosystem in Co-Creation*, ICA, https://www.ica-it.org/images/publications/Combating-misinformation.pdf. [71]

Lesher, M., H. Pawelec and A. Desai (2022), "Disentangling untruths online: Creators, spreaders and how to stop them", *Going Digital Toolkit*, OECD, Paris, https://goingdigital.oecd.org/data/notes/No23_ToolkitNote_UntruthsOnline.pdf. [2]

Matasick, C., C. Alfonsi and A. Bellantoni (2020), "Governance responses to disinformation : How open government principles can inform policy options", *OECD Working Papers on Public Governance*, No. 39, OECD Publishing, Paris, https://doi.org/10.1787/d6237c85-en. [16]

McCallum, K. (ed.) (2021), *Director General Ken McCallum gives annual threat update 2021*, MI5 - The Security Service, London, https://www.mi5.gov.uk/news/director-general-ken-mccallum-gives-annual-threat-update-2021. [29]

Meco, L. and K. Wilfore (2021), *Gendered disinformation is a national security problem*, https://www.brookings.edu/techstream/gendered-disinformation-is-a-national-security-problem/. [5]

Mezzofiore, G. (2018), "No, Emma Gonzalez did not tear up a photo of the Constitution", *CNN*, https://edition.cnn.com/2018/03/26/us/emma-gonzalez-photo-doctored-trnd/index.html (accessed on 5 October 2021). [50]

Miguel, R. (2021), *The battle against disinformation in the upcoming federal election in Germany: Actors, initiatives and tools*, EU DisinfoLab, https://www.disinfo.eu/publications/the-battle-against-disinformation-in-the-upcoming-federal-election-in-germany-actors-initiatives-and-tools/#_ftn14 (accessed on 25 February 2022). [13]

Minister for Communications, Urban Infrastructure, Cities and the Arts (2022), *New disinformation laws*, https://minister.infrastructure.gov.au/fletcher/media-release/new-disinformation-laws. [32]

Minister of State for Digital and Culture (2021), *Draft Online Safety Bill*, Presented to Parliament by the Minister of State for Digital, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf. [40]

Ministry of Culture of Latvia (2021), *Kultūras ministrijas medijpratības kampaņa "Zini, kā neuzķerties!" sasniegusi teju 895 000 unikālo lietotāju*, https://www.km.gov.lv/lv/jaunums/kulturas-ministrijas-medijpratibas-kampana-zini-ka-neuzkerties-sasniegusi-teju-895-000-unikalo-lietotaju. [23]

MIT Open Documentary Lab (2020), "Brandi Collins-Dexter, Jane Lytvynenko & Karen Hao | Deepfakes, parody, and disinformation", http://opendoclab.mit.edu/presents/brandi-collins-dexter-jane-lytvynenko-karen-hao-deepfakes-parody-disinformation/. [51]

Mon Opinion (2021), "Concertations - MyOpinion", https://monopinion.belgium.be/processes (accessed on 3 March 2022). [22]

Nadler, J. and D. Cicilline (2020), *Investigation of Competition in Digital Markets*, United States of America: Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary, https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf?utm_campaign=4493-519 (accessed on 3 March 2022). [62]

Nelson, M. (2017), "What is to be done? Options for combating the menace of media capture", in Schiffrin, A. (ed.), *In the Service of Power: Media Capture and the Threat to Democracy*, pp. 143–162, National Endowment for Democracy, Washington, DC, https://www.cima.ned.org/wp-content/uploads/2017/08/CIMA_MediaCaptureBook_F1.pdf#page=151. [72]

Newman, N. et al. (2018), *Reuters Institute Digital News Report – 2018*, Reuters Institute for the Study of Journalism, http://www.digitalnewsreport.org/. [74]

OECD (2022), *OECD Handbook on Competition Policy in the Digital Age*, OECD, Paris, https://www.oecd.org/daf/competition-policy-in-the-digital-age/. [59]

OECD (2021), *An International Collaboration to tackle Misinformation with Behavioural Insights*. [18]

OECD (2021), *Lobbying in the 21st Century: Transparency, Integrity and Access*, OECD Publishing, Paris, https://doi.org/10.1787/c6d8eff8-en. [46]

OECD (2021), *OECD Report on Public Communication: The Global Context and the Way Forward*, OECD Publishing, Paris, https://doi.org/10.1787/22f8031c-en. [4]

OECD (2021), "Transparency reporting on terrorist and violent extremist content online : An update on the global top 50 content sharing services", *OECD Digital Economy Papers*, No. 313, OECD Publishing, Paris, https://doi.org/10.1787/8af4ab29-en. [77]

OECD (2020), "Youth and COVID-19: Response, recovery and resilience", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/c40e61c6-en. [12]

OECD (2019), *Going Digital: Shaping Policies, Improving Lives*, OECD Publishing, Paris, https://doi.org/10.1787/9789264312012-en. [58]

OECD (2018), *OECD Regulatory Policy Outlook 2018*, OECD Publishing, Paris, https://doi.org/10.1787/9789264303072-en. [35]

OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://doi.org/10.1787/sti_in_outlook-2018-en. [69]

OECD (2017), "Trust in peer platform markets: Consumer survey findings", *OECD Digital Economy Papers*, No. 263, OECD Publishing, Paris, https://doi.org/10.1787/1a893b58-en. [34]

OECD (2014), *Annual Report on the OECD Guidelines for Multinational Enterprises 2014: Responsible Business Conduct by Sector*, OECD Publishing, Paris, https://doi.org/10.1787/mne-2014-en. [65]

OECD (2012), "Recommendation of the Council on Regulatory Policy and Governance", *OECD Legal Instruments*, OECD/LEGAL/0390, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0390. [7]

OECD (2011), *OECD Guidelines for Multinational Enterprises, 2011 Edition*, OECD Publishing, Paris, https://doi.org/10.1787/9789264115415-en. [66]

OECD (forthcoming), *Committee on Digital Economy Policy (CDEP) Ministerial Meeting Theme 1: Digital enablers of the global economy*. [17]

OECD (forthcoming), *Committee on Digital Economy Policy (CDEP) Ministerial Meeting Theme 2: Building better societies*. [3]

OECD (forthcoming), *Principles of Good Practice for Public Communication Responses to Mis- and Disinformation*. [8]

Prime Minister Jacinda Ardern (2022), *Prime Minister Ardern Commencement Address to Harvard University*, https://news.harvard.edu/gazette/story/2022/05/jacinda-arderns-forceful-warning-democracies-can-die/. [41]

Privacy International (2021), *The UN Report on Disinformation: a role for privacy*, Privacy International, https://privacyinternational.org/news-analysis/4515/un-report-disinformation-role-privacy (accessed on 22 September 2021). [67]

Roozenbeek, J. and S. van der Linden (2019), "Fake news game confers psychological resistance against online misinformation", *Palgrave Communications*, Vol. 5/1, https://doi.org/10.1057/s41599-019-0279-9. [15]

Schiffrin, A. (2017), "How Europe fights fake news", *Columbia Journalism Review*, https://www.cjr.org/watchdog/europe-fights-fake-news-facebook-twitter-google.php. [76]

Sen, D. (2021), "Explained: Why is it becoming more difficult to detect deepfake videos, and what are the implications?", https://indianexpress.com/article/explained/explained-deepfake-video-detection-implications-7247635/ (accessed on 5 October 2021). [52]

Stamos, A. et al. (2019), "Combatting State-Sponsored Disinformation Campaigns from State-aligned Actors", in McFaul, M. (ed.), *Securing American Elections: Prescriptions for Enhancing the Integrity and Independence of the 2020 U.S. Presidential Election and Beyond*, Stanford Cyber Policy Center, p. 48–49, https://fsi.stanford.edu/publication/securing-american-elections-prescriptions-enhancing-integrity-and-independence-2020-us. [33]

Statt, N. (2019), "China makes it a criminal offense to publish deepfakes or fake news without disclosure", *The Verge*, https://www.theverge.com/2019/11/29/20988363/china-deepfakes-ban-internet-rules-fake-news-disclosure-virtual-reality. [57]

Stiglitz, J. (2017), "Toward a taxonomy of media capture", in Schiffrin, A. (ed.), *In the Service of Power: Media Capture and the Threat to Democracy*, pp. 9–18, National Endowment for Democracy, Washington, DC, https://www.cima.ned.org/wp-content/uploads/2017/08/CIMA_MediaCaptureBook_F1.pdf#page=151. [73]

Stupp, C. (2019), "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case", *The Wall Street Journal*, https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 (accessed on 5 October 2021). [49]

Suarez-Alvarez, J. (2021), "Are 15-year-olds prepared to deal with fake news and misinformation?", *PISA in Focus*, No. 113, OECD Publishing, Paris, https://doi.org/10.1787/6ad5395e-en. [25]

The Economist (2009), *Tipping point*, The Economist, https://www.economist.com/news/2009/04/20/tipping-point. [43]

US Congress (2019), *H.R.3230 - 116th Congress (2019-2020): Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019*, Congress.gov, Library of Congress, https://www.congress.gov/bill/116th-congress/house-bill/3230 (accessed on 12 October 2021). [56]

US Government Office of the President (2015), *Executive Order - Promoting Private Sector Cybersecurity Information Sharing*, https://obamawhitehouse.archives.gov/the-press-office/2015/02/13/executive-order-promoting-private-sector-cybersecurity-information-shari. [31]

van Huijstee, M. et al. (2021), *Tackling deepfakes in European policy*, European Parliamentary Research Service, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf. [48]

Wardle, C. and H. Derakshan (2017), *Information Disorder: Towards an interdisciplinary framework for research and policy making*, http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf. [6]

Notes

[1] European Commission (2020, p. 18[39]), European Democracy Action Plan: making EU democracies stronger, COM(2020) 790 final.

[2] See also (Lesher, Pawelec and Desai, 2022[2]) for a typology of untrue content online, including contextual deception, propaganda and satire.

[3] https://stratcomcoe.org/

[4] Public communication is distinct from political communication, which is linked to elections or political parties, and is understood as the government function to deliver information, listen and respond to citizens in the service of the common good (OECD, 2021[4]).

[5] The definitions developed by governments may be informed by existing work; see for example Lesher, Pawelec and Desai (2022[2]), "Disentangling untruths online: Creators, spreaders and how to stop them", *Going Digital Toolkit Note*, No. 23, https://goingdigital.oecd.org/data/notes/No23_ToolkitNote_UntruthsOnline.pdf

[6] Order of the Lithuania's Government, No.955, 26 August 2020.

[7] For more, see OECD Centre for Educational Research and Innovation (CERI)'s 21st Century Children project, in particular "21st Century Children: Digital Risks and Resilience".

[8] Recent data from PISA showed that an average of 54% of students in OECD countries reported being trained at school on how to recognise whether information is biased or not. Among OECD countries, more than 70% of students reported receiving this training in Australia, Canada, Denmark, and the United States. However, less than 45% of students reported received this training in Israel, Latvia, the Slovak Republic, Slovenia, and Switzerland (Suarez-Alvarez, 2021[25]).

[9] https://medialukutaitosuomessa.fi/mediaeducationpolicy.pdf

[10] For more information, see: https://www.oecd.org/digital/vtrf/

[11] https://www.acma.gov.au/report-government-adequacy-digital-platforms-disinformation-and-news-quality-measures

[12] For more information, see: https://www.coons.senate.gov/news/press-releases/coons-portman-klobuchar-announce-legislation-to-ensure-transparency-at-social-media-platforms

[13] For an overview of TVEC-related transparency reporting of leading online content-sharing services, see: OECD (2021[77]), "Transparency reporting on terrorist and violent extremist content online: An update on the global top 50 content sharing services", *OECD Digital Economy Papers*, No. 313, OECD Publishing, Paris, https://doi.org/10.1787/8af4ab29-en.

[14] See also the 2020 Santa Clara Principles 2.0 and the recommendations on transparency laid out in the Forum on Information & Democracy's Working Group on Infodemics 2020 Policy Framework (Forum on Information and Democracy, 2020[64]).

[15] https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

[16] https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2585

[17] For more information, see: https://digital-strategy.ec.europa.eu/en/policies/online-platforms

[18] See also: https://www.accc.gov.au/system/files/Final%20legislation%20as%20passed%20by%20both%20houses.pdf

[19] For example, the UN Guiding Principles on Business and Human Rights and the ILO Tripartite Declaration of Principles Concerning Multinational Enterprises and Social Policy.

[20] For more information, see: https://futureofmediacommission.ie/