EUROPEAN
COMMISSION

Brussels, 10.9.2020
SWD(2020) 180 final

COMMISSION STAFF WORKING DOCUMENT

Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement

Disclaimer

This document is a European Commission staff working document for information purposes. It does not represent an official position of the Commission on this issue, nor does it anticipate such a position.

Contents

Assessment of the Code of Practice on Disinformation.

Achievements and areas for further improvement

1. Introduction: overview of the Code of Practice

This Staff Working Document sets out the key findings of the Commission services' assessment of the implementation and effectiveness of the Code of Practice on Disinformation (the Code)[1] during its initial 12-months period of operation. This assessment is based on (i) the self-assessment reports submitted by the Code's signatories[2]; (ii) the monitoring report provided by the European Regulators Group for Audiovisual Media Services (the ERGA Report)[3]; (iii) a study procured by the Commission from an independent consultancy, Valdani, Vicari and Associates (the VVA Study)[4]; and (iv) the Commission's Report on the 2019 elections to the European Parliament (the Elections Report).[5] A summary of these contributions is attached as an **Appendix.**

One of the signature initiatives originating out of the Commmision's April 2018 *Communication on tackling online disinformation: A European Approach* (the April 2018 Communication),[6] the Code sets out principles and commitments for online platforms and the advertising sector that its signatories have agreed to implement, on a voluntary basis, to counter online disinformation in the EU. Drafted by a Multistakeholder Forum convened by the Commission[7], the Code includes a Preamble, a statement of Purposes, and a set of 15 Commitments prefaced by explanatory comments that refer to the objectives of the April 2018 Communication, detail the commitments' scope and purposes, and provide context. The Code's commitments are organised under five pillars:

A. Scrutiny of ad placements
B. Political advertising and issue-based advertising
C. Integrity of services
D. Empowering consumers
E. Empowering the research community

[1] https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

[2] https://ec.europa.eu/digital-single-market/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019

[3] https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf

[4] https://ec.europa.eu/digital-single-market/en/news/study-assessment-implementation-code-practice-disinformation

[5] https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1123

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

[7] The Multistakeholder Forum consistied of a "Working Group" composed of the major online platforms and associations from the advertising sector as well as a "Sounding Board" composed of representatives of the media, academia and civil society. The Working Group drafted the Code, and the Sounding Board provided advice and an Opinion on the Code. The Code was published on 26 September, along with the Opinion of the Sounding Board. https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.

Signatories are required to identify which commitments they adhere to, in light of their relevance to the products or services they provide, and to set out the specific policies and actions they intend to pursue to implement the Code[8].

On 16 October 2018, the Code's initial signatories provided their formal subscriptions to the Code[9]. The initial signatories were Facebook, Google, Twitter and Mozilla as well as the trade association representing online platforms (EDIMA) and trade associations representing the advertising industry and advertisers (the European Association of Communications Agencies (EACA), IAB Europe, and the World Federation of Advertisers (WFA) as well as WFA's Belgian national association, the Union of Belgian Advertisers). Since the Code's inception, additional signatories have subscribed. These include Microsoft and TikTok as well as the French, Czech, Polish and Danish national associations affilated with EACA[10].

2. Monitoring of the Code of Practice

The Code requires signatories to report on the implementation of their commitments, in the form of annual self-assessment reports, and to cooperate with the Commission in assessing the Code, including the provision of information upon request and responding to questions. The signatories further committed under the Code to selecting an objective third-party organisation to review the annual self-assessment reports and to evaluate the level of progress made against the commitments.

The December 2018 *Joint Action Plan against Disinformation* (the Action Plan)[11] set out an intensive monitoring programme for the Code and called upon the European Regulators Group for Audiovisual Media Services (ERGA) to assist the Comission with the monitoring. The monitoring programme included a targeted intermediate monitoring phase, whereby the Commission sought to verify that effective policies pertinent to the integrity of the electoral processes were in place before the May 2019 elections to the European Parliament. The results of the intermediate monitoring phase are summarised in the June 2019 *Progress Report on the Action Plan*.[12] The Action Plan further charged the Commission to carry out a comphrensive assessment of the Code at the conclusion of its initial 12-month period of application; this Staff Working Document presents the results of the Commission's assessment.

Finally, on 10 June 2020, the Commission and the High Representative issued a *Joint Communication on Tackling COVID-19 disinformation*[13] (the June 2020 Communication), which focuses on the immediate response to disinformation around the coronavirus pandemic, taking stock of steps taken by platforms during the outbreak of the crisis and setting out concrete follow-up actions. This

[8] The Code is intended to apply within the framework of existing laws of the EU and its Member States, and is not to be construed in any way as replacing or interpreting the existing legal framework. Code, at p. 2.

[9] https://ec.europa.eu/commission/news/code-practice-fight-online-disinformation-2018-oct-16_en

[10] These national associations are, respectively, Association des Agences Conseils en Communication (AACC), Stowarzyszenie Komunikacji Marketingowej /Ad Artis Art Foundation (SAR), Asociace Komunikacnich Agentur (AKA) and Kreativitet & Kommunikation. In addition, The Code's most recent signatory is Goldbach Audience (Switzerland) AG, an ad sales business that works with publishers and operates primarily in the Swiss German-language market.

[11] https://ec.europa.eu/commission/publications/action-plan-disinformation-commission-contribution-european-council-13-14-december-2018_en

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

[13] https://ec.europa.eu/info/sites/info/files/communication-tackling-covid-19-disinformation-getting-facts-right_en.pdf

Communication *inter alia* called upon the Code signatories to make available monthly reports on their policies and actions to address COVID-19 related disinformation and encouraged other stakeholders who were not signatories to the Code to participate in this monitoring process on a voluntary basis.

For the purposes of the present assessment, the actions implemented by the signatories of the Code in response to COVID-19 related disinformation technically fall outside the scope of the 12-month period of evaluation of the Code's performance, but they will be taken into account as an important contextual element.

3. Commission's assessment

Based on the lessons learned from the Commission services' monitoring of the Code, as well as the inputs referenced above, it is possible to identify positive results of the current self-regulatory framework as well as shortcomings that need to be addressed. This document, while highlighting the Code's main achievements, focuses in particular on the Code's shortcomings, with a view to stimulating reflection on improvements that may be required and strengthening the dialogue with stakeholders.

3.1 Achievements of the Code

As reflected in the key conclusions of the ERGA Report and the VVA Study, there is a consensus that the Code is an important and necessary first step towards achieving the goals of the Commission's April 2018 Communication and the Action Plan. A description of the measures taken by the Code signatories can be found in the Commission analysis of the self-assessment reports covering the Code's first year of operations[14] and in the June 2020 Communication.

In essence, the Code has provided a framework for a structured dialogue between relevant industry actors, the Commission, and ERGA authorities, and greater transparency of platforms' policies against disinformation within the EU. This framework has set general policy objectives, identified relevant requirements for appropriate measures, and enabled public disclosure of information relating to the implementation of those measures, thereby contributing to increase the platforms' accountability. This represents significant progress over the situation prevailing before the Code's entry into force.

Pursuant to the Code, the signatories have introduced new policies aimed at achieving the agreed objectives, modified or clarified accordingly their terms of service, and engaged in supporting activities, such as collaborations with fact-checkers and media literacy initiatives.

As regards pillar A of the Code (scrutiny of ad placements), signatory platforms have enforced policies to prevent their services from being used to spread misrepresentative or misleading advertisements and have blocked or suspended ad accounts of "imposter websites", i.e. sites that misrepresent their identity or purpose, or scrape content from other sources in order to generate income from ad placements. This has effectively contributed to reducing monetisation incentives for actors that disseminate disinformation online for economic gain. For instance, in March and April 2019, Facebook took action against over 600.000 ads each month in the EU which violated its policies

[14] Analysis Code of Practice Annual Reports, at: https://ec.europa.eu/digital-single-market/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019

on low quality or disruptive content, misleading or false content and circumvention of its systems; between September 2018 and August 2019, Google reported 314.288 actions taken against EU-based Google Ads accounts for violations of its Google Ads Misrepresentation policy, and 55.876 actions for violations of its Google Ads Original Content policy; Twitter rejected 11.307 ads for violation of its Unacceptable Business Practices Policies and 10.639 ads for violation of its Quality Ads Policy in the period from January 2019 through August 2019.

Moreover, in order to facilitate the development of adequate brand safety tools for advertisers, the WFA, one of the Code signatories, launched the Global Alliance for Responsible Media in June 2019, involving a wide range of stakeholders from the digital advertising ecosystem. The goal of this Alliance is to "work collaboratively to identify actions that will better protect consumers online, working towards a media environment where hate speech, bullying and disinformation is challenged."[15]

As regards pillar B of the Code (transparency of political and issue-based advertising), the main signatory platforms introduced, for the first time in the EU, systems enhancing the transparency and public disclosure of political advertising in the run-up to the 2019 European Parliament elections. These systems included a requirement for all political ads to be clearly labelled as sponsored content, clearly identifying candidates, political parties or sponsors. These policies, which initially applied to the European elections, were also extended to national elections in the Member States. Moreover, the systems put in place under pillar B included the creation of online repositories[16] (ad libraries) intended to contain all political ads served, while making available application programming interfaces (APIs) enabling users and researchers to perform customised searches. On this basis, it has been possible for interested stakeholders to gather detailed information about the volume and budget of political ads served by different political advertisers in the Member States, as well as information about number of impressions[17] and basic targeting criteria (age, gender, and location). For example, between March and September 2019, Facebook served some 444.000 political ads in the EU27, totalling around 31.5 million euros of political ads spend; in the same period, Google successfully verified 376 advertisers to run election ads in the European elections and labelled more than 185.000 election ad creatives. Microsoft updated its advertising policies in April 2019 to prohibit globally ads for election-related content, political candidates, parties, ballot measures and political fundraising. On Twitter, by the end of May 2019, there were 27 certified political campaign accounts, of which 21 ran ads and accounted for some 23.253.153 impressions and a total of 98.531 euros in revenues. It is relevant to note that in October 2019 Twitter decided to ban all political ads.

As regards pillar C of the Code (integrity of services), signatory platforms took action against accounts using manipulative techniques to artificially amplify the reach and impact of false or misleading information. Notably, signatory platforms started providing information on their efforts to remove or prevent the creation of fake accounts and the use of malicious bots and other spamming techniques, including coordinated disinformation campaigns by hostile actors. Platforms have mostly relied on artificial intelligence to automatically detect and block hundreds of millions of fake accounts[18]. They also gradually increased the quality of disclosed information about coordinated

[15] https://wfanet.org/knowledge/item/2019/06/18/Global-Alliance-for-Responsible-Media-launches-to-address-digital-safet

[16] https://www.facebook.com/ads/library, https://transparencyreport.google.com/political-ads?hl=en

[17] The number of impressions corresponds to the number of times an ad has been displayed online.

[18] 99.6% of fake accounts actioned by Facebook in the second quarter of 2020 were found and flagged by the company before users reported them Source: https://transparency.facebook.com/community-standards-enforcement#fake-accounts

inauthentic behaviour and influence operations. Actions against this kind of behaviour resulted in several thousands of accounts, pages and groups, both from third countries and from within the EU, being removed from the different platforms. Facebook disabled 2.19 billion fake accounts in the first quarter of 2019 and between January and October 2019 removed some 7.606 Accounts, Pages and Groups engaging in coordinated inauthentic behaviour; between September 2018 and August 2019, Google removed over 10.842.500 YouTube channels for violation of its spam, misleading, and scams policy and more than 56.500 channels for violation of its impersonation policy; and between January and August 2019, Twitter actively challenged 126.025.294 accounts platform-wide.

As regards pillar D of the Code (empowering consumers), signatory platforms took a broad range of actions including investments in technology to give prominence to trustworthy information sources on their content ranking and recommender systems, while making it easier for users to find diverse perspectives about topics of public interest. For instance, Facebook notifies users when they share content that was fact-checked and rated as "false" or "mixture" and makes it easier for users to view information, via a Context Button, about websites and publishers they see on Facebook. Via the "Full Coverage" feature in Google News, users can access context and diverse perspectives about news stories from a variety of publishers, and in September 2019 Google announced ranking updates that give more prominence in Search to articles identified as significant original reporting, which will stay longer in a highly visible position; The "Microsoft News" service partners with over 1.000 news sources worldwide, which are all vetted by Microsoft to ensure that the service only shows licensed reputable content. At the same time, the platforms' collaboration with the fact-checking community has provided users with additional possibilities to critically assess information accessed online, and enabled the development of new features giving users more contextual information about fact-checked websites or webpages, with the aim to reduce the spread of false narratives online. Moreover, new tools have been provided for users to flag potential instances of disinformation and be warned about content that has been fact-checked and rated as false or misleading. Platforms have also promoted the development of trustworthiness and credibility indicators for online sources (e.g. through the Trust Project[19] and the Credibility Coalition[20]) and participated in the work of the Journalism Trust Initiative.[21]

In particular, since the outbreak of the COVID-19 crisis, signatory platforms have ensured due prominence to information from public health authorities, the World Health Organisation (WHO) and professional media, while reducing the distribution of content fact-checked as false or misleading, or removing content directly harming public health, safety and security, in violation of their terms of service. For instance, the Facebook and Instagram Info hubs directed more than 2 billion people to resources from health authorities, including the WHO, through their COVID-19 Information Centre. Microsoft News created COVID-19 information hubs in 53 markets across the globe, with an experienced team editing content from more than 4.500 trusted news brands. YouTube reviewed over 100.000 videos related to dangerous or misleading coronavirus information and removed over 15.000 videos in March and April 2020, while Twitter took robust action against content conflicting with guidance provided from public health authorities, including labelling tweets containing potential harmful misleading information related to COVID-19.

[19] https://thetrustproject.org/

[20] https://credibilitycoalition.org/

[21] https://jti-rsf.org

As regards pillar E of the Code (empowering the research community), platforms implemented a number of policies and tools intended to provide researchers and the fact-checking community with access to platform data. For instance, as a result of these policies, Twitter disclosed the first comprehensive archive of state-backed information operations on Twitter in October 2018 and made new datasets available in January, June, August and September 2019, providing access to more than 30 million Tweets. Researchers in 15 EU countries accessed these datasets over 20,000 times. Google released a large dataset on visual deep-fakes to facilitate the development of detection methods for synthetic videos. In April 2018 Facebook launched a partnership with Social Science One (SS1), a group of 83 academic researchers, to share data with the academic research community while maintaining stringent privacy protections.

These endeavours have allowed for unprecedented public scrutiny of the measures taken by the Code signatories to counter disinformation within the EU, although the quality of their reporting varies across platforms and is not consistent across the five pillars of the Code, as explained in section 3.2.3.

3.2 Shortcomings of the Code

The available evidence also suggests that the Code presents a number of shortcomings. These can be grouped in four broad categories: (i) inconsistent and incomplete application of the Code across platforms and Member States, (ii) lack of uniform definitions, (iii) existence of several gaps in the coverage of the Code commitments, and (iv) limitations intrinsic to the self-regulatory nature of the Code.

3.2.1 Inconsistent and incomplete application of the Code across platforms and Member States

While some disparities reflect objective differences in the platforms' respective technical attributes and services offered, the Commission's review of the self-assessment reports reveals inconsistencies that cannot be justified on this basis.

(a) Scrutiny of ad placements (Pillar A)

One of the Code's purposes is to improve the scrutiny of ad placements and disrupt the economic drivers that contribute to the dissemination of disinformation online. The self-assessment reports indicate efforts by the platforms to better scrutinise ad placements and reduce advertising revenues to purveyors of disinformation. However, the reported policies pursue a range of objectives, some of which are not specifically tailored to tackle disinformation as defined in the Code (e.g. the restriction of misleading advertising, unsupported commercial claims, deceptive business practices). Moreover, and critically, they fail to distinguish between:

- measures aimed at scrutinising paid-for placements of socially influential content (i.e. issue-based advertising) **on the platforms' own services** (e.g. on YouTube videos, Facebook's newsfeed, or Twitter's timeline); and
- measures aimed at restricting ad placements **on third-party websites** that are used as vectors for disinformation campaigns.

As regards measures addressing **the placement of issue-based ads on platforms' own services,** no evidence has been provided demonstrating a consistent implementation of specific restrictions on ad accounts sponsoring verifiably false or misleading information. This shortcoming may be attributable, in part, to the absence of a definition agreed among signatories, as explained in section 3.2.2, or well-developed understandings of the concept of "issue-based advertising" as opposed to commercial advertising[22]. It can also be the result of an inadequate collaboration with independent fact-checking organisations to enable effective action against issue-based ads determined to be false or misleading.[23].

More generally, a more effective scrutiny of ad placements on platforms' own services would require a better integration of fact-checking and research activities in order to enhance knowledge about sources, vectors and targets of disinformation campaigns with a view to enabling a more pro-active approach, and therefore a more comprehensive and timely response, towards the placement of misleading issue-based ads.

As regards measures to limit **ad placements on third-party websites that purvey disinformation**, the platforms' enforcement appears to have focused on the low-hanging fruit of so-called "imposter websites", i.e. sites that misrepresent their identity or purpose. It is unclear the extent to which, if at all, this scrutiny has been effective in addressing ad placements on websites that do not misrepresent themselves but persistently purvey disinformation. As noted, reports from civil society organisations suggest that the placement of advertising on third-party disinformation websites continues unabated[24]. Moreover, according to the ERGA Report, monitoring in Germany and Italy demonstrated that "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 ad placement."[25] The VVA Study found that "The Code does not have a high enough public profile to put sufficient pressure for change on the platforms in this area."[26]

Failure to achieve progress in this area seems to be the consequence of a lack of effective and joined-up participation by relevant stakeholders, including platforms, ad exchanges, fact-checking organisations and, notably, advertisers[27]. Possible options for increasing transparency and enabling a better scrutiny of ad placements have been identified by some stakeholders, but none of them have been operationalised yet. These options include:

- The development and regular update of lists of websites identified by fact-checkers as systematic purveyors of disinformation and the removal or temporary suspension by platforms and ad exchanges of the accounts concerned (a "black list" approach aimed at establishing a risk assessment for relevant ad accounts);
- The development and regular update of lists of websites identified by fact-checkers as occasional purveyors of disinformation, and the provision of tools enabling advertisers to selectively exclude such websites from hosting their ads (a "grey list" approach); and

[22] ads about social issues such as e.g. economy, health or environmental policies

[23] Commitment No. 1 engages relevant signatories to disrupt advertising and monetisation incentives "for relevant behaviours, such as misrepresenting material information about oneself or the purpose of one's properties. These policies can include, for example, the … limiting of paid placements, and could potentially take place in partnership with fact-checking organisations" (emphasis added).

[24] *Global Disinformation Index, Research Brief: Ad Tech Fuels Disinformation Sites in Europe – The Numbers and Players* (March 2020), at: https://disinformationindex.org/wp-content/uploads/2020/03/GDI_Adtech_EU.pdf.

[25] ERGA Report, at p. 24. The national authorities from two Member States – Germany and Italy – monitored national-level implementation of commitments under pillar A.

[26] VVA Study, at p. 39.

[27] Code, Section (I), para.(ii); see also Code, pillar II.A, 3rd Recital. ("The Signatories recognise that "all parties involved in the buying and selling of online advertising and the provision of advertising services need to work together … to effectively scrutinise, control and limit the placement of advertising on accounts and websites belonging to purveyors of Disinformation.")

- Ex ante approval by ad-placement service providers of websites selling advertisement space, possibly based on trustworthiness indicators agreed with advertisers (a “white list” approach).

It should also be noted that the aggregated reporting from the trade association signatories representing the advertising sector does not provide clarity on the extent to which brand safety practices are evolving to encompass the control of advertising placements next to disinformation content.

(b) Transparency of political and issue-based advertising (Pillar B)

The evidence provided to the Commission shows persisting insufficiencies in the implementation of policies designed to ensure an adequate level of transparency for political and issue-based ads.
First, the **identification and public disclosure of issue-based ads** has not been properly and systematically ensured, as the signatories of the Code have followed different approaches. Moreover, none of the signatories has provided specific information and verification tools to evaluate the actual impact of the transparency measures adopted in this area.

Second, the signatory platforms have not considered measures to ensure that **labels remain visible** when ads are organically shared[28].

Third, as noted by the ERGA report and the VVA Study, while all online platforms have created ad-libraries accessible through APIs to facilitate the running of queries by researchers and civil society, concerns remain regarding **the limited functionalities of the APIs, the completeness of the repositories, and the quality of searchable information**. These concerns include platforms’ discretion in altering or restricting access on a unilateral basis, as well as inconsistencies regarding the disclosure of sponsors’ identity, amounts spent and targeting criteria.

Fourth, criticism has been levelled regarding the absence of **uniform registration and authorisation procedures for political ads**. Such criticism involves alleged systemic loopholes enabling political advertisers to circumvent the registration requirements, as well as potential obstacles to the organisation of pan-European campaigns by European political groups.

Additionally, questions remain on the monitoring and proper application of electoral laws online as well as other relevant legislation in the context of political advertising. The European Democracy Action Plan, planned for the end of 2020, will look into solutions to ensure greater transparency on paid political advertising at EU level.

(c) Integrity of services (Pillar C)

The self-assessment reports demonstrate that platforms have put in place policies to counter the use of manipulative techniques and tactics on their services, including measures to address spammy or inauthentic behaviour, fake accounts and malicious, bot-driven activity.

However, as reporting on such manipulative behaviour is provided at aggregated and global level, it is not possible to evaluate precisely the impact and relevance for the EU of the policies at issue. In particular, the level of transparency at Member State level is still inadequate as regards actors, vectors,

[28] ERGA notes that when political ads are shared on Facebook, “the ‘Paid for by’ disclaimer vanishes …. [which] is an important limitation to the effectiveness of the system.” ERGA Report, at p. 19.

targets, content, delivery mechanisms and propagation patterns of messages intended to manipulate public opinion. More transparency is also required about the levels of user engagement with detected disinformation campaigns.

In addition, the platforms' reporting on coordinated inauthentic behaviour has mainly focused, so far, on foreign actors. However, civil society and research organisations have pointed to an increasing role played by domestic actors seeking to manipulate public opinion from within the EU[29].

Timely analysis of evolving threats and trends of disinformation campaigns and more effective strategic communication to increase societal resilience cannot be properly pursued in the absence of more accurate and detailed information.

(d) Empowering consumers (Pillar D)

The platforms' self-assessment reports include information on a number of tools, deployed during the first 12 months of the Code, intended to improve findability of trustworthy content, provide context on information sources and warning labels, discover alternative viewpoints, and flag possible disinformation cases. Section 3.1 includes some examples of these tools. Besides, during the COVID-19 pandemic, the main online platforms have experimented with new tools and methods aimed at raising users' awareness regarding false, misleading, unsubstantiated or disputed content relating to the virus[30]. However, the reports do not address sufficiently a number of important concerns.

First, the information provided by the signatories so far is insufficient to establish whether these tools are fully and equally available across the EU, in all languages[31].

Second, no data has been made available to demonstrate the extent to which such tools are effective to increase user engagement with trustworthy information sources, enhance critical thinking, and promote civic behaviour online.

Third, in the absence of a transparent and systematic analysis of the effectiveness of such tools, media literacy initiatives, fact-checking efforts or journalistic responses to relevant disinformation campaigns risk being untimely or sub-optimal.

Fourth, there is no dedicated, user-friendly and uniform procedure available on all platforms for users to flag possible disinformation cases and be adequately informed about the outcome of their actions.

Fifth, when a piece of content has been evaluated as false or misleading by independent fact-checking organisations, different platforms have implemented different approaches aimed at raising users' awareness, including targeted warnings to users having viewed such content, but no indicators have been provided to enable an independent evaluation of their comparative performance.

[29] T. Davis, S. Livingston & M. Hindman, Suspicious Election Campaign Activity on Facebook: How a Large Network of Suspicious Accounts Promoted Alternative Für Deutschland in the 2019 EU Parliamentary Election (GWU 2010). The GWU study found, for example, that one medium-sized party in Germany, whose support never in the polls never exceeded 15 % during the months leading up to the elections to the European Parliament, received 86 % of total shares and 75 % of all comments on German political Facebook — four times the comments, and about six times the shares, of all other political parties combined.

[30] https://techcrunch.com/2020/03/16/facebook-reddit-google-linkedin-microsoft-twitter-and-youtube-issue-joint-statement-on-misinformation/?guccounter=2

[31] ERGA points out that during the assessment period "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 or translated yet for the European market". ERGA Report, at p. 26.

Sixth, while several platforms have collaborated with stakeholders to develop trustworthiness indicators, as highlighted by the VVA Study, no detailed information is available on the integration of those indicators in their search services and recommender systems.

(e) Empowering the research community (Pillar E): Cooperation with fact-checkers

The Code expressly envisions the collaboration between the platforms and fact-checking organisations in view of various objectives, including prioritised online display of authentic content, tracking of disinformation content, and demonetisation of disinformation websites. In general, the level and form of such collaboration vary considerably the across platforms and Member States, as does the platforms' follow-up with respect to content that has been fact-checked.

Effective empowerment of the research community requires consistent use of fact-checking services across platforms, with full coverage of all EU Member States and languages. Apart from multi-bilateral agreements with selected fact-checking organisations, online platforms have not considered other cooperative models such as open and non-discriminatory collaborations with independent fact-checking initiatives that fulfil relevant ethical and professional standards. Moreover, due prominence to fact-checks on online platforms' services would likely help increase their impact on audiences and, therefore, enhance citizens' awareness.

(f) Empowering the research community (Pillar E): Access to platforms' data for researchers

The Code aims at enabling privacy-compliant access to data for fact-checking and research activities. Both the ERGA report and the VVA Study signal that this goal has not been achieved. Despite encouraging willingness of platforms during the COVID-19 crisis to enter into licencing agreements with academic institutions, it is a shared opinion amongst European researchers that the provision of data and search tools required to detect and analyse disinformation cases is still episodic and arbitrary, and does not respond to the full range of research needs.

This issue concerns first the quality of the datasets and APIs that should be made available to the research community at large in order to acquire a better understanding of sources, vectors, methods and propagation patterns of false narratives having the potential to affect democratic debates and processes in the EU. Second, it also concerns the collaborative models developed so far by certain platforms with the academic community, which are based on discretionary, multi-bilateral arrangements, rather than on open and non-discriminatory approaches empowering a larger, multi-disciplinary community of researchers to carry out the appropriate detection and analysis activities.

The European Digital Media Observatory (EDMO) established in June 2020 constitutes the first initiative to overcome these shortcomings.[32] EDMO, with the support of the scientific community, is expected to contribute to defining the necessary requirements to enable access to anonymised or aggregated datasets and APIs for research purposes. EDMO will ensure that access to platform users'

[32] https://ec.europa.eu/digital-single-market/en/european-digital-media-observatory

data is granted in compliance with the General Data Protection Regulation (GDPR)[33] to avoid the identification of users and limit the purposes for which datasets may be used. In line with the GDPR, if, in exceptional circumstances, the processing of personal data is unavoidable, access to such data should only be made available pursuant to an appropriate legal basis for processing and safeguarded through appropriate technical and organisational security measures, including purpose limitation and data minimisation.

3.2.2 Lack of uniform definitions

As underlined in both the ERGA Report and the VVA Study, the Code would benefit from further scoping of certain key concepts and more precise definitions of certain operational terms. This would need to be well articulated with the existing regulatory framework and legal requirements applicable in the Member States.

The Code incorporates the definition of "disinformation" set out in the April 2018 Communication[34] and consistently employed in the Commission's subsequent statements of policy on the topic. However, the COVID-19 "infodemic"[35] has highlighted the need to further clarify additional concepts and differentiate more precisely between various forms of false or misleading content and manipulative behavior intended to amplify its dissemination online in order to enable the framing of appropriate responses by the platforms and other relevant stakeholders.

The June 2020 Communication discussed a number of examples that emerged in the context of the coronavirus infodemic: false claims and misleading healthcare information (such as 'it does not help to wash your hands' or 'the Coronavirus is only a danger to the elderly'); conspiracy theories that may harm social cohesion, lead to public violence or create social unrest (for example, conspiracies and myths about 5G installations spreading COVID-19 and leading to arson attacks on masts); disinformation blaming a particular ethnic or religious group for the spread of COVID-19 and eventually assuming the character of illegal hate speech; and targeted influence operations and disinformation campaigns by foreign actors seeking to improve their own image, undermine democratic debate, or exacerbate social polarisation in the EU.

The June 2020 Communication pointed to the need for a better scoping of the disinformation phenomenon through the articulation of certain adjacent concepts. These include[36]:

- "misinformation," understood as the dissemination of false or misleading content without the intention to deceive or cause public harm, or make economic gain; and
- "influence operations," understood as information campaigns by third-country actors that employ false or misleading information in combination with manipulative online techniques to

[33] Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, 4.5.2016.

[34] "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. Public harm comprises threats to democratic political and policy-making processes as well as public goods such as the protection of EU citizens' health, the environment or security. Disinformation does not include reporting errors, satire and parody, or clearly identified partisan news and commentary." April 2018 Communication, at pp.3-4.

[35] As defined and used by the WHO, "infodemics are an excessive amount of information about a problem, which makes it difficult to identify a solution. They can spread misinformation, disinformation and rumours during a health emergency. Infodemics can hamper an effective public health response and create confusion and distrust among people." https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b 4

[36] See Section 1 of the June 2020 Communication, p.4

interfere in EU or Member State electoral or policy-making processes, where the intention to deceive the public may be presumed.

There is also a range of operational terms that could benefit from definitions or common understandings that are more precise. These include:

- "Scrutiny of ad placements": the term is employed, in the context of commitments under pillar A, to encompass both the review of advertisements for display on a platform's own services and the scrutiny of third-party sites where advertising is placed by ad networks;
- "Political advertising": loosely defined under pillar B of the Code[37], the term has been delineated in varying ways by the platforms themselves and is defined differently at Member State level, or is not defined;
- "Issue-based advertising": while undefined in the Code, the platform signatories committed under pillar B to develop a definition and devise approaches toward public disclosure of such advertising, as explained in section 3.2.1; only Facebook did so;
- "Inauthentic behaviour": in the context of pillar C, the term may be used interchangeably with "inauthentic coordinated behaviour" or "online manipulation" and concerns user conduct aimed at artificially amplifying the reach or perceived public support for particular content. The conduct concerned includes the use of "fake accounts," "fake engagement" and "malicious bots," terms which would benefit themselves benefit from uniform definition and application; and
- "Indicators of trustworthiness" for news sources: in order to empower consumers, platforms commit under pillar D to develop such indicators in collaboration with other stakeholders in the news ecosystem and to prioritise trustworthy information in automated content ranking systems.

The lack of common understandings of the scope of fundamental concepts and of uniform definitions of key operational terms inhibits the effective implementation of measures by the signatories and impedes the monitoring, evaluation and comparison of the Code's implementation and effectiveness across platforms and Member States. It may also inhibit further take up of the Code insofar as potential signatories may be uncertain about the scope of commitments they would be undertaking by signing up to the Code.

3.2.3 Areas not covered so far by the Code

The experience gained through the monitoring of the Code shows that the scope of its commitments may be too narrow. The following issues have emerged as problematic features of the online ecosystem that remain unaddressed, or as clear-cut omissions in the Code affecting the proper oversight and evaluation of platforms' policies.

(a) Manipulative online behaviour falling outside the scope of the Code

Under Commitment 5, relevant signatories "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". Due

[37] The Code commits relevant signatories to "enable public disclosure of political advertising (defined as advertisements advocating for or against the election of a candidate or passage of referenda in national and European election elections)..." Code, Commitment No. 3.

to the narrow scope of this provision, the platforms' reporting has focused predominantly on actions taken against coordinated inauthentic behaviour, fake accounts and misuse of bots, a finding highlighted in the VVA Study[38]. The platforms' reporting addresses fake engagement techniques aimed at inflating the perceived popularity (or unpopularity) of certain content (such as fake or purchased followers, likes, dislikes, views) only very broadly, if at all.

However, recent studies[39] suggest that disinformation campaigns often leverage a wider array of manipulative techniques, which obviously include the use of fake accounts, account takeovers and bot-driven amplification, but also encompass hack-and-leak operations, impersonation, the creation of fictitious groups or fake parody accounts, the theft of existing artefacts, deep fakes, the purchase of fake engagements and the involvement of influencers. The vagueness of the Code's commitments in this respect creates serious risks of incomplete action against disinformation.

(b) Micro-targeting of political advertising

Identifying and responding to the policy concerns and preferences of the public is the essential purpose of democratic electoral processes, and the targeting of different voter groups with political messages crafted to appeal to them is a longstanding practice of political campaigns. However, micro-targeting of online political advertising has revolutionised political campaigns by enabling a narrow segmentation of voters based on collated personal data and sophisticated psychological profiling techniques, as well as the use of differentiated and customised political messages, disseminated with great speed, precision and reach[40].

Micro-targeting raises concerns, given the lack of public transparency of political content, which may blur the political lines of electoral programmes, contribute to increased polarisation of the public, encourage voter suppression (for example through negative messaging focused on decreasing turnout for rival campaigns), or the risk that mainstream voters are marginalized from the flow of information in political campaigns;[41]. Micro-targeting prevents adequate scrutiny of political content by journalists and fact checkers and timely response by political actors potentially impacted by the content.

The micro-targeting options offered to political advertisers are established by the platforms' respective advertising policies, which vary among platforms and have been evolving. Microsoft, for example, prohibits political advertising globally, including on LinkedIn[42]. After the Code's first annual reporting period, Twitter instituted a global ban on political advertising[43] and Google announced that it would limit targeting options for election ads on its services to the basic demographic criteria of age, gender, and general location.[44] Facebook, by contrast, has opted against limiting micro-targeting

[38] VVA Study, at p. 49.

[39] C.R. Walker and S-J. Terp, *Misinfosec: Applying Information Security Paradigms to Misinformation Campaigns* at https://drive.google.com/file/d/1QQzKNzC3b0cEnr6f8aM7kG_n_2j2SN06/view; J.F Gray and S-J. Terp, *Misinformation: We're Four Steps Behind Its Creators,* at https://cyber.harvard.edu/

[40] Personal data used in micro-targeting comprises voter profile data derived from official records; voting data collected by political partiers via "offline" canvassing and online interactions; and consumer data amassed for commercial purposes through online sales and marketing campaigns.

[41] See, e.g., IDEA, *Digital Microtargeting: Political Party Innovation Primer* (19 June 2018), at: https://www.idea.int/publications/catalogue/digital-microtargeting .

[42] https://www.linkedin.com/legal/ads-policy#4

[43] https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content.html; https://twitter.com/jack/status/1189634360472829952

[44] https://blog.google/technology/ads/update-our-political-ads-policy/

options on the grounds that such restrictions, if any, should be industry-wide and set by regulation rather than by the platforms' commercial policies.[45]

The use of personal data for micro-targeting is subject to, *inter alia¸* the GDPR. Moreover, as part of the September 2018 electoral package, the Commission published specific guidance[46] on the application of protection law in the electoral context, including EU data protection principles and compliance requirements applicable to micro-targeting. The European Cooperation Network on elections is currently investigating the issue in depth. This work will inform the European Democracy Action Plan, which will look into the issue of micro-targeting in political campaigns to ensure greater transparency on paid political advertising. The Code presently does not prohibit micro-targeting or restrict the range of targeting criteria that platforms may offer with respect to paid-for political content[47], although one of the objectives set out for the Code in the April 2018 Communication was "restricting targeting options for political advertising."[48] Recent research shows that the vast majority of the public are opposed to the micro-targeting concerning certain content (including political advertising) or based on certain sensitive attributes (including political affiliation). Further reflections in this area will be pursued without prejudice to any future policy on micro-targeting of commercial ads.

(c) Fairness in online political advertising

Political advertising during electoral periods is heavily regulated in the EU, for both the broadcasting and the press media sectors. Most Member States have rules aimed at maintaining a level playing field for parties and candidates and ensuring that elections are not dominated by a narrow range of interests. There are two basic approaches to achieve this goal: (i) the imposition of spending limits for advertising on traditional media; and (ii) the allocation of free or subsidised political advertising space or time on traditional media, according to principles intended to afford fair access to all contenders.[49]

However, social media are largely not covered by national measures applicable to broadcast and press media, and there are no rules at EU level establishing spending limits for political advertising or addressing fair access to media for political parties or candidates participating in the elections to the European Parliament. As campaigns are increasingly turning to online advertising, using social media and personal data to target voters[50]– a trend that is particularly pronounced outside of continental

[45] https://about.fb.com/news/2020/01/political-ads/

[46] Commission guidance on the application of Union data protection law in the electoral Context, 19-20 September 2018, (COM(2018) 638 final). The guidance provides clarity to actors involved in election processes - national electoral authorities, political parties, data brokers and data analytics companies, social media platforms and online advertising networks – and calls on the national data protection authorities to make full use of their strengthened powers under the GDPR to monitor the situation and address possible data protection breaches.

[47] Rather, the Code's commitments require signatories (under pillar B) to provide for the transparency and public disclosure of political and issue-based advertising and (pursuant to Commitment No. 11) to provide tools that help consumers understand why they have targeted to receive particular advertising.

[48] April 2018 Communication, at p. 7.

[49] The principles for providing access and allocations of airtime or slots include equal access (candidates or parties are allocated the same amount of airtime, irrespective of popular support); proportionality (candidates or parties are allocated time according to objective criteria such as previous election results or seats held in the legislature); and mixed access (a minimum amount of time is allocated to all contenders, and supplementary time is provided on a proportional or other basis).

[50] By one account, political parties and organisations across Europe spent at least €100 million to advertise on Facebook and Google for their election campaigns in 2019. https://worldacquire.com/2019/12/30/which-political-parties-in-europe-spent-the-most-on-online-political-advertising-in-2019/

Europe[51] – the shift of spending from offline to online media has the potential to undermine the effectiveness of existing rules ensuring the fairness of the democratic processes.

At present, the main online platforms have adopted different approaches to online political advertising. Facebook, Instagram, Google and Snapchat allow all sponsored political communications (i.e. political and issue-based ads) on their services, but under different definitions and conditions. Twitter has banned political advertising but continues to serve issue-based ads. Other platforms, such as LinkedIn, Pinterest and Reddit, do not allow political advertising and it is unclear to what extent they serve issue-based ads.

This raises concerns about the fair representation of political parties and candidates on social media platforms[52] and how, for example, safeguards applying to the broadcasting sector under national laws should be translated into the digital environment.[53] The issue of online application of laws relevant in the electoral context and their modernisation is addressed in the work of the European Cooperation Network on Elections. The European Democracy Action Plan will look into solutions to ensure greater transparency of paid political advertising as well as clearer rules on the financing of European political parties.

(d) Key Performance Indicators and data for monitoring and oversight of the Code

Under the Code, the signatories committed to present regular reports on the effectiveness of their policies. The Code signatories considered that the loose notion of "Key Performance Indictors" (KPIs) employed in the Code was necessary to accommodate differences in their respective business models. However, in practice the approach taken has revealed substantial deficiencies, owing in part to the lack of a common reporting structure and a consistent understanding of monitoring and evaluation needs.

ERGA criticised in particular the platforms' tendency to provide data aggregated for the whole EU, which made it difficult to evaluate the impact of the Code across the EU Member States. Moreover, as pointed out in the VVA Study, the Code fails to set out clear and measurable KPIs enabling cross-platform comparisons and objective measurements.

In line with the principles of better regulation[54], a set of KPIs is required to allow for proper monitoring of the implementation of the policies put in place by the signatories, as well as for assessing the Code's overall impact on disinformation.

To this end, two classes of KPIs are pertinent. The first class refers to **service-level indicators** capable of measuring the results of the policies implemented by the signatories, while accounting for the specificities of the different digital services they provide. Ideally, service-level indicators include benchmarks and targets and make use of samples of identified disinformation websites to verify the

[51] In the 2017 national elections in the UK, online ad spending made up 43% of total spending by political campaigners, rising from less than 2% in 2014. During the 2016 US presidential campaign, the Trump campaign spent more than 40% of its ad budget on online ads, and some estimates forecast that federal candidates in the 2020 US election will spend some $1.2 billion to run ads on Facebook, Google and other platforms. https://qz.com/1773107/the-rise-of-us-and-uk-online-political-ads-in-three-charts/

[52] The Commission has informed online platforms and other Code signatories that it is reflecting on possible ways to adapt existing rules to ensure fair representation of political parties on social media platforms. The issue has also been referred to by ERGA in its discussions with signatories of the Code.

[53] The issues of online application of laws relevant in the electoral context and their modernisation are addressed in the work of the European Cooperation Network on Elections.

[54] https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en

evolution over time of user engagement with such sites, and variations in advertising revenues flowing to them. Such indicators are also useful to monitor the level of platforms' responsiveness to users' complaints and to flagging by independent fact-checking organisations (e.g. evolutions in the number of fact-checks referenced on their services and levels of user engagement with such fact-checks), as well as platforms' ability to detect and disclose cases of manipulative behaviour affecting the integrity of their services.

The second class refers to **structural indicators,** which measure the prevalence of disinformation at a general level and illuminate the extent to which the Code has had impacts on the phenomenon. As suggested by VVA, a meaningful way to assess the effectiveness of the Code's policies would be to apply structural indicators to representative national samples of users in order to determine the sources from which users access news and to assess the ratio of authoritative sources of such news as compared to sources that are purveyors of disinformation.

A proposal for concretely structuring such indicators has been elaborated in the VVA Study, a summary of which can be found in Appendix.

The operationalisation of KPIs for an independent monitoring and oversight of the Code requires the availability of robust, raw data from platforms. This would require setting up a methodology defining the basic data requirements and access rights as well as a procedure for platforms to deliver such data in a standardised manner, on a timely basis and in compliance with the General Data Protection Regulation (GDPR).

3.2.4 Limitations inherent to the self-regulatory nature of the Code

(a) Limited participation

A self-regulatory instrument, the Code only applies to online platforms and advertising sector stakeholders that agree to subscribe to its commitments and does not replace the need to comply with existing international and national legal requirements. Further, signatories select, based on their own assessments, which of the Code's commitments they will adhere to and identify the policies and actions they will carry out to implement those commitments.

The Code's flexibility may encourage relevant stakeholders to take up the Code and implement policies that address particular aspects of disinformation which, in the words of the Code, "correspond to the product and/or service they offer, their role in the value chain, their technical capabilities and their liability regimes as provided under EU Law." [55] Altogether, the Code's platform signatories – in particular, Facebook (including Instagram[56]), Google, and Twitter as well as Microsoft (which subscribed to the Code in May 2019) and Tik Tok (which subscribed to the Code in June 2020) are the main online platforms active in the EU. Moreover, the platform signatories, with the exception of Twitter,[57] have elected to adhere to all of the Code's commitments.

[55] Code, at pp. 1-2.

[56] As set out in Facebook's Annual Self-Assessment Report, Instagram applies certain policies implementing Facebook's commitments under the Code.

[57] In its subscription document, Twitter did not undertake to comply with two of the Code's commitments: Commitment No.7, which requires investment in means to help users make informed decisions about news that may be false and to efforts to support the

Nevertheless, as ERGA has pointed out, the voluntary nature of the Code establishes an inherent "regulatory asymmetry" as between Code signatories and non-signatories.[58] Code signatories incur compliance costs and are obligated to take measures to protect the integrity of their services that may encourage malicious actors to migrate towards platforms that are not Code signatories. This possibility materialised during the COVID-19 pandemic, which witnessed the use of certain non-signatory platforms to disseminate disinformation around the crisis.[59]

More generally, as ERGA notes, instant messaging platforms (such as WhatsApp and Messenger, which are not Code signatories) are efficient means to spread disinformation since they enable the sharing of content within closed groups with large numbers of users as well as the transfer of content from one closed group to another.[60]

Moreover, the take-up of the Code by the online advertising sector has been very limited. While a number of national-level trade associations subscribed to the Code during its first year of operations, there has been a failure so far of any brands or other corporate actors from the advertising sector operating in the EU to join the Code. In June 2019, the Global Alliance for Responsible Media was launched under the auspices of the WFA with the goal of "work[ing] collaboratively to identify actions that will better protect consumers online, working towards a media environment where hate speech, bullying and disinformation is challenged."[61] As the Alliance features notable stakeholders from across the advertising ecosystem, this initiative should be further pursued, with a view to developing appropriate transparency measures for advertisers and tools that enable them to manage their preferences and protect their brands against reputational harms.

(b) Oversight, monitoring and enforcement

The existing self-regulatory framework does not establish an independent oversight mechanism for monitoring the completeness and impact of the signatories' actions in tackling disinformation. Rather, the Code provides for a system of peer review whereby signatories commit to submit self-assessment reports and to meet regularly to analyse the Code's implementation and functioning, and, if they choose, to propose follow-up actions. However, it can support the exercise of oversight function by competent authorities in line with applicable regulatory frameworks.

The Code provided for the review and evaluation of the signatories' self-assessment reports by an objective third-party organisation (TPO) selected by the signatories. Ultimately, however, the signatories were unable to engage an appropriate TPO and thus give effect to even this modest degree of outside review. Further, while the Code commits signatories to cooperate with the Commission in evaluating their reports, it provides no mechanisms institutionalising cooperation with the competent Member State authorities.

Similarly, the Code does not establish mechanisms for ascertaining the signatories' compliance with Code commitments and attributing consequences in case of breach. Instead, the sole sanction for non-

development and implementation of trustworthiness in collaboration with the news ecosystem; and Commitment No. 8, which requires investments in means to prioritise relevant, authentic and authoritative information in automatically ranked distribution systems.

[58] ERGA Report, at pp. 44 and 52.

[59] In earlier months of the crisis, Tik Tok emerged as vector for disinformation on the pandemic, with a surge of videos elaborating conspiracy theories linking the spread of COVID-19 to 5G deployment, Bill Gates and the WHO. https://www.buzzfeednews.com/article/laurenstrapagiel/pandemic-conspiracy-theorists-disinformation-tiktok

[60] ERGA Report, at p. 44, note 56.

[61] https://wfanet.org/knowledge/item/2019/06/18/Global-Alliance-for-Responsible-Media-launches-to-address-digital-safety

compliance is expulsion. The effectiveness of this remedy relies on the aversion of signatories to incur the reputational damage of expulsion, as well as on the collective goodwill and readiness of the signatories to repudiate a peer.

(c) Protection of fundamental rights and mechanisms for redress

The platform signatories have emerged as gatekeepers in the online information ecosystem, as users increasingly turn to their services to access, impart and share information, and to engage in public affairs. While the Code acknowledges the importance of upholding fundamental rights, in particular the freedom of expression it does not set out procedures to ensure in practice the protection of these rights in the pursuit of actions addressing disinformation. Also, the Code does not provide for regular reporting on these matters to ensure transparency and accountability.

Notably, there is no requirement for complaint procedures or other remedies to prevent or redress the erroneous treatment of content (e.g. demotion) or unwarranted actions against users (e.g. suspension of accounts) which platforms consider to be in violation of their policies on disinformation. It should be noted that the need to ensure the protection of fundamental rights through adequate complaint procedures and redress mechanisms arises in various contexts of information society services provision, not only in relation to disinformation.

4. Conclusions

The Code of Practice on Disinformation is a unique and innovative tool in the fight against online disinformation. The Code has been acknowledged worldwide, including by some international partners (e.g. Australia, Canada), as a good example of structured cooperation with online platforms to ensure greater transparency and accountability, as well as a useful framework to monitor and improve platforms' policies on disinformation. The Code has also prompted concrete actions and policy changes by the platforms aimed at countering disinformation.

However, this overall assessment highlights that, in order to ensure a complete and consistent application across platforms and Member States, the Code should be further improved in several areas by providing commonly-shared definitions, clearer procedures, more precise commitments as well as transparent key performance indicators and appropriate monitoring, all taking into account applicable regulatory frameworks. Further efforts should also be made to broaden the participation to other relevant stakeholders, in particular from the advertising sector.

Moreover, a more structured model for cooperation between platforms and the research community should be developed. At present, it remains difficult to precisely assess the timeliness, comprehensiveness and impact of platforms' actions, as the Commission and public authorities are still very much reliant on the willingness of platforms to share information and data. The lack of access to data allowing for an independent evaluation of emerging trends and threats posed by online disinformation, as well as the absence of meaningful KPIs to assess the effectiveness of platforms' policies to counter the phenomenon, is a fundamental shortcoming of the current Code.

A structured monitoring programme may constitute a pragmatic way to mobilise the platforms and secure their accountability. The programme for monitoring disinformation around COVID-19 foreseen in the June 2020 Communication will be an opportunity to verify the adequacy of such an approach and prepare the ground for further reflection on the best way forward in the fight to disinformation.

The information and findings set out in this assessment should support the Commission's reflections on pertinent policy initiatives, including the European Democracy Action, as well as the Digital Services Act, which will aim to fix overarching rules applicable to all information society services.

APPENDIX TO STAFF WORKING DOCUMENT

Summary of the reports underpinning the assessment of the Code of Practice

The following sections present a brief overview of (i) the annual self-assessment reports from the signatories of the Code of Practice on Disinformation (the Code); (ii) a monitoring Report by European Regulators Group for Audiovisual Media Servicew (the ERGA Report); (iii) the Study commissioned from the independent consultancy, Valdani Vicari & Associati (the VVA Study); and (iv) the Commission's 2019 Report on the elections to the European Parliament (the Elections Report). It should be noted that, in breach of the Code, the signatories failed to provide a report from a third-party organisation evaluating progress in achieving the Code's commitments.

1. Annual Self-Assessment Reports of the Signatories

In October 2019, the Commission received annual self-assessment reports from the signatories detailing policies, processes and actions undertaken to implement their respective commitments during the Code's first year. The Commission published the signatories' self-assessment reports, along with a summary analysis highlighting particular areas of progress and identifying gaps and shortcomings.[62]

Briefly, the self-assessment reports indicate comprehensive efforts by the signatories to implement their commitments during the Code's first year. Reported measures vary in terms of speed and scope across the five pillars of the Code. In general, measures to empower consumers (pillar D) and to empower the research community (pillar E) lag behind those that were subject to the Commission's targeted intermediate monitoring phase. The latter measures concern the disruption of advertising and monetisation incentives for purveyors of disinformation (pillar A), the transparency of political and issue-based advertising (pillar B), and measures to ensure the integrity of services against inauthentic accounts and behaviours (pillar C). There are also differences in the scope of measures taken by each platform to ensure the implementation of their commitments under the five pillars. Finally, there are differences across Member States as regards the deployment of the respective policies by the platforms.

The reports also indicate some intensification of joint efforts between platforms and other stakeholders, including fact-checkers, researchers, civil society and national authorities. These have aimed at improving the resilience of platforms' services against various forms of meddling and media manipulation and at diluting the distribution of disinformation.

Finally, the reports indicate that trade associations signatories have raised awareness over the past year and advocated in favour of take-up of the Code among their members. However, apart from Microsoft and TikTok, no additional platform stakeholders have subscribed to the Code, signalled their intention to adhere to its principles and commitments or otherwise engaged in dialogue to provide information on how they intend to resolve issues of disinformation through their services. It is also notable that brands or other corporate actors from the advertising ecosystem operating in the EU have so far failed to join the Code.

In the Commission's summary analysis, it was observed that the Code, as a self-regulatory standard, has provided an opportunity for greater transparency into the platforms' policies on disinformation, as well as a framework for structured dialogue to monitor, improve and effectively implement those

[62] https://ec.europa.eu/digital-single-market/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019

policies. However, the analysis of the reports also shows that the provision of data and search tools to the research community is still episodic and arbitrary and does not respond to the full range of research needs. Moreover, cooperation with fact-checkers across the EU is sporadic and a genuine coverage of all Member States and EU languages is still not in sight, showing the need for further efforts towards a mechanism allowing truly independent organisations to cooperate with the platforms (including via relevant and privacy-compliant access to datasets for research purposes).

2. ERGA Monitoring Report

As called for in the Action Plan against Disinformation (the Action Plan), the European Regulators Group for Audio-visual Media Services (ERGA) has assisted the Commission in monitoring the implementation of the Code at Member State level and in assessing the Code's overall effectiveness. These activities were carried out in two phases by a special ERGA Task Force. The first phase focused on the platforms' implementation of commitments concerning political and issue-based advertising during the run-up to the May 2019 European elections (pillar B).[63] The second phase focused on the implementation of commitments relating to empowering consumers and the research community (pillars D and E) over the course of the Code's first year of operations, with some Member States carrying out additional monitoring on pillars A and C. [64]

Broadly, ERGA's main conclusions are that the Code is "a unique and innovative tool[65]" in the fight against online disinformation, and should be regarded "as a substantial step" in building a structured cooperation between platforms and EU and national audio-visual regulators.

However, ERGA emphasises significant weaknesses that need to be addressed:

- First, the Code lacks uniform definitions and sufficiently precise commitments, which makes it difficult to monitor and quantify the effectiveness of the relevant implementing measures, or draw comparisons across platforms and Member States;
- Second, the Code's commitments have not been implemented consistently across the Code's five pillars, and in some cases not at all, or not in all Member States;
- Third, the Code does not provide for a transparent oversight mechanism as the signatories' self-assessment reports disclose only aggregate EU-level data, which limits the possibilities for a truly independent and objective verification; and
- Fourth, the Code does not apply to all platforms active in the EU, but only to its signatories, who retain discretion as to the scope of their commitments. Moreover, there is no coercive mechanism to ensure effective compliance.

To overcome these shortcomings, ERGA sets out a number of recommendations, including:

- The development of common definitions of relevant terms (e.g., "political advertising," "issue-based advertising," "fake news," "manipulation") to ensure a consistent approach across the EU;

[63] Thirteen Member States participated in the first-phase monitoring: Belgium, Croatia, Cyprus, France, Hungary, Italy, Ireland, Latvia, Luxembourg, Poland, Slovakia, Sweden and Spain. The results of the monitoring are summarised in *Report of activities carried out to assist the European Commission in the intermediate monitoring of the Code of Practice on disinformation,* published in June 2019: https://erga-online.eu/wp-content/uploads/2019/06/ERGA-2019-06_Report-intermediate-monitoring-Code-of-Practice-on-disinformation.pdf

[64] For the second phase, 13 Member States participated in the monitoring of Pillars D and E. In addition, Italy monitored Pillars A and C and Germany monitored Pillars A, B and C. Pillar B was also monitored by the NRAs in three countries where elections were foreseen: Hungary, Poland and the United Kingdom. The results of the second-phase monitoring are summarised in *ERGA Report on Disinformation: Assessment of the Implementation of the Code of Practice,* published in May 2020: https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf

[65] ERGA report, p. 3, http://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf

- Harmonised mechanisms for the implementation of the Code's commitments, including, for example, common labels for political ads, common registration procedures for sponsors of political ads, and common procedures for issuing warnings about fact-checked information;
- The provision by the platforms, at specified intervals, of datasets, monitoring tools and country-specific information enabling national authorities to carry out their own specific queries and analyses and conduct regular and proper monitoring;
- The drafting of guidelines aimed at improving, rationalising or harmonising the platforms' (i) relationships with fact-checkers, (ii) reactions to consumer complaints and flagging; (iii) involvement in media literacy campaigns; and (iv) relationships with researchers, including potential opportunities for cooperation through EDMO; and
- Further efforts to increase the number of platform signatories to the Code.

Finally, ERGA's overarching recommendation is that a "shift from the current flexible self-regulatory approach to a more co-regulatory one" is required to improve the Code and address the shortcomings identified. This could be achieved through "the introduction of a formal backstop mechanism to deliver the required monitoring and enforcement elements" and provide for "clear reporting obligations, more harmonised procedures and appropriate timeframes."

ERGA further notes that co-regulatory backstops exist in other fields at Member State level, which are grounded in EU and national law and overseen by an independent regulatory authority. These examples might provide inspiration for an EU-wide co-regulatory system ensuring platforms' transparency and accountability. In particular, ERGA recommends the Commission to pursue "a holistic approach" to online content regulation, which might culminate in a Digital Services Act setting out overarching principles for information society services, along with a dedicated legal instrument addressing disinformation in a more flexible and granular way.

3. VVA Study on Implementation of the Code

To support its evaluation, the Commission commissioned a study by the independent consultancy Valdani, Vicari and Associates.[66] The VVA Study collected evidence based on structured interviews with key stakeholder groups (Code signatories, non-signatory platforms, national audio-visual regulatory authorities, academia, civil society organisations) identifying points of consensus and areas of disagreement with respect to the Code's:

- **effectiveness**, with each of the Code's five pillars reviewed individually;
- **efficiency**, in terms of benefits achieved versus administrative burdens;
- **relevance** in spurring stakeholder action to address disinformation;
- **coherence** with other EU interventions in adjacent areas[67];
- **EU added value**, in relation to initiatives taken at Member State level; and
- **sustainability** as regards the longevity of outcomes produced.

In addition, the VVA Study provides contextual background on the evolution of the disinformation phenomenon and the challenges posed in the online environment, as well as a summary of legislative

[66] Study for the Assessment of the Code of Practice against Disinformation (SMART 2019/0041): https://ec.europa.eu/digital-single-market/en/news/study-assessment-implementation-code-practice-disinformation. The Study is based on data gathered through, *inter alia,* structured interviews and other consultations with the Code's platform and trade association signatories, non-signatory platforms, and a cademic and other experts on disinformation; online surveys disseminated to the national regulatory authorities active in ERGA and to experts and professionals on the topic of disinformation; academic literature and specialised journalism addressing the topic; and a review of the Code signatories' self-assessment reports.

[67] The other EU initiatives addressed are the E-Commerce Directive, the Audio-visual Media Services Directive, the Copyright Directive, the General Data Protection Regulation, the Directive on the Security of Network and Information Services, and the Code of Conduct on Illegal Hate Speech.

and non-legislative initiatives at Member State level addressing disinformation and adjacent concerns. It also develops four case studies on key topics[68], including proposals for possible KPIs.

The VVA Study's overall conclusions is that the Code should not be abandoned as it has established a common framework to tackle disinformation. Its aims and commitment areas are highly relevant, while its implementation has produced overall positive results. It constitutes a first and crucial step to curb online disinformation and shows European leadership on an issue that is international in nature.

However, the Code has also substantial shortcomings relating to its self-regulatory nature, notably the lack of uniform implementation standards and insufficient clarity around its scope and some of its key concepts.

In essence, the VVA Study recommends that the implementation of the Code should continue, while its effectiveness could be strengthened through the introduction of clearer definitions as well as a mechanism for action in case of non-compliance. It considers that the Code should be complemented by a co-regulatory backstop providing for appropriate enforcement mechanisms, sanctions and redress mechanisms.

The VVA Study proposes two sets of KPIs. The first set would measure disinformation at a general (structural) level, which would measure disinformation and the platforms as a group. The second set would cover the five Pillars and measure the individual performance of each of the platform signatories.

Structural indicators can be tested against any platform service (i.e. these indicators should be generic enough to compare the performance of all types of platforms) and/or website with disinformation content. At the same time, these can also show the impact of the Code at the general level. This would aid the monitoring of whether disinformation is gaining in influence, staying stable or declining. These KPIs should measure the prevalence of disinformation online. This means, for example, monitoring the sources from which users currently access news and the ratio of such sources being authoritative or purveyors of information. In order to monitor the prevalence of disinformation (e.g. how many purveyors of disinformation there are) within the EU, representative samples of the population would be set up in each Member State. Participants would agree to have their online interaction with news and disinformation content monitored in line with GDPR requirements. The VVA Study suggests the following structural indicators:

Proposed KPIs	Description
Prevalence of authoritative and disinformation sources in accessing news	Number of authoritative and disinformation sources through which the sample groups access news and in what language
Direct versus indirect access to authoritative and disinformation source	Whether the sample accesses authoritative/disinformation content directly (e.g. via browser) or indirectly through social media and in what language
Engagement with websites that are not mainstream news outlets and not purveyors of disinformation (i.e. not in the top 100 websites identified as such)	Within the sample, how many people engage (access, share, comment, etc.) with news published on non-authoritative websites that cannot be considered purveyors of disinformation and in what language
Level of engagement with top 100[69] websites identified as purveyors of disinformation	Within the sample, how many people engage with the top 100 websites identified as purveyors of disinformation by fact-checkers in their native and other languages

[68] The topics addressed are: (i) issue-based advertising; (ii) Key Performance Indicators; (iii) partnerships between platforms with researchers and data requests; and (iv) partnerships between platforms and fact-checkers.

[69] This number would be proportionate to the population of each Member State, i.e. smaller Member States could consider the top 30 to 50 websites.

Service-level indicators should measure the progress of each platform in countering disinformation. The aim would be to find out not just the quantity of disinformation that is present on the platforms, but to focus on its actual impact by better understanding how disinformation propagates through the society, how it affects users. The service-level KPIs could also make use of the sample groups created for the structural level KPIs to collect some control data. The study suggests the following list of service-level indicators:

Code Pillar	Proposed KPIs	Description
General level	Direct investment spent on identifying/deactivating disinformation content as percentage of turnover	Quantify investments of platforms in fighting disinformation
	Direct investment spent on identifying/deactivating disinformation content per user as percentage of total turnover per user	Quantify investment of platforms in fighting disinformation in terms of human resources
	Number and types of tools available per MS	Qualitative assessment of whether all tools platforms developed in connection to combating disinformation are available in all MS
	Ratio of total donations made to NGOs, civil society, etc to support initiatives combating disinformation (e.g. media literacy campaigns) against total turnover	Quantify investment of platforms into independent initiatives to combat disinformation
	These indicators aim to provide an overarching picture into the importance of combating disinformation. As the platforms have different business models, they should be grouped based on the services provided when comparing the results.	
Pilar I: Scrutiny of ad placements	Total turnover of advertising operators from advertisement placement	Revenue created during a certain time period
	Total of foregone (lost) revenue due to certain accounts being closed due to being purveyors of disinformation	Estimate of revenue lost due to closure of certain accounts due to their link with disinformation
	Total advertisement revenue from top 100 websites identified as purveyors of disinformation	The amount of money that flows to the most prominent purveyors of disinformation due to advertisement placement
	The data for these indicators could be collected, for example, on a semester basis comparing the evolution. These indicators would show the development with regard to demonetisation of purveyors of disinformation, in monetary terms but also in number of ad operators/accounts and how the revenue developed over time in comparison with the overall market.	
Pillar II: Political and issue-based advertising	Number of mislabelled political and issue-based advertising	The prevalence of false positives and negatives in the ad libraries
	Ratio of total turnover of issue-based advertising with revenue lost due to accounts closed down due to breach of issue-based advertising policies	The prominence of issue-based labelling and its impact on revenues
	Ratio of number of labelled political advertising against number of political advertising that lost its labelling due to further engagement (e.g. sharing) by platform users per genuine and inauthentic users	To estimate the redistribution of political advertising due to further engagement by genuine and inauthentic users. This is to first look at how many advertisements lose their labelling due to further engagement but also what is the impact on the engagement with the advertising once it lost its labelling. In case a platform does not have a working definition of an inauthentic user, a common set of principles could be set up[70]
	Ratio of engagement with labelled political advertising against engagement with political advertising that lost its labelling due to further engagement (e.g. sharing) by platform users per genuine and inauthentic users	
	These KPIs are based on the assumption that an auditing system will be put in place to ensure that all platforms are to provide the dedicated independent body with a dataset for all paid for content to spot false negatives/positives. These indicators would show development regarding completeness of the ad libraries and political and issue-based policies particularly regarding engagement with labelled and mislabelled content.	
Pillar III: Integrity of services	Ratio (estimate) of inauthentic accounts/users that remained alive/active after creation against all active accounts	What proportion of all accounts are considered to be fake/inauthentic that are not caught by the platforms policies before their creation
	Ratio of all engagement (e.g. posts, likes, comments, shares) inauthentic accounts/users have had with genuine users before being detected and deactivated due to a breach of platform policies against all engagement inauthentic accounts/users have had with other inauthentic accounts/users before being detected and deactivated due to a breach of platform policies	Measure of the actual impact of inauthentic accounts/ users. Can be used to monitor the role of inauthentic accounts/ users in promoting disinformation
	Ratio of directly contracted employees (or FTE) tasked with identifying/deactivating disinformation content as percentage of total number (or FTE) of staff	Quantify investments of platforms in fighting disinformation
	Number of fake accounts/fake users deactivated following their report by a genuine user	Measure magnitude of the problem of fake accounts/fake users

[70] Using, for example, the method to identify suspicious behaviour elaborated in T. Davis, , S. Livingston, M. Hindman. Suspicious Election Campaign Activity on Facebook: How a Large Network of Suspicious Accounts Promoted Alternative Für Deutschland in the 2019 EU Parliamentary Elections. Available at: https://smpa.gwu.edu/sites/g/files/zaxdzs2046/f/2019-07-22%20-%20Suspicious%20Election%20Campaign%20Activity%20White%20Paper%20-%20Print%20Version%20-%20IDDP.pdf

	These indicators are looking into the impact inauthentic accounts and users have on spreading disinformation across platforms. These indicators can also be further assessed/analysed in connection with structural indicators, particularly, the prevalence of top 100 websites identified as purveyors of disinformation.	
Pillar IV: Empowering consumers	Ratio of complaints submitted by users about disinformation content against number of complaints followed up	Effectiveness of platforms to follow up with their users
	Ratio of the number of pieces of content reviewed by fact-checkers against total number of pieces of new content created on a monthly basis	Approximation of the “safety” of content on platforms
	Ratio of number of users who have used tools designed to improve empowering consumers (e.g. tools that provide context to content, or that show why consumers are shown certain content) against all instances these tools were available	The effectiveness of tool designed to empower consumers
	Number of users that interacted with disinformation content produced by inauthentic accounts/users that were notified when such content was removed	How effective the platforms are on informing users that they have interacted with a disinformation content
	These indicators aim to assess the effectiveness of the tools the platforms provide to consumers in combating disinformation. As a proxy, these indicators should provide information on how reliable the algorithms put in place by platforms are to combat disinformation and to empower their consumers.	
Pillar V: Empowering the research community	Ratio of number of academic/research organisations that enter into relevant arrangements with platforms against number of data requests received	Effectiveness of cooperation of platforms and the research community
	Ratio of donations made to academic/research organisations for research projects against total investment into combating disinformation	Quantify investment of platforms into research supporting fight against disinformation
	With these indicators, the aim is to investigate how effective the cooperation with the research community is.	

4. The Commission 2019 Elections Report

On 19 June 2020, the Commission published its report on the conduct of the 2019 European elections[71], which sets out key lessons learned and potential follow-up measures, including under the European Democracy Action Plan and the 2020 EU Citizenship Report. It looks in particular at the participation in elections, assesses ways towards a transparent electoral process with a truly European dimension, summarises work done under the Electoral Package[72], and addresses disinformation and election manipulation.

Complementing the specific transparency commitments for online platforms under the Code of Practice, the Electoral Package included a Recommendation[73] that invited Member States to encourage the transparency of online political advertising by calling upon European and national political parties, foundations and campaign organisations to take measures that enable citizens to recognise paid-for communications online, the amount of money spent, its source and the means by which such communications are targeted to them. In this regard, the Elections Report finds that:

- Some Member States reported having transparency rules on paid political communications, though difficulties remain with access to data and capability to conduct monitoring and thus enforcement for online conduct.
- Some Member States reported that they are now considering legislation on transparency of online political advertising, to introduce or to modernise online transparency rules; and
- National and European political parties reported that they took steps to comply both with the applicable legislation and the requirements set by online platforms in their terms of service.
- However, they generally did not take additional transparency measures (e.g. listing their online political advertisements on party websites or providing information on online ad spend) as called for in the Recommendation.
- National and European political parties expressed their wish for clarification of the rules on transparency, the correct approach to be taken by platforms as regards EU-wide campaigning and the provision of open access online databases for parties to make their transparency disclosures.

As regards the authorisation procedures for online political advertising established by online platforms, some political parties found the procedures complicated, slow and insufficiently transparent, including problems of late approvals and editorial censorship.[74]

More generally, while certain online platforms engaged directly with some Member States, the Elections Report highlights the need for more equal treatment of different national electoral authorities and more operational interactions to support their oversight functions, including better access to data, identification contact points and possibilities for information exchanges in the languages of the Member States concerned.

[71] Commission Communication, Report on the 2019 elections to the European Parliament, at: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1123

[72] The Electoral Package consisted of a Recommendation on election cooperation networks, online transparency, protection against cybersecurity incidents and fighting disinformation campaigns; guidance on the application of the EU's General Data Protection Regulation; and a legislative amendment tightening the rules on European political party funding. https://ec.europa.eu/commission/presscorner/detail/en/IP_18_5681

[73] Commission Recommendation of 12 September 2018 on election cooperation networks, online transparency, protection against cybersecurity incidents and fighting disinformation campaigns in the context of elections to the European Parliament, C(2018) 5949 final, at : https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-cybersecurity-elections-recommendation-5949_en.pdf

[74] Several European parties cited policy changes introduced by Facebook in March 2019, which require advertisers to be separately authorised in each country in which political advertising was being run, as presenting an obstacle to paid-for political communications at a European level.

As follow-up actions, the Elections Report states that the Commission will *inter alia:*

- Continue to engage with the online platforms, including in the context of the Code of Practice on disinformation and the Code of Conduct on countering illegal hate speech online[75];
- Together with the relevant parties such as the European Cooperation Network on Elections, the Rapid Alert System and other relevant EU networks, develop further actions aimed at establishing a joint approach and common standards;
- Further examine issues linked to the effective application of data protection rules in elections, including transparency requirements and the use of micro-targeting techniques and algorithms;
- Launch specific actions aimed at further empowering citizens through initiatives to support media literacy and critical thinking;
- Further support media freedom, media pluralism and journalism, including through better protection of journalists and in the context of the transformation of the media sector and the digital future;
- Strengthen citizens' awareness and societal resilience, including through the establishment of the European Digital Media Observatory; and
- Examine the impact of new technologies and techniques on free and fair elections through a study.

[75] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct