



Combatting COVID-19 disinformation on online platforms

16 July 2020

Disinformation and misinformation about COVID-19 is quickly and widely disseminated across the Internet, reaching and potentially influencing many people. This policy brief derives four key actions that governments and platforms can take to counter COVID-19 disinformation on platforms, namely: 1) supporting a multiplicity of independent fact-checking organisations; 2) ensuring human moderators are in place to complement technological solutions; 3) voluntarily issuing transparency reports about COVID-19 disinformation; and 4) improving users' media, digital and health literacy skills.



Key messages

- As the COVID-19 global pandemic continues, many countries are emerging from confinement, and as a result are focusing on how to keep people safe and healthy and prevent a “second wave”. A key aspect of this effort is ensuring the accurate and timely delivery of health-related information.
- In the past six months, an outbreak of disinformation (i.e. false or misleading information, deliberately circulated to cause harm) about COVID-19 has spread quickly, widely and inexpensively across the Internet, endangering lives and hampering the recovery. As effective new treatments and vaccines become available, disinformation could hinder uptake and further jeopardise countries’ efforts to overcome the COVID-19 pandemic.
- Online platforms are a key channel for this disinformation, but they also play an important role in limiting its circulation. Many online platforms have taken bold actions during the pandemic, including stepping up their support of, and reliance on, independent fact checking organisations as well as on automated content moderation technologies to reinforce their efforts to detect, remove, and otherwise counter false, misleading and potentially harmful content about COVID-19. Some platforms have also banned ads for medical masks and respirators.
- Platforms should be encouraged to continue and enhance these practices in support of successful “re-openings” while ensuring that users’ rights to privacy and freedom of expression are preserved. The latter concern requires an eye for contextual nuance that is highly challenging for algorithms. Therefore, more human moderators are also needed to complement automated approaches.
- Online platforms are also a key channel for distributing accurate information about COVID-19, but they should not be expected to act alone, either in their efforts to make accurate information available or to counter disinformation. Co-operation and co-ordination among companies, governments, national and international health authorities, and civil society is crucial.
- People need digital and health literacy skills to navigate and make sense of the COVID-19 content they see online, to know how to verify its accuracy and reliability, and to be able to distinguish facts from opinions, rumours and falsehoods.
- The practices and collaborations undertaken by online platforms in response to disinformation about COVID-19, including possible future actions to enhance transparency, offer a strong foundation for addressing other forms of disinformation.

The COVID-19 crisis stems from an “infodemic” as well as a pandemic

As the coronavirus (COVID-19) pandemic continues, so does the volume of related news information, leading to what has been called an “infodemic”. While the issues of misinformation (the spread of false information, regardless of whether there is an intent to deceive) and disinformation (the deliberate spread of false or misleading information with an intent to deceive) have been widely studied and discussed, COVID-19 throws into sharp relief the challenges and consequences of this important policy area.

Disinformation and misinformation about COVID-19 is quickly and widely disseminated across the Internet, reaching and potentially influencing many people. There is no single source, with different actors driven by largely dissimilar motives producing and propagating false and deceptive information to further their own goals and agendas. For example, some people are using online platforms to spread conspiracy theories, claims that COVID-19 is a foreign bioweapon, a partisan sham, the product of 5G technology, or part of a greater plan to re-engineer the population. Others are spreading rumours of supposedly secret cures such as drinking diluted bleach, eating bananas or turning off one’s electronics. Others are using the COVID-19



pandemic for financial benefit, selling test kits, masks and treatments on the basis of false or deceptive claims about their preventive or healing powers.

The harmful effects of disinformation about COVID-19 cannot be overstated. [Data](#) from Argentina, Germany, Korea, Spain, the United Kingdom and the United States show that about one in three people say they have seen false or misleading COVID-19-related information on social media. [Research](#) has also shown that COVID-19 disinformation is disseminated significantly more widely than information about the virus from authoritative sources like the World Health Organization (WHO) and the United States Centers for Disease Control and Prevention. By calling into question official sources and data and convincing people to try bogus treatments, the spread of dis- and misinformation has led people to ingest fatal home cures, ignore social distancing and lockdown rules, and not to wear protective masks, thereby undermining the effectiveness of containment strategies.

The harmful effects of disinformation, however, go beyond public health concerns. For example, in the United Kingdom, the false [claim](#) that radio waves emitted by 5G towers make people more vulnerable to COVID-19 has resulted in over 30 acts of arson and vandalism against telecom equipment and facilities, as well as around 80 incidents of harassment against telecom technicians. In the Netherlands, 15 such acts of arson have been recorded. Similarly, in Australia, the European Union and the United States, the spread of disinformation framing minorities as the cause of the pandemic has fuelled animosity against ethnic groups, motivating a rise in discrimination and incidents of violence in what has been characterised as “[coronaracism](#)”.

Understanding how COVID-19 disinformation spreads is essential for crafting effective responses

People increasingly access information, news and content through the Internet on news aggregator sites, social media platforms, and video-sharing platforms. Unlike traditional media such as TV news or newspapers, online platforms automate and personalise content by leveraging data concerning users’ past online activity, social connections, location, etc. Attention grabbing headlines with sensationalist content can attract even the savviest Internet users and studies have shown they tend to generate [more user engagement](#). As a result, content personalisation algorithms can repeatedly expose people to the same or similar content and ads even on the basis of disinformation. Moreover, some platforms present news content alongside non-news, ads and user-generated content. This can make it difficult for users to distinguish reliable news. Unless measures are implemented to remove, de-emphasise or otherwise limit the circulation of false and misleading information, it can spread quickly.

Another important consideration is the extent of the influence exerted by different sources of disinformation. COVID-19 disinformation moves top-down, from politicians, celebrities and other prominent figures, as well as bottom-up, from ordinary people. However, the impact of these two sources of disinformation differs dramatically. [Empirical research](#) shows that top-down disinformation constitutes only 20% of all misleading claims about COVID-19, but generates 69% of total social media engagement. Conversely, whilst the majority of COVID-19 disinformation on social media is created from the bottom-up by ordinary users, most of these posts appear to have significantly less engagement. Accordingly, influential public figures bear more responsibility in efforts to address disinformation claims and rumours about COVID-19.

Initiatives from online platforms to tackle COVID-19 disinformation

As scientists and governments continue to work on a treatments and vaccines, it is imperative that online platforms, governments and national and international health organisations work together to curb the spread of false and misleading information about COVID-19.



Some platforms have taken important steps, such as directing users to official sources when searching for COVID-19 information, banning ads for medical masks and respirators, and reinforcing their efforts to detect and remove false, misleading and potentially harmful content related to COVID-19, including by terminating online shops or removing listings that make false or deceptive claims about products preventing or curing COVID-19. Moreover, Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter and YouTube have published a [joint statement](#) on their collaboration with government healthcare agencies to combat fraud and disinformation about COVID-19. In general, there are three main types of collaborative efforts between platforms and public health authorities:

- **Highlighting, surfacing and prioritising content from authoritative sources.** Platforms like Facebook, Instagram, TikTok and Pinterest are redirecting users to information from the WHO in response to searches for information on and hashtags associated with COVID-19. Similarly, Google launched a one-stop-shop [COVID-19 microsite](#) and an “SOS Alert”, which directs people searching for “coronavirus” to news and other content from the WHO. YouTube features videos from public health agencies on its homepage and highlights content from authoritative sources in response to searches for information on COVID-19. Twitter, in turn, features a [COVID-19 event page](#) with the latest information from trusted sources on top of users’ timelines. Snapchat has also partnered with the WHO to create filters and stickers that provide guidance on how to prevent the spread of the virus.
- **Co-operation with fact-checkers and health authorities to flag and remove disinformation.** Facebook co-operates with third-party fact checkers to debunk false rumours about COVID-19, label that content as false and notify people trying to share such content that it has been verified as false. [Facebook](#) partnered with the International Fact-Checking Network (IFCN) to launch a USD 1 million grant programme to increase their capacity and has been removing content flagged by public health authorities, including “claims related to false cures or prevention methods — like drinking bleach cures the coronavirus — or claims that create confusion about health resources that are available”. Likewise, reports note [Google](#) donated USD 6.5 million to fact-checkers focusing on coronavirus. In turn, [Twitter](#) broadened the definition of harm on its platform to address content that goes directly against guidance from authoritative sources of global and local public health information.
- **Offering free advertising to authorities.** Facebook, Twitter and Google have granted [free advertising credits](#) to the WHO and national health authorities to help them disseminate critical information regarding COVID-19.

There is no easy way to solve the COVID-19 disinformation problem entirely or permanently

The initiatives above are important, as they both facilitate access to accurate, reliable information about COVID-19 and possibly save lives. However, they can go only so far. Two developments highlight that the COVID-19 disinformation crisis does not have an easy fix:

- **Enhanced reliance on automated content moderation.** As a result of the pandemic and lock-down measures, online platforms like Facebook, Google and Twitter have faced a [shortage](#) of human moderators and consequently increased their reliance on automated monitoring technologies to flag and remove inappropriate content. However, these technologies cannot catch every instance of disinformation. Furthermore, acting as a content moderator may have legal implications for platforms under existing communications legislation.¹ Accordingly, to reduce the likelihood of missing important misleading or false claims and rumours, they tend to be programmed to err on the side of caution. That raises the risk of false positives. Thus [YouTube](#) has noted that, as a consequence of increased reliance on automated moderation systems, “users and creators may see increased video removals,

¹ See, for example, Section 230 of the United States’ Communications Decency Act ([47 U.S.C. § 230](#)).



including some videos that may not violate [its] policies”. Indeed, there have been [multiple reported incidents](#) of automated monitoring systems flagging COVID-19 content from reputable sources as spam. Therefore, there is a risk that, to suppress COVID-19 disinformation, content moderation without adequate human oversight may limit the availability of, and accessibility to, reliable information about COVID-19. In addition, without proper transparency, accountability and due process, automated content moderation is more likely to have a chilling effect on free speech, thus adding a layer of complexity to the infodemic.

- **Banning ads that exploit the crisis.** [Facebook and Instagram](#) banned ads suggesting that a product is a guaranteed cure or that it prevents people from contracting COVID-19 as well as ads and commerce listings for masks, hand sanitisers, surface disinfecting wipes and COVID-19 testing kits. [Twitter](#) implemented comparable measures under its Inappropriate Content policy. Similarly, [Google and YouTube](#) prohibit any content, including ads, that seeks to capitalise on the pandemic, and on this basis they have banned ads for personal protective equipment. These measures deter scammers seeking to sell fake COVID-19-related products and help to protect consumers against price gouging, but they may also make it marginally more difficult for people to find and buy hygiene products online. In any case, dis- and misinformation remain difficult to stamp out altogether without infringing on users’ freedom of expression and/or privacy rights. Furthermore, some online platforms and content producers are still profiting from the ads that are displayed on the misleading and false content that escapes the platforms’ moderation efforts. For example, the [Tech Transparency Project](#) found several misleading coronavirus videos – promising cures and preventative treatments - on a major video sharing platform featuring ads from various advertisers, and [EUvsDiSiNFO](#) recently concluded that “major platforms continue to monetise disinformation and harmful content on the pandemic [...] e.g. by hosting online ads on pages that misrepresent migrants as the cause of the virus, promote fake cures or spread conspiracy theories about the virus.” Furthermore, after conducting experiments on several social media platforms, the [EU DisinfoLab](#) concluded that “the moderation process of [...] debunked conspiratorial content has been slow and inconsistent across platforms [...]”.

A necessary first step towards a more effective, long-term solution is to gather evidence systematically. Regular transparency updates from online platforms about the COVID-19 disinformation that is showing up and being viewed, and how the platforms are detecting and moderating it, would help researchers, policymakers, and the platforms themselves to identify ways to make improvements. The [European Commission](#) recently called on platforms to voluntarily issue such transparency reports on a monthly basis. If that became a requirement, and particularly if it grew to include other jurisdictions, timely international, multi-stakeholder co-operation and co-ordination on a common reporting standard could be very valuable to reduce inefficiencies and expense. Furthermore, a common approach across companies and countries could facilitate global, cross-platform comparisons. The OECD is an ideal forum for that type of project and is already leading such an [effort](#) with regard to voluntary transparency reporting on terrorist and violent extremist content online.

Improvement will likely also need to include online platforms taking on a more watchful role by investing in more and improved vetting and fact-checking. This could involve hiring more reviewers, providing additional support to multiple, external fact checking organisations, continuing to develop automated content moderation systems, and further prioritising quality, reliable content. In this scenario, misleading and fake content could continue to be uploaded, but the speed and magnitude of its dissemination would be [greatly reduced](#).

However, the platforms cannot be alone in this effort. Governments, public health organisations, international organisations, civil society, media organisations and tech companies must join forces to combat disinformation. That includes working together to improve people’s media, digital and health literacy skills. If users are able to accurately identify the source of what they read, who wrote or paid for it, and why the information they see is shown to them, they are less likely to be persuaded and manipulated by misleading claims and rumours.



Key policy recommendations

The spread of COVID-19 disinformation is an intricate issue that warrants co-operation, co-ordination and trust amongst online platforms, governments and national and international health organisations, as well as a mixture of carefully-balanced measures. Although a number of responses are already taking place, more can be done. The following considerations are intended to guide stakeholders in their efforts to arrive at ever more effective solutions to COVID-19 disinformation.

Support a multiplicity of independent fact-checking organisations. Disinformation about the coronavirus calls into question the actions, recommendations and competence of governments, doctors, scientists and national and international health organisations. Because it damages their credibility, disinformation also hampers their ability to debunk or refute false and misleading information, including the very information that casts them in a negative light in the first place. Independent fact-checkers are able to provide unbiased analysis of information whilst helping online platforms identify misleading and false content; governments and international authorities can help by supporting and relying on their analyses to restore public trust. Facebook's and other platforms' increased reliance on, and financial assistance to, fact-checkers are examples to be followed not only by other tech companies, but by the public sector, as well. Platforms might also consider labelling content that has successfully passed fact-checks by two or more independent fact-checking organisations with a trust mark. Such trust marks have been [shown](#) to be effective in increasing consumers' trust in the e-commerce context, so they might also be a useful tool for increasing trust in online information about COVID-19.

Ensure human moderators are in place to complement technological solutions. Whereas automated monitoring systems are an important tool to detect and remove COVID-19 disinformation, content moderation also requires human intervention, especially when more nuanced decisions are required. This is a complicated problem with no easy solution, particularly during the pandemic, because while sending content moderators back to work too soon would be an unacceptable public health risk, making them working from home gives rise to privacy and confidentiality concerns. Nevertheless, online platforms should look for alternatives to have access to the required content moderation workforce. For those platforms with adequate resources, hiring (more) moderators as full-time staff may help.

Voluntarily issue transparency reports about COVID-19 disinformation. Regular updates from online platforms about the nature and prevalence of COVID-19 disinformation, and the actions that platforms are taking to counter it, would enable better, more evidence-based approaches by both public and private sector stakeholders. If multiple jurisdictions consider requiring such reports, international co-operation and co-ordination through multi-stakeholder processes will be needed to avoid a fragmented array of reporting standards that would needlessly raise costs and complicate global, cross-platform comparisons.

Improve users' media, digital and health literacy skills. People need the skills to navigate and make sense of what they see online safely and competently, and to understand why it is shown to them. This includes knowing how to verify the accuracy and reliability of the content they access and how to distinguish actual news from opinions or rumours. To this end, collaboration between platforms, media organisations, governments and educators is critical, as are efforts to improve health literacy. The recent [partnership](#) between the European Union, UNESCO and Twitter to promote media and information literacy amid the COVID-19 disinformation crisis is a laudable initiative that should be replicated by other platforms and relevant stakeholders.



Further reading

Moreira, L. (2018), "Health literacy for people-centred care: Where do OECD countries stand?", *OECD Health Working Papers*, No. 107, OECD Publishing, Paris, <https://doi.org/10.1787/d8494d3a-en>.

OECD (2020), "Protecting online consumers during the COVID-19 crisis", OECD, Paris. <http://www.oecd.org/coronavirus/policy-responses/protecting-online-consumers-during-the-covid-19-crisis-2ce7353c/>.

OECD (2020), "Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new coronavirus", OECD, Paris, <https://www.oecd.org/coronavirus/en/#policy-responses>.

OECD (2019), *An Introduction to Online Platforms and their Role in the Digital Transformation*, OECD Publishing, Paris, <https://doi.org/10.1787/53e5f593-en>.

This paper is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.

www.oecd.org/sti – sti.contact@oecd.org –  [@OECDInnovation](https://twitter.com/OECDInnovation) – <http://oe.cd/stinews>

