ERGA REPORT ON DISINFORMATION: ASSESSMENT OF THE IMPLEMENTATION OF THE CODE OF PRACTICE

ERGA

European Regulators Group for Audiovisual
Media Services

The Code of Practice on Disinformation ("The Code") is a unique and innovative tool in the fight against online disinformation. By signing the Code and thus voluntarily accepting obligations that are not part of the legal framework, the signatories demonstrated a commitment to the EU approach to governance of the digital environment. During 2019, the Code's signatories implemented actions to deliver on their commitments under all five pillars of the Code and engaged with the EU and national institutions with the common goal of countering online disinformation. The **Code, therefore, should be regarded as an important step in the process of building a new relationship between its signatories, the EU** and National AV Regulators. Nevertheless, the work carried out by ERGA in 2019, and presented in this report, shows that **the Code has significant weaknesses that need to be addressed if it is to achieve its objectives**.

Firstly, there is a need for greater transparency about how the signatories are implementing the Code. The Code relies on self-reporting but lacks a mechanism through which the information from these reports can be independently verified. The information provided by the platforms is generally aggregated for the whole EU, which makes it difficult to evaluate the impact of the Code across the EU. This difficulty is amplified at a national level where language, societal and cultural factors make it the most relevant sphere for monitoring the impact and effectiveness of the Code.

Secondly, the measures of the Code are too general in terms of content and structure. To some extent, this is caused by the character of the instrument and, therefore, understandable. However, it provides space for the signatories to implement measures only partially or, in some cases, not at all. There is also a lack of uniformity in the procedures (and the definitions) adopted by the different platforms.

Thirdly, the number of signatories of the Code is limited. Although the current signatories are the main online platforms active in the EU, significant platforms/tools such as Tik-Tok, WhatsApp and Messenger are missing.

Therefore, ERGA believes **that steps are required to increase the effectiveness of the measures of the Code itself and also the oversight\reporting structures if it is to evolve into an effective tool in** combating disinformation.

For this reason, **ERGA encourages the Code's signatories and the EU Commission to improve the Code and its measures** by requiring that all of the platforms comply with the same obligations in a uniform manner (whenever possible taking into account the specificity of the individual platforms) and adopt more precise definitions, procedures and commitments, as well as measurable key performance indicators (KPIs). There is also a need for a set of provisions that apply to a broader number of online platforms active in Europe, as well as a need for provisions allowing the Commission (and the National Regulatory Authorities - NRAs, if delegated) to carry out specific monitoring activities, especially at the national level, and to adopt enforcement tools to ensure the compliance to the rules. The above-mentioned background suggests that **moving from the current self-regulatory model to more structured co-regulation may prove to be more effective to counter disinformation online**.

Based on the detailed summary of the outcomes of ERGAs monitoring activity (build on the national monitoring reports by the participating NRAs), carried out during 2019, **ERGA proposes in this report a set of recommendations,** based on three levels of intervention aimed at:

Improving the monitoring of the existing Code's commitments:

o   to ensure a consistent approach towards these issues/principles in the whole EU a set of relevant definitions should be drafted,

o   to improve the provision of information by the platforms by making available datasets, data monitoring tools and Country specific information (in a structure proposed by ERGA and by the Commission and similar for all the platforms) allowing the NRAs to monitor the commitments of the Code,

o   ERGA to draft sets of guidelines concerning the relationship between the platforms and the fact-checkers; platforms' reactions to consumers complaints and flagging; the media literacy campaigns in each Country and lastly improve the relationships between online platforms and researchers,

o   create intensive cooperation between ERGA and the new European Digital Media Observatory.

Expanding the existing Code's commitments:

o   address the problem of lack of uniformity by ERGA analysing further the commitments and compare the way the platforms implement them and then make recommendations aimed at harmonising the implementation of these commitments,

o   formally identify specific moments of the year in which the platforms would provide data on the implementation of the Code that includes Country-specific information, so to allow the ERGA to conduct a regular and proper monitoring activity,

o   increase the number of platforms signing the Code.

Exploring new (more effective) tools to counter disinformation:

o   To imporve the existing model of the Code points toward a more structured form of regulation. To this effect, a **shift from the current flexible self-regulatory approach to a more co-regulatory one would be required**. Such a system would involve the evolution of the current self-regulatory Code to provide for more consistency in its formulation and in its implementation and the introduction of a formal backstop mechanism to deliver the required monitoring and enforcement elements. This should include a mechanism to incentivise industry players to take part in a self (or co)-regulatory structure. The ERGA experience in 2019 indicates that the effective development and implementation of the Code requires such a framework. To achieve this, operative rules should be put in place. These should consist of clear reporting obligations, more harmonised procedures and appropriate timeframes. **This is the solution that ERGA recommends to enhance the relationship with the platforms**.

o   Ideally, **all the platforms[1] which distribute content in Europe should be engaged in this co-regulation procedure** and should then be subject to the co-regulatory obligations. Should this not be the case, the **EU institutions might explore the possibility of adopting a more conventional regulatory approach**. With the current review of the regulatory framework that should culminate with the announced Digital Services Act (DSA), ERGA sees the value in a holistic approach to governance of online content regulation. In this overall framework, **the DSA-package should create a framework that would also include the basis for the effective fight against disinformation (liability regime). In addition, a dedicated legal act is needed to address the problem more directly and in greater depth**. Such a separate instrument (e.g. a regulation) would ensure not only a level of detail of provisions and comprehensive coverage of stakeholders but also the legislative speed required given the threat the current information crisis presents to European democracies.

---

[1] Or at least all the platforms which have a relevant size

TABLE OF CONTENT

ABOUT ERGA

National regulators in audiovisual media services have, not exclusively, the role of implementing the rules set by the European legislative framework – the Audiovisual Media Services Directive (AVMSD). The European Regulators Group for Audiovisual Media Services brings together the leaders or high level representatives of these regulatory bodies to advise the Commission on the implementation of the EU's Audiovisual Media Services Directive (AVMSD). The responsibilities of ERGA Members (i.e. statutory regulatory bodies responsible for applying a legal framework) vary significantly. In some cases, NRAs have a level of involvement in the development of voluntary codes, whereas others have a very clear delineation between aspects of statutory regulation.

1. INTRODUCTION

The phenomenon of disinformation is not new, it has always existed in the linear and traditional media environment. However concerns about disinformation have increased in the digital age due to disruptive impact of the internet and new communications technologies on the dissemination of information across the globe. Instead of broadcasting a single, coherent message to the general public, the algorithms used by the social platforms offer the opportunity to tailor the type of information and messages that are conveyed to specific portions of the population. Differentiating the messages/information depending on the gender, the social class, the geographical area, the age, the political views or the economic status of the recipients gives a much higher chance to negatively influence democratic processes and societal debates.

The European institutions have recognised that disinformation is an evolving challenge and that the approach to intervention in this field is a sensitive topic, especially given the rights and principles at stake (in particular, the principles of freedom of expression and freedom of information).

The profileration of false news, guided by profit-oriented and/or political purposes, that is accompanying the the recent outbreak of, and response to, COVID-19, is only an example of how the information manipulation strategies pose serious threats to the formation of public opinion and of how important debunking such news can be to protect the democratic values and counter the attempts to incite hatred and violence[2].

For this reason, the European institutions have tried to counter the spread of disinformation online in recent years by adopting several measures[3]. On 26 April 2018, the European Commission (hereafter: Commission) adopted a Communication on *"Tackling Online Disinformation: a European Approach"* [4]. The Communication delineates the challenges online disinformation present to our democracies and outlines five clusters of actions for private and public stakeholders that respond to these challenges. The outlined actions include (inter alia) the development of a self-regulatory code of practice on disinformation for online platforms and the advertising industry in order to increase transparency and better protect users; the creation of an independent European network of fact-checkers to establish common working methods, exchange best practices and achieve the broadest possible coverage across the EU.

In May 2018 the Commission convened the Multistakeholder Forum on Disinformation[5] to draft a self-regulatory Code of Practice on Disinformation. The Forum consisted of a "Working Group" composed of the major online platforms and representatives of the advertising industry and

---

[2] In particular, the circulation on social media and online platforms of false information about the COVID-19 outbreak remains a substantial concern. Most of the media in Europe started addressing the COVID-19 issue in 2020, thus well beyond the period in which ERGA had carried out its monitoring activities (see section 2 of this report). Nevertheless, since this report focuses on the activities implemented by the signatories of the Code of Practice on disinformation to comply with the Code's provisions, it is worthwhile highlighting that some relevant media outlets (mainly from the United States, such as the New York Times and the Washington Post) emphasize that social media companies have begun taking disinformation on the COVID-19 issue "seriously, setting policies and algorithmic tools to mute toxic speech".

[3] In particular, the Commission set up in late 2017 a High-Level Expert Group to advise on this matter. The Group delivered its report on 12 March 2018. The Commission also launched a broad public consultation process, comprising online questionnaires that received 2,986 replies, structured dialogues with relevant stakeholders, and a Eurobarometer opinion poll covering all 28 Member States. A more practical result was the creation of the Website https://euvsdisinfo.eu/, which provides facts checking against some fake news appearing in the media.

[4] http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51804

[5] https://ec.europa.eu/digital-single-market/en/news/meeting-multistakeholder-forum-disinformation.

major advertisers, as well as a "Sounding Board" composed of representatives of the media, academia and civil society. The Working Group was tasked with drafting the Code, and the Sounding Board was tasked with providing advice and issuing an Opinion on the Code. The Code was published on 26 September, along with the Opinion of the Sounding Board[6].

The Code comprises a Preamble, a statement of Purposes, and a set of 15 commitments prefaced by explanatory comments that reference the Communication's objectives, detail the commitments' scope and purposes, and provide context. The commitments are organised under five Pillars[7]:

A. Scrutiny of ad placements (aimed at demonetizing online purveyors of disinformation)

B. Transparency of political advertising and issue-based advertising (aimed at making sure that political adverts are clearly identified by the users)

C. Integrity of services (aimed at identifying and closing fake accounts and using appropriate mechanisms to signal bot-driven interactions)

D. Empowering consumers (aimed, on the one hand, at reducing the risks of social media 'echo chambers' by making it easier for users to discover and access different news sources representing alternative viewpoints and, on the other hand, to plan and execute media literacy campaigns against disinformation)

E. Empowering the research community (aimed at by granting researchers access to platforms' data that are necessary to continuously monitor online disinformation)

Signatories to the Code are required to identify which of these commitments they will adhere to and how, in light of their relevance to the products or services they provide. The signatories also commit to cooperating with the Commission in assessing the Code, including providing information upon request and responding to questions.

On 16 October 2018, the Code's initial signatories, Facebook, Google, Twitter and Mozilla as well as the trade association representing online platforms (EDIMA) and trade associations representing the advertising industry and advertisers (EACA, IAB Europe, WFA and UBA), provided their formal subscriptions to the Code. Microsoft joined as well on 22 May 2019, becoming the 13th signatory of the Code.

* * *

As regards the terminology used, as highlighted in the previous report published by ERGA on Disinformation[8], many national stakeholders and institutions use the widespread expression of "fake news", while another interesting approach was adopted by the Council of Europe, which preferred to talk about "information pollution" and identifies three types of information disorders:

---

[6] https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.
[7] https://ec.europa.eu/digital-single-market/en/news/draft-code-practice-online-disinformation
[8] The report is available at this Webpage: http://erga-online.eu/wp-content/uploads/2019/01/ERGA-2018-07-SG1-Report-on-internal-plurality-LQ.pdf

Misinformation ("when false information is shared, but no harm is meant"), disinformation ("when false information is knowingly shared to cause harm") and malinformation ("when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere").

This report will refer to this phenomenon as "disinformation", since this is the term used by the European Commission and by the Code of Practice. In the aforementioned Communication of the European Commission entitled *"Tackling online disinformation: a European Approach"* in particular, *"disinformation" is "understood as verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm"*.

2. DEFINING THE BOUNDARIES OF ERGA'S INVOLVEMENT

The Joint Communication adopted on 5 December 2018 by the European Commission and the European External Action Service (also known as "Action Plan against disinformation[9]") assigned to the European Commission, with the help of the European Regulators Group for Audio-visual Media Services (ERGA), the task to monitor the implementation of the five Pillars (A Scrutiny of ad placements, B Political advertising and issue-based advertising, C Integrity of services, D Empowering consumers and E Empowering the research community) of the Code of Practice. The topic of disinformation is highly relevant for ERGA as an association of national regulators as it implies a potential danger to the democratic media system and democratic society.

The Action Plan against Disinformation was accompanied by the European Commission's Report on the implementation of the Communication "Tackling online disinformation: a European Approach[10]". The Action Plan required ERGA to assist the Commission in monitoring the implementation of the Code and assessing its effectiveness. In consultation with the EU Commission, ERGA decided to implement this task over two phases in 2019. The first phase aimed at monitoring the implementation of the Code's commitments regarding the integrity of the electoral processes during the 2019 European elections. The second phase was aimed at providing an assessment on the implementation of the commitments of all the five Pillars of the Code after an initial 12-month period (i.e. in October 2019). To co-ordinate this assignment ERGA created a specific Task Force as part of Sub-Group 1 led by AGCOM Italy.

As regards the first phase, ERGA sought to monitor, in as many Member States as possible, the implementation by Facebook, Google and Twitter of their commitments in the Code of Practice related to Pillar B, focusing on "transparency of political advertising" during the European Elections in May 2019. Between January and May 2019, the above-mentioned platforms issued monthly reports on the implementation of their actions under the Code most pertinent to the European elections. These monthly reports were regularly published by the Commission[11] and were examined by ERGA.

The outcome of ERGA's phase one monitoring activities, along with the conclusions, was summarized in the "*Report of the activities carried out to assist the European Commission in the intermediate monitoring of the Code of practice on disinformation[12]*", hereinafter also

---

[9] https://ec.europa.eu/digital-single-market/en/news/action-plan-against-disinformation

[10] https://ec.europa.eu/digital-sigle-market/en/news/communication-tackling-online-disinformation-european-approach

[11] The reports have been published on the following Webpages:
  - January: https://ec.europa.eu/digital-single-market/en/news/first-results-eu-code-practice-against-disinformation
  - Feb: https://ec.europa.eu/digital-single-market/en/news/second-monthly-intermediate-results-eu-code-practice-against-disinformation
  - March: https://ec.europa.eu/digital-single-market/en/news/third-monthly-intermediate-results-eu-code-practice-against-disinformation
  - April: https://ec.europa.eu/digital-single-market/en/news/fourth-intermediate-results-eu-code-practice-against-disinformation
  - May: https://ec.europa.eu/digital-single-market/en/news/last-intermediate-results-eu-code-practice-against-disinformation

[12] http://erga-online.eu/wp-content/uploads/2019/06/ERGA-2019-06_Report-intermediate-monitoring-Code-of Practice-on-disinformation.pdf

referred to as the *"ERGA Intermediate Report"* adopted by the 11th ERGA Plenary Meeting, held on 20-21 June 2019, in Bratislava, Slovakia.

As regards the second phase of the monitoring activities, and in compliance with the commitments of the Code of Practice, the Code's signatories presented a Self-Assessment Report (hereinafter referred to as "SAR") of the implementation of the measures of the Code after the initial 12-month period. ERGA in consultation with the European Commission decided to analyse these reports and provide an evaluation of their content in the context of activity at a country level. This approach was consistent with that adopted for Phase one and recognised the importance of implementation at an individual country as well as EU level. In addition, under the Code of Practice, signatories also committed to select an objective 3rd party organization (hereinafter referred to as "TPO") to review the annual self-assessment reports submitted by them, and evaluate the level of progress made against the commitments, which would include accounting for commitments signed on to. ERGA was also asked to analyse and evaluate the report of the TPO.

Following discussions with the Commission in July 2019, it was agreed that ERGA's analysis in Phase two should focus specifically on Pillars D and E of the Code, respectively the empowerment of consumers and the empowerment of the researchers.

In order to start collecting information and fulfil this task, on July 25 AGCOM, the NRA leading the ERGA Task Force, organized a workshop aimed at gathering views from researchers and academia. The ERGA Task Force members were invited to the workshop, along with the Code's signatories. In addition, ERGA organised a workshop on September 27th and also the EU Commission organized a workshop on October 18th with researchers, academia and the online platforms to discuss issues relevant to the Code.

The three workshops showed that some online platforms (mainly Facebook) had been used by unidentified groups of people who attempted to manipulate the public opinion during the latest EU electoral campaigns. This happened in Germany, but also in Italy and in other countries. Some ERGA Task Force members concluded that the "phase two" monitoring activities should be expanded -if possible- to update the results in relation to pillar B and examine issues relevant to pillars A and C.

The ERGA Task Force therefore agreed that the "phase two" monitoring should comprise two different levels of monitoring:

- <u>LEVEL A</u> monitoring would focus on pillars D and E of the Code, as required by the EU Commission. The 13 NRAs which committed to carry out this monitoring were asked to verify, at national level, the correctness of the information provided by the SAR that were going to be delivered by the platforms. The approach to be adopted for this monitoring included *(inter alia)* making contacts with the fact-checking organizations, universities, civil society organizations, associations of consumers and journalists, to better assess the way in which the activities reported in the SAR had actually been carried out by the online platforms with regards to pillars D and E of the Code in the various Countries.

- <u>LEVEL B</u> monitoring would instead be aimed at assessing the implementation of the other 3 pillars of the Code. This level of monitoring required far more time and resources, and it was carried out only by very few NRAs: Italy monitored pillars A and C, Germany pillars A, B and C while pillar B was monitored by the NRAs located in the Countries where political elections were foreseen (Hungary, Poland and the United Kingdom). This latter monitoring activity (pillar B) was very useful to assess whether the platforms had made any progress in dealing with political advertising after the publication of the ERGA Report.

It was also agreed that the monitoring would focus on the Google, Facebook and Twitter. Some NRA's also looked at the information available for Microsoft but as this was very limited (Microsoft joined the signatories of the Code of Practice only on May 2019) it is not a focus for attention in this Report.

2.1 The ERGA monitoring activity during phase 2

As stated above, the main sources of information for the ERGA monitoring activity were supposed to be the SAR published by the Code's signatories and the report from the TPO. The NRAs planned to gather external information and verify the way in which the activities reported in the SAR had actually been carried out by the online platforms in the relevant country.

The SARs were delivered and published[13] on October 29, 2019 together with a brief summary/ analysis published by the EU Commission[14], and contained very little Country specific data. Because of that, there was very limited information that the NRAs could verify through their monitoring activity conducted at national level based on the SARs.

In addition, the TPO that was supposed to assess the SARs was never appointed, in breach of the provisions of the Code of Practice.

In order to address the lack of data in the SARs, on November 11, 2019, ERGA and the EU Commission sent a letter to the platforms with specific questions to be answered at Country level for the 13 Countries that were participating in the monitoring exercise. In addition, meetings with the platforms to discuss their responses were organised on December 3rd.

---

[13] The SARs are available at this webpage: https://ec.europa.eu/digital-single-market/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019

[14] For the brief analysis from the EU Commission, refer to the following link: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62698. In this document, the Commission acknowledges that the signatories have made comprehensive efforts to fulfil their commitments over the last 12 months. The Code led to higher transparency regarding the platforms' policies against disinformation and the ability to monitor structured dialogues. However, further serious steps by individual signatories and the community as a whole are still necessary.
The Commission also observes that the reported actions taken by the platforms vary significantly in terms of scope and speed. In general, actions to empower consumers and the research community lag behind the original commitments (as evidenced prior to the European Parliament elections in May 2019). Furthermore, there are differences across the Member States as regards the deployment of the respective policies for the various commitments included in the Code.
Although cooperation between the platforms and stakeholders (e.g., fact-checkers, researchers, and civil society organisations) improved, the provision of data and search tools is still episodic and arbitrary and does not respond to the demands of researchers for independent scrutiny. More efforts are also needed to establish sound cooperation with truly independent organisations.
Lastly, the Commission notes that no other IT platforms and advertising companies/services operating in the EU have joined the Code.

Unfortunately, the data emerging from this process was generic and -once again- contained very little Country specific information.

The NRAs that committed to carry out monitoring activities conducted an autonomous evaluation, using the limited information provided by the platforms and adding data that they collected from relevant third parties (e.g. civil society groups, the associations of consumers and journalists, academics, researchers and fact-checkers). Also, where relevant and possible, the NRAs proactively (including through using of individual user accounts) explored the tools and other available resources from the platforms and their availability in the individual countries. In addition, some useful reports from renowned experts were considered[15] .

In order to ensure that the monitoring activity would be implemented in a coherent manner, a very detailed list of questions based on the Code were drafted (see the complete list in Annex 1 to this Report); thirteen NRAs participated to Phase two and managed to answer all the questions that were circulated, making specific reference to the findings concerning each of the three platforms (see the summary of the answers in Annex 2 to this Report).

---

[15] For example, the reports from the researchers Trevor Davis and Rebekah Tromble, from the Dublin University, from the Istituto per la ricerca sociale, the report from the workshop on "Removing barriers to digital platform transparency across Europe", held in Bruxelles on October 18, 2019, the statement from the European Advisory Committee of Social Science One of December 11, 2019

3. SUMMARY OF THE MAIN FINDINGS OF PHASE 2 MONITORING

This section will present the main findings of the ERGA Task Force "phase two" monitoring activity, pillar by pillar.

3.1 – Pillar A, Scrutiny of Ad Placements

The Code of practice on disinformation envisages ways to improve the scrutiny of advertisement placements to reduce revenues of the purveyors of disinformation. The goal of this pillar of the Code, is to reduce the incentives to produce deceptive, illegal, or otherwise questionable content.

The NRAs involved in the monitoring of the compliance to the provisions of this pillar were the Italian AGCOM and the German Association of State Media Authorities (DLM), but the total lack of data on this issue made the task of the two entities extremely difficult: unlike Pillar B, the platforms did not provide any dataset or repositories concerning Pillar A. There was very little information about the activities implemented by the platforms at national level in the SARs and the platforms did not provide the detailed answers that ERGA had requested in its letter.

DLM assigned the monitoring activity to an external researcher who analysed advertising placed on disinformation/sensationalist websites. By placing the advertisement on such websites, the advertisers contributed (unintentionally) to the monetarization of disinformation.

In general, all platforms have implemented a broad range of policies to disrupt advertising and monetization incentives for relevant behaviours, such as misrepresenting material information about oneself or the purpose of one's properties. Those policies intended to regulate the presentation of advertisements on platforms but without information on all the advertisements presented on a platform at on a national level, it was not possible to conduct a reliable assessment of the implementation of those policies.

In addition, the monitoring showed that in both Germany and Italy the platforms had not enabled engagement with third party verification companies and had not provided advertisers with necessary access to client-specific accounts to enable them to monitor the placement of ads and make choices regarding where ads are placed. On the contrary, the random monitoring (carried out through individual users' accounts) showed ads that were placed on webpages whose content might even conflict with the advertisers' policy. According to the findings of DLM's report, the ad placement (which is opaque for advertisers) on websites that disseminate disinformation leads to a monetisation of disinformation. This is contrary to the commitment "to improve the scrutiny of advertisement placements to reduce revenues of the purveyors of disinformation".

Once again, it is worthwhile highlighting that the monitoring activities that Italy and Germany could carry out were extremely limited due to the total lack of Country specific data. In the future, if ERGA has to carry out any meaningful monitoring, the platforms will have to provide a reliable set of data, disaggregated at national level.

3.2 – Pillar B, Political Advertising and Issue-Based Advertising

According to the provisions of Pillar B, all advertisements should be clearly distinguishable from editorial content whatever their form, and regardless of the medium used. An advertisement should be presented in such a way as to be readily recognizable as a paid-for communication or labelled as such, when it appears in a medium containing news or editorial matter. The Code's Signatories commit to enable public disclosure of political advertising which could include actual sponsor identity and amounts spent, and to use reasonable efforts towards devising approaches to publicly disclose "issue-based advertising".

In Phase 2 monitoring information in relation to this pillar was provided by the regulators in Hungary, Poland, Germany and the United Kingdom.

As it was already reported in the ERGA intermediate Report, a key difficulty in monitoring the implementation of Pillar B is the fact that EU Countries have different definitions of political advertising and most of them do not define "issue-based" advertising. In the Code, political advertising is defined as advertisements advocating for or against the election of a candidate or passage of referenda in national and European elections. The platforms also formulate their own definitions, which differ from one another[16]. While both Facebook and Google distinguish between paid political advertisements and commercial ads, only Facebook provides a working definition of issue-based ads and publicly disclose them among political ads in its Ad-Library. Twitter was monitored in Phase 2 only by the Polish NRA KRRiT (see section 3.2.2 of this report), because it introduced a prohibition on political advertising in November 2019.

The results of this Phase 2 monitoring did not differ significantly from the Phase 1 outcomes as presented in the ERGA Intermediate Report. This indicates that there was little or no improvement in the platforms' procedure or activity since May 2019 in the aforementioned country. The monitoring shows that the platforms were making efforts by publishing repositories that included different organised data sets, which allowed some searching, filtering and analysis activities of relevant advertisement. However, these repositories were not the databases themselves, but a searching tool, with pre-defined filters, that allow the general public to access data and information that the platforms had previously filtered and organized[17]. For example, none of the platforms provide the basic population of advertisements, therefore it's not possible to assess if all political ads are correctly labelled as such. Also in Facebook's API the data is insufficient and many political ads in the API could not be found in the platforms and vice versa. Therefore it was difficult to monitor whether the data was reliable or not.

---

[16] Facebook:https://de-de.facebook.com/business/help/167836590566506?id=288762101909005,
Google:https://support.google.com/adspolicy/answer/6014595?hl=de,
Twitter: https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content.html

[17] For example, the Facebook search tool was showing the list of political and non-political ads, while the Google search tool was providing only the ads that Google had labelled as "political"; Twitter instead provided the political ads for the whole of Europe, not filtered by Country, and the list of ads was extremely limited (only one ad was available at the time of the meeting).

DLM's contracted expert found that the bulk of paid content is not reported via the platform disclosures. It also found that, in Facebook, if a political ad is shared by a user, the "Paid for by" disclaimer vanishes because the content is seen as organic by Facebook. This latter finding is very interesting, as it shows an important limitation to the effectiveness of the system.

All of the NRAs found that the API and/or datasets available provide little information other than the advertiser, the date, and the ad itself. For example, results are not presented by country or over an extended period, and there is no information on reach or targeting: as a matter of fact, very little data (if any) is available on micro targeting of the users.

Broadly speaking, the quality of data in the political ads archives did not improve, as requested by the Intermediate ERGA Report of June 2019, and once again ERGA highlights that, in order to carry out any meaningful monitoring, a reliable API and/or a detailed dataset is needed.

If the NRAs are to carry out an acceptable assessment of the platforms' activities on transparency of political ads, the unconditioned, unmanaged and unfiltered access to the raw database is needed, containing all the advertisements of the platform (including the political ads, the non-political ads, the ads that have been published in accordance to the new procedures adopted by the platforms and those that have not been published). Being aware that, for some platforms, the number of ads in this database could be excessively high, ERGA suggests that the platforms provide access to an extract of the database (in machine-readable form), showing all the advertisements run during a specified period. In this way the ERGA Members would be able to run queries and filter the data available on these extracts so as to make random checks and assess the effectiveness of the process put in place by the platforms. ERGA's request for data was submitted through the EU Commission twice, in May and October 2019, but no additional data was ever delivered by the platforms. Without this data it is difficult to see how any future monitoring would be of value.

Below is a more detailed description of the monitoring activities carried out during the national elections held in 2019 in Hungary, in Poland and in the UK.

3.2.1 Outcome of pillar B monitoring during the local elections in Hungary

In view of the local elections of October 2019, that were accomplished on the basis of the EU election monitoring, the Hungarian regulator **NMHH** investigated 816 Facebook ads and the available Google ads (because of the small amount of Google ads, and the lack of Twitter ads, its monitoring focused primarily on Facebook). Briefly, the results of the monitoring are the following:

- On the Facebook News feed advertisements do not include a "political ad" label oppose to the Ad Library, so users can not simply detect whether they see a political ad or a non-political one. Users need to check the "paid by" label or the content of the ad itself to recognise the political nature of the advertisement, which is one of the biggest transparency issues in Facebook's advertising practise.

    o Some ads in the Ad Library, which are published by political organizations or clearly have political content, are not labelled as political advertisements. It is not unambiguous whether it is a technical issue of the Ad Library or the ad were not properly labelled by the advertiser and Facebook identified as such but didn't mark it.

- Regarding sponsors, in some cases a private individual is shown in the "Paid by" section of an ad. In these cases the connection between the advertisement and the financial supporter is not verifiable, and in some cases, these ads also have unverified sponsors (possibly under fake name).

- Regarding Facebook's Ad Library, it does not show that ads that have been removed by Facebook, were prohibited due to the topic (lack of disclaimer) of the ad, or lack of proper advertiser information (lack of sponsor information). Also there are several ads that have been blocked due to a conflict with Facebook's advertising policies, but it is not apparent which point of the policy is violated.

    o On the Facebook News Feed some obviously political advertisements from non-political organisations are not labelled as such. These advertisements are present neither in the Ad Library nor in the Ad Library Report. There is no transparent data regarding the relabelling and importation of these ads into the Ad Library as political ads.

    o In case of Google due to the lack of ads, at the time of the monitoring no representative data could be recorded. During this period on the total spend on political advertising in Hungary is visible. Since the amount of advertisements has grown a bit, we could monitor them from the user's perspective. We can say that the visible advertisements fulfils the requirement of transparency as the sponsors' identity and the amount spent on the given ad is available.

    o Google, however, struggles with technical issues, as the Hungarian report includes -mostly political- advertisements from abroad and the number of available ads is small and cannot be verified.

3.2.2 Outcome of pillar B monitoring during the elections in Poland

During the electoral campaign preceding the national elections of October 13, 2019, the National Broadcasting Council of Poland (**KRRiT**) monitored around 80 political ads that were presented by Facebook and Google[18]. In order to ensure comparable results, the same group of researchers which carried out the Phase 1 monitoring and the same procedures (e.g. the monitoring activity was conducted during a 5 day sample) were deployed for Phase 2.

The KRRiT monitoring team did not identify any significant improvement in the platforms' activity or procedures with regards to the transparency of political advertising between the May and October elections, and stated that:

1. Several electoral spots appeared only on YouTube, absent in repository despite being clearly marked as sponsored by political parties. Therefore, it was impossible to get any information on the funding or the target group in relation to this advertising. Indeed, during the sample period Google's repository did not include enough ads to support sufficient analysis.

2. Several political ads from other European countries were included in the Polish part of the Google repository.

3. Inconsistencies identified in May in relation to data in the Facebook repository and Excel Ad Library Report remained.

4. In the case of Twitter, the repository was disabled for two weeks preceding the polling day. It was reinstated immediately after polling day. At the same time, a lot of Twitter political ads circulated on the Internet during the electoral campaign, clearly marked as sponsored by political parties.

---

[18] Twitter's repository was not working properly during electoral campaign

3.2.3 Outcome of pillar B monitoring during the elections in UK

Prior to the United Kingdom exiting the European Union on 31 January 2020, **OFCOM** participated in the monitoring work of Sub-Group 1. During the electoral campaign preceding the national elections of 12 December 2019, OFCOM monitored the implementation of commitments made by Google, Facebook and Twitter under Pillar B of the Code of Practice. The key findings were the following:

1. Transparency of targeting criteria: Information on who had been exposed to an ad, their degree of engagement, and information on the targeting criteria used is relatively limited for both researchers and users. In its sampling of ads on Facebook and YouTube, OFCOM encountered instances of ads appearing to be targeted at specific voting constituencies which suggests the possibility that more sophisticated targeting criteria are available to advertisers than are captured in the platforms' ad libraries. **Google**'s transparency tools rely on information provided by advertisers during self-verification, and the sponsoring entity behind an ad is indicated to users only as the URL an ad links to. This allowed for ads paid for by political parties to appear as having been paid for by an opposing party. **Twitter**'s Ad Transparency Center is not searchable by topic or keyword making it difficult to monitor their policy for issue-based advertising. Additionally, its Ad Transparency Center uses a 7-day retention policy which is an impediment to researchers. **Facebook**'s Ad Library at the time of this review did not appear to filter ads for Messenger, although it is possible that the Ad Library was accurately reporting that there were no political ads being served on Messenger.

2. Is political advertising clearly distinguishable/recognizable: All three platforms appear to rely on self-verification of political advertising to trigger the transparency tools they have committed to under the Code. This means there is a risk of political ads not properly self-verified being served to users without sufficient transparency and also that ads which are not properly classified are not included in the resources made available to researchers. The process for subsequent review/classification of ads after they have been published is not transparent. The onus is on advertisers to declare their ads as political or issue based. This means that advertisers may fail to declare their adverts, or bad actors may even attempt to avoid being identified, meaning they do not appear with a 'paid for' disclaimer or in the ad archives.

3. Disclosure of sponsor's identity: Twitter: No observations given ban on political advertising. **Google**: The identity of the sponsoring entity is not, strictly-speaking, provided to the user encountering an ad. Instead, the ad suggests it is sponsored by the URL it links to. Generally, these URLs contain the name of the sponsoring entity, and more often than not, the content of the ad itself also refers to the sponsoring entity. In such cases the content is clearly distinguishable. However, this is problematic when ads link to a URL which does not include the name of the sponsoring entity, nor does the ad content. **Facebook**: The vast majority of ads sampled did identified sponsors, but the amount of information provided by advertisers varied significantly.

4. Progress made on issue-based advertising: Facebook's policy for issue-based ads also relies on self-verification, creating risk of at least one unintended consequence: a) That issue-based ads are more likely to go unlabelled and be missing from the archive entirely; b) That 'non-polemic' ads, which are neutral about or tangential to the listed topics, are incorrectly labelled by advertisers, users or Facebook itself as 'issue-based'. The second risk raises significant concern for the protection of free expression, which is highlighted under Commitment 4 of the Code. Further, Facebook's policies do not make clear how an advertiser could contest instances where an ad is taken down and entered into the archive erroneously. **Google**'s Transparency Report does not contain ad previews for those ads served by a third party, nor does it contain ad previews for ads found to be in violation of Google's ad policies. This obstacle to researchers is compounded by the fact that OFCOM could find no explanation of why ads had been removed – an important factor in assessing the efficacy of the Code. **Twitter**'s ban on political advertising extends to issue-based advertising. It applies to "ads that refer to an election or a clearly identified candidate" and "ads that advocate for legislative issues of national importance", regardless of whether an election is national, regional or local. However, the Twitter Ads transparency center only allows users to search ads by advertiser. The fact that it is not possible to search by issue significantly limits the usefulness of this tool, makes it difficult to ascertain how effective the policy on issue-based advertising is in practice, and makes it impossible to readily check that Twitter has fulfilled its transparency obligations.

5. Are ad libraries user-friendly? In general the ad archives of all platforms are reasonably intuitive, but there are some issues to report: **Google:** The Google Transparency Report does not appear to distinguish between ads that run over a selected video as a 'trailer' (i.e. pre-roll), and YouTube videos with purchased prominence either in search results or in the recommended videos sidebar. While the format is implied by the length of video, it is impossible to verify this by the Transparency Report. This could be important as 100k-1M impressions of a 40 minute-long video will have different implications if this was advertised as a 'trailer' or if this was a normal video for which prominence in recommendations or sidebar was purchased. The general question of whether/how videos with purchased prominence are included as ads in the Transparency Report is an interesting point to clarify with Google. **Facebook**: During the monitoring period, on 10 December, it was reported that around 60,000 ads had gone missing from Facebook's ad archive. This illustrates the risks of relying purely on data curated by platforms for monitoring and assessing compliance with the Code[19]. **Twitter**: The Ads transparency center does not allow users to search ads by issue, which limits its usefulness. Further, ads are only searchable for a seven-day period. Finally, the Ads transparency center does not provide any information other than the advertiser, the date, and the ad itself (i.e. when searching for non-political ads). For example, results are not presented by country or over an extended period, and there is no information on reach or targeting.

---

[19] See for instance https://uk.reuters.com/article/uk-facebook-advertising/thousands-of-uk-political-ads-wentmissing-from-facebook-archive-searches-idUKKBN1YE2HZ

3.3 – Pillar C, Integrity of Services

According to the provisions of Pillar C, the Code's signatories recognise "the importance of intensifying and demonstrating the effectiveness of efforts to close fake accounts" as well as the importance of establishing "clear marking systems and rules for bots to ensure their activities cannot be confused with human interactions". To that end, they commit to put in place clear policies regarding identity and the misuse of automated bots on their services and to enforce these policies within the EU.

The DLM's contracted expert found that during the 2019 European Parliament Elections in Germany, Facebook was manipulated by a large network of fake accounts. In this context it noted that a political party which had commanded less than 13% of the actual vote had allmost 85% of the inteactions of political parties on the platform (e.g. likes, comments, posts). The researchers conclude that this shows that the threats of manipulation by domestic and foreign interests exist and that the necessary data to investigate thoroughly remains sparse and difficult to obtain.

The monitoring carried out by DLM and AGCOM shows that the platforms have put in place clear policies regarding the identity and the misuse of automated bots and on what constitutes impermissible use of automated systems. There are lots of statements from the platforms, in their SARs but also in public advertisements, highlighting that hundreds of accounts have been removed. However, none of the platforms -in their SARs or in the answer to the ERGA information request- has provided a breakdown on the number of bots disabled for malicious activities and fake accounts identified and removed per Country. The limited information on Pillar C does not reflect the importance that the topic of disinformation entails. Regulators must be able to scrutinize the capacity of platforms to recognize and eliminate fraudulent activities[20] .

In conclusion, just like pillar A and unlike pillar B, the platforms did not provide any dataset or repositories concerning pillar C: there was very little information about the activities implemented by the platforms at national level in the SARs and the platforms did not provide the detailed answers that ERGA had requested in its letter sent on November 11, 2019.

---

[20] This has been recently affirmed also by a ground breaking decision recently adopted by the Court of Rome (see the webpage https://www.corriere.it/politica/19_dicembre_12/casapound-tribunale-roma-ordina-riattivazione-pagina-facebook-3a5f9a86-1cc9-11ea-9d5e-8159245f62dc.shtml), which decided in favour of the CasaPound association after the deactivation of its Web-pages and profiles by Facebook on 9 September 2019. According to the judge, *«the prominent importance assumed by Facebook service is evident in reference to the implementation of the fundamental principles of the system such as that of pluralism of political parties (49 Const.). Therefore any political party or affiliated organization that is not present on Facebook is effectively excluded from the Italian political debate».* In this way, the Court of Rome confers to social media platforms a role that goes far beyond that of "private operator" and compares it to a public service to which everyone has the right to access, unless illegal conduct has been demonstrated.
If confirmed in its second instance (Facebook brought the case in front of the Court of Appeal), this decision might have very serious consequences: for example, it would not allow Facebook or any other social media platform to delete the accounts of certain users (linked to political bodies) in proximity to the elections or during electoral campaigns, even if these accounts are in clear breach of the platform's internal rules, because the deletion would "exclude them from the political debate".
However, it is also worthwhile highlighting that, on Feb 23, 2020, well after the conclusion of the ERGA monitoring period, another Italian tribunal decided differently on a similar case (Facebook had removed the accounts of another association, Forza Nuova, linked to the extreme right wing parties); additional information on this latter decision may be found on this webpage: https://roma.corriere.it/notizie/cronaca/20_febbraio_24/facebook-legittima-rimozione-pagine-forza-nuova-2a91bf78-5728-11ea-b89d-a5ca249e9e1e.shtml.

As pointed out earlier, any future meaningful monitoring of this Pillar requires the provision by the platforms of specific data, disaggregated at national level. The DLM's contracted expert states that all the platforms currently offer far more extensive data to their commercial partners than they are willing to provide to researchers and the regulators/monitors. This means that the data can be made available and that there are solutions that may overcome the obstacles reported by the platforms (i.e. data security and privacy).

3.4 – Pillar D, Empowering consumers

The aim of Pillar D is to dilute the visibility of disinformation by improving the findability of trustworthy content and by making it easier for users to discover and access different news sources representing alternative viewpoints and by providing them with easily-accessible tools to report disinformation. At the same time, with this Pillar, the platforms commit to undertake media literacy campaigns aimed at making the general public aware of the threats of disinformation.

The monitoring of the compliance to the provisions of this pillar was conducted by 13 NRAs, which gathered information from the SARs, from third party reports[21] and from their meetings with civil society organizations, associations of consumers and journalists, universities, researchers and fact-checkers (or from their websites). Also, whenever relevant and possible the NRAs were proactively (including through using of individual user accounts) exploring the tools and other available resources from the platforms and their availability in the individual countries.

The outcome of this monitoring shows that consumers of all the platforms are supported primarily in two ways:

- Firstly, immediately via the interface of the platforms through labelling and links to additional information, and

- Secondly, more broadly, through media literacy or fact-checking initiatives.

In order to **improve the transparency of advertising**, Facebook for example has upgraded its policies, relevant services and tools improving transparency of paid messages, by offering features such as the "Context" button, "Why do I see this ad" etc. The Context Button is presented as an information button on the right side of news. It provides background and information on the publishers and links that appear in the News Feed but to add context information is not mandatory, and then information is not always available.

All the platforms encourage market acceptance of tools that help consumers understand why they are seeing particular advertisements and they inform the users on how their data is used by them. The "Why am I seeing this ad" feature on Facebook or the "Why this Ad" feature on Google allow users to collect more detailed information about targeting activity of commercial and political ads. Although these tools share vague information regarding targeting criteria such as users' preferences, it may remain unclear to the public why a specific user is targeted, since

---

[21] For example, the reports from the Dublin University, from the Istituto per la ricerca sociale, the report from the workshop on "Removing barriers to digital platform transparency across Europe", held in Bruxelles on October 18, 2019

limited information is provided by the platform.

However, several NRAs recognize that these tools are easy to use, but not always easy to spot. Furthermore, it is not clear how often these features are used by consumers, what actions were taken by them or by each company based on these potential complaints. According to several NRAs in order to accurately assess the level to which consumers are empowered to understand, report and impact information on these platforms, it is necessary for the company signatories to report the national totals of complaints received via these tools as well as the corresponding actions taken by them to highlight this content as problematic.

In order to **reduce the spread of misinformation**, Facebook set up a procedure to report false messages and this was a focus for attention by the Slovak regulator, CBR. This NRA reported that although some users may be designated as a "trusted flagger", there is no feedback or information from Facebook about the results of flagging[22]. Several ERGA members (Ireland, Germany, Italy, Portugal etc.) recognize that some content has quickly and reliably been demoted or even deleted (in the case of illegal content). However, in other cases the demotion (or deletion, in case of illegal content) takes longer or does not happen at all. The decision is not always clear, satisfactory or transparent. On YouTube, Google indicates that the new changes related to this have been experimented within the UK, Ireland and «other European markets», but it does not reveal which European markets. CBR reports that in Slovak Republic there is no tool for reporting fake news in Google News and Search[23]. Google reported that YouTube removes content that violates its policies as quickly as possible. In cases where content comes close to violating the platforms' policies, including attempts to spread harmful misinformation, YouTube reduces recommendations for that content. Twitter doesn´t have the direct option to report false news in the case of individual pieces of content. There is a possibility to report an account on Twitter on the basis of it being "suspicious or containing spam[24]".

The problem of **availability of the platforms tools in the various Countries** has been signalled by many NRAs. It is also worthwhile highlighting that a large number of products described by the signatories have been developed for the USA or published in the USA and have not been adapted and translated yet for the European market. For example, at the time of the ERGA monitoring the measure described by Facebook named *"Off-Facebook Activity[25]"* had only

---

[22] According to CBR, there seems to be a big difference of treatment among the types of content that are flagged: hate speech usually generates a reaction from Facebook in the least time, while it is not clear what happens to content that has been marked (even by trusted flaggers) as a fake news / disinformation.

[23] There is a possibility to report "Spam or Misleading" video content on YouTube, although this could be potentially confusing to users as there is no clear category for disinformation content.

[24] Although Microsoft is not covered in this report it is relevant to note that it helps its customers to evaluate the quality of the news they encounter on Internet through their partnership with NewsGuard Technology, an organization led by journalists and entrepreneurs. However, NewsGuard has only already been launched in Italy, Germany, France, and the UK.

[25] Off-Facebook Activity is a tool through which on can see and control the data that other apps and websites share with Facebook. For example, one can:
- See a summary of the information other apps and websites have sent Facebook through our online business tools, like Facebook Pixel or Facebook Login;
- Disconnect this information from your account if you want to; and
- Choose to disconnect future off-Facebook activity from your account. You can do this for all of your off-Facebook activity, or just for specific apps and websites.

been announced and had not started in most of the 28 EU Countries. For example, in Ireland the NRA could not easily identify the Breaking News Top Shelf on YouTube while searching for content related to topical Irish news stories. Some of the products, technologies and programs mentioned by the platforms in the SARs are designed to **help people make informed decisions when they come across online news that may be wrong**. For example, Google Search is trying to prioritize "reliable" content over content that does not seem reliable: users are provided with authoritative sources when searching for health-related information such as 'vaccines'. In Ireland the NRA found that Twitter prioritised references to the Irish Health Service Executive (HSE) when searching for content such as 'vaccines' but not to other authoritative sources of information on issues such as 'global warming'[26]. However, the "prioritization tools" are not always available in all the EU Countries (for example CBR reports that the tools "Breaking News" and "Top News" are not available in Slovakia, although Google states the opposite in its report) and in any case, when other contested terms such as "global warming" or "immigration" were searched for by the monitors, authoritative sources were not prioritized at the top of the search results thus highlighting inconsistencies within this mechanism.

The signatories provide users with **tools to obtain information about the use of their own data** and about control options. Information about the use of data and about the possibilities of control is usually not immediately visible. From the consumer's point of view, the use of control options requires a high level of competence. For example, in Italy, users can modify the use of their data in the section "Your information on Facebook" but this function has not been introduced in Italian language yet.

---

[26] https://www.bai.ie/en/new-report-on-political-social-media-ads-identifies-inconsistencies-in-datasets-and-definitions/

3.4.1: Planning and implementation of media literacy campaigns

As far as **media literacy** is concerned, in several Countries the Code's signatories have entered into various partnerships with media companies and educational organizations to plan and execute media literacy campaigns[27]. For example, the ERGA monitoring proved that Facebook and Google organized trainings for journalists and for politicians (usually belonging to national political parties/organizations rather than local ones). Facebook, in particular, organized in Italy a digital literacy training program[28] that was conceived and created by one of the leading

[27] Below is a selection of the initiatives lauched by the platforms to promote media and information literacy
Facebook
  • Last autumn, Facebook launched the Digital Literacy Library (DLL) in 45 languages around the world. The DLL is a collection of ready-to-use lessons from the Youth and Media team at the Berkman Klein Center for Internet & Society at Harvard University, aimed to help young people between the ages of 11-18 to think critically and share thoughtfully online. The lessons address thematic areas such as privacy and reputation, identity exploration, security, safety and wellbeing, and more. This library could be a resource for educators but it should be implemented with training activities for teachers.
  • Poland: Facebook worked with the Digital Youth Forum in Poland on a three-day educational event for youths that aimed to promote a safe, informed and innovative use of new technologies as an alternative to a risky online behavior. 400 youths, between ages 13-17, participated in the Forum and over 10,000 students from 148 schools all over Poland also followed the event online.
  • Italy: Facebook reported that it collaborated with Freeformers (a company that trains workforces in digital skills) and over 20 in-country NGO's and training organizations to deliver a Digital Skills Training Programme to 75,000 citizens across seven European countries. The associated webpage with this campaign lists just six participant countries (Italy, Germany, France, Spain, Poland and the UK) .
  • Portugal: GeraZão is a Media Literacy program launched in Portugal by Facebook in October 2019 to promote digital literacy, especially the good use of the social media amongst younger people. The program has the support of official partners (DGE, SeguraNet, Centro de Internet Segura and FCT).
Google
  • Be Internet Citizens campaign – a joint initiative from Google with Family Online Safety Institute, ConnectSafely and iKeepSafe – has a site that provides resources to teach teenagers about media literacy, critical thinking, and digital citizenship, with the aim of encouraging young people to have a positive voice online.
  • Spain: the Spanish Government and Google have announced a joint program to train 30,000 young people between the ages of 14 and 16 in critical thinking and media literacy. The project "(In) form" has three fundamental elements:
  • Experiential videos: Some journalists will present a variety of situations related to their profession on how they select, analyze, contrast, write, express and discuss informative contents.
  • Gamified training: students will be divided in teams comprised of between one and four people to complete the missions of the game during school hours. The game is composed by four levels, which correspond to the four skills that they have be acquired for developing critical thinking when information are consumed.
  • Info_Influencers Contest: The main objective is the creation of content with the support of media professionals.
Twitter
  • Ireland: Twitter has been a member of Media Literacy Ireland since early 2018 and has contributed to the Working Group responsible for the delivery of the #BeMediaSmart campaign, even though they are not officially a member of said Working Group. #BeMediaSmart is an Irish campaign in association with the Broadcasting Authority of Ireland (BAI). As contribution, Twitter offered Ads for Good grants up to €5,000 each to nonprofit/charitable members of MLI to support the #BeMediaSmart campaign online.
  • German: #WeDeserveBetter campaign, which aims to promote digital empathy, tolerance and respect in debates and public discourses. The campaign should raise awareness of the normalisation of hate speech in public and political discourse. As part of its ongoing commitment, Twitter supported a global net-work of security partners on Safer Internet Day 2019 by expanding its campaigns and promoting advertisements. Twitter launched a special emoji for Hashtags #SaferInternetDay and #SID2019, which are available in twelve languages to promote conversations about a safer Internet.
[28] The programme's aim was to help young people better understand the disinformation phenomenon. The initiative was supported by AGCOM and promoted by "Generazioni Connesse", the Italian Safer Internet Centre coordinated by the Ministry of Education. The events were carried out at Binario F, a space created in the Rome train station, where businesses, families, academics, NGOs and publishers/media were gathered to improve their digital skills. But this training programme was attended only by 40 students representatives of Roman schools, who were asked to become the "ambassadors" of this important topic in their schools: at the end of the training, the students received a training kit that they could use to convey the skills acquired to their schools and broaden as much as possible the impact of this training activity.

national experts on the phenomenon of online disinformation[29]. However, these campaigns involve only a tiny fraction of the total population (mainly journalists, politicians and school teachers), usually concentrated in the main cities. Ireland is possibly an exception to this trend: in that Country, each of the four signatories made efforts to promote digital media literacy both through their interaction with Media Literacy Ireland (MLI) and *online* initiatives. Even in this case, however, as highlighted by the French CSA and several other NRAs, in the absence of any data on the uptake and impact of these initiatives provided by the platforms it is not possible for the monitors to evaluate their efficiency.

The importance of media literacy in developing critical thinking and countering phenomena such as disinformation cannot be highlighted enough. Technological advances have provided new powerful means to producers and distorted information to reach an ever-wider audience. Social networks, in fact, are able to attract and engage with millions of people by giving them a platform where disinformation can quickly spread. For these reasons, Media literacy campaigns have assumed an essential role within a society. Education and critical thinking are regarded as an essential skill for citizens to effectively engage on social media platforms: as stated by several experts, *«Media and information literacy includes a set of competences to search, critically evaluate, use and contribute information and media content wisely; knowledge of one's rights online; understanding how to combat online hate speech and cyberbullying; understanding of the ethical issues surrounding the access and use of information; and engage with media and ICTs to promote equality, free expression, intercultural/ interreligious dialogue, peace, etc.»*[30]. This definition attempts to synthesize all the key elements and fully covers the objectives of media education. This means that the purpose of Media literacy is to enable people to have the skills, knowledge and understanding to make full use of the opportunities presented by both traditional and new communications services.

In an online environment where the possibility to direct content regulation diminishes, the need for a media-literate public increases. Children and adults need to be equipped with the necessary information and critical analysis to understand contents on social media, to work out what is accurate and trustworthy, and what is not. Media literacy is not a skill to be obtained and consistently maintained, but rather is an ever-changing process that requires constant reflection and adaptations. This has to be an ongoing process because the people's needs for media literacy play a vital role in the formulation of public policy, as well as providing to organizations and agencies the evidence they need in order to best target their initiatives on the field.

For this reason ERGA believes that the efforts made by the Code's signatories to foster media literacy initiatives should be part of a more systematic campaign and should address a much bigger part of the population in the whole national territory. For example, in the case of

---

[29] Mr. Walter Quattrociocchi, the coordinator of the Data Science and Complexity Lab at the University Ca' Foscari in Venice.

[30] Grizzle, A. 2015. Measuring media and information literacy: Implications for the sustainable development goals

projects directed to school students, it seems fundamental to cooperate with the Ministries of education to spread the information and courses to improve critical thinking in all the schools of the territory. Specific documents and materials on the disinformation topic, such as videos, brochures or recorded lessons or on line lessons, produced by experts in an attractive way for students, should be delivered and/or showed to all the schools in the national territory. In general, cooperation with the NRAs and with civil society and associations on the planning and conducting phases of the campaigns could be crucial.

3.5 - Pillar E, Empowering the Research Community

The aim of Pillar E is to enable privacy-compliant access to data for fact-checking and research activities. To that end, the Code's signatories commit to support good faith independent efforts to track disinformation and understand its impact and to *"cooperate by providing relevant data on the functioning of their services, including data for independent investigation by academic researchers and general information on algorithms"*[31].

As it occurred for pillar D, the monitoring of the compliance with the provisions of pillar E was conducted by 13 NRAs, which gathered information from the SARs, from third party reports[32] and from their meeting with civil society organizations, associations of consumers and journalists, universities, researchers and fact-checkers.

The monitoring showed that Facebook and Google regularly provide funding to academic researchers for projects that the companies deem important but this possibility is not equally widespread among all EU Countries.

In addition, many NRAs confirm that specific events/discussions and partnerships with research and academic institutions remain episodic and largely inadequate to support any rigorous analysis and monitoring of online disinformation trends.

---

[31] Code of Practice on disinformation, section II.E, page 8

[32] For example, the reports from the researcher Rebekah Tromble, from the Dublin City University, from the Istituto per la ricerca sociale, the report from the workshop on "Removing barriers to digital platform transparency across Europe", held in Bruxelles on October 18, 2019, the statement from the European Advisory Committee of Social Science One of December 11, 2019

3.5.1 The relationships between the platforms and the fact-checkers

As regards the initiatives deployed by the Code's signatories in supporting independent fact-checkers, **Facebook** is the only platform that has established a contractual relationship (the so-called Third-Party Fact-Checking Program) with fact-checkers in different Countries that are certified through the International Fact-Checking Network (IFCN) of Poynter Institute in Florida. In Countries where the fact-checking organizations are not certified by the IFCN, Facebook has not established any contractual agreement. In addition, Facebook committed to ensure due prominence in its Newsfeed feature to all fact-checks coming from all professional fact-checking organisations (including in-house fact-checking teams of professional media outlets).

The fact-checkers partnering with Facebook follow specific procedures: every fact-checker partnering with Facebook gets access to a customized dashboard, presenting a list of the links (statements, posts, pictures, videos etc. from Facebook and Instagram properties) to fact-check. Any link/piece of content could be included in the dashboard and reported by a Facebook/Instagram user or by Facebook itself through an AI-based system; every fact-checker could insert a link/piece of content in the dashboard by its own[33].

A combination of algorithms and human review is used by Facebook to prioritise what content is referred to every fact-checker. Representatives from Facebook explained that the platform does not prioritize the content sent to fact-checkers on the basis of news importance or the potential to cause public harm. Apparently a certain priority is given to the topic of "vaccines", but the monitoring from a number of NRA's indicate that other topics such as "immigration", "global warming" are treated similar to any other piece of content. The links are not categorized by issue or other specific categories, and every fact-checker can choose the links to fact-check.

Once the link has been fact-checked, if the content is false or partly false (Facebook provides 9 rating options[34]), the fact-checker writes a related article and publishes it in its webpage. AGCOM, CSA and other NRAs report that in their Countries the fact-checker contracted by Facebook receives a lump sum for every published article, regardless of the nature and content

---

[33] The French CSA highlighted that a fact-checkers in France has published information on how this collaboration works. Here is some additional information:
- The fact-checkers are paid by Facebook for this fact-checking activity on the basis of the number of articles they fact-check;
- The fact-checkers working with Facebook have to be part of the International Fact-Checking Network (it implies that they accept a code of transparency and independence);
- In practice, every fact-checker partner has access to a tool with hundreds of shared links, photos and videos on Facebook. For every link, the media has access to an estimate number of shares. The media can label it ("misleading title", "true", "fake", "opinion", "satire"...). For every label, the fact-checker has to link an article published by their media and explaining why the content on Facebook is "fake", has a "misleading title"... ;
- Why do the fact-checkers see some links ? Because it was reported by users or because it was detected by the algorithms of Facebook. All the fact-checkers of the same country have access to the same list of content;
- The fact-checking collaboration evolves. The fact-checker can now add links of articles that are fake to the databases;
- The results of the fact-checking is also used by Facebook to train its algorithm.

[34] See https://www.facebook.com/help/publisher/182222309230722

of the link they report[35]; in other cases, as reported by the Polish KRRiT, the fact-checkers seem to get a monthly salary, not directly connected with the number of published articles. In Ireland there was no clear structure of payment between Facebook and the factchecker and monitoring indicated a very low and sporadic level of activity.

When the contracted fact-checker reports that a link/piece of content is false, Facebook reduces its "visibility" by up to 80% on the users' news feeds and shows a "warning label" (which says that "that link has been fact-checked" and links it to the fact-checked article) to any user who intends to share that link.

The NRAs that managed to meet with the fact-checkers report a number of problems connected to the Third-Party Fact-checking Program:

- first of all, the "warning label" appears only when the link/content is shared, while it does not appear when the content is received or watched for the first time[36];

- it is not clear whether the circulation of the link reported as false is limited also within the public or private groups;

- from September 2018[37] also videos can be fact-checked; however, the percentage of videos fact checked appears considerably smaller that the percentage of written information, and for sure the fact-checkers do not address the problem of deep-fake videos, which is likely to be one of the major challenges in the field of disinformation in the coming years;

- neither the general public nor the fact-checkers partnering with Facebook know what happens once the links/content analysed by the fact-checkers have been recognized as false: it is not clear whether and when Facebook intervenes with the "warning label" (in some cases, apparently, Facebook is very quick in publishing the label, in 39 to spread disinformation through political advertising is a highly sensitive issue, especially since Facebook changed its policies to exclude politicians' ads from fact-checking in

---

[35] Some fact-checkers reported that a minimum of three fact-checks and a maximum of fifty (50) fact-checks should be guaranteed every month. This information however has not been confirmed by all the fact-checkers.

[36] This was the situation when the monitoring was carried out. More recently, Facebook started placing an "overlay" (see picture below) that darkens the content from the moment in which the it is declared false; the "overlay" says "this information is false, as verified by independent fact-checkers" and contains a box stating "discover why", which links to the fact-checker's webpage. However, the "overlay" has been introduced only recently and it is not known whether it is available in all the EU Countries or not. In particular, it is not known whether it is available in Countries where Facebook has not contracted independent fact-checkers.

[37] See https://about.fb.com/news/2018/09/expanding-fact-checking
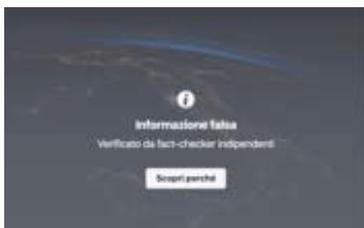
September 2019[39] (a change classified as a dangerous loophole by the journalists that were consulted by the NRAs). For this very reason, in November 2019, the Dutch news website NU.NL, a professional fact-checking organization, withdrew from the contractual agreement with Facebook because of Facebook's refusal to fact-check statements by politicians[40].

The opinion of the NRAs that managed to meet with the fact-checkers is that the Facebook Third-Party Fact-checking Program might be greatly improved by adopting more transparent and publicly available guidelines aimed at solving the aforementioned problems and by ensuring the prioritization of the topics that are particularly relevant for the public opinion[41].

NRAs that carried out the monitoring activity did not receive information from Facebook about the implementation of its commitment to ensure due prominence in Newsfeed to all fact-checks coming from all professional fact-checking organisations. In principle, the latter organisations remain free to fact-check any link/piece of information without contractual restrictions, but it was impossible to assess whether this measure is adopted in an efficient manner and whether all fact-checks are given due prominence in Newsfeed.

As opposed to Facebook, **Google** has not created contractual partnerships with any fact-checking organization and does not "demote" fake content that was debunked by fact-checking organizations, but decided to support fact-checking mostly through the use of dedicated tools and by providing some training for journalists. This shows how the signatories of the Code of Practice comply in a very different manner (implying very different levels of effectiveness) with the same obligation. The main tools launched by Google are a Fact-Check Explorer Tool, which links a searched term or name to independent fact-checker articles[42], and the IFCN project FactCheckEU.info, bringing together the European signatories of IFCN's Code of Principles to counter misinformation in the European Union at a continental scale ahead of the European Parliament elections of May 2019. In addition, some NRAs have learned about some specific tools Google is sharing with fact-

---

[38] For additional information on the Facebook directive not to fact-check the politicians' posts, see the Facebook websites https://www.facebook.com/help/publisher/182222309230722?locale=cs_CZ and
https://it-it.facebook.com/help/publisher/182222309230722, respectively in Czeck language and in Italian, according to which "Posts and ads from politicians are generally not subjected to fact-checking ", but also the Facebook Self Assessment Report , according to which "...posts from politicians will not be subject to ratings from our third-party fact-checking partners. However, when a politician shares previously debunked content including links, videos and photos, we will demote that content, display related information from fact-checkers, and reject its inclusion in advertisements. Fact-checkers will continue to fact-check content about politicians and we will take action by reducing the reach of these false stories and informing people with additional context in-product. Any ads from politicians must still comply with our Community Standards and Advertising policies, including new standards that require registration and transparency".

[39] See the Webpage https://about.fb.com/news/2019/09/elections-and-political-speech/, in which Nick Clegg, Facebook VP of Global Affairs and Communications states that "Facebook exempts politicians from our third-party fact-checking program. [...] This means that we will not send organic content or ads from politicians to our third-party fact-checking partners for review".

[40] See this Webpage on Dutch news: https://www.dutchnews.nl/news/2019/11/nu-nl-pulls-out-of-facebook-fact-checking-service-citing-policy-toward-politicians/

[41] It would be enough, for example, to prioritize the topics that Facebook has identified in its own issue-based advertising policies.

[42] https://toolbox.google.com/factcheck/explorer

checkers in order to correctly label fact-checks in Google Search: the MarkUp Tool (also known as *ClaimReview*) is part of an open ecosystem coupled with programmatic and policy layers, open to all publishers who may signal their use of fact-checking by using a dedicated html (mark-up) by adding structured data that contains information about the fact-check (i.e. what claim was assessed, who made the claim, what was the verdict of the fact-check). The mark-up then allows the Google search engine to display this information in the search results. Since adding a ClaimReview mark-up requires some knowledge of coding, Google developed a Fact-check Markup Tool for easily adding this structured data to fact-checks. Collecting information about the effectiveness of these fact-checking tools and services has been very difficult. In particular, it is difficult to understand to which organisations, and by which criteria, Google makes these tools available[43]. No information was collected about specific fact-checking tools on YouTube.

Twitter is not collaborating with fact-checking organizations on a regular basis. Twitter partnerships are limited to associations, organizations and research centres tracking and promoting democratic values in the platform. However, even without official support by Twitter, a wide range of journalistic fact-checking organizations are active on the platform.

As a general remark, it seems clear that the initiatives deployed by the Code's signatories in supporting independent fact-checkers are diverse. Although Facebook showed significant efforts in contracting fact-checkers to detect fake news, its Third-Party Fact-checking Program requires substantial improvements. In any case, since not in all the EU Countries Facebook engaged fact-checkers (possibly because it was difficult to find fact-checkers belonging to renown international organizations in all of the EU Member States) and, most of all, since the other platforms did not follow the same path, probably another solution should be found in order to make the efforts of the various platforms more uniform and less dependent on contracts signed with a specific organization in one or few Countries.

---

[43] Facebook, for example, allows only fact-checkers signatories of the International Fact-Checking Network Code of Practice to get access to its Third-Party Factchecking Program

3.5.2 The relationships between the platforms and the research community

Coming to the initiatives deployed by the Code's signatories in supporting the research community, only **Facebook** launched an *ad hoc* program aimed at partnering with academics and sharing privacy protected datasets. In particular, in April 2018, Facebook launched Social Science One, a very ambitious programme involving a commission of 83 academic researchers and a group of funders, with the goal of building a fair and transparent procedure to share the platform's data with academic research community. One year later, in April 2019, Facebook announced a new set of research projects that will look into social media's impact on democracy. The projects provided access to "privacy-protected Facebook data" to more than 60 researchers from 30 academic institutions across 11 Countries, in an attempt to help conduct research into a range of topics related to election campaign in Europe. To support these projects, Facebook built a first-of-its-kind data sharing infrastructure to provide researchers access to Facebook data in a secure manner that protects people's privacy. The selected researchers have gained access to data through the following tools:

1. CrowdTangle: allows researchers to track the popularity of news items and other public posts across social media platforms;

2. URLs shares Data Set: aggregated and anonymized list of posts and web page addresses (URLs) that have been shared (with *"public"* privacy settings[44]) at least 100 times in the past two years (2017-2019);

3. Ads Library API: provides data on political ads and issue-based ads.

Nevertheless, the NRAs monitoring shows that, despite the efforts and the provisions of these tools, at the moment Facebook initiatives in the field of disinformation and political advertising research are limited to few partners. It is still difficult for any academic or researcher to get access to the data. For example, not all countries received access to all data resources. While the NRA in Ireland found that APIs were accessible and that CrowdTangle was made available to some researchers, they could not identify an example of data provided to an Irish institution via Social Science One. Moreover, no initiatives aimed at fostering discussions within academia, the fact-checking community and members of the value chain have been deployed, except for some meetings behind closed doors, with participation limited to partners in fact-checking and academic research programs:

• The CrowdTangle API has been available to many journalists and media companies for use in tracking public posts' performance, but to date, only a limited number of scholars have been given access. Besides, CrowdTangle's utility is limited, as it provides only aggregated data and does not allow researchers to explore the comments and replies

---

[44] The dataset only contains posts e links that were shared publicly: when they were shared, in other words, the option "share publicly" or "share with everybody" had been chosen; the posts and the links that were shared only with specific uses, on the contrary, are not included in the "URLs shares Data Set", for obvious privacy reasons.

associated with public posts[45].

- Data within the URLs Data Set is aggregated, it is not possible to scrutinize manipulation effects on single users: important questions about the individual use and effects of social media cannot be studied with such data.

- Also, the access to the Ads Library API has been reported by researchers to be unstable, complicated and un-reliable: ad libraries do not offer micro-targeting information but only data on audience reach in broad categories, and the researchers cannot verify whether the findings based on ad libraries are accurate.

The research team from University of Urbino Carlo Bo in Italy is one of the 8 academic teams that had access to Facebook data. The researchers have recently published a report on coordinated and inauthentic link sharing behaviour in Italy's 2018 General Election and 2019 EU Election. AGCOM reported that Facebook had organized a meeting with representatives from the Social Science One in its headquarters in the Silicon Valley, with the aim to share more details about the usage of the shared datasets and find solutions to the limits of the datasets shared by Facebook (in particular URLs shares dataset), due to the fear of being exposed again to data breaches similar to the Cambridge Analytica case. Some NRAs (e.g. the Irish BAI and the Italian AGCOM) reported that Social Science One was trying to find ways to extend the access to the datasets to other researchers and scholars and that researchers and academic teams in their Countries were getting access to some of these datasets. After the initial excitement, however, Social Science One's members also begun to experience difficulties in getting access to data. For this reason, on December 11, 2019, the members of the European Advisory Committee of Social Science One issued a public statement complaining about the lack of an adequate data access from Facebook. The Committee also highlighted the extremely limited scientific value of the URL light data set received, and expressed its frustration about the overall relationships with Facebook. The Committee noted concern about the frequent delays and obstacles from both within and beyond the company that are undermining the innovative model of partnership between academic researchers and the private sector launched by Social Science One. Surprisingly, on February 2020, well beyond the end of the ERGA monitoring period on which this report is based, Facebook provided Social Science One with a remarkably large dataset, resulting from processing approximately an

---

45 Rebekah Tromble, The Digital Platforms' Responses to Pillar 5 of the Code of Practice on Disinformation, The George Washington University, DC
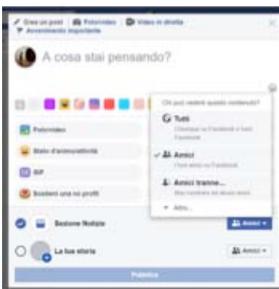
exabyte of raw data from the platform. The dataset contains more than 10 trillion numbers that summarize information about 38 million URLs shared worldwide more than 100 times publicly on Facebook (between 1/1/2017 and 31/7/2019). It also includes characteristics of the URLs (such as in which country they were shared and whether they were fact-checked or flagged by users as hate speech) and the aggregated data concerning the types of people who viewed, shared, liked, reacted to, shared without viewing, and otherwise interacted with these links. This dataset will be now made available to academic researchers, through Social Science One, that has immediately launched a request for proposal open to scholars and researchers involved in academic research. According to Social Science One itself, this dataset will enable social scientists to study some of the most important questions of our time about the effects of social media on democracy and elections with information to which they have never before had access[46]. This statement will be verified by ERGA in its future monitoring activities[47].

In its SAR, also Google reported several efforts aimed at allowing researchers to access data:

- the platforms are supporting EU-level research such as the Oxford University's Reuters Institute for the Study of Journalism reports;

- it also published a Political Ads Transparency Report, and made its Ad Library available for download in CSV format;

- an interesting project and tool, supporting independent research and fact-checking at the same time, is the Data Commons project. Its "Open Knowledge Graph" and its Graph Browser integrate data from various sources such as different census or official statistics: users can access data about a city, county or organization, aggregated from different sources. Currently, however, available data stems mostly from the USA.

Similarly to what has been said for Facebook, however, in reality the researchers consulted by the NRAs (such as the Italian AGCOM) stated that it is still difficult for any academic or researcher to get access from Google to useful raw data for his researches in disinformation field. On the contrary, Google seems to be making fewer efforts as compared to the other Code's signatories, because it neither provides data access tools nor offers an open API data access; Google, besides has not launched specific programs or partnerships (such as Facebook's Social Science One) and has not shared any YouTube datasets[48]. The datasets mentioned in the SAR look appealing for common users, but enable only a limited range of scientific research projects.

Since 2006, **Twitter** has been one of the few online platforms which made available APIs

---

[46] see the webpage https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one.

[47] NRAs should get access to the datasets made available to scholars and academic researchers by platforms, in order to evaluate their effectiveness and their compliance to scholars' needs, even thanks to the support of third-party verification committees

[48] German researchers claim that they consider data from YouTube as vital since the platform dominates the German market for video sharing and is becoming increasingly popular among younger audiences, being used also for information purposes.

to researchers and developers. Twitter's APIs are a unique data source for academics that is used around the world in a wide range of fields, from disaster management to political science, every day. All of Twitter API data is public - no private user data is included and no email addresses, IP data etc. are shared. Furthermore, Twitter has recently disclosed an archive of state-backed information operations on its platform that is periodically updated. The datasets have proved significant to researchers within the EU and were accessed over 20 thousand times by independent as well as institution-affiliated researchers across the EU. Twitter has shared datasets on election integrity, too. On the contrary, no initiatives aimed at fostering discussions within academia, the fact-checking community and other stakeholders have been organized by Twitter in any EU Country.

In general terms, the NRAs monitoring shows enormous difficulties for the researchers to get access to data: according to the research promoted by the German DLM, the main challenges in the relationship among platforms and researchers is the imbalance with regard to data control and data access: while platforms such as Facebook, Twitter and YouTube deliver large quantities of personal data to the advertisers and their respective businesses that can be incorporated into specific business products such as targeted advertising, researchers have only limited access to such data. Platforms also often prevent individual users from deliberately passing on their personal data to researchers thus severely limiting the possibilities to independently study disinformation processes. Most researchers consulted by the NRAs – including researchers from the chosen projects within Social Science One – however agree on the following critical points:

- The problem of lack of useful, measurable and researchable data is the main problem flagged by all the reports from the researchers and the scientific community, which state openly that in spite of the promises of the platforms, scientific research remains extremely difficult to conduct: the platforms do not share crucial data points, including data on ad targeting and user engagement with disinformation. Scholars seeking a better understanding of disinformation are therefore left without adequate sources of platform data. As a result, the most important questions about the extent and impact of micro-targeting and disinformation remain unanswered.

- Interestingly enough, the researchers share the view of the ERGA intermediate report that the ad libraries provided by Facebook, Google, and Twitter in response to the Code of Practice were inadequate to support in-depth systematic research into the spread and impacts of disinformation in Europe[49]:

    – first, the ad libraries do not offer useful (micro) targeting information; they instead focus on audience reach in broad categories such as gender, age, and region. Micro-targeting criteria used by political advertisers and by advertisers in general

---

[49] Cases from the Czech Republic (Pavel Havlicek, Association for International Affairs), Germany (Simon Kruschinski, Johannes Gutenberg University Mainz & Jörg Haßler, Ludwig Maximilians University Munich), Italy (Gaia Giombelli, Istituto per la Ricerca Sociale), and the UK (Dommett) each showed the ad libraries to be incomplete, inconsistent, and difficult to use for research purposes.

should become fully transparent to the public and targeted individuals,

– second, researchers cannot verify that findings based on the ad libraries are accurate. Any results researchers present to the public are based on trust that the platforms have provided complete, accurate data.

- Not all projects need the same data. They stated that accessible data should be defined by the specific research interest and not by a company granting access on its own terms.

- There is no possibility to assess the quality of the data. Since they cannot create own data samples, they cannot trace biases. So, researchers are dependent on data provided by Facebook without the possibility to directly observe and test its quality.

The expert[50] contracted by the German DLM to carry out the monitoring of compliance with the provisions of pillars E, examined recent scholarly research on two issues at the heart of the Code of Practice – online political ad micro-targeting and disinformation – and sought to assess the extent to which this research has been enabled and supported by Google, Facebook, and Twitter. The report delivered by DLM stated that very little scholarly research on online political ad micro-targeting and disinformation has been based on data found in Facebook's, Google's, and Twitter's respective ad archives and that even the more advanced academic-platform partnership, Facebook's Social Science One, is stuck due to different problems[51]. The submission from DLM provides several recommendations (that fed, together with all the national monitoring reports and other examined sources, into the general reflection of ERGA recommendations in the following chapter):

- As part of their public ad archives, the platforms should provide more precise data on ad spending and impressions.

- The platforms should also provide more precise targeting data in the ad archives. This should include direct targeting data, as well as information about categories targeted indirectly through custom audience and lookalike features.

- For sensitive categories (e.g., race or political ideology), audience reach data might be substituted for targeting data. Alternatively, sensitive targeting data could be reported to regulatory authorities, with researchers given the opportunity to access the data under controlled conditions.

- The platforms should preserve deleted ad content, including content removed for violation of ad policies, for analysis by researchers.

---

[50] Rebekah Tromble, Associate Professor at the George Washington University's School of Media and Public Affairs

[51] Several researchers state that the data provided under the Social Science agreement does not allow for sufficient data access. If the Social Science One is not able to fulfil its promise of enabling secure data access for independent research, scientists propose to invest in alternative models with data access being enforced by political actors. Alternative solutions proposed by the researcher we consulted rely on an independent intermediary institution to negotiate data access with the platforms

- The platforms should provide formal analyses identifying their specific concerns regarding data sharing for independent academic research under GDPR. Such analyses will provide a starting point for resolving areas of ambiguity and uncertainty.

- In turn, Data Protection Authorities should offer formal guidance on permissible data sharing practices under GDPR.

- Regulatory authorities should begin to require that the platforms share data for research purposes. The types and amounts of data should remain flexible, with priorities set based on public interest as defined by the regulatory authorities, in consultation with both the platforms and scholars. The platforms' proprietary interests should not be neglected, but these should be balanced against the public's interest in platform transparency.

- The establishment of "safe harbors" should be promoted, to the aim of supporting independent scholarly research carried out on platform data. Models from the health and medical sectors, as well as the government statistics offices, could be consulted.
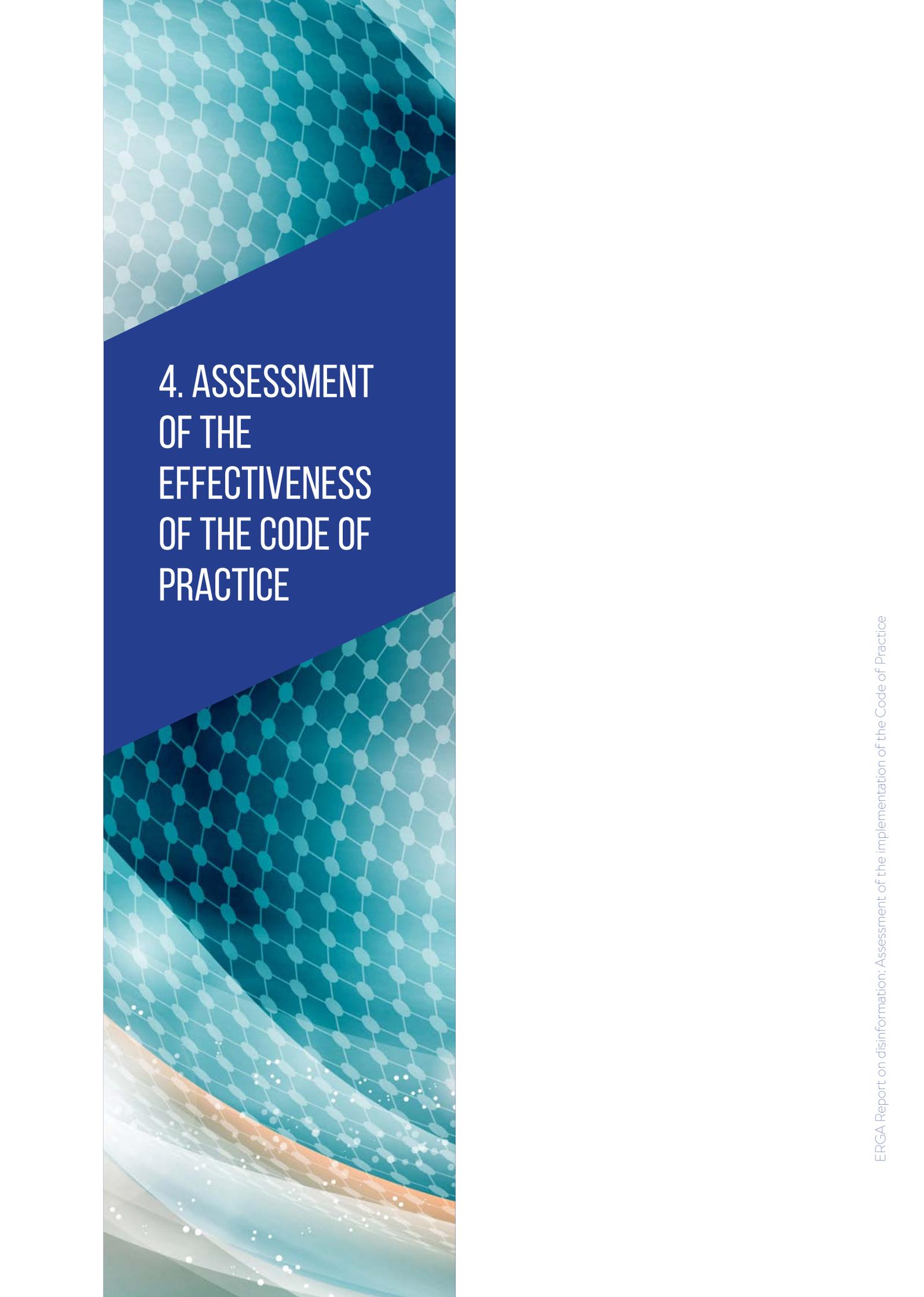
4. ASSESSMENT OF THE EFFECTIVENESS OF THE CODE OF PRACTICE

As stated in Pillar C of the Action Plan against Disinformation, the tasks of ERGA as regards the Code of Practice are not limited to the monitoring activities aimed at verifying the compliance to the provisions of the Code, but involve also a very important activity aimed at assessing the general effectiveness of the measures of the Code[52].

At the outset questions were raised about some NRAs' legal competence to undertake the task assigned to ERGA by the European Commission, however their suitability for this task was never questioned. All the NRAs have solid monitoring and enforcement powers and, broadly speaking, are skilled in evaluating the appropriateness of a regulatory framework to oversee market developments and to foster the adoption of new rules. In addition, several NRAs had already started adopting non-traditional and innovative tools[53] in an attempt to regulate (*rectius*, co-regulate) the online media and platforms. The NRAs of the ERGA Task Force, therefore, carried out an assessment of the effectiveness of the Code across 2019 in parallel with the monitoring activities.

Preliminarily, it is worthwhile highlighting that the **Code of Practice is a unique and innovative, tool in the fight against online disinformation**. In no other environments/regulatory frameworks have the online platforms agreed to comply with obligations similar to those included in the Code. Even in the US, where often the measures implemented by the platforms against hate speech and fake news are tested and adopted earlier than anywhere else, the platforms are not subject to specific obligations[54] and provide these measures in a voluntary manner (in contrast with the more structured[55] system of the Code). The Code also establishes a cooperation relationship between the platforms and the EU Commission, which then extends to ERGA, in assessing and reporting on the functioning of the Code. By signing the Code and thus voluntarily accepting obligations that are not stemming from the legal framework, the platforms have shown considerable respect towards the EU institutions and have committed to work together with the EU in an attempt to improve their internal rules and procedures.

During 2019, the Code's signatories also showed a clear commitment to making progress in implementing the 5 Pillars of the Code. The provisions of data repositories concerning the political ads, the publications of the Self-Assessment Reports, the availability to meet with the EU Commission and ERGA, the measures adopted, at national level, to promote media literacy and to counter fake news, are all clear efforts aimed at working together with the EU and national institutions with the common goal to counter disinformation. The perception that many NRAs had during its work in 2019 is that the Code's signatories are trying to avoid being perceived as supporting an environment where fake news and disinformation can proliferate unimpeded.

---

[52] Pillar 3 of the Action Plan, at page 9, states that: "the Code of Practice envisages that the signatories will provide a full report after twelve months. These reports should include complete data and information to enable a thorough assessment by the Commission. On this basis, the Commission, assisted by independent expertise and with the help of the ERGA, will assess the overall effectiveness of the Code of Practice".

[53] such as self-regulation and co-regulation, which are broadly used, for example, to foster protection of minors

[54] This was verified when the ERGA monitoring was carried out

[55] Even though the participation into the Code of practice is always voluntary, nevertheless the Code introduces a set of obligations that the platforms commit to comply with

For these reasons, the Code should be regarded as an important step in the process of building the relationship between its signatories, the EU and NRAs.

Nevertheless, the work carried out by ERGA in 2019, shows that **the Code has significant weaknesses** that need to be addressed if it is to achieve its objectives:

1. First of all, as regards the effectiveness of the Code's measure, it is clear that these **measures are too general in terms of content and structure**: this is understandable to a certain extent, because the Code's signatories are very different from each other, but the result of this genericity is that it is difficult to assess the measures' efficiency: the Code lacks definitions and sufficiently precise obligations and is generally based on principles that are difficult to monitor, to compare and to quantify. For example:

   o there is only a general definition of political ads;

   o There is no definition for issue-based ads.

2. Secondly, the ERGA monitoring activities carried out in phase 1 and phase 2 showed that the measures of **the Code's five pillars are not always implemented** (or not always implemented effectively) by the platforms. For example:

   o not all the political ads were labelled as such;

   o little (if any) data was provided about the activities carried out by the platforms at national level;

   o very little information was provided about the accounts removed and the activities aimed at complying with pillars A and C in each Country.

3. Thirdly, **there is a need for greater transparency about how the signatories are implementing the Code.** The Code relies on self-reporting but lacks a mechanism through which the information from these reports can be independently verified. The information provided by the platforms is generally aggregated for the whole EU, which makes it difficult to evaluate the impact of the Code across the EU. This difficulty is amplified at a national level where language, societal and cultural factors make it the most relevant sphere for monitoring the impact and effectiveness of the Code.

4. Lastly, even when the measures are implemented, there is a **serious problem of lack of uniformity** in the procedures (and the definitions) adopted by the different platforms. For example:

   o the procedures adopted by Facebook to <u>identify the sponsors of the political ads</u> were different from the procedures adopted by Google or by Twitter;

   o Facebook is the only Code's signatory who adopted measures aimed at ensuring transparency about <u>issue-based advertising</u>, while Google and Twitter did not report any activity on this matter; this means that the issue based advertising is somehow regulated by one platform but not by the others;

o Facebook contracted some facts checkers to the aim of contrasting fake news, while Google and Twitter did not. In other words, Facebook adopted a <u>procedure to react against the fake news</u> signalled by the contracted facts checkers in some countries, while the other platforms have adopted other procedures, which are very different (for example the timing of the reaction, the degree of demotion/de-ranking and so on);

o the procedures with which the platforms react to the consumers' flagging are very diverse and not transparent.

These are only few examples that demonstrate an obvious lack of uniformity and lack of common approach on how the platforms meet the commitments of the Code. This can partially be explained by the nature of self-regulation that allows for possibilities for signatory specific measures. But this is also due to the fact that **the five pillars of the Code do not include obligations that are valid for all the signatories** and do not provide tools for monitoring the compliance in a co-ordinated manner.

Due to the aforementioned issues (vague definitions, ineffective implementation of the measures, lack of uniformity and common approach to compliance) and also the absence of effective enforcement ERGA believes that **steps are required to increase the effectiveness of the measures of the Code itself and also the oversight\reporting structures if it is to evolve into an effective tool in combating disinformation**. For this reason **ERGA encourages the Code's signatories and the EU Commission to improve the Code and its measures** by requiring that all of the platforms comply with the same obligations in a uniform manner (whenever possible taking into account the specificity of the individual platforms) and adopt more precise definitions, procedures and commitments, as well as measurable KPIs.

Furthermore it should be noted that the scope of the **Code of Practice is limited as it does not apply to all the online platforms** active in the EU, but only to those who have signed it. Even if its measures were effective (and this is not always the case), their reach would not cover the platforms which did not sign the Code, thus creating a regulatory asymmetry. For example, although the current signatories are the main online platforms active in the EU, significant platforms such as TiK-Tok and communication tools such as WhatsApp and Messenger are missing[56].

There is a need for a set of provisions that apply to a way broader number of online platforms active in Europe, as well as a need for provisions allowing the Commission (and the NRAs, if delegated) to carry out specific monitoring activities, also at national level, and to adopt enforcement tools to ensure the compliance to the rules. The above mentioned background suggest that move **from the current self-regulatory model to more structured co-regulation may prove to be more effective to counter disinformation online**.

---

[56] Even though they may be defined as "instant messaging tools" more than "platforms", WhatsApp and Messenger allow the users to share content in closed groups that may contain an enormous amount of people, and offer simple functions to transfer messages/content from one group to another, thus making the content very easily viral. From this point of view, therefore, instant messaging tools may become very efficient tools to spread disinformation. These services are therefore fully concerned by the objective pursued by the Code of Practice, but they are even more difficult to monitor.
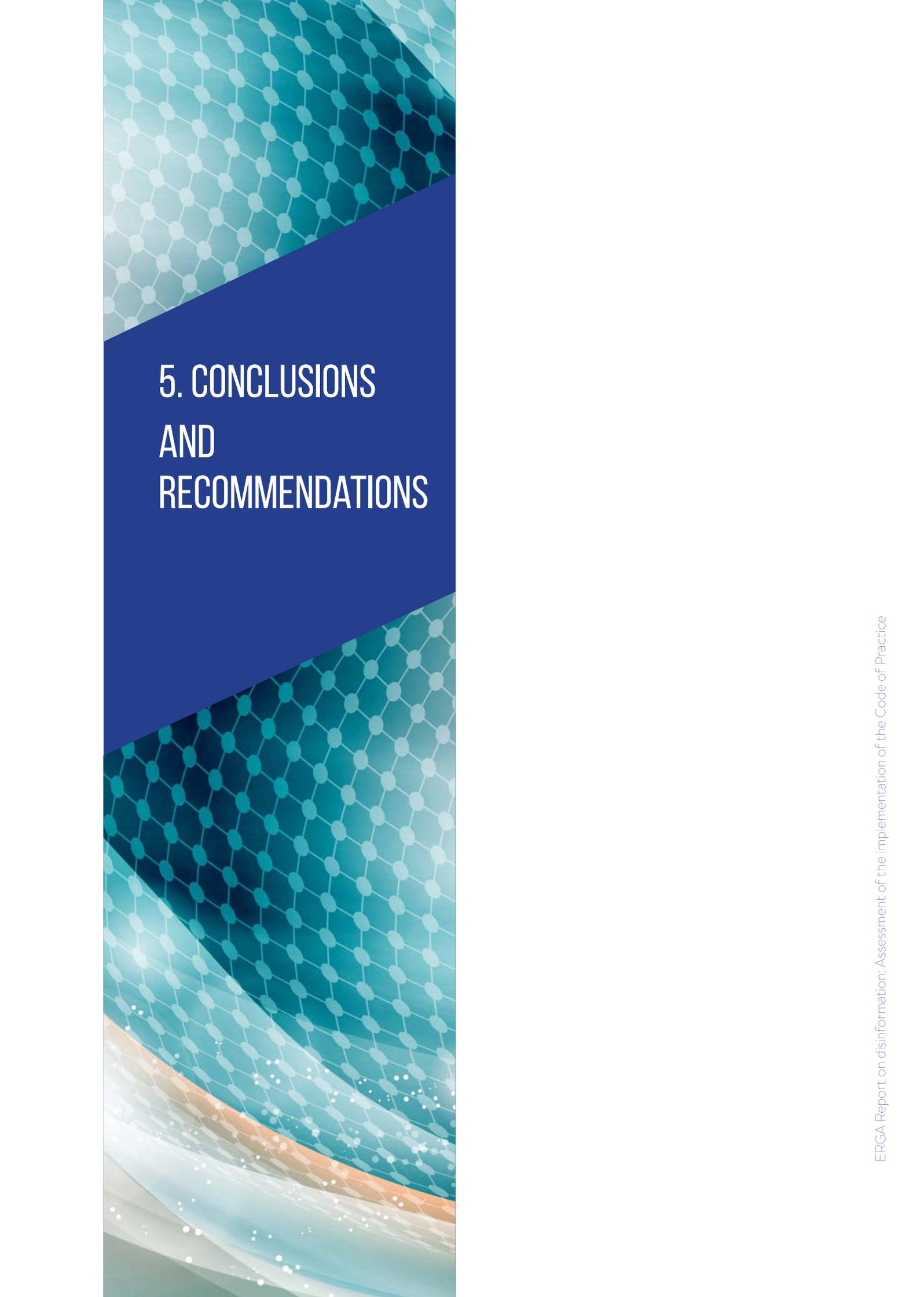
5. CONCLUSIONS AND RECOMMENDATIONS

As already stated in the ERGA *"Report of the activities carried out to assist the European Commission in the intermediate monitoring of the Code of practice on disinformation"*, published in June 2019, and highlighted in section 4 of this Report, the Code of Practice is a unique, very innovative tool. In no other regulatory environment/ framework have online platforms agreed to comply with obligations similar to those included in the Code[57].

Therefore the Code has established a cooperation relationship between the platforms and the EU Commission, which then extends also to ERGA in assessing and reporting on the functioning of the Code. By signing the Code and voluntarily accepting obligations that are not stemming from the legal framework, as well as by making serious efforts in order to provide data repositories concerning the political ads, to publish the Self-Assessment Reports, to put in place activities aimed at promoting media literacy and countering fake news, the platforms have shown considerable respect towards the EU institutions and have committed to work together with the EU in an attempt to improve their internal rules and procedures. For these reasons, the Code should be regarded as a substantial step in the process of building the relationship between its signatories and the EU.

The work carried out by ERGA in 2019, nevertheless, shows that **the Code has significant weaknesses that need to be addressed if it is to achieve its objectives**.

After completing the monitoring activities regarding Pillar B of the Code of Practice (focusing on political advertising) during the first part of 2019 (phase One), the ERGA Task Force decided to focus its attention on the other pillars of the Code (phase Two).

The main sources of information for ERGA's monitoring activities were expected to be the Self-Assessment Reports published by the Code's signatories and the report from the TPO. The NRAs also planned on gathering external information to verify the effective implementation of activities reported in the SAR by the online platforms.

The SARs were published on October 29, 2019 and contained very little Country specific data. Therefore there was very little that the NRAs could verify through their monitoring activity conducted at national level.

In addition, it emerged that the no TPO was appointed to assess the content of the SARs, in breach of the provisions of the Code of Practice.

In order to make up for this lack of data, the participating NRAs decided to conduct an autonomous evaluation, using the limited information available on the SARs and adding data that could be collected from relevant third parties e.g. civil society, consumer associations , journalists, academics, researchers and fact-checkers. Also, whenever relevant and possible, the NRAs were proactively (including through using of individual user accounts) exploring the

---

[57] This was verified when the ERGA monitoring was carried out

tools and other available resources from the platforms and their availability in the individual countries. In addition, some relevant expert reports were considered. Once again, most NRAs had to face the problem of having limited remit, and resources to engage in monitoring online platforms. Nevertheless, many ERGA members recognised the importance of this assignment and offered a valuable contribution, understanding that the new communications technologies are posing serious challenges to the traditional way of regulating (and monitoring) the audio-visual sector, and that these challenges may be tackled only through forward-looking approaches and innovative methodologies.

In total 13 NRAs decided to engage in monitoring the implementation of pillars D and E in line with the approach presented above. Three NRAs agreed to monitor the compliance with the provisions of Pillar B again, focusing on political advertising, during the electoral campaigns preceding the elections in their Countries (Hungary, Poland and the United Kingdom), while 2 NRAs decided to try to monitor compliance to pillars A and C.

As stated in section 3 of this Report, the **results of this intense monitoring activity were as follows**:

- As regards Pillars **A to C**, the main challenge for ERGA in completing its monitoring tasks was the lack of data. This was neither unprecedented nor unforeseen given the experience in Phase One and presented in the Report published on June 2019. This demonstrated that while the data provided by the platforms on the transparency of the political advertising could be meaningful for individual users, it was insufficient to support an effective monitoring activity, since the online repositories presented some information in aggregate or summary form that was lacking the required level of detail.

  This was also the position when dealing with pillars A and C, for which the platforms did not provide (and were not supposed to do so) the API and/or dataset that was provided for Pillar B. In this case ERGA was supposed to gather information for its analysis from the Self-Assessment Reports published by the platforms after 12 months from the adoption of the Code and on the Report that was supposed to be drafted by the Third Party Organization. Unfortunately, the Self-Assessment Reports did not contain disaggregated data concerning specific Countries and the Third Party Organization wasn't appointed. Consequently, its report was never prepared. In addition, the request for information that was addressed directly to the platforms by ERGA (through the Commission) did not produce useful results.

  The problem of lack of data for Pillars A and C was overcome by referring to third party experts who have resources and skills to partly address the gap. However, this situation cannot be sustained in the long run. It is important that the institution which is given the monitoring role is also given adequate tools, information and autonomy to carry out this task. It is crucial that monitors have the ability to create their own queries, filters and analysis tools which should be directed towards the raw, unfiltered and unmanaged data

in the platforms databases, at least over a defined period of time.

- **For Pillar B**, in particular, only Google and Facebook provided access to an API/dataset allowing the NRAs of the Countries where national elections were held to monitor the transparency of political ads. Twitter's repository was not working properly during the Polish electoral campaign[58], and then the platform decided not to allow political advertising any longer starting from November 2019. The result of the monitoring showed little improvement if compared to the results of the monitoring carried out for the EU elections in May, in spite of the recommendations provided by ERGA in its Intermediate Report (June 2019).

  Another problem concerning in particular Pillar B but not limited to it, is the fact that the ERGA Member States have different definitions (or sometimes no definition at all) of concepts like "political advertising" and "issue-based advertising". This is an issue that was raised by the platforms and that needs to be addressed in order to allow a proper monitoring of the platforms' activity and to find solutions that are in line with all the national legal frameworks.

- As regards Pillars D and E, the problem of lack of data from the platforms (again, the Self-Assessment Reports published by the platforms contain little, if any, information that could be verified and measured at national level) was partially overcome by the efforts of the NRAs: they proactively explored the tools and other available resources from the platforms and contacted autonomously the civil society organizations, consumer protection associations, universities, researchers and fact-checkers in order to gather data on the activities implemented by the platforms in their Countries to comply with the provisions of these pillars of the Code.

  o For **Pillar D**, the result was that the platforms are making an evident effort to invest in products, technologies and programs (especially in search, feeds, or other automatically ranked distribution channels) to help people make informed decisions when they encounter online news that may be false (Google Search, for example, is trying to prioritize "reliable" content), to encourage market uptake of tools that help consumers understand why they are seeing particular advertisements and to improve critical thinking and digital media literacy. These efforts, nevertheless, are not made in the same way in all the Countries and, when they are made, they are conducted in a scattered manner: as regards media literacy in particular, they are involving only a tiny fraction of the total population (mainly journalists, politicians and school teachers), usually concentrated in the main cities. These efforts should be part of a more systematic campaign and should address a much bigger part of the population, in the whole national territory.

  o For **Pillar E**, the result is even less uniform: one thing that should be highlighted is that Facebook had contracted fact-checking organizations (all of which are part of renown

---

[58] For more information on the availability of Twitter's ad repository during the Polish electoral campaign, see section 3.2.2 of the report

international organizations of fact-checkers) to detect fake news and to be able to react promptly. The reaction takes the form of a "warning label" that is shown to the users when they try to share the content that was identified as "fake news". However, Facebook did not contract fact-checkers in all the EU Countries (possibly because it was difficult to find fact-checkers belonging to renown international organizations in all of the EU Member States) and, in any case, the other platforms did not follow the same path. Besides, the relationship between Facebook and the contracted fact-checkers could be improved (for example, it should somehow involve also the deep fakes and the statements from politicians) and should contain some directives aimed at ensuring the prioritization of the topics that are particularly relevant for the public opinion.

Although Facebook showed significant efforts in contracting fact-checkers to detect fake news, another solution should be found in order to make the efforts of the various platforms more uniform and less dependent on contracts signed with a specific organization in one or few Countries.

Pillar E also contains the commitment from the platforms to *"take the necessary measures to enable privacy-compliant access to data for fact-checking and research activities" and to "cooperate by providing relevant data on the functioning of their services, including data for independent investigation by academic researchers and general information on algorithms".* The contacts made by the ERGA NRAs with the universities and the researchers show clearly that the platforms provided very little (if any) access to data for independent investigations. To their defence, the platforms argue that they cannot provide freely access to data because of privacy and data security reasons, but these reasons are not fully convincing and, in any case, it is evident that a solution must be found.

Based on the aforementioned outcome of its monitoring activity carried out during the whole 2019, **ERGA proposes a set of recommendations**, based on three different levels of intervention:

1. Recommendations aimed at improving the monitoring of the existing Code's commitments;

2. Recommendations aimed at expanding the existing Code's commitments;

3. Recommendations aimed at exploring new (more effective) tools to counter disinformation.

5.1 Recommendations aimed at improving the monitoring of the existing Code's commitments

This set of recommendations has been identified with the aim to improve the monitoring of the commitments deriving from the Code that are currently valid and have been agreed upon by the platforms: ERGA would like to avoid facing the same problems that were detected in 2019 and to improve its monitoring activity. ERGA believes that these recommendations if implemented would contribute to bring the relationship between the platforms and the EU institutions a step higher.

1. The first recommendation is that a **set of relevant definitions should be drafted** on e.g. political ads, issue-based ads, fake news/manipulation of information and so on. The same set of definitions might be used also by the online platforms when dealing with these issues. This effort should help ensure a consistent approach towards these issues/ principles in the whole EU.

2. The second recommendation aims at improving the provision of information by the platforms: as said multiple times in this report, the data provided by the platforms so far has been insufficient and its validity could not be verified. This was due -inter alia- to the fact that ERGA and the EU Commission had asked for information through letters that the platforms considered insufficient; to solve this problem, instead of asking for data with ad hoc letters, ERGA recommends that **platforms make available datasets[59]**, data monitoring tools and Country specific information allowing the NRAs to create their own queries, filters and analysis tools which should be directed towards the raw, unfiltered and unmanaged data in the platforms databases, at least over a defined period of time (i.e. every 6 months for a week that is randomly chosen by ERGA and the Commission). **The structure of this data should be proposed by ERGA and by the Commission** and should be similar (with limited pre-agreed flexibility in the case of structural differences between platforms) for all the platforms[60].

3. The third recommendation is that some **sets of guidelines should be drafted** with the aim to:

   – Improving and rationalizing the **relationship between the platforms and the fact-checkers** (see Section 3.5.1);

   – Improving and harmonizing the **platforms' reactions to consumers complaints and flagging** (see Section 3.4);

   – improving the **media literacy campaigns** in each Country, for example involving the regulators, running Webinars targeting different groups of people, publishing short ads in which people who are readily recognizable by Internet users (i.e. YouTubers

---

[59] For pillar 2, the platfomrs might make available an API as they did on the occasion of the electoral period before the EU elections, in May 2019

[60] For example, as highlighted by the Hungarian NMHH, platforms should provide a searchable and filterable report (or database) with all banned political advertisements altogether with the reason of the ban..

and influencers) explain the concept of disinformation, and implementing additional training initiatives (see Section 3.4.1);

– improving the **relationships between online platforms and researchers** in accordance to the proposals made by the DLM's report (see Section 3.5.2).

4. The last recommendation, recognizing the importance of the work done by fact-checking organisations at national level, is to explore the opportunity for ERGA to **build a cooperation with these organizations and with the new European Digital Media Observatory (EDMO)**. EDMO will be operational during 2020 and it will serve as a hub for fact-checkers, academics and researchers to collaborate with each other and actively link with media organisations and media literacy experts, and provide support to policy makers[61]. In particular, it might support ERGA and the EU Commission in monitoring and analysing disinformation campaigns that may achieve Europe-wide reach and help solve the issue of the relationship among platforms and fact-checkers: once the Observatory through is fact-checking members has fact-checked the news and identified malicious content, then the platforms may be informed so that they can react in a uniform manner.

5.2 Recommendations aimed at expanding the existing Code's commitments

Section 4 of this Report has shown that the measures currently included in the Code are not as effective as they should be, for a number of reasons. This situation can hardly be reconciled with the objective of protecting a democratic society which is endangered by politically motivated disinformation. The existence of this danger is confirmed by the current analyses and the numerous discussions with stakeholders in recent months and by the over-abundance of false news, guided by profit-oriented and/or political purposes, that has accompanied the the recent outbreak of and response to COVID-19 in all the platforms available in Europe (not only those that signed the Code)[62]. It is therefore necessary to consider revising the Code of Practice.

In the following, ERGA would like to provide some initial recommendations which would require the agreement of the EU Commission and of the platforms, to the aim of expanding the commitments of the Code and making them more effective. This is without prejudice to any further ERGA recommendations aimed at expanding the existing Code's commitments in the future.:

1. The first recommendation is that the problem of lack of uniformity addressed in Section 4 should be solved. To that end, ERGA might analyse further (additionally to the work in 2019) the commitments and compare the way the platforms implement them and then make recommendations aimed at **harmonizing the implementation of these commitments**. For example:

---

[61] https://ec.europa.eu/digital-single-market/en/news/commission-launches-call-create-european-digital-media-observatory

[62] as discussed in the introduction to this report.

- all platforms will have to <u>define and react to issue-based ads</u> (currently only Facebook does this);

- all platforms will have to <u>label political ads in a similar manner</u> and possibly with the same wording;

- the <u>procedures to register the political ads sponsors</u> should be as similar as possible in all the platforms;

- as far as possible, all platforms will have to adopt the <u>same procedure to react to fact-checked news</u> (same timing, same degree of demotion/de-ranking, same reactions to each fact-checked news).

2. The second recommendation is that (beyond the recommendation 2 aimed at improving the monitoring of the existing Code's commitments above) the Code should formally (taking into account the experience from the proposed enhanced monitoring) identify **specific moments of the year in which the platforms should provide data** on the implementation of the Code (i.e. twice a year) that include **Country specific information**, so to allow the NRAs to conduct a regular and proper monitoring activity.

3. The third recommendation is, of course, to make any possible effort to **increase the number of platforms signing the Code,** in order to avoid regulatory asymmetries, while ensuring that the code is flexible enough to adapt to differences of resources or nature of services.

5.3 *Recommendations aimed at exploring new (more effective) tools to counter disinformation*

As was stated in this report multiple times, the elements of the Code that should be improved in order to make the Code more effective are the non-compulsory feature and the excessive genericity of the commitments (which allows the platforms to decide not to comply with some obligations, or to do so in ways that sometimes show no consistency/uniformity at all, thus potentially generating confusion among the users), the difficulty in verificating their implementation and the lack of enforcement measures.

If the model of the Code of Practice has to be improved in a more effective manner, then the solution of the abovementioned issues clearly points toward a more structured form of regulation. To this effect, a **shift from the current flexible self-regulatory approach to a more co-regulatory one would be required**. Essentially this would involve an evolution of the current self-regulatory Code to provide for more consistence in its formulation and in its implementation and the introduction of a formal backstop mechanism to deliver the required monitoring and enforcement elements.

The important part of every co-regulatory system in general is a mechanism that incentivises

industry players to take part in a self-regulatory structure. Therefore, there usually is a parallel system of rules, to which the entities by default belong if they do not take-up the self-regulatory obligations.

Existing backstop mechanisms are already functioning in other areas on a member state level and these tend to be grounded in EU and Member States legislation that provides for a state-founded, albeit often independent, authority. In the context of EU-wide initiative, like the one represented by the Code of Practice, there would have to be other routes to put this element into practice. The ERGA experience in 2019 indicates that the effective development and implementation of the Code requires such a framework. As a connection between the elements outlined above, the **operative rules should be put in place. These should consist of clear reporting obligations, more harmonized procedures and appropriate timeframes**.

This is the solution that ERGA recommends to enhance the relationship with the platforms.

Ideally, **all the platforms which distribute content in Europe should be engaged in this co-regulation procedure** and should then be subject to the co-regulatory obligations. Should this not be the case, however, in order to solve the problem of the regulatory asymmetry, under which the Code's commitments are valid only for the platforms that signed the Code of Practice but not for the platforms which are not involved by the Code, the **EU institutions might explore the possibility of adopting a more conventional regulatory approach** (statutory regulation), as it is already happening in some EU Member states (i.e. France, Germany). Such an instrument should ensure a common approach towards the implementation of the Code's measures, provide a legal basis for enforcement and ensure that the required access to valid data is granted.

With the current review of the regulatory framework that should culminate with the announced Digital Services Act, ERGA sees the value in a holistic approach to the governance of online content regulation. In this overall framework, **the DSA-package should create at least a framework that would also include the basis for the effective fight against disinformation (liability regime). In addition, a dedicated legal act is needed to address the problem more directly and in greater depth**. Such a separate instrument (e.g. a regulation) would ensure not only a level of detail of provisions and comprehensive coverage of stakeholders but also the legislative speed required given the threat the current information crisis presents to European democracies.

ANNEX 1

Questions for pillar A – Scrutiny of ad placements:

1. Have the platforms deployed policies and processes in your Country aimed at disrupting advertising and monetization incentives for relevant behaviours, such as misrepresenting material information about oneself or the purpose of one's properties?

2. Have the platforms restricted advertising services or limited paid placements in your Country?

3. Have the platforms promoted and/or included the use of brand safety and verification tools in your Country?

4. Have the platforms enabled engagement with third party verification companies in your Country?

5. Have the platforms provided advertisers in your Country with necessary access to client-specific accounts to help enable them to monitor the placement of ads and make choices regarding where ads are placed?

6. Have the platforms assisted and/or allowed advertisers in your Country to assess media buying strategies and online reputational risks?

Questions for pillar B – political advertising:

The NRAs located in Countries were elections have taken place or are going to take place after the publication of the ERGA Report may try to answer again to the questions already provided during monitoring phase one, formulated on the basis of the precise wording of the Code of Practice as regards transparency of political ads:

1. What is the degree of transparency of the political and issue-based advertising?

2. Is the platform adopting appropriate and efficient measures to enable users to understand why they have been targeted by a given advertisement?

3. Is the platform adopting appropriate and efficient measures to ensure that political ads are clearly distinguishable and are readily recognisable as a paid-for communication or labelled as such?

4. Is the sponsor's identity publicly disclosed?

5. Are the amounts spent for the political ads publicly disclosed (at least in price ranges)?

6. What progress has been made on the commitment to publicly disclose "issue-based advertising"? Have platforms provided a definition of issue-based ads and complied with it?

Additional questions (already provided for during monitoring phase one):

7. Are the definitions of political ads and issue-based ads adopted by the platforms consistent with requirements set out in the legislation of the NRA's Member State?

8. Is the "registration/ authorization" procedure for advertisers of political ads effective?

How effective is the process to identify, and remove or re-label political adverts that did not meet the verification requirements?

9. Is the archive presented in a user-friendly manner (e.g. is it searchable and analysable) and does it contain all the required information as defined in questions 2, 4 and 5?

Final question (not previously asked during phase one):

10. Was there any improvement in the platforms' procedure or activity since the latest monitoring, whose results are published in the ERGA report?

Questions for pillar C – Integrity of services:

1. Have the platforms put in place clear policies regarding identity and the misuse of automated bots?

2. Have they enforced these policies in your Country? Please provide any type of data available. In particular:

    a. What is the number of bots disabled for malicious activities in your Country on a monthly basis since January 2019?

    b. If possible, provide additional info on the reason and the completeness of the disabling.

3. Have the platforms put in place policies on what constitutes impermissible use of automated systems and made this policy publicly available and accessible to the users?

4. Have they enforced these policies in your Country? Please provide any type of data available. In particular:

    a. What is the number of fake accounts identified and removed in your Country on a monthly basis since January 2019?

    b. If possible, provide additional info on the reason and the completeness of the accounts' removal

    c. What is the number of posts, images, videos or comments acted against for violation of platform policies in your Country on a monthly basis since January 2019?

Questions for pillar D: empowering consumers

1. Are the products, technologies and programs identified by the platforms in the SAR effective in your Country in helping people to make informed decisions when they encounter online news that may be false, including by supporting efforts to develop and implement effective indicators of trustworthiness in collaboration with the news ecosystem?

2. Are the products, technologies and programs identified by the platforms in the SAR to prioritize relevant, authentic and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels effective in your Country?

3. Are the products, technologies and programs identified by the platforms in the SAR

effective in your Country that make it easier for people to find diverse perspectives about topics of public interest?

4. Is there evidence that the activities and initiatives identified in the SAR in relation to partnerships with civil society, governments, institutions, and other stakeholders to support efforts aimed at improving critical thinking and digital media literacy were implemented and effective?

5. Did the platforms encourage market uptake of tools in your Country that help consumers understand why they are seeing particular advertisements (i.e. why they have been targeted by a particular ad)? Are these tools easy to see and to access?

6. Do the platforms inform the users in your Country on how their data are used by them? Do the platforms give tools to consumers and inform them on how they can control and personalize the use of their data?

7. Have the platforms provided tools for users in your Country to report false news?

8. Have the platforms organized specific events, or workshops, campaigns, videos or any other tool in your Country to inform consumers or journalists about the means to counter disinformation?

9. Please provide an evaluation on the adequacy of the activities carried out by the Code's signatories to empower consumers.

Questions for pillar E: empowering the research community

10. Are the initiatives identified in the SARS effective in supporting good faith independent efforts in your Country to track Disinformation and understand its impact, including the independent network of fact-checkers?

   • Have the platforms partnered with a fact-checking organization in your Country?

   • Have the platforms supported the fact-checkers community in your Country?

   • Based on a reasonable sample of the activity can you conclude if the fact-checking arrangements are effective in your Country?

11. Have the platforms shared privacy protected datasets, undertaken joint researches, or otherwise partnered with academics and civil society organizations in your Country?

12. Are the initiatives identified in the SARS effective in encouraging (researchers in the field of disinformation and political advertising in your Country?

13. Are the initiatives identified in the SARS effective in meeting the commitment to organized events in your Country to foster discussions within academia, the fact-checking community and members of the value chain?

14. Please provide an evaluation on the adequacy of the activities carried out by the Code's signatories to empower the research Community.

ERGA Report on disinformation: Assessment of the implementation of the Code of Practice
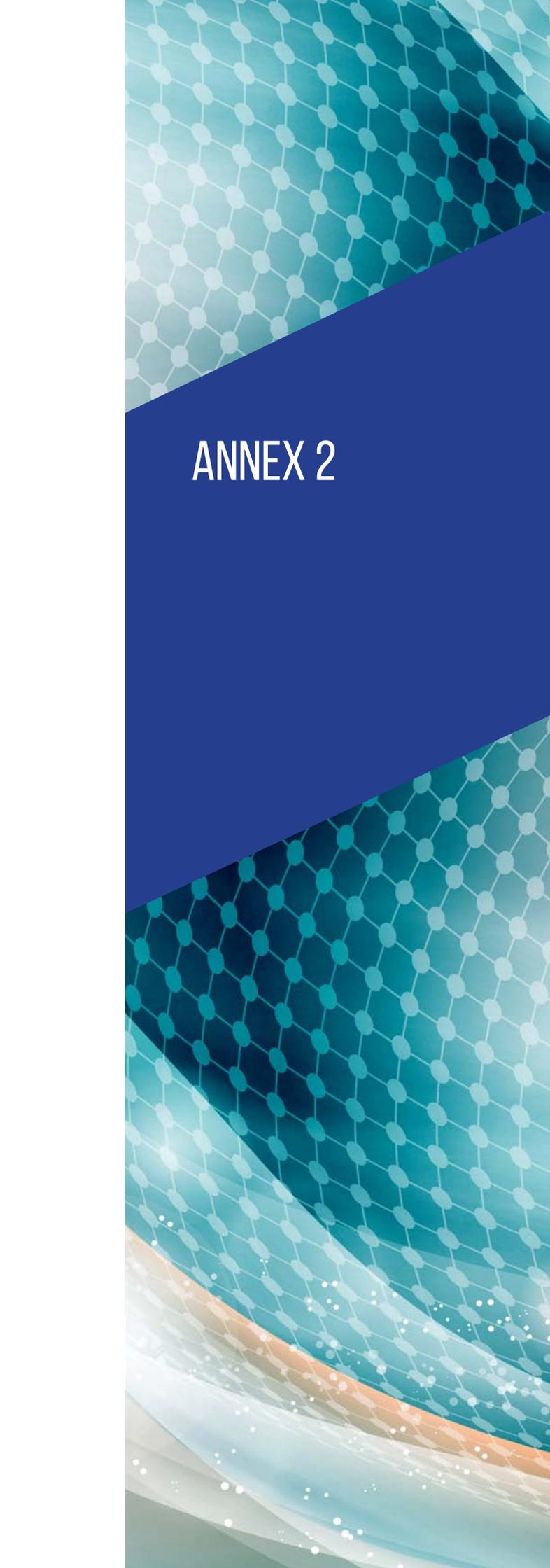
ANNEX 2

In the following pages the answers to the 14 questions of Pillar D and E are summarized. The answers were provide by the participating NRAs after the monitoring activities carried out between November and December 2019.

Questions for monitoring level A Pillar D: empowering consumers

Q1 Are the products, technologies and programs identified by the platforms in the SAR effective in your Country in helping people to make informed decisions when they encounter online news that may be false, including by supporting efforts to develop and implement effective indicators of trustworthiness in collaboration with the news ecosystem?

Facebook

- **positive answers:** Facebook introduced three measures 1) Users are notified if they have shared or are about to share information that has been fact-checked and given the result of 'false' or 'mixture' with enhanced warnings for false videos and photos, 2) the provision of explanatory articles written by fact-checkers alongside the fact-checked content in the newsfeed, and 3) the Context Button feature which allows users to view more information about the websites, publishers and images they see on the platform[64].

- **negative answers: several NRAs recognize that tools are easily accessible, but not easy to find and see in all cases. In addition, many of these features have not been implemented in several countries[65]:** "context button" and "related articles" have not been implemented yet in order to ensure their full effectiveness. "False news" are one of the possible answers presented in the "Find support or report post" hidden on the three-point menu provided with the posts shared in the News Feed. In this case, it may be argued that although this is an available option, its visibility to users is not immediate (such as are for example the options: "like", "comment" or "share"). This means that its use requires that users explore these features opportunities[66].

- One NRA recognizes that there is no label for false information, but the context button is available, it is easy to find and provides a lot of information[67]. One NRA states that they haven't encountered these features during the monitoring. Nonetheless, Facebook asserts that the context button has been implemented[68].

In addition, Facebook did not report on initiatives to place a label directly on content that has been fact-checked and given the result of 'false' to alert people to the trustworthiness of a post in conjunction with the news ecosystem. Facebook announced on Oct 21st 2019 that it

---

[63] Slovakia, Ireland, Hungary, Germany, Portugal, France, Poland, Cyprus, Slovenia, Italy, Croatia, Latvia
[64] Ireland, Germany, Latvia
[65] Croatia, Poland, Portugal, Sweden, Ireland, Italy, France, Cyprus, Hungary
[66] Croatia, Portugal, Italy, France, Cyprus, Poland
[67] Latvia
[68] France

will clearly label content that has been labelled 'false' or 'partly false' in an effort to protect the 2020 US elections. It is not clear how this will be implemented or distributed across the EU member states[69].

In 5 EU Member States there is no designated certified third-party fact-checker recognized by Facebook; the Slovak regulator, CBR, states that this means unavailability of the features described in Facebook reports that depend on fact-checked content, for example the notifications before sharing fact checked content, labels for video/photo misinformation. In these Countries users may still report posts and news as "fake": a Slovak NGO, in particular, obtained the status of "trusted flagger". In any case, Facebook does not provide any feedback to the users (not even to the trusted flagger) about the results of their labelling the content as fake.

Google

- **positive answers:** there are two Fact Checking Tools: Fact Check Explorer and Fact Check Markup Tool. The products are easily accessible, easy to use and reliable from the consumer's point of view. In addition, Google is developing transparency standards that help to assess the quality and credibility of journalism[70] easily.

- **negative answers:** two NRAs state that the tools listed in the SAR seem to be not implemented[71]. One NRA states that there are no indicators of false news on YouTube[72]. Two NRAs claim that there are no media from their country involved in the Trust Project mentioned in the SAR[73].

Twitter

- **positive answers:** one NRA recognizes that on Safer Internet Day 2019 Twitter supported their global network of safety partners. Some of the partners are NGOs acting in the field of Freedom of Expression and could have worked on the issue of disinformation (e.g.: CDT, FLIP...)[74].

- **negative answers:** several NRAs state there are no indicators on the interface of platforms regarding the quality of the content published on Twitter[75].

[69] Ireland
[70] Croatia, Poland, Italy, Portugal
[71] Cyprus, Latvia
[72] Latvia
[73] Croatia, France,
[74] France
[75] Croatia, Portugal, Ireland, Germany, Hungary, Latvia

Q2 – Are the products, technologies and programs identified by the platforms in the SAR to prioritize relevant, authentic and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels effective in your Country?

Facebook

- **positive answers:** some NRAs recognize that Facebook has improved his machine learning capabilities. Machine learning models identify articles' links which might be false. We can use the predictions model to prioritize the links showing the third-party fact-checkers[76].

- **negative answers**: many NRAs state that the solutions described by Facebook are not consistent when they come to combat disinformation directly. Professionally detected disinformation is "de-ranked", but continues to appear in the news-feed. For example, the distribution of vaccine misinformation has been de-ranked in the News Feed. However, a search of other contested terms such as 'global warming' did not reveal authoritative content providers as prioritised at the top of the feed and so it cannot be verified that the processes employed by the platform to direct users to authoritative sources are sufficient and comprehensive[77].

One NRA[78] mentioned several problems in this area, that were highlighted by the organisations consulted during the monitoring:

- the software Newshwhip shows that often, on a given topic, less authoritative content seems to be more successful and popular;

- there was also an example of a channel where numerous pieces of content have been already deleted by the platform, but this seemed not to have any significant impact on its popularity;

- several legitimate accounts appeared to have been blocked in the past based on coordinated flagging of accounts of local NGOs/personalities fighting against disinformation;

- some disinformation stories keep resurfacing even though they have been debunked in the past;

- photos of famous personalities have been used by some illegitimate accounts to increase their reach and number of friends; the claims by these personalities did not lead to any result, while the illegitimate accounts remained unblocked by Facebook.

Google

- **positive answers**: several NRAs s recognize that "Top news shelf on YouTube" are implemented and easily accessible[79].

---

[76] Portugal, France, Poland
[77] Ireland, Germany, Hungary, Latvia
[78] Slovakia
[79] Poland, Italy, France, Hungary, Cyprus

- **negative answers**: several NRAs recognize that "Breaking news" and Top news shelves on YouTube, Publisher transparency and Information panels providing topical context are not available[80].

Twitter

- **negative answers**: some NRAs recognize that there are no indicators on the interface of platforms regarding the quality of the content published on Twitter[81].

Q3 – Are the products, technologies and programs identified by the platforms in the SAR effective in your Country that make it easier for people to find diverse perspectives about topics of public interest?

Facebook

- **negative answers**: one NRA recognizes that it has no data to conclude on the effectiveness of products, technologies and programs in facilitating access to different perspectives on topics of public interest. The SAR doesn't stress clearly any tool aiming the objective to facilitate access to different perspectives on topics of public interest, it can be considered that the "related articles" feature can contribute to this greater diversity[82].

- several NRAs recognize that measures reflect greater transparency in relation to what content consumers are seeing on Facebook as a result of their own activities and preferences but they do not address the provision of diverse perspectives to users about topics of public interest[83]. The 'Context' button is the feature that may supplement such feature; however, this feature is not active in some Countries[84].

Google

- **positive answers:** some NRAs recognize that Google tools (full Coverage in Google News) are very helpful to find diverse perspectives on the same topic[85]. One NRA' state that "full coverage" is only working for foreign news[86].

- **neutral assessment**: One NRA recognizes that Google offers products that can make it easier for people to find different perspectives on topics of public interest. There is no data available in the SAR on the success and effectiveness of these products[87].

Twitter

- **positive answers**: some NRAs recognize that Twitter lets the choice to the users to "see the most popular tweets" first or to "see the more recent tweets" first. The content seen by the users might therefore be different and more diverse[88]. However, it is the only tool

---

[80] Croatia, Portugal, Ireland, Cyprus, Latvia, France, Germany, Hungary
[81] Croatia, Portugal, Ireland, Hungary, Latvia
[82] Portugal
[83] Croatia, Ireland, France, Hungary, Italy, Latvia , Germany
[84] Slovakia, Germany
[85] Croatia, Hungary, Portugal, Cyprus, France
[86] Latvia
[87] Germany
[88] France, Latvia

described in the SAR.

- **negative answers**: some NRAs recognize that while Twitter's response suggests that the algorithm does do this, there is no way to verify if this is the case[89].

Q4 - Is there evidence that the activities and initiatives identified in the SAR in relation to partnerships with civil society, governments, institutions, and other stakeholders to support efforts aimed at improving critical t hinking and digital media literacy were implemented and effective?

Facebook

- positive answers: several NRAs confirm that specific events/discussions aimed at improving critical thinking and digital media literacy were implemented[90].

  Media Information Literacy initiatives can take many forms, but whatever the context, the scale, or the event, platforms should take a common and integrated strategy in order to promote a specific literacy media plan for each European country. In this regard, an interesting initiative was organized by Facebook in Italy.

  In Italy, Facebook recently carried out an experiment in collaboration with Walter Quattrociocchi using a participatory format. The program called "Memedia: media literacy nel mondo dei meme" was carried out at Binario F in Rome. It was attended by 40 students' representatives of Roman schools, with the task of becoming ambassadors of this important topic in their respective schools. Students worked together on the production of content, starting from a meme and finishing with communication strategies related to controversial news stories. Through the simulation of real-life case scenarios, students get to experience and understand the dynamics that govern the digital world and the platforms we all use. By reflecting on these mechanisms, the students become much more aware of their impact and effects.

- **negative answers**: some NRAs recognize that signatories have provided incomplete information on activities and initiatives related to partnerships with civil society or other stakeholders to support efforts to improve critical thinking and digital media literacy. The effectiveness of the implementations cannot be assessed[91]. In other cases, based on the feedback from stakeholders, some NRAs indicate that the current efforts are not sufficient and seem to be part of marketing strategies that do not have any practical effect[92].

- One NRA' states that Facebook has launched websites about how Facebook works, how ads work and how to know if the information is false, but no one of these websites are available in their national language[93].

---

[89] Ireland, Germany
[90] Ireland, France, Portugal, Poland, Slovakia
[91] Croatia, Sweden, Italy, Germany, Latvia, Hungary
[92] Slovakia
[93] Latvia

Google

- **positive answers**: several NRAs confirm that specific events/discussions aimed at improving critical thinking and digital media literacy were implemented[94].

- **negative answers**: five NRAs recognize that signatories have provided none or incomplete information on activities and initiatives related to partnerships with civil society, governments, institutions and other stakeholders to support efforts to improve critical thinking and digital media literacy[95].

Twitter

- **positive answers:** some NRAs confirm partnership with civil society, governments, institutions, and other stakeholders to support efforts aimed at improving critical thinking and digital media literacy but it is difficult to know how effective the initiatives are[96].

- **negative answers:** three NRAs recognize that signatories have provided none or incomplete information on activities and initiatives related to partnerships with civil society[97].

Q5 – Did the platforms encourage market uptake of tools in your Country that help consumers understand why they are seeing particular advertisements? (i.e. why they have been targeted by a particular ad)? Are these tools easy to see and to access?

Facebook

- **positive answers:** several NRAs confirm that "Why am I seeing this ad?" And "Ad preferences" tool are implemented and easy to see and access[98].

- **negative answers:** some NRAs claim that the visibility of this feature is not immediate. The feature "Why am I seeing this ad?" does not appear on the post itself; rather, to access this information the user is required to click on the three dots on the top right-hand corner of the post[99]. One NRA stated that it did not detect any particular campaigns or activities carried out by Facebook on this matter (in addition to few reported popup windows for the user on the availability of the feature "Why am I seeing this ad?"[100].

Google

- **positive answers:** several NRAs confirm all tools are implemented and easy to see and access[101].

- **negative answers:** Two NRAs recognize that the visibility of this option is not immediate[102].

---

94 Poland, Ireland, Portugal, France, Hungary
95 Croatia, Italy, Cyprus, Germany, Slovakia
96 Ireland, Italy, Cyprus, France, Germany
97 Croatia, Portugal, Hungary,Latvia
98 Croatia, Poland, Portugal, Italy, Cyprus, France, Germany, Sweden, Latvia
99 Portugal, Ireland, Hungary
100 Slovakia
101 Croatia, Poland, Italy, Cyprus, Hungary, France, Germany
102 Portugal, Latvia

Twitter

- positive answers: several NRAs confirm all tools are implemented and easy to see and access[103].

- **negative answers:** some NRAs recognize that the visibility of this option is not immediate. The information is sparse and offer two broad statements as to why a user might see an advert[104].

Q6 - Do the platforms inform the users in your Country on how their data are used by them? Do the platforms give tools to consumers and inform them on how they can control and personalize the use of their data?

Facebook

- **positive answers:** several NRAs confirm that users can easily find information on the use of data by Facebook and modify the use of their data in the section "Your information on Facebook"[105].

- **negative answers:** many NRAs recognize that the functions "Your information on Facebook" and "Off-Facebook Activity" are not introduced yet[106].

Google

- **positive/negative answers:** several NRAs recognize that Google clearly and repeatedly points out that the data protection information should be noted by the user. From the consumer's point of view, the use of control options requires a high level of competence[107].

Twitter

- **positive/negative answers:** twitter also provides structured information on the use of personal data and possibilities for control and personalisation. Several NRA' confirm that the use of control options requires a high level of competence[108].

Q7 - Have the platforms made tools for users in your Country to report false news?

Facebook

Facebook has set up a way to report false messages. From the point of view of consumer protection, there is a fundamental criticism of the signatories' reporting options. Several ERGA members recognize that some content has quickly and reliably been deleted; others take longer or are not deleted at all. The decision is not always comprehensible, satisfactory or transparent[109].

---

[103] Croatia, Ireland, Italy, Cyprus, France, Germany, Latvia
[104] Portugal, Ireland,
[105] Croatia, Ireland, Portugal, Germany, Latvia, Poland, Sweden, France
[106] Ireland, Italy, Cyprus
[107] Croatia, Poland, Germany, Hungary, Portugal
[108] Germany, Portugal, Ireland, Hungary, Latvia, France, Croatia
[109] Ireland, Germany, Italy, Portugal

Google

- **positive answers:** some NRAs confirm that Google implemented tools to report false news[110].

- **negative answer:** one NRA explains that there is a possibility to report "Spam or Misleading" video content on YouTube[111]. One Slovak organization, which is trusted flagger for YouTube, proactively finds and regularly reports videos with problematic content on YouTube. The problem seems to be that YouTube, as opposed to Facebook, only seldom removes flagged videos by this trusted flagger. There is also a problem with reporting comments under the videos (when there is simply no feedback from the platform, so the flagger has to keep track of the reported comments and regularly checks if they have or have not been taken down).

- One NRA recognizes that it is not possible to report false information on Google[112].

Twitter

- **negative answers:** Two NRAs recognize that in the option "Report the ad" or "Report a tweet" where one can report the problem for various reasons, disinformation is not mentioned among the "reasons"[113]. One NRA states that there is no information on this in the SAR[114].

Another NRA confirms that Twitter doesn´t have the direct option to report false news in the case of individual pieces of content. There is a possibility to report an account on Twitter on the basis of it being "suspicious or containing spam" and in the second rollout window there is a possibility to specify that "the account is false". But there does not seem to be a possibility to report an individual post rather than an account for this category of problematic content as there is for other types of content like hate speech or self-harm[115].

Q8 - Have the platforms organized specific events, or workshops, campaigns, videos or any other tool in your Country to inform consumers or journalists?

Facebook

- **positive answers:** several NRAs confirm that Facebook organized some specific events and workshop to deliver a Digital Skills Training Programme[116].

- **negative answers:** one NRA claims that no specific events, workshops, campaigns, videos or any other tools took place on the topic how to counter disinformation for consumers or journalists[117]. In another case, in Slovakia, Facebook stated that -in view of the Elections- it had contacted the Slovak Election Commission to ensure proper

---

[110] Croatia, Ireland, France, Portugal, Poland
[111] Slovakia
[112] Latvia
[113] Croatia, Latvia
[114] Germany
[115] Slovakia
[116] Croatia, Ireland, Sweden, Italy, France, Portugal, Germany, Poland, Hungary
[117] Croatia

channels of information and to train it about protecting the integrity of the elections and fight the spread of misinformation on the platform. According to statements and overall impressions of journalists and representatives of NGOs that took part in meeting and seminars organized by Facebook, however, these events were more PR oriented rather than focusing on actual working and examining of the disinformation and fake news on social media and the approach was from the point of global strategy not specifically aimed on the local focus.

Google

One NRA states that there have been some events organized, but according to the feedback from the consultations with stakeholders they seemed to be a part of PR strategy and had not any practical effect on the ongoing problems[118].

Twitter

- **positive/negative answers:** many NRA' confirm that specific events/discussions, and partnerships with research and academic institutions remain episodic and largely inadequate to support any rigorous analysis and monitoring of online disinformation trends[119].

Q9 - Please provide an evaluation on the adequacy of the activities carried out by the Code's signatories to empower consumers

According to one NRA, many of the platforms have provided a range of tools and information portals such as the capacity to report information as problematic. However, it is not clear how often these features are employed by Irish consumer's, what actions were taken by them or what actions each company took on the basis of these potential complaints. In order to accurately assess the level to which consumers are empowered to understand, report and impact information on these platforms, it is necessary for the company signatories to report the national totals of complaints received via these tools as well as the corresponding actions taken by them to highlight this content as problematic, remove it entirely or to address the publishers of this content[120].

Another NRA states that some of the products, technologies and programs mentioned by the platforms in the SARs are designed to help people make informed decisions when they come across online news that may be wrong. However, some of the products and technologies have only recently been introduced or are hardly advertised, so their reach and effectiveness are limited. The signatories provide tools to help consumers understand why they see certain ads. These tools are easily accessible. It is not easy to see them in all cases. Therefore, no statements can be made about their use and thus their effectiveness[121].

---

[118] Slovakia
[119] Positive:Ireland, Germany, France; Negative: Croatia, Portugal
[120] Ireland
[121] Germany

On the same point, other NRAs state that several of the measures presented in the SARs by the platforms are in reality non available in its Country. For example, there appears to be no evidence of activities, relating to the commitments of empowering consumers, improving critical thinking, and planning and conducting media literacy campaigns[122].

Two NRAs states that the activities carried out by Facebook and Google seem to be adequate, but the extent in which they are implemented is rather narrow and refers only to some categories of people and to activities carried out in very few cities[123]. One of these two NRAs, in particular, highlights that the ordinary Internet users don't seem to be aware of available tools, which in some cases are not so easy to find. Some of the tools are available only in English, which might cause problems for certain groups of consumers[124].

One NRA[125] states about the measures adopted by the three platforms that based also on the consultations with local stakeholders:

- of all the three platforms, Facebook is the most active and advanced. But in general the main issue seems to be that there is no designated certified third-party fact-checker recognised by Facebook. This in consequence means unavailability of several of the main features as described in Facebook reports since they depend on the availability of local fact-checked content. The absence of a local fact checker might be also having impact on the indicators that inform users about the trustworthiness of the content;

- many of the features described by Google were available (Full Coverage features in Google News, "rich snippets" and "dedicated tags"). But some of these only help the user to a limited extent to understand whether the content is trustworthy, less trustworthy or even fake. Besides, some of the features (Breaking News and Top News) are not available, although Google states the opposite in its report. As for initiatives and partnerships on the local level regarding media literacy, none of the measures referred to in the SAR by Google is focusing on specific Countries. Moreover there is no tool for reporting fake news in Google News and Search. The user only can improve what Google News shows for her/him. In practice, there is a possibility to report "Spam or Misleading" video content on YouTube, although this could be potentially confusing to users (also feedback from the stakeholders) as there is no clear category for disinformation content;

- on Twitter is impossible to report "false news", but only suspicious content (or spam)

Questions for pillar E: empowering the research community

Q10 – Are the initiatives identified in the SARs effective in supporting good faith independent efforts in your Country to track Disinformation and understand its impact, including the independent network of fact-checkers? Have the platforms supported the fact-checkers

---

[122] Croatia
[123] Poland and Italy
[124] Poland
[125] Slovakia

community in your Country? Have the platforms partnered with a fact-checking organization in your Country? Based on a reasonable sample of the activity can you conclude if the fact-checking arrangements are effective in your Country?

Facebook

- **negative answers:** Facebook is not collaborating with fact checking organizations in Cyprus and in Slovakia. It does not have a fact checker partner in Sweden, either[126]. Several NRAs reported a problem with the exclusion of politicians' posts and ads from debunking and lacking statistics about the dissemination of disinformation and of content flagged as false.

- **positive answers:** several NRAs confirm that Facebook collaborates with fact checking organizations actively in their Country[127].

Google

- **positive answers:** some NRAs confirm that Google collaborates with fact checking organizations actively[128].

- **negative answer:** Two NRAs confirm that there's currently no fact-checker operating in collaboration with Google[129].

Twitter

- **Positive/negative answers:** twitter is not collaborating with fact checking organizations on a regular basis. However even without official support by Twitter, a wide range of journalistic fact-checking organizations are active on the platform[130].

As for any specific initiatives to encourage research or to foster discussion with local academia or fact-checking community among the events that the company mentions in its report, there are none related to Slovakia and the CBR consultation did not show any information about this kind of activity organized by Twitter in Slovakia.

Q11 – Have the platforms shared privacy protected datasets, undertaken joint researches, or otherwise partnered with academics and civil society organizations in your Country?

Facebook

- **positive answers:** many NRAs confirm that Facebook strongly invested in building for the necessary infrastructure by installing a team and providing three access points for Facebook data: Crowd Tangle, Ad Library API and URLs Data Set.[131]

  Based on the reports from NGOs, academics and institutions are using tools that have been developed for scraping data from Facebook (although this might be problematic

---

[126] The Swedish fact-checker partner was Viralgranskaren until recently. Viralgranskaren was part of the Swedish newspaper Metro who went bankrupt. Facebook is now looking for a new partner to collaborate with
[127] Croatia, Poland, Portugal, Ireland, Italy, France, Germany
[128] Poland, Italy, France, Portugal
[129] Cyprus, Hungary
[130] Germany
[131] Ireland, Portugal, Sweden, Italy, France, Hungary

due to the fast-changing nature of the settings on the platform that impacts external collection of data) and they are using also official Facebook data sets to certain extent (Crowd Tangle, and external services like News Whip). However, there was no experience with Facebook individually sharing privacy protected databases[132].

- **negative answers:** one NRA confirms that researchers have received access to the Facebook Ad Library API providing data on political or issue-based advertisements published on the platform. However, there are significant concerns in relation to the comprehensiveness of the data provided through the API, its limited functionality and user friendliness which limit its usefulness for effective independent analysis and monitoring[133].

Google

- **negative answers:** some NRAs recognize that datasets have serious limitations, enabling only a limited range of scientific research projects[134].

Twitter

- **negative answers**[135]**:** one NRA confirms that all the information can be found in the main website, but only in English. If one needs to gather more details or data, the user has to write from his/her e-mail address[136].

Q12 - Are the initiatives identified in the SARS effective in encouraging researchers in the field of disinformation and political advertising in your Country?

Facebook

- **negative answers:** some NRAs confirm a lack of initiatives aimed at encouraging researchers in the field of disinformation and political advertising[137].

Google

- **negative answers:** several NRAs confirm a lack of initiatives aimed at encouraging researchers in the field of disinformation and political advertising[138].

Twitter

- **negative answers:** several NRAs confirm a lack of initiatives aimed at encouraging researchers in the field of disinformation and political advertising[139].

Q13 - Are the initiatives identified in the SARs effective in meeting the commitment to organized events in your Country to foster discussions within academia, the fact-checking community and members of the value chain?

---

[132] Slovakia
[133] Ireland, Croatia, Cyprus, Poland, Hungary
[134] Poland, Italy, Croatia, Cyprus, Hungary
[135] Croatia, Ireland, France, Poland, Hungary
[136] Hungary
[137] Croatia, Poland, Portugal, France, Italy
[138] Croatia, Poland, Portugal, Ireland, France, Germany, Hungary
[139] Croatia, Ireland, France, Germany, Hungary

Facebook

- negative answers: some NRAs haven't identified any initiatives specifically organized to foster discussions within academia, the fact-checking community and members of the value chain[140].

Google

- positive answers: several NRAs confirm that Google organized some specific events and initiatives to foster discussions within academia and the fact-checking community[141].

- negative answers: some NRAs haven't identified any initiatives specifically organized to foster discussions within academia and the fact-checking community[142].

Twitter

- positive answers: one NRA confirms that Twitter organized some specific events and initiatives to foster discussions within academia and the fact-checking community[143].

- negative answers: some NRAs haven't identified any initiatives specifically organized to foster discussions within academia and the fact-checkers[144].

Q14 – Please provide an evaluation on the adequacy of the activities carried out by the Code's signatories to empower the research Community

According to one NRA, the signatories to the Code, Facebook, Twitter, Google all reported on the introduction of several policies, tools and training programmes to empower the research community and to provide them with access to data. In evaluating the adequacy of these initiatives within Ireland, the researchers found that although some progress have been made in this regard, there exists significant room for improvement across the four platforms. The most significant data sources provided to the research and fact-checking community in the run up to the European elections in Ireland, was in from of online libraries or archives of political adverts made publicly available on the platforms in the run up to these elections. However, as reported upon in the Elect Check 2019 report, the data available within these repositories has significant limitations for the purposes of comprehensive research and monitoring. Particularly so in relation to the nature of the targeting practices of online advertisers and the amounts spent by them[145].

Another NRA states that the platforms definitely made some steps forward in favour of transparency and against disinformation, but these actions are far not as effective and well-published as they should be. "There are no warnings or signs related to disinformation, we have no specific knowledge of companies working with domestic fact-checking sites, and we have not found any evidence to that effect. In addition, we have only a few information about

---

[140] Croatia, Portugal, Ireland, Poland
[141] Ireland, Italy, France, Portugal
[142] Croatia, Poland, Cyprus, Germany, Hungary
[143] Ireland,
[144] Croatia, Portugal, Hungary
[145] Ireland

events, programs, or grants held in Hungary"[146].

Two NRAs highlight that active collaboration with fact-checking organizations are most prominent within the actions undertaken by Facebook. However, the personal resources of the fact checker are limited, and retrospective debunking has in general only limited potential to stop the dissemination of disinformation due to the delay in publication. Retrospective debunking has other important functions such as improving the resilience of the users and by making disinformation visible. Google supports fact checking mostly through tools such as the Fact Check Markup Tool and by providing training for journalists. Twitter is not collaborating with fact-checking organizations on a regular basis. The imbalance between researchers and platforms with regard to data control and data access is one of the main challenges: while platforms such as Facebook, Twitter and YouTube deliver large quantities of personal data to their respective businesses that can be incorporated into specific business products such as targeted advertising, researchers have only limited access to such data[147].

One NRA explains that the interviewed researchers indicate lack of sufficient activity from the Code's signatories: the privacy protected datasets are not offered for the scientists, while there was no big event organized to foster discussion within academia[148].

One more NRA confirms that "Facebook has partnered with fact-check organization from Croatia". As regards Google, "we can only point out their partnership with IFCN, and Political Ads Transparency Report (Google data) which can be downloaded as a CSV and is published as public data". Eventually, as regards Twitter, "we can only point out Twitter API which enables academics and researchers to conduct their own investigations. Other than that, there were no activities to empower the research community carried out by the three platforms in Croatia"[149].

Another Regulator, comparing the activities carried out by the three platforms, states that:

- Facebook is providing an opportunity to get research grants for academics, but for NGOs and Slovak academics it is difficult to compete with others on the global level due to limited resources and some structural limitations in size and capacity on the Slovak level;

- Google is currently not partnering with fact-checking organization in Slovakia. This is probably because no such organization is a member of Poynter's International Fact-Checking Network (IFCN). Another big issue, according to the stakeholders, is the lack of quality and transparency in the process of content moderation. Google does not provide any data on this issue concerning Slovakia. NGOs and research community in Slovakia claim that supporting the work of researchers who explore the issues of disinformation is not sufficient;

Twitter is currently not partnering with fact-checking organization in Slovakia. Also based on the feedback from the consulted stakeholders there was no significant experience with

---

[146] Hungary
[147] Germany and Italy
[148] Poland
[149] Croatia

reporting problematic content on this platform as the penetration on the Slovak market is quite low and it was reported that the quality of content on this platform remains relatively high. As for any specific initiatives to encourage research or to foster discussion with local academia or fact-checking community among the events that the company mentions in its report, there are none related to Slovakia and our consultation did not show any information about this kind of activity organized by Twitter in Slovakia[150].

---

[150] Slovakia