EDMO

European Digital Media Observatory

Assessing the Signatories' Covid-19 Reports

/

Executive Summary

M 18

Project number:	SMART 2019/1087
Project Acronym:	EDMO
Project title:	European Digital Media Observatory
Start date of the project:	01/06/2020
Duration of the project:	30
Period covered by the periodic description:	From 03/2020 to 05/2021
Periodic description:	..
Date of submission of the periodic report:	29/11/2021
Project website address:	https://edmo.eu/

Table of Contents

1 Executive Summary: Assessing the CoP Signatories` COVID-19 Reports[1]

Maren Beaufort, Centre for Media Pluralism and Media Freedom

1.1 Introduction & Research Issue

The COVID-19-pandemic has been accompanied by an unprecedented 'infodemic'[2], which – like the pandemic itself – is a global challenge. The European Commission speaks of a 'flood' of information about the virus[3], which is often false or inaccurate and spreads quickly over social media. The World Health Organization (WHO) warns that this inundation can create confusion and distrust, as well as 'undermine an effective public health response'. With its Communication of 10/06/2020 titled 'Tackling COVID-19 disinformation – Getting the facts right', the Commission therefore established a COVID-19 monitoring programme for six signatories of the Code of Practice on Disinformation: Facebook, Google, Microsoft, Mozilla, TikTok and Twitter. These signatories were asked to report monthly on their strategies and actions to combat COVID-19 disinformation, focusing on the following:

- *initiatives to promote authoritative content at EU and Member State level.* Platforms should provide data on the actions taken to promote information from national and international health agencies, national and EU authorities, as well as professional media.

- *Initiatives and tools to improve users` awareness.* Platforms should provide data about the implementation of their policies to inform users when they interact with disinformation.

- *Manipulative behaviour*. Platforms should report all instances of social media manipulation and malign influence operations or coordinated inauthentic behaviour detected on their services. Platforms should also cooperate with EU Member States and institutions in order to facilitate the assessment and attribution of disinformation campaigns and influence operations.

- *Data on flows of advertising linked to COVID-19 disinformation.* The Signatories of the Code should provide data, broken down by Member State where possible, on policies undertaken to limit advertising placements related to disinformation on COVID-19 on their own disinformation

---

[1]   Maren Beaufort, Centre for Media Pluralism and Media Freedom, Thanks to Konrad Bleyer-Simon and Roberto Sequerra Koogan Breitman for their participation.

[2]   The term has been used and described by the WHO as follows: 'infodemics are an excessive amount of information about a problem, which makes it difficult to identify a solution. They can spread misinformation, disinformation, and rumours during a health emergency. Infodemics can hamper an effective public health response and create confusion and distrust among people.' https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4

[3]   The European Commission's daily media monitoring points to hundreds of thousands of COVID-19 related articles (https://ec.europa.eu/jrc/en/covid-19-media-surveillance?solrsort=ds_created%20desc) and millions of mentions on social media worldwide.

on their own services. Platforms and advertising network operators should also provide such data on policies to limit advertising placements on third-party websites using disinformation around COVID-19 to attract advertising revenues. To this end, platforms should consider using trustworthiness indicators developed by independent fact-checkers/researchers or other stakeholders (e.g., GDI, Newsguard). Many platforms have already responded to the Commission's call in this regard by removing millions of misleading advertisements for illegal or unsafe products. [4]

● *Fact-checkers.* The Commission also invites the platforms to expand and intensify their cooperation with fact-checkers and to actively provide organisations in all EU Member States – as well as in the EU's neighbourhood – with access to their fact-checking programmes in all languages.

● *Vaccines.* In December 2020, along with reported progress towards the distribution of COVID-19-vaccines, the European Commission has asked the platforms to include in their monthly monitoring reports a specific section highlighting actions taken to fight disinformation and misinformation around COVID-19-vaccines.

Since the declaration of the global pandemic, the expertise of researchers and experts on public health issues has been in greater demand than ever. Previous experiences with the COVID-19-crisis have highlighted in particular the importance of promoting information from reliable sources, that decisions should be based on the advice of scientists and health professionals, and that it is essential to maintain a democratic debate. They also highlight the challenge of dealing with a changing online environment.

Empowering citizens to critically analyse online information is therefore key to addressing disinformation. This requires the promotion of citizens' media and information literacy and would support their empowerment. This implies, for consumer protection issues, access to resources in consumers' own languages.

As a pillar of cooperation and exchange of information between the major platforms and the European Commission, a reporting system on the COVID-19-actions has been set up. It obliges the platforms to provide regular information on their effectiveness and on new trends until the end of the COVID-19-crisis, as well as to coordinate with actors at EU-level and worldwide (e.g., WHO).

---

[4]

In May 2021, the European Regulators Group for Audiovisual Media Services (ERGA) published an interim report stating that signatories had intensified their efforts on COVID-19-disinformation and that the overview of actions in the reports appeared useful. There was criticism that the lack of country-specific data hindered monitoring and evaluation of the effectiveness of the reported initiatives. In September 2021, the DCU Institute for Future Media, Democracy and Society commissioned by the Broadcasting Authority of Ireland presented its report, 'Assessing the implementation of EU Code of Practice on disinformation in relation to COVID-19', which provides a systematic analysis of the transparency reports submitted by Facebook, Google, Microsoft, Mozilla, TikTok, and Twitter. The researchers analysed 47 reports submitted between August 2020 and April 2021. The research involved manual coding to identify individual actions and an automated textual analysis to identify themes and patterns. In addition, the report presents findings from case studies investigating the implementation of the reported COVID-19-policies on Facebook and TikTok. In the end, nine recommendations were made, which mainly refer to the need to improve the reporting format, as the lack of standardisation of this system had been criticised.

According to the European Commission's June 2020 Joint Communication on 'Tackling COVID-19 disinformation', the platforms should also agree with the European Digital Media Observatory (EDMO) and relevant stakeholders on a framework to provide privacy-compliant access to the relevant data in order to maximise the analytical capacities with regard to the disinformation phenomenon. This would allow a better understanding of disinformation threats and trends.

The purpose of this analysis, 'Assessing the Signatories` Covid-19 Reports', is – in accordance with Task 5 of the EDMO-project and in view of the framework required – to evaluate to what extent the COVID-19-reporting of the signatories is useful as a basis for assessing the impact of the actions of the European Commission (on the one hand in relation to the objectives of the 'Tackling COVID-19 disinformation' document and on the other hand in relation to the Code of Practice on Disinformation). The following sections provide a summary of the research methodology and the results of the analysis. The report concludes with specific recommendations.

1.2   Methodology to Analyse Signatories` Reporting

i.   Reports

The analysis is based on the 57 so-called 'Covid-19 Reports' from Facebook (11), Google (11), Microsoft (11), Mozilla (2), TikTok (11), and Twitter (11), submitted between August 2020 and June 2021 as part of the 'Fighting COVID-19 Disinformation Monitoring Programme', reports which are freely accessible

online. The platforms were asked to submit a baseline report in August 2020, in which they announced the actions implemented by the end of July 2020. Thereafter, updates were requested monthly. The reporting period of the disinformation monitoring programme, which was originally intended to end in December 2020, was extended to June 2021.

ii. Coding

The reports differ considerably in form and content, so that manual coding was necessary for the systematic analysis. The main focus of this analysis was the qualitative assessment of reports, although quantitative data was also collected and evaluated. This data is also available for further evaluation. In addition to collecting key data on the form of the reports, individual action was used as a reference point for the content analysis and categorised as follows, using a specially created codebook:

- Action (text extract for description; new initiative: yes/no/extension; cooperation: yes/no, details; policy relevant: yes/no/details)
- Regulatory scope (far-reaching/corporate policy changing; specific/punctual)
- Geographical Scope/roll-out & audiences (global, EU/non-EU, member states, languages, specific target group(s), reach of these target group(s)
- Temporal scope
- Content category (own content; original third-party content; comment section; other)
- Target
- Use of artificial intelligence (AI): yes/no/details
- Reflection on legal frameworks and fundamental rights (in particular, data protection, Article 10 of the European Convention on Human Rights and the principle of freedom of expression): yes/no
- Monitoring the impact (engagement metrics: platform, user (e.g., clicks, views, impressions), partner, others; aggregate level: temporal, geographic, others; independent audit: yes/no, details; regular check-ups: yes/no, details)
- Forecast (text extract for description, quantitative key figures: yes/no, details)

Several aspects were examined. Firstly, whether and to what extent each of these specifications is reflected in the reporting on each action. Secondly, we relied on the approach of a target performance analysis to access compliance with the requirements of the European Commission's information request of 10/06/2020, 'Tackling COVID-19 disinformation – Getting the facts right' and with those imposed on the platforms by the Code of Practice on Disinformation. Thirdly, to what extent control mechanisms had been installed that went beyond the reporting instrument in question. Regarding the

quantitative indicators, it was examined whether and to what extent the reports contain clear, regularly collected quantitative key figures that can be compared over time. Regarding the qualitative data, we checked to what extent these were reported as clear and comparable qualitative indicators. The coding was done by three independent coders. About 20% of the data was counter-coded.

1.3   Results

The 57 'Covid-19-reports' differ significantly in form and content. This is due to the following: 1) a lack of a standardised reporting format; 2) the platforms' reaction to challenges in a completely different manner and from a different understanding of their roles. So far, there are no requirements from the European Commission that would amend tendencies towards the individualisation of the reports.

Even if the actions taken by the platforms are not the subject of this analysis (the *reporting* of the platforms on these actions is the subject), the different approaches and business models are partly responsible for the heterogeneous style from which diverse ambiguities have arisen:

- *What actions are required?* The requirements for the platforms formulated in the 'Tackling Covid-19 disinformation' document refer to partial aspects of the Code of Practice on Disinformation but are intended to reflect the necessities that have arisen in particular as a result of the COVID-19-pandemic. Thus, the actions are implemented from different approaches: on the one hand, the COVID-19-response is integrated into existing initiatives as one among other 'sensitive events' (from the Google report: '... our Sensitive Events Policies do not allow ads that potentially capitalize on or lack reasonable sensitivity towards a sensitive event'. 'This experience comes as a complement to pre-existing work on Google Search and Google News to recognize sensitive events and contexts, and our systems are designed to elevate authoritative sources for those classes of queries'). On the other hand, COVID-19 acts as a motivation or even justification for the implementation of more general initiatives. Also, some reported actions are only marginally associated with the COVID-19-pandemic (from the Twitter report: 'Addressing the wider impact of Covid-19: fighting gender-based violence, fighting eating disorders, fighting child sexual abuse', etc.).
- *What is the role of platforms?* It is unclear what role the platforms play, both in the context of the COVID-19-pandemic and in the light of the entire disinformation phenomenon. Regarding COVID-19, the platforms report both on actions to combat the pandemic itself (from the Facebook report: 'We released new visualizations and datasets publicly, as well as a new survey to help combat the COVID-19-pandemic. These include: [...] A new map showing travel patterns

between countries and states to help researchers and NGOs understand how long-distance travel continues to impact the spread of COVID-19.'), as well as those to combat disinformation about the pandemic (numerous actions of all platforms). Furthermore, regarding the disinformation phenomenon, the role of the platforms and the ways in which they are expected to act is not clearly defined (nor are the roles clearly distinguishable from one another). This is because the platforms act as a.) providers of content and b.) as providers and operators of an infrastructure that sorts and prioritises the content of others, in any case subjecting it to their own terms and conditions. Finally, platforms often additionally act (de facto) as c.) their own auditors. In the reports, actions from all three levels are listed together, although Facebook, for example, argues against this biased practice several times in the course of this project: 'We also believe that no company should grade its own homework. Last year, we committed to undertaking an independent, third-party audit of our content moderation systems to validate the numbers we publish, and we'll begin this process this year.' In fact, the platforms are being asked to expand and intensify their cooperation with fact-checkers. The background here, however, is likely to be the need for accountability to independent third parties.

● Furthermore, the actions reported are extremely heterogeneous, both in terms of their subject, matter, and scope (regulatory, geographical, temporal or target group and audience reach) and in terms of their output (see examples).

● Finally, the decision to rely on a self-regulatory framework implies the development of many different individual solutions. Thus, in tension between legal frameworks, fundamental rights (especially data protection rules) and with an eye toward the desired protection against disinformation, each platform has developed its own terms and conditions, which have been used to implement the activities required by the European Commission (especially concerning prioritisation/removal/labelling of content). Consequently, all platforms act according to own rules, some of which are not made available for evaluation by independent third parties or the public. Moreover, as these rules are not standardised, they cannot be effectively compared.

● Similarly, self-regulatory freedoms result in different terminologies or even definitions for the same phenomena.

All of these ambiguities flow unfiltered into the reporting, unless they are delineated, defined, and differentiated earlier in the process.

To get an idea, some excerpts of how the platforms report their activities are shown:

Twitter

'Twitter released a specific COVID-19 API endpoint into Twitter Developer Labs to enable approved developers and researchers to study the public conversation about COVID-19 in real-time. This is a unique dataset that covers many tens of millions of Tweets daily and offers insight into the evolving global public conversation surrounding an unprecedented crisis. Making this access available for free is one of the most unique and valuable things Twitter can do as the world comes together to protect our communities and seek answers to pressing challenges.' Or 'Throughout the pandemic we've granted 183 Ads for Good grants to non-profits across 55 countries.' Or 'In early December we provided training to the two German health organisations regarding COVID-19: Robert Koch-Institut and Bundeszentrale für gesundheitliche Aufklärung on safety and security [on our service].'

TikTok

'Our in-app information hub directs users to trusted information from respected experts when they search for vaccine information' or 'A vaccine tag is applied to all COVID-19 vaccine videos and a banner redirects the user to verifiable, authoritative sources of information.'

Mozilla

'The October and November 'Debates and Dialogues' event on misinformation amassed a viewership of 45,000 from around the globe.'

Microsoft

'For instance, as illustrated in the screenshot below, we redirect any LinkedIn member that entered a search for the term 'corona-virus' to a link labelled 'Know the facts about coronavirus,' which appears first in the list of search results. By clicking on this link, members are directed to LinkedIn's own official page on the coronavirus, which displays current information and broadcasts from verified international sources, primarily the World Health Organization. The storylines on this page are available in 8 languages.'

Google

'Preserving the integrity of our advertising platforms during the COVID-19 pandemic is a continuation of the work that we do to minimize content that violates our policies and stop malicious actors.' or 'In 2019, we blocked and removed globally 2.7 billion bad ads—that's more than 5,000 bad ads per minute. We also suspended nearly 1 million advertiser accounts for policy violations.'

Facebook

'Facebook has implemented a range of policies and products to deal with misinformation on our platform. These include adding warnings and more context to content rated by third-party fact-checkers, reducing their distribution, and removing misinformation that may contribute to imminent harm.' Or 'To better apply warning labels at scale, we are developing new AI technologies to match near duplications of known misinformation at scale. When fact-checkers have identified a piece of misinformation, our AI would help spot copies of cropped or altered content and content that convey the same meaning but look different.'

The reports themselves differ considerably in form and content and are therefore very difficult to compare with each other. In the following, some meaningful impressions will be presented in an overview:

1) ... of formal representation:

Structure. While Facebook, Microsoft, and Google in particular report in a comparable structure over periods of several months, following closely the content requirements of the European Commission, the reports of Twitter and TikTok are structured differently from month to month.

Presentation. While most reports offer a combination of a wide variety of actions as continuous text and provide only selective quantitative data, TikTok divides the reports into a qualitative part and a quantitative part, such that each provides comparable data in the form of tables. At the same time, the other platforms partly refer to external own data sources (which were not part of this analysis).

Length. While TikTok gets by with about 1,000-3,000 words, Twitter`s and Google's reports measure between 5,000 and 10,000 words long (other platforms ranges fall in between).

2) ... of the presentation of the content:

Actions and their regulatory, temporal, and geographical scope or audience reach. Essentially, the reports have in common that they piece together very different initiatives in a narrative way. In all reports, both the own approaches and the projects done in cooperation with other organisations are communicated in a comprehensible way (with about a quarter of all actions supporting content from WHO, national health authorities or other third-party organisations; the highest proportion of which can be found in the reports of Tiktok and Twitter). However, the initiatives differ greatly from each other:

Some go hand in hand with the development or transformation of corporate policies, are rolled out worldwide and implemented for a long time and are seen as guiding corporate policy (including rule-setting for removing and prioritising content, for example from the Microsoft report: 'As we also noted in our August report, for specific queries, Bing will point users to special COVID-19-'information hubs' on Microsoft News. The articles in these information hubs, currently available in 39 markets globally, are sourced from over 4500 trusted news brands worldwide.' Or, from the Twitter-report: 'Starting on 1 March 2021, we will apply labels to Tweets that may contain misleading information about COVID-19-vaccines, in addition to our continued efforts to remove the most harmful COVID-19 misleading

information from the service. Machine-learning and automated language processing takes time to be effective. As such, we are beginning with English-language content first and use this same process as we work to expand to other languages and cultural contexts over time. (Spanish and Portuguese will follow). Labels will appear in the user's set display language and may link to curated content and official public health information or the Twitter Rules. Our goal with these product interventions is to provide people with additional context and authoritative information about COVID-19.') At the same time, the very same section of text also reports about very specific actions with a short time horizon and limited scope (from the Twitter report: 'On 8th May 2020, we held a workshop on crisis response on Twitter for the European Emergency Number Association (EENA) and its members, around 40 participants from all over Europe' 'On the framework of the Safer Internet Forum, Twitter participated in a panel to discuss the fight against online child abuse together with representatives from the European Commission, InHope, Europol and law enforcement.'). Sometimes this *regulatory scope* remains unclear.

The geographical scope or audience reach differs from initiative to initiative for all platforms, ranging from very limited target groups to individual member state markets to global initiatives. Sometimes the part that is relevant for the EU is in some way differentiated or highlighted. It is not always clear from the reports whether the respective services were also rolled out in all the languages of the member states that were specified as the geographical scope for an initiative. Almost half of all initiatives did not specify the geographical scope, or the audiences reached. In some cases, the announcement of a rollout in certain states was delayed by monthly reports and it was not clear from these when the rollout took place.

Comparable ambiguities were also found in the *temporal scope:* When an action was actually implemented and to what extent (and for how long) cannot usually be determined by the reports. In addition, many of the activities are reported more than once, sometimes, repeatedly, even, without clearly indicating that the same process was already part of an earlier report. Mostly these indications include the implementation periods but not consistently.

In many cases, it remains unclear which *content categories* the activities refer to: to own content, to original content of third parties, to comments or others. In most cases, the comments by users are not considered in the activities. This is a major weakness in the platforms` reporting, especially as often we see cases of pile-ons and coordination in such attempts to discredit authoritative information or experts. The prominence and reach of some of these authoritative

accounts and public figures is exploited through such disinformation-containing replies to saw doubts and provide visibility of disinformation, without fear of that being moderated.

Except for Mozilla, all platforms report on initiatives and activities that are implemented using artificial intelligence (especially regarding prioritising and removing content), but only selectively provide insights into the extent to which the AI is used. (In the Twitter report: 'From 1 March 2021 to 31 March 2021, we found 945 Promoted Tweets in violation of our COVID-19 advertising policy. They were halted and removed from promotion according to our policy. It is estimated that about 94% of the violating content during that time was detected by our automated systems'). Details on the systems used and how they work remain mostly unclear, and there are no reflections on the legal framework.

The impact on or compliance with data protection aspects or other legal provisions (such as Article 10 of the European Convention on Human Rights and the principle of freedom of expression) were barely dealt with in the reports, which is astonishing, given that many of the reported initiatives move in precisely this area of tension between legal protection provisions and protection against the threat of disinformation.

It is therefore not meaningful to compare this data in an undifferentiated manner or across the reporting period, or even across platforms. It should be noted, however, that this is due not only to the reporting, but also to the heterogeneity of the actions themselves and their very different regulatory, temporal, and geographical scope or audience reach.

Performance target analysis: It therefore makes more sense to consider the reported actions in relation to the requirements of the European Commission. In response to the European Commission's information request 'Tackling COVID-19 disinformation', Google, Microsoft and Facebook met the objectives very precisely. Twitter also met the targets, although many reported initiatives were only marginally in line with them. TikTok registered comparatively few initiatives, but, in line with the requirements. Furthermore, when looking at the extent to which the reported actions met the objectives of the Code of Practice on Disinformation, the common answer is: 'yes', with the very foci required in the 'Tackling COVID-19 disinformation' document and that are in line with the requirements of the Code.

Monitoring the impact. A reporting system that could be used as a basis for any assessment would require data preparation, disclosure, and transparency regarding the impact of the activities undertaken, one that is useful to both factual and temporal comparisons (on different aggregate

levels). None of the reports submitted met this requirement. Although highlights of the results were quantified in about one-third of the reported actions, (in rare cases, we also found data accessible to a direct comparison over time or countries, as in the cases of Google, Tiktok and Facebook in particular, but the reports provided mostly aggregated examples, if any), these were neither contextualised nor presented in a way that would have allowed an assessment of their significance for the online media environment. Although the European Commission requested data for the EU and at the member state level, this request was not met. (From the Google report: Our established News Lab training for journalists has continued, our free workshops are being facilitated live in Dutch, English, French, German, Italian, Polish, Portuguese and Spanish. As of April 2021, 22.3k Europeans have taken part in live sessions covering digital verification and data journalism skills since the start of the pandemic.' Or from the Facebook report: 'Our COVID-19 Information Centre on Facebook is available globally, including all 27 EU member states. More than 120 million people globally, including over 15 million people in the EU, visited the COVID-19 Information Centre during the month of March 2021'). Furthermore, the platforms hardly provide any detailed insight into the technical implementation of their activities (Artificial Intelligence). Cooperation with fact-checker organisations is reported by the platforms, but mostly without much detail. TikTok at least names the fact-checking organisations they cooperate with ('TikTok's fact-checking organisations in the context of COVID-19 disinformation include AFP, LeadStories and Sciverify, all of whom are members of the IFCN (International Fact-Checking Network).'), Facebook provides a link to their network, Microsoft emphasises their collaboration with Newsguard, and Google offers more details: 'It is possible to explore the fact-checks that we index by visiting https://toolbox.google.com/factcheck/explorer. Overall, fact checks published by fact-checking organizations from EU27 Member States appear in Google Search over 6m times a week on average, adding up to about 33m(million) impressions throughout March 2021.'

What is needed here is the disclosure of meaningful, quantitative data, oriented to comparable benchmarks and regularly made available or, where possible, accessible in real time (engagement metrics from platform, user, and partner perspective) at different aggregate levels (EU-wide, member state level, and, where it makes sense, down to content and individual level). This would make the impact of each individual initiative assessable and comparable from different perspectives. Finally, the data would have to be reported in a way that would also allow subsequent analyses by independent third parties – statistically analysable, not as part of the prose text, regularly or in real-time and in a comparable form.

Forecast: This area of analysis was the weakest. All platforms reported future plans from time to time. (From the Twitter Report: 'In the future, we will try to expand or update these policies to include misleading information around different thematic areas.') But this aspect was rarely found in any of the reports, or without details of the regulatory, temporal, and geographic scope of potential implementation.

Overall, the analysis of the 'COVID-19-reports' of Facebook, Google, Microsoft, Mozilla, TikTok and Twitter therefore provided many valuable insights into the work of the platforms. It makes sense from the perspective of the European Commission to demand self-reports on the activities and initiatives of the platforms. However, some changes are needed to make this reporting useful as a basis for an assessment of the implementation of the European Commission's objectives and for the framework (on a data protection compliant access to all relevant data) demanded by EDMO. With this, EDMO, together with other stakeholders, should improve the monitoring of initiatives addressing the disinformation phenomenon. The following recommendations are intended to assist in the improvement process.

2   Recommendations

Related to reporting in the narrower sense:

1. **Standardised guidelines for report structuring and presentation.** Standardised, formal *guidelines are required for the structuring and presentation* of the reports. In addition to a uniform structure, the content should be closely aligned with that of the European Commission. This also implies a *standardised set of rules (key figures)* on which activities are reported and how. There is a need for standardised, formal guidelines for the presentation of *quantitative data.*

2. **Systematic and complete presentation of regulatory, temporal, and geographical scope or audience reach.** To ensure better comparability, a systematic and complete presentation of the reported activities regarding their *regulatory, temporal, and geographical scope or audience* reach is required (including information on available languages and detailed EU or member state reference). Platforms should also report details on the *legal framework* and fundamental rights of the respective initiative.

3. **Artficial Intelligence (AI).** Platforms should report details on the use of artificial intelligence (AI) and its implication for all processes (e.g., appeals) as well as the relation to the legal framework.

4. **Content categories should be distinguished / Comments should be addressed.** The *content categories* to which the activities relate should be distinguished from one another and it should

15

be clear from the reports which activities relate to platforms' own content, to original third-party content or to comments or other content categories. Since *comments* are, at present, mostly excluded from the activities, it is recommended that they be compliant with all legal requirements.

5. **Only new initiatives or significant changes to existing ones.** The emphasis of reports should be on *new initiatives*. Activities should not be reported more than once, unless extensions or changes to existing initiatives took place. In this regard, a change of the monthly rhythm into quarterly reporting is strongly recommended.

6. **Reference to the requirements of the European Commission.** All reported activities should have a direct *reference to the requirements* of the European Commission and reflect this with a short *performance-target comparison*. In addition, it would be useful if a reflection on the connection to the Code of Practice on Disinformation for all associated initiatives (also in a standardised form) were included.

7. **Forecasts are recommended.** Activities with a longer time horizon should be reported together with a *forecast* (standardised guidelines sufficient).

8. **Report the impact and results of activities in a standardised way / Accessibility for factual, temporal, or cross-country comparisons and statistical evaluations.** The platforms should be obliged to *report the impact and results* of their activities in a standardised way, and to make them accessible for factual, temporal, or cross-country comparisons and statistical evaluations. This requires the definition of quantitative key performance indicators that include the perspectives of all stakeholders and thus: engagement metrics from platform, user, and partner perspective on different aggregate levels (where it makes sense, on content and individual level, on organisation level and on member state level and EU-wide). The data must be set in relation to meaningful, independently created and continuously updated benchmarks (e.g., 'out of 50,000 labelled posts, 500 were removed' versus '500 posts were removed'). This information must also be made available regularly and entirely; alternatively, real-time access could be ensured for regular audits by independent third parties. Data protection aspects and the respective legal framework should be reflected upon.

9. **Ongoing independent auditing of all processes to create transparency (including data access).** In addition to self-reporting, it is recommended that the platforms be obliged to carry out an ongoing independent *auditing of* all processes and to create the necessary transparency (including data access). Also, cooperation with fact-checking organisations should be reported in detail and in a standardised manner.

Further information, related to clarifications needed earlier in the process (objectives, roles, common definitions):

1. **Clear objectives within an area.** The platforms should not be confronted with different, only partially overlapping objectives. They should have the possibility to coordinate their activities in a meaningful way and to address the disinformation phenomenon with a holistic strategy.

2. **Clear role(s) of platforms to address the disinformation phenomenon.** There is a need for differentiation according to which role the platforms take in launching initiatives (as content producers or as operators of infrastructures that sort, prioritise and subject third-party content to their terms and conditions or as their own auditors). In relation to COVID-19, it should be clearly differentiated whether the platform is setting an initiative to support member states or authorities in their efforts against the COVID-19-pandemic per se, or to address disinformation about the COVID-19-pandemic.

3. **Harmonisation of rules.** The platforms act in accordance with their own specially developed terms and conditions, which differ from one another. Regarding these, there is a need for the greatest possible transparency and a comprehensible reflection on the respective legal framework. In the medium and long term, *harmonisation of these different sets of rules is recommended*. This could be realized with a co-regulatory approach.

4. **Transparent AI (Artificial Intelligence).** The platforms work with different systems and programming using *artificial intelligence.* In relation to this, there is a need for the *fullest possible transparency and a reflection on the legal framework.*

5. **Harmonised definitions and terms.** Each of the platforms has its own overall approach, which has led to different uses of key terms and different definitions of the same phenomena. In the medium and long term, *harmonising these different terms and definitions* is recommended, possibly as a first step and with a co-regulatory approach.

With all this, we must bear in mind that we are moving from a media order to a communication order and that the workings of the Internet are not comparable to the old media order. It is necessary to build governance structures, to think in different communication spaces and to grade the regulation – and with this, the reporting – accordingly.