STUDY FOR THE “Assessment of the implementation of the Code of Practice on Disinformation”

Final Report

Authors
Iva Plasilova
Jordan Hill
Malin Carlberg
Marion Goubet
Richard Procee
Assessment of the implementation of the Code of Practice on Disinformation

Directorate-General for Communications Networks, Content and Technology
Internal identification

Contract number: SMART 2019/0041

EUROPEAN COMMISSION
Directorate-General for Communications Networks, Content and Technology
Directorate I — Media Policy
Unit I.4 — Media Convergence and Social Media
Contact: CNECT-I4@ec.europa.eu

European Commission
B-1049 Brussels
STUDY FOR THE “Assessment of the implementation of the Code of Practice on Disinformation”

Final Report
Assessment of the implementation of the Code of Practice on Disinformation

Directorate-General for Communications Networks, Content and Technology
# TABLE OF CONTENTS

**EXECUTIVE SUMMARY** ...................................................................................... 3  
**CHAPTER 1: INTRODUCTION** ............................................................................. 7  
  Objectives and scope ................................................................................ 7  
**CHAPTER 2: METHODOLOGY** ............................................................................. 9  
  Research tools ....................................................................................... 10  
  Analysis ................................................................................................ 12  
  Limitations............................................................................................. 12  
**CHAPTER 3: THE CONCEPT OF DISINFORMATION** .............................................. 14  
  Emergence of disinformation and its definitions .......................................... 14  
  Current and future challenges .................................................................. 19  
**CHAPTER 4: THE CODE OF PRACTICE ON DISINFORMATION** ......................... 24  
  Context ................................................................................................. 24  
  Aim and Scope of the Code of Practice ...................................................... 27  
  Commitments of the Code of Practice........................................................ 29  
**CHAPTER 5: STUDY FINDINGS** ......................................................................... 32  
  Overall findings ...................................................................................... 32  
  Effectiveness.......................................................................................... 36  
  Efficiency ............................................................................................... 60  
  Relevance .............................................................................................. 63  
  Coherence ............................................................................................. 65  
  EU added value ...................................................................................... 69  
  Sustainability ......................................................................................... 71  
**CHAPTER 6: INITIATIVES AT MEMBER STATE LEVEL** ........................................... 75  
  Implementation of non-legislative measures to combat disinformation ........ 75  
  Implementation of combined non-legislative and legislative measures to combat disinformation ................................................................. 78  
  Other measures in some Member States ................................................... 80  
  Conclusions on Member States’ measures to combat disinformation .......... 81  
**CHAPTER 7: METRICS AND KEY PERFORMANCE INDICATORS** ....................... 89  
  Structural indicators ............................................................................... 91  
  Service-level indicators ........................................................................... 92  
  Data provision and assessment ................................................................ 95  
**CHAPTER 8: CONCLUSIONS** ............................................................................ 96  
  Relevance .............................................................................................. 96  
  Effectiveness.......................................................................................... 96  
  Efficiency ............................................................................................... 97  
  Coherence ............................................................................................. 98  
  EU added value ...................................................................................... 98  
  Sustainability ......................................................................................... 99  
  Priorities for improving the Code .............................................................. 99  
**ANNEX 1: CASE STUDIES** .............................................................................. 101  
  Case study on Issue-Based Advertising .................................................... 102  
  Case study on Key Performance Indicators ............................................. 108  
  Case study on partnerships with researchers - Data requests ...................... 111
EXECUTIVE SUMMARY

This Executive Summary presents key findings and the conclusions of the study “Assessment of the implementation of the Code of Practice on Disinformation, SMART 2019/0041”. The study was commissioned by the Directorate-General for Communications Networks, Content and Technology (DG CNECT) of the European Commission and it was carried out by VVA Economics & Policy with the support of DisinfoLab.

A. Objectives and methodology

The overarching study objective is to support the European Commission’s evaluation of the Code of Practice’s effectiveness. The assessment focuses on the 13 current Signatories of the Code of Practice on Disinformation (online platforms and business associations).

The study analyses the standard terms of service and the specific policies and tools adopted by the online platforms to implement the commitments of the Code in the first year of implementation of the Code (October 2018 to October 2019). For this reason, ongoing efforts made by the Signatories to combat disinformation in light of the COVID-19 pandemic are not covered.

The study methodology is documented in the Evaluation Plan (Annex 4). The methodology was designed specifically to be replicable so that it can form the basis for future assessments.

B. Findings and conclusions

The study's overall conclusion is that the Code of Practice has produced positive results.

There is a consensus among stakeholders that the Code of Practice is needed. Since disinformation continues to be a widespread problem, the Code, its aims and activities are considered to be highly relevant.

Furthermore, stakeholders consulted for the study also agreed that disinformation is a topic where the EU has an added value and where it should continue to lead and coordinate action. Despite differences in stakeholder views with regards to the effectiveness of self-regulation, there is widespread acknowledgement that the Commission is right in pursuing a dialogue with the social media platforms.

There is also acknowledgement that the Code constitutes a first and crucial step in the global fight against disinformation. In this sense, the Code shows European leadership on an issue that is international in nature.

When it comes to effectiveness, the study identifies a range of achievements. Firstly, the Code has established a common framework under which to agree on and implement activities to tackle disinformation. In doing this, the Code has set a foundation on which further activities can be built. Indeed, the Code – and the preparatory activities carried out before its establishment – has contributed to the debate on disinformation, raised awareness, and provided guidance to stakeholders (e.g. civil society, policymakers, the media and publishing sector) and to the Signatories.
Secondly, the discussions facilitated by the work of the Code have also contributed towards a specific set of actions and measures at EU and national levels and it has improved cooperation between policymakers and the Signatories to combat disinformation:

- For instance, the Code has established a platform for negotiation that has produced concrete results in the form of regular monitoring of Signatory activities and continuous action to combat disinformation activities. The monitoring processes report on change over time, which allows for better transparency of social media platforms during elections and other political campaigns.

- In addition, the Code has also led to increased reflection among Member States with regards to activities to understand and combat disinformation. Some Member State authorities are planning/undertaking activities relevant to the Code, e.g. planning disinformation strategies and preparing (better) monitoring of the phenomenon.

In addition to these general conclusions, the table below summarises specific key findings on the Code’s five Pillars.

<table>
<thead>
<tr>
<th>Pillar</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillar 1: Scrutiny of advertisement placements</td>
<td>The Code has prompted Signatories to put in place new, or to enhance existing measures to scrutinise ad placements. However, available data are not sufficiently detailed to be assess the effectiveness of these measures. There is an inconsistent understanding of the details and implications of the Code under this pillar, which (partly) explains the lack of data. Currently, the Code does not have a high enough public profile to put sufficient pressure for change on platforms. Future iterations of the Code should refer to click-baiting as a tool used in disinformation and specifically ad placements.</td>
</tr>
<tr>
<td>Pillar 2: Political advertising and issue-based advertising</td>
<td>While efforts have been made by platforms in the area of political and issue-based advertising, there is still room for improvement. For instance, measures regarding political advertising have been more effective than measures regarding issue-based advertising. However, there are widely different views (also among the experts consulted for the study) regarding the scope of ‘issue-based’ advertising, with national culture being a crucial factor influencing understanding of the scope of concept.</td>
</tr>
<tr>
<td>Pillar 3: Integrity of services</td>
<td>The Signatory platforms have put in place tools and policies to combat inauthentic behaviour and malicious actors as agreed by the Code, while non-Signatory stakeholders consulted seem to have less knowledge on the impacts of such tools under this pillar. To better understand the ongoing interaction and development of inauthentic behaviours and malicious actors, the focus within this pillar could also be on the reach and influence of these aspects. Studies and experiments could be conducted into how such actors and their associated behaviour contribute to the spread of disinformation and how effective the platforms’ tools and policies are in preventing this spread.</td>
</tr>
<tr>
<td>Pillar 4: Empowering consumers</td>
<td>Although there is no convergence of opinion regarding stakeholders’ assessment of Pillar 4, most of the signatory platforms have a range of tools in place for empowering consumers. However, these tools have not (yet) been consistently rolled out across all Member States and reporting on their impact is inconsistent. Most stakeholders identified see an increased consumer awareness on the topic of disinformation, but it is difficult to establish the extent to which this can be attributed to the Code of Practice.</td>
</tr>
<tr>
<td>Pillar 5: Empowering</td>
<td>Efforts have been made by the Signatories to support and encourage good faith research into disinformation. However, these initiatives should be further developed to be more effective. Most researchers consulted for this study indicated that access to</td>
</tr>
</tbody>
</table>
Assessment of the implementation of the Code of Practice on Disinformation

<table>
<thead>
<tr>
<th>Pillar</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>the research community</td>
<td>data is still limited, or that databases that are made available are not user-friendly. In addition, there is a general lack of trust between researchers and platforms.</td>
</tr>
</tbody>
</table>

The main criticism of the Code relates to its self-regulatory nature, lack of uniformity of implementation – evidenced by the unevenness of progress made under the specific Pillar – monitoring, and lack of clarity around its scope and some of the key concepts.

First, as mentioned already above, it is a voluntary document, and as such there are no means to enforce the commitments of the Signatories nor do the 13 signatories cover all relevant stakeholders. This has led to **at times fragmented implementation across the various Pillars, across platforms and across Member States.** In particular, in relation to Pillars 4 and 5 that are focused on the relationship between Signatories and consumers and between Signatories and the research community respectively. Further time and effort will be required to ensure that all five Pillars are implemented as effectively as possible and in a uniform manner across the different stakeholders and geographies.

Second, the study shows that there remains a need for a common understanding of key concepts. Indeed, **disinformation is a topic that is not very clearly defined and it can at times be interpreted subjectively.** To combat this lack of clarity and foster a harmonised approach, it is important that the action that are agreed upon are as concrete as possible to facilitate the definition of intended results and key performance indicators and support implementation and monitoring. For example, it appears to be the case that measures regarding political advertising have seen stronger development than issue-based advertising because stakeholders are unclear about the remit of issue-based advertising which lacks a common definition. Similarly, to ensure a common language and improve future evaluation exercises a joint terminology needs to be agreed among stakeholders in the near future.

**C. Future considerations**

As the study has shown, the Code remains relevant, it has led to positive results, and it provides value added at a European level.

For this reason, **the Code should not be abandoned, and its implementation should continue.** However, the effectiveness of the Code can be strengthened in the following ways:

1. Continued efforts to debate the Code’s strengths and weaknesses with the Signatories, non-Signatories and wider stakeholders. These debates could focus on agreeing on terminology and definitions of key terminology, as well as discuss and assess the current scope of the Code and how current weaknesses can be addressed.

2. A mechanism for action in case of non-compliance of the Code’s Pillars could be considered. To that effect, the European Commission should consider proposals for co-regulation within which appropriate enforcement mechanisms, sanctions and redress mechanisms should be established.
3. Further support to evaluation and monitoring of the Code is needed; this study also provides methodological recommendations that can be used to further improve this process (proposed key performance indicators are documented in Chapter 7).

4. Consider strengthening the practical implementation of the current requirements of the Code, which entails that signatories should implement activities to the same standard across Member States. Ensuring that the common standards for the platforms to adhere to in each Member State are also enforced would ensure better consistency in the implementation of the Code across the Union and help create more bargaining power towards the platforms. For example, if a tool is not implemented in all MS at the same time, the platforms should communicate a provisional calendar when users in each of the MS can expect to be able to use these tools. This is important given that a large part of the threat comes from state actors outside of Europe.
CHAPTER 1: INTRODUCTION

This document forms the Final Report for the “Assessment of the implementation of the Code of Practice on Disinformation, SMART 2019/0041”. This study was commissioned by the Directorate-General for Communications Networks, Content and Technology (DG CNECT) of the European Commission and was carried out by VVA Economics & Policy with the support of DisinfoLab.

This document is structured as follows:

- **Chapter 2** provides a brief description of the study methodology. This description is supported by complementary material, including the study’s Evaluation Plan, provided in the report Annexes.

- **Chapter 3** deals with the wider context of the study subject, outlining a summary of the emergence of the phenomenon that is ‘disinformation’, along with current and future challenges.

- **Chapter 4** provides an introduction to the Code of Practice.

- **Chapter 5** outlines the findings according to the evaluation criteria of the study, starting with Effectiveness.

- **Chapter 6** gives an overview of Member State initiatives that aim to tackle disinformation.

- **Chapter 7** presents the study’s proposed metrics and key performance indicators (KPI) to strengthen the foundation for future monitoring and evaluation activities of the Code.

- **Chapter 8** summarises the study’s conclusions.

- **Supporting Annexes** provide further evidence and information on the study activities.

**Objectives and scope**

The overarching objective of this study was to support the evaluation of the Code of Practice’s effectiveness being performed by the European Commission. A key task was to gather data and information and analyse the implementation of the Code of Practice by its Signatories.

More specifically, the study had three operational objectives:

1. The development of an Evaluation Plan to measure and assess the effectiveness of achieving the Code’s objectives by its Signatories.
2. The implementation of the Evaluation Plan once approved by the Commission.
3. A qualitative and quantitative assessment of the effectiveness of the implementation of the Code of Practice in 2019. Where relevant, the study indicates areas where the Signatories can improve their implementation of the Code and its objectives.

The study focused on the 13 entities that are currently the Signatories of the Code of Practice on Disinformation. These include online platforms and business associations.
Assessment of the implementation of the Code of Practice on Disinformation

The results of the study analyse the standard terms of service and the specific policies and tools adopted by the online platforms to implement the commitments of the Code.

The study focuses on the first year of implementation of the Code of Practice, i.e. October 2018 to October 2019. Therefore, the current efforts made by the Signatories to combat disinformation in light of the COVID-19 pandemic are not analysed.
CHAPTER 2: METHODOLOGY

This section provides a brief description of the methodology designed and deployed for the study. It describes the main research tools used as well as provides a summary of the analysis undertaken in order to reach the study conclusions. The final section outlines some (inherent) limitations of the study.

The project roadmap is outlined in Figure 1 (below) and shows the steps in which the study was undertaken. It was divided into four main tasks:

Task 0: Inception and conceptualisation. This task finalised the study design, timeline and saw the study team undertake preparatory work, focusing on the evaluation framework and conceptual approach. The study team undertook a small number of scoping interviews during this task.

Task 1: Development of an evaluation plan: The purpose of task 1 was to design a clearly defined and comprehensive evaluation plan that would inform the study’s assessment of the effectiveness of the Signatories in achieving the objectives of the Code. The output of task 1 was a brief report which built on the inception phase (Task 0) in terms of developing the evaluation questions framework, intervention logic of the Code and relevant indicators. Task 1 was informed by an evidence review and by a pre-consultation interview programme which explored methodological questions (rather than evaluative ones). Two external experts were also consulted and asked to provide quality assurance at this stage of the study.

Task 2: Implementation of the evaluation plan: Task 2 constituted the main data collection activities of the study and covered mapping the measures and policies of the Signatories, creation of an online survey, selection of case studies and interview programme. These are further described below in section 2.1.

Task 3: Suggestions for future development: The key findings and conclusions are presented in this report, which constitutes the main output of the study.
The main research tools of the study were deployed as part of Task 2. The data collection and analysis followed the methodology developed under Task 1. The research tools covered the following activities:

Table 1: Overview of the study research tools

<table>
<thead>
<tr>
<th>Study Task</th>
<th>Objective</th>
<th>Approach and activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping of Signatory measures and policies</td>
<td>The main objective to carry out in-depth desk research exercise in order to provide a detailed mapping of the measures and policies implemented by the Signatories of the Code of Practice</td>
<td>The mapping of activities and policies covered: 1) Mapping of implementation of the Code of Practice; and 2) mapping of European and national measures to counter disinformation.</td>
</tr>
<tr>
<td>Interview programme</td>
<td>The interview programme complemented the desk research and gathered qualitative input on the effectiveness of the implementation of the Code.</td>
<td>The study team conducted 27 interviews for this task of the study. The interview programme targeted a wide range of stakeholder categories. A complete list of the individuals consulted can be found in Annex 3. List of stakeholders contacted.</td>
</tr>
<tr>
<td>Online survey</td>
<td>Two online surveys were designed and launched to gather input from wider stakeholders and experts, including from national level authorities.</td>
<td>One online survey targeted national audiovisual regulatory authorities active in ERGA. The second survey targeted academic and other experts on disinformation. The first survey was disseminated with the help of DG CNECT while the second survey was disseminated with the help of DisInfoLab. The survey to the national authorities generated 15 complete responses. The survey to the expert public generated 20 complete answers.</td>
</tr>
</tbody>
</table>
**Assessment of the implementation of the Code of Practice on Disinformation**

<table>
<thead>
<tr>
<th>Study Task</th>
<th>Objective</th>
<th>Approach and activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development of four case studies</td>
<td>The case studies were developed to support the conclusions with contextual analysis on specific disinformation-related topics.</td>
<td>Four case studies were developed on the following topics: 1. Issue-based advertising 2. KPIs/Metrics 3. Data requests 4. Partnerships with fact-checkers</td>
</tr>
</tbody>
</table>

**Mapping of Signatory measures and policies**

The study team examined the content of the monitoring reports submitted by the Signatories and EC assessments of the monitoring. This exercised provided a critical view of progress made of the Code’s commitments on a pillar-by-pillar basis. This task was an ongoing exercise throughout the study however a greater focus on the mapping preceded the interview programme. This enabled those conducting the interviews to tailor the questions, taking into account the specific commitments and documented progress of each platform.

**Interview programme**

In total 27 semi-structured interviews were carried out by the study team. A complete list of the organisations consulted can found in Annex 2. Extensive preparations were undertaken by the study team with the support of the Commission. Semi-structured interview guides were developed for:

- Academic and other experts on the topic of disinformation
- Non-Signatory platforms, i.e. social media platforms that have not signed up to the Code of Practice)
- Signatory associations
- Signatory platforms

The guidelines were circulated in advance of the interview and the minutes collected for the synthesis analysis of the overall findings.

**Online survey**

Two on-line surveys were also designed and launched to collect further data. The target groups for the surveys were: i) National audiovisual authorities active in ERGA, and ii) experts and professionals with an interest and knowledge in the topic of disinformation.

The first survey was disseminated via the European Commission. The second survey was disseminated via the mailing list of the NGO DisInfoLab.

The survey questionnaires were designed using a mix of qualitative and quantitative questions. Draft questionnaires of the surveys can be found in the Evaluation Plan appended to this report.
The study team undertook follow-up interviews with a select number of survey respondents to follow-up on some of their initial responses.

**Development of four case studies**

The study team developed four case studies to complement the above data collection tools. The case study topics are: i) Issue-based advertising; ii) KPIs/Metrics; iii) Data requests; and iv) Partnerships with fact-checkers. These were selected to provide additional qualitative evidence on some of the key issues of the Code that could not easily be documented elsewhere or deserved special attention. The purpose of the case studies is therefore to provide further context, map progress made, and highlight challenges to date.

Each case study was developed through desk research and 1-3 additional interviews, carried out with relevant stakeholders. The case studies are presented in Annex 1.

A more complete overview of the research tools can be found in the Evaluation Plan.

**Analysis**

The first activities undertaken were the desk research and the interview programme. A preliminary/initial analysis of the desk research findings and the interview programme was first undertaken in December 2019 and January 2020.

Once information was also available from the survey results (launched in January 2020) and from the case studies (undertaken in February/March 2020), all information gathered was triangulated according to source and evaluation criteria. All findings were stored in a database for final analysis and cross-tabulation.

Preliminary findings were discussed at multiple points of the study with the Steering Group and preliminary analysis was also shared with three external experts (two former members of the High-Level Expert Group and DisinfoLab) who provided extensive feedback and quality assurance throughout the various stages of the analysis.

**Limitations**

Although the initial expectations of the study in terms of quality and scope were met, there are some inherent limitations in the research that should be highlighted for context and transparency.

Firstly, the evaluation concerned an intervention designed to address a recent (and increasingly growing) phenomenon – disinformation – where there is a limited existing knowledge base. Although the study undertook an evidence review to consider the status quo of different interpretations of disinformation, there are still issues pertaining to this topic where there is a lack of evidence or agreement. Grey areas (e.g. the relationship between click bait and disinformation spreading) are flagged up throughout the report as areas where conclusive evidence is still lacking.
The Code of Practice on Disinformation is the first of its kind and the first attempt to self-regulate the phenomenon. The study found that, although many stakeholders were very well-informed, others had only peripheral knowledge of the Code and its activities. Although monitoring reports are available documenting the Code’s activities, there were also very limited written analyses on the performance of the Code to date on which the study could draw.

Finally, the Signatories of the Code were consulted as part of the study. They were an important target group for retrieving information, but they also occupy a subjective position since they are the Code’s main implementors. The study team undertook several mitigating actions to avoid bias in the analysis of data gathered. This included rigorously ensuring the quality of research tools, pre-consultations with stakeholders as well as a comprehensive interview programme to allow for a widest range of input as possible, and careful triangulation and validation of primary and secondary data used by the study.

The study team also made repeated attempts to contact non-signatory platforms with a request for an interview in order to provide a broader and more balanced assessment.
CHAPTER 3: THE CONCEPT OF DISINFORMATION

This section describes the wider context of disinformation as a concept. The text presented below is based on the desk research and literature review undertaken as part of the study. This chapter first describes the recent emergence of disinformation. The latter section highlights some of the key challenges posed by this phenomenon.

Emergence of disinformation and its definitions

In the last 10 to 15 years societal discourse and the spread of information have changed dramatically, triggered by the disruptive technological innovation of the internet, and social media more specifically. Through such processes, dominant one-to-many authoritative media channels – based on institutionally-embedded editorial functions – have been complemented, challenged and in many ways replaced by many-to-many social media platforms and a plethora of websites and fora for news aggregation and opinion dissemination.

In this context, concepts such as “fake news” have emerged in the political debate. Although these terms are often used ambiguously,1 available data does suggest a growing concern about the truthfulness and reliability of news and information spread through social media. In 2014, the rapid spread of disinformation online was identified by the World Economic Forum as one of the 10 main trends in the modern society.3

Widespread social media use has resulted in a faster spread of disinformation, notably during election periods, which has contributed to an increase in, and evolution of, propaganda campaigns. The rise of social media use means differentiated political messages can now be delivered much more easily and cheaply down to the level of the individual voter. Individual-level data is collected from users when they connect via computers or mobile applications and this data is then used during political campaigns.4

The challenge with reigning in the spread of fake news is to ensure processes of societal opinion formation which accurately and sincerely reflect the opinions of the population, whilst at the same time preventing the spread of malicious, inaccurate, and wilfully damaging misinformation as well as covert attempts to prevent democratic behaviour.

News, by definition, is not false. It is the narratives that can be false and, although advertised as news and containing parts of texts copied from newspapers or websites of the same kind, incorporates false, inaccurate or misleading content. This false

1 Ott 2016 The age of Twitter: Donald J. Trump and the politics of debasement http://www.tandfonline.com/doi/abs/10.1080/15295036.2016.1266686
information is intentionally designed, presented and promoted to cause public harm or make a profit. Thus, such narratives may not even contain information with illegal content, such as hate speech, incitement to violence, terrorism, child sexual abuse, all of which are subject to regulatory remedies under European or national law. Instead, these kinds of narratives are potentially harmful to the formation of public opinion and, therefore, for the support of a democratic society.\(^5\)

**There are many definitions of disinformation.**\(^6\) According to a Reuters report of 2017 on digital news, it is difficult to precisely define what disinformation or “false” news is because it is often applied to three distinct categories:\(^7\):

1) News that is made up or “invented” to make money or to discredit others;
2) News that has a basis in fact, but is “spun” to suit a particular agenda; and
3) News that people do not feel comfortable hearing about or do not agree with.

However, many authors have tried to define the phenomenon by considering its diverse characteristics. “False” news can, for example, be classified according to various criteria such as the source of the news (as in Russian agents or Macedonian click-baiting teenagers\(^8\)), the content (factually incorrect or distorted views), the spreading method (targeted advertising, bots, social networks) and the intention (to influence elections, to divide and stoke discontent, or to earn money).\(^9\) Alcott & Gentzkow (2017) defines more precisely “false” news as “intentionally and verifiably wrong or false news produced for the purpose of earning money and/or promoting ideologies”. Their definition explicitly excludes “slanted” news, conspiracy theories, rumours and “false statements by politicians”. They argue that there is a market for verifiably false news because:

- It is cheaper to produce false than accurate news
- it is costly for consumers to distinguish between accurate and false news, and;  
- consumers may enjoy reading fake news because it confirms their beliefs.\(^10\)

In contrast, the phenomenon can also be defined via the definition of the quality of a news item. Quality news should: i) be accurate and reliable, ii) help to understand complex issues, iii) communicate strong viewpoints and opinions, and iv) provide amusing and entertaining content.\(^11\) Despite the numerous definitions of the term “false news” and its frequent use in the media, scholars have argued that the term is inadequate


\(^7\) https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

\(^8\) https://www.bbc.com/news/magazine-38168281


\(^11\) https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf
to describe the complex phenomena of mis-/dis-information. As the researcher Ethan Zuckerman states: “it is a vague and ambiguous term that spans everything from false balance (actual news that does not deserve our attention), propaganda (weaponized speech designed to support one party over another), and disinformation (information designed to sow doubt and increase mistrust in institutions).”

The vagueness of the term is also noted in a study by Tandoc et al. published in August 2017, which examined 34 academic articles that used the term “fake news” between 2003 and 2017. The authors noted that the term has been used to describe several different phenomena over the past 15 years: News satire, news parody, fabrication, manipulation, advertising, and propaganda. The term ‘fake news’ is thus used in a variety of contexts and, for many researchers, it does not reflect the complexity of the phenomenon. Similar problems with the term ‘fake news’ were identified by the European Commission’s High-Level Expert Group on Fake News and Online Disinformation (HLEG). In its final report, the HLEG found the term ‘fake news’ to be “inadequate to capture the complex problem of disinformation which involves not necessarily “fake”, but fabricated content and practices going beyond the conventional “news”. The 2018 EC Communication on Tackling online disinformation: a European Approach also captures the additional complexities in understanding disinformation in a policy context (rather than an academic one). The EC Communication refers to disinformation 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.”

Finally, the term ‘fake news’ is found to be misleading due to it being appropriated by some politicians to dismiss any content they regard as disagreeable. The HLEG preferred the term ‘disinformation’ over ‘fake news’ and, throughout the HLEG report, refers to ‘disinformation’ as: “All forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit.” Among scholars the term ‘disinformation’ is often used but is sometimes mentioned alongside the term ‘misinformation.’ Both terms are therefore not consistently used. The

---

12 Wardle C. and Derakhshan H., Information disorder: definitions in “Understanding and addressing the disinformation ecosystem”, 2017, p. 6
13 Idem
15 COM (2018) 236 final
16 High level Group on fake news and online disinformation, A multi-dimensional approach to disinformation, 2018, p. 10
17 Ibid., p. 3
two terms are also used interchangeably and treated as if there were no difference (Fox, 1983; Losee, 1997). Sometimes, one is used as a variation of the other (Zhou & Zhang, 2007), but the reasoning behind that relationship is not always clear.18

While the Oxford online dictionary and the Collins English Dictionary list ‘misinformation’ as a synonym for ‘disinformation’, the Merriam–Webster and Oxford Living Dictionaries make subtle distinctions between the two definitions. The EU's interinstitutional terminology database IATE (InterActive Terminology for Europe)19 specifically notes that disinformation should not be confused with misinformation, defined in IATE as “information which is wrong or misleading but not deliberately so”.20 Yet, the European Parliament resolutions have used ‘misinformation’ and ‘disinformation’ interchangeably.21 One approach to distinguishing the two terms is provided by the definitions of Wardle & Derakshan (2017). They clearly distinguish between three dimensions of harm and falseness, and introduced three types of “false news”:

(1) Misinformation when false information is shared, but no harm is meant;  
(2) Disinformation when false information is knowingly shared to cause harm; and  
(3) Malinformation when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere.22

Table 2: Categorisation by Wardle & Derakshan of “false news”

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misinformation</td>
<td>When false information is shared, but no harm is meant</td>
<td>During the 2016 US presidential elections, a tweet about a ‘rigged’ voting machine in Philadelphia was shared more than 11 000 times. It was later established that the original tweet was a mistake made by a voter who had failed to follow the instructions exhibited on the voting machine.23</td>
</tr>
<tr>
<td>Disinformation</td>
<td>When false information is knowingly shared to cause harm</td>
<td>During the 2017 French presidential elections, a duplicate version of the Belgian newspaper Le Soir was created, with a false article claiming that Emmanuel Macron was being funded by Saudi Arabia.24</td>
</tr>
<tr>
<td>Malinformation</td>
<td>When genuine information is shared to cause harm</td>
<td>Examples include intentional leakage of a politician’s private emails, as happened during the presidential elections in France.25</td>
</tr>
</tbody>
</table>

The inability of scholars to agree on a single terminology is also because disinformation is not a new phenomenon at all; it is as old as the newspaper industry. The first occurrence of disinformation was reported in the 16th century. An Article from the New York Review of Books describes that Pietro Aretino tried to manipulate the pontifical election of 1522 by writing wicked sonnets about all the candidates (except the favourite of his Medici patrons) and pasting them for the public to admire on the bust of a figure known as Pasquino near the Piazza Navona in Rome. The “pasquinade” then developed into a common genre of diffusing nasty news, most of it false, about public figures. The Article then states that pasquinades were succeeded in the 17th century by a more popular genre, the “canard,” a version of false news that was sold in the streets of Paris until the end of the 18th century. Canards were printed broadsides, sometimes set off with an engraving designed to appeal to those regarded as particularly impressionable. During the French Revolution, the engravers inserted the face of Marie-Antoinette on the old copper plates, and the canard took on new life, this time as intentionally fake political propaganda. Although its impact cannot be measured, the article states that it certainly contributed to the pathological hatred of the queen, which led to her execution on 16 October 1793.

The specific term of ‘fake news’ however only emerged around the end of the 19th century due to the relative novelty of the word ‘fake’. According to the website of the Merriam–Webster Dictionary, ‘fake’ was rarely used as an adjective prior to the late 18th century, and before that point, the most common term in use was ‘false news’. The phenomenon has risen with the digital transformation of news from offline to online distribution and the rise of social media as a news distribution channel. The phenomenon gained particular momentum and global visibility during the final months of the 2016 US presidential election, when viral “false news” across the political spectrum received more engagement on Facebook than real news.

The data from Google Trends and the Web of Science show that searches for “false” news and the number of peer reviewed papers including the term ‘false’ or ‘fake’ news have grown exponentially since November 2016. Yet, for 2016, ‘post-truth’ was designated as the word of the year, signalling the crisis of journalism and the growing presence of misinformation and disinformation. ‘Fake news’ was designated as the term of the year by the Collins English Dictionary in 2017. In the same year the World Economic Forum addressed the problem of “misinformation”, “disinformation” and “propaganda” in its risk assessment, also pointing at the danger posed by the decreasing trust in institutions.

Despite the academic debate around the various terms used to describe a false information, this document will use the term ‘disinformation’ throughout the final report.

---

26 https://www.1843magazine.com/technology/rewind/the-true-history-of-fake-news
30 https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook
on the Evaluation of the Code of Practice on Disinformation, as it is the term used and referred to throughout the Code. Furthermore, this study was commissioned by the European Commission and ‘disinformation’ is the term preferred by the EC Communication on disinformation, and subsequently adopted by the HLEG report.

**Current and future challenges**

The 2016 Brexit vote in the United Kingdom and the tumultuous U.S. presidential election in the same year highlighted how the digital age has affected news and cultural narratives. In this context, social media became the most attractive vehicle for disinformation. A study by Allcott and Gentzkow, showed that the use of online social media is rising sharply (e.g. the number of users on Facebook grew from 500,000 in 2010 to nearly 2 billion in mid-2017), while the trust in mass media continues to decline.

Indicated by a poll conducted during the American elections of 2016, between 1998 and 2016 there was a continuing decline of “trust and confidence” in the mass media “when it comes to reporting the news fully, accurately, and fairly.” This decline was particularly marked among Republican voters compared to Democrat voters, with a particularly sharp drop among Republicans in 2016. Allcott and Gentzkow explain that the declining trust in mainstream media could be both a cause and a consequence of fake news gaining more traction. Social networks have become an important source of political news and information in general.

Furthermore, the research from Allcott and Gentzkow showed that the largest share (41.8%) of traffic to disinformation sites comes from social networks, while legitimate news sites are most reached by direct browsing (48.7%), and social media there plays only a minor role (10.1%). In contrast to the social media networks, the role of other digital platforms, such as search engines and messaging applications, is less thoroughly researched and thus less understood.

Investigations by the Guardian newspaper have reported Google Search results on a particular political topic being dominated by the blogs promoting similar extreme viewpoints and pushing credible news sources out of the first page. Google later

---


35 Constine, J., Facebook now has 2 billion monthly users... and responsibility, 27 June 2017, https://techcrunch.com/2017/06/27/facebook-2-billion-users/


acknowledged struggling with people who ‘try to game the system’ in order to bolster ‘low quality’ content and ‘fake news’. The actors behind the disinformation campaigns can introduce the misleading content onto the platforms and make use of services available to amplify their messages.

Research also suggest that people are much more likely to share false stories. False news reached an audience of 1,500 people six times quicker than a true story. Political false news spread faster than business, terrorism-related, sports and entertainment false news. This study found that it is humans (rather than bots) who are more likely to spread false news. Robots made little distinction between false stories and true stories. One reason could be that the reaction of humans to false news is notably different to that of true news. False news inspires strong emotions such as fear, disgust and surprise while the true stories inspired joy, sadness and anticipation. This is leading many commentators to the conclusion that “social media seems to systematically amplify falsehood at the expense of the truth” which is supposed to be “a dangerous moment for any system of government premised on a common public reality.”

Widespread social media use has resulted in a faster spread of disinformation, notably during election periods, which has contributed to an increase in and evolution of propaganda campaigns. Freedom House documented prominent examples of disinformation around elections or referendums in at least 16 of the 65 countries assessed in 2017. The Oxford Internet Institute reports that in 2017 they “found evidence of formally organized social media manipulation campaigns in 48 countries.”

Differentiated political messages can now be delivered much more easily and cheaply down to the level of the individual voter through social media. Individual-level data is collected on users when they connect on computer or mobile applications and this data is then used during political campaigns. In March 2018, it was disclosed that user data from 87 million Facebook users – including that of 2.7 million EU citizens – had been improperly shared with the controversial political consultancy company Cambridge Politicalpropaganda. Also see: Ghosh D. and Scott B., #Digitaldeceit. The technologies behind precision propaganda and the Internet. Harvard Kennedy School, January 2018, p. 20; Meserole C. and Polyakova A., Disinformation Wars, 25 May 2018, https://foreignpolicy.com/2018/05/25/disinformation-wars/  
42 Ibid.  
44 Freedom house, Manipulating Social Media to Undermine Democracy, November 2017  
46 Ibid. Pg. 3  
Analytica (which used the data to micro-target and mobilise voters in the US and the UK). 48

The data collected can also be used by hostile actors to spread disinformation and to target specifically some parts of the population, according to individual characteristics. 49

At a US Senate subcommittee hearing in October 2017, it was claimed that Russian operatives had, for example, specifically targeted patriotically minded adult Texans with a political advertisement purporting to be from a Texas-based organisation, which contained false claims against the then Presidential candidate Hillary Clinton. 50 These sophisticated and targeted political advertisements used Facebook’s own targeting technology, which enabled a hostile actor to quickly and cheaply reach audiences in a Western country during election periods.

Domestic and foreign actors are using disinformation to sow distrust and create societal tensions, which can have serious consequences for security. 51 Actors can use it to manipulate policy, societal debates and behaviour in areas such as climate change, migration, public security, health, and finance. As a result, disinformation can diminish trust in science and empirical evidence. 52 Disinformation also erodes trust in institutions and in digital and traditional media and harms democracies by hampering the ability of citizens to take informed decisions. Disinformation is often used as a boon for radical and extremist ideas and activities.

Media literacy is widely acknowledged as being a key mitigation factor for disinformation. Media literacy initiatives often take the form of awareness campaigns organised by research centres or individual researchers in schools and high schools explaining the concept of disinformation and how to combat it. This can also include workshops and conferences directed at the general or expert public and often involves partnerships between research centres and platforms or research centres with other institutes.

Fact-checking is another approach to combating disinformation that can contribute to building trust, accountability and transparency in online news sources. In traditional media, this is mitigated by journalistic principles and ethical standards. It is a soft, non-regulatory approach mostly driven by non-governmental and civil society organisations. 53 Online platforms have been notably active in this area i.e. Facebook and the Journal.ie partnered to identify disinformation in advance of the May 2018 Irish Referendum. 54

---

51 https://www.rusemb.org.uk/press/2029
53 Irish Government, First report of the interdepartmental group on security of Ireland’s electoral process and disinformation, June 2018
54 https://www.siliconrepublic.com/companies/facebook-and-the-journal-team-up-on-fact-checking-project
Assessment of the implementation of the Code of Practice on Disinformation

Platforms have also taken other measures, including those to remove fake accounts and to prohibit certain types of advertising on their platform.\textsuperscript{55}

In order to prevent or at least minimise the spread of disinformation ahead of the European Parliament elections in 2019, the EU invested in many activities to address the challenges of disinformation. The Code of Practice on Disinformation represents a key element of this effort. The Code of Practice was signed in October 2018 and it aims at achieving the objectives set out by the Commission's 2018 Communication\textsuperscript{56}, notably through securing a wide range of measures related to the industry’s self-regulatory commitment. Adherence to the Code of Practice is thus voluntary and the initiative to participate in, adhere to, and act on it is put on industry.

Despite the adoption of an Action Plan to fight the phenomenon of disinformation, the results of the literature review suggest that the EU will continue to face new types of challenges related to disinformation in the future.

In the future, humankind will notably need to face the constant rise of technologies; and their variety will not facilitate the control of disinformation. A report published by the European Parliament found new services’ reliance on personal data will only intensify in the future, as will the capacity to imitate reality through, for example, augmented reality (AR) or virtual reality (VR), which enables falsification to become unrecognisable by both human and machine control.\textsuperscript{57}

The same report states that machine learning, advanced demographic analytics, the internet of things, voice and facial recognition “will further increase the vulnerability of humans to erosion of privacy and having their personal data involuntarily exposed”. New and strong regulatory intervention will be needed to avoid that human psychological traits are exploited for disinformation or more dangerous purposes.

More recently, in February 2020, the European Commission published a package of three documents setting out its digital agenda. Firstly, a White Paper on Artificial Intelligence and secondly, a European data strategy. These two documents as well as the Commission’s approach to platform regulation are set out in the document “Shaping Europe’s digital future” covering three priorities: Technology that works for the people, a fair and competitive digital economy and open, democratic and sustainable society as well as a commitment to act as a global leader on the topic.\textsuperscript{58,59}

More concrete legislative proposals on these topics are expected by the end of 2020, when the Commission will publish a package for the Digital Services Act.

\textsuperscript{55} Code of practice on Disinformation Summary of the signatories’ first reporting – January 2019
\textsuperscript{56} COM/2018/236 final
\textsuperscript{57} European Parliament - IPOL_STU(2019)608864_EN-Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States
\textsuperscript{58} https://ec.europa.eu/commission/presscorner/detail/en/ip_20_273
CHAPTER 4: THE CODE OF PRACTICE ON DISINFORMATION

This section first describes the background to the Code of Practice on Disinformation, including the wider context in which the Code was developed. It then outlines the aims and scope of the Code.

Context

The European Union, with the then upcoming European elections on the horizon and concerns of outsider interference in the democratic process, stepped up its efforts in the area of combating and preventing the spread of disinformation in 2018. In January 2018, the European Commission set up a high-level group of experts (the HLEG referred to in chapter 3), to advise on policy initiatives to tackle the spread of disinformation. The group was composed of 40 representatives of social media platforms and media organisations and of citizens, civil society organisation and experts such as journalists and academics. This expert group triggered and supported the idea of a European Code of Practice "reflecting the respective roles and responsibilities of relevant stakeholders, especially online platforms, media organisations, fact-checking and research organisations."

It met for the first time on 15 January 2018 and gathered opinions on actions that could be taken at EU level to give citizens effective tools to identify reliable and verified information and adapt to the challenges of the digital age. The main deliverable of the HLEG was a report designed to review best practices in the light of fundamental principles, and suitable responses stemming from such principles. This report also suggested key principles for the elaboration of the Code of Practice, and how it should be implemented by the Commission the relevant stakeholders:

1. Platforms should adapt their advertising policies, including adhering to “follow-the-money” principle, whilst preventing incentives that lead to disinformation, such as to discourage the dissemination and amplification of disinformation for profit. These policies must be based on clear, transparent, and non-discriminatory criteria;

2. Platforms should ensure transparency and public accountability regarding the processing of users’ data for advertisement placements, with due respect to privacy, freedom of expression and media pluralism;

3. Platforms should ensure that sponsored content, including political advertising, is appropriately distinguished from other content;

---

60 DG Connect, 2018, Report of the independent High-level Group on fake news and online disinformation, A multi-dimensional approach to disinformation
4. Platforms should take the necessary measures to enable privacy-compliant access to data for fact-checking and research activities;

5. Platforms should make available to their users advanced settings and controls to empower them to customise their online experience;

6. Platforms should, in cooperation with public and private European news outlets, where appropriate take effective measures to improve the visibility of reliable, trustworthy news and facilitate users’ access to it;

7. Where appropriate, trending news items should, if technically feasible, be accompanied by related news suggestions;

8. Platforms should, where appropriate, provide user-friendly tools to enable users to link up with trusted fact-checking sources and allow users to exercise their right to reply;

9. Platforms that apply flagging and trust systems that rely on users should design safeguards against their abuse by users; and

10. Platforms should cooperate, for instance by providing relevant data on the functioning of their services including data for independent investigation by academic researchers and general information on algorithms in order to find a common approach to address the dissemination and amplification of disinformation.

The HLEG stressed that the best responses were likely to be ‘those driven by multi-stakeholder collaborations’ and it would be more effective to ‘minimize legal regulatory interventions’, and avoid the ‘politically dictated privatization of the policing and censorship of what is and is not acceptable forms of expression.’

In April 2018, the Commission published a Communication, *Tackling Online Disinformation: a European Approach*. This Communication outlined actions to be taken to tackle online disinformation. Building upon the Key Principles established by the HLEG, the Communication called upon industry to draft of a Code of Practice for platforms and the advertising sector with the following specific objectives:

1. Significantly improve the scrutiny of advertisement placements, notably in order to reduce revenues for purveyors of disinformation, and restrict targeting options for political advertising;

2. Ensure transparency about sponsored content, in particular political and issue-based advertising; this should be complemented by repositories where comprehensive information about sponsored content is provided, such as the
actual sponsor identity, amounts spent and targeting criteria used. Similar mechanisms should be put in place so that users understand why they have been targeted by a given advertisement;

3. Intensify and demonstrate the effectiveness of efforts to close fake accounts;

4. Facilitate users' assessment of content through indicators of the trustworthiness of content sources, based on objective criteria and endorsed by news media associations, in line with journalistic principles and processes, transparency regarding media ownership and verified identity;

5. Dilute the visibility of disinformation by improving the findability of trustworthy content;

6. Establish clear marking systems and rules for bots and ensure their activities cannot be confused with human interactions;

7. Empower users with tools enabling a customized and interactive online experience so as to facilitate content discovery and access to different news sources representing alternative viewpoints; provide them with easily accessible tools to report disinformation;

8. Ensure that online services include, by design, safeguards against disinformation; this should, for example, include detailed information on the behaviour of algorithms that prioritise the display of content as well as development of testing methodologies; and

9. Provide trusted fact-checking organisations and academia with access to platform data (notably via application programming interfaces), while respecting user privacy, trade secrets, and intellectual property; this will enable them to better understand the functioning of related algorithms and better analyse and monitor disinformation dynamics and their impact on society.

In April 2018, the Commission convened a multi-stakeholder forum to draft a Code of Practice addressing the objectives set out in the Communication. The Code was open for signature and has been in operation since October 2018. The Code made the objectives of the HLEG “operational” through the development of principles and commitments set out under five pillars:

1. Need to scrutinise ad placements;

2. Need to increase the transparency of political and issue-based advertising;

3. Need to ensure the integrity of the services provided by online platforms;

4. Need to empower consumers; and

5. Need to empower researchers.
The Commission Communication called upon platforms to decisively step up their efforts to tackle online disinformation and considered that self-regulation could contribute to these efforts, provided it is effectively implemented and monitored. The Code was thus adopted following a self-regulatory approach. As part of the Code’s Annex, the platforms provided examples of measures already in place before Code’s implementation that could be considered a good practice in the fight against disinformation. There are also a considerable number of online platforms which are not Signatories of the Code. While the Code of Practice can be considered as a precedent in unifying several online platforms, advertisers and the advertising industry in their efforts to address the issue of disinformation, there are also several initiatives at EU and national level that focus on this particular area. Other EU measures to combat disinformation include, for example, restrictive measures against cyber-attacks or sanctions for the European political parties that abuse data protection rules to attempt to influence the outcome of European elections.

The Code of Practice forms part of the wider EU agenda to tackle the issue of disinformation, as set out in the EU Action Plan against Disinformation adopted in December 2018, which presents 10 concrete actions in the following four areas:

- Improve detection, analysis and exposure of disinformation;
- Stronger cooperation and joint responses to disinformation;
- Mobilise private sector to tackle disinformation; and
- Raise awareness and improve societal resilience.

Aim and Scope of the Code of Practice

The aim of the Code of Practice is for the Signatories to actively contribute to tackling the spread of disinformation while at the same time maintaining the right to freedom of expression and an open Internet. To this end, the Signatories expressly recognise the importance of the objectives for the Code set out in the Communication, which are already quoted above.

These are overarching objectives to ensure that the Signatories can put feasible measures in place which are aligned to their business models. The overarching objectives should be reflected in the roadmaps and action plans produced by each Signatory as well as in the best practices included in the Annex to the Code.

---

62 For example, Snapchat, TikTok, Reddit, Yahoo, Seznam, VKontakte, Viber, Facebook services other than the main platform such as WhatsApp and Instagram (although it is not a signatory service to the Code, some policies that Facebook introduced also apply to Instagram) and smaller, national platforms.
63 In May 2019, the Council adopted a Decision and a Regulation establishing a framework enabling the EU to impose sanctions or other restrictive measures (e.g. asset freeze, travel ban) against cyber-attacks which constitute an external threat to the Union or its Member States. As part of the electoral package, the Council and the Parliament amended the Regulation on the European political parties and foundations at European level, which entered into force in March 2019.
On top of these, within each of the five commitment areas, the Code recognises more specific requirements. These include:

I. Scrutiny of ad placements
   a. The need to "significantly improve the scrutiny of advertisement placements, notably in order to reduce revenues of the purveyors of Disinformation".
   b. Use of commercially reasonable efforts to implement policies and processes; not to accept remuneration from, or otherwise promote accounts and websites which consistently misrepresent information about themselves.
   c. All parties involved in the buying and selling of online advertising and the provision of advertising-related services need to work together to improve transparency across the online advertising ecosystem and thereby to effectively scrutinise, control and limit the placement of advertising on accounts and websites belonging to purveyors of Disinformation.
   d. Refinement of already widely used brand safety tools to successfully continue to meet this challenge, in recognition of the nature of this content.

II. Political advertising and issue-based advertising
   a. Transparency about political and issue-based advertising should be ensured also with a view to enabling users to understand why they have been targeted by a given advertisement.
   b. Approaches to issue-based advertising developed should be reflective of the European market for political and issue-based advertising and take note of the European Commission Recommendation on election cooperation networks, online transparency, protection against cybersecurity incidents and fighting disinformation campaigns in the context of elections to the European Parliament.

III. Integrity of services
   a. Ensuring that online services include and promote safeguards against Disinformation.
   b. Before launching new services, the signatories consider implementing and promoting safeguards against misrepresentation.
   c. Review existing services to ensure that such safeguards are likewise implemented, to the extent possible.
   d. Intensify and demonstrate the effectiveness of efforts to ensure the integrity of services connected to accounts whose purpose and intent is to spread disinformation whose specifics should be assessed and determined by the relevant signatory.
e. Consistently with Article 8 of the European Convention on Human Rights, signatories should not be prohibited from enabling anonymous or pseudonymous use of accounts and services.

IV. Empowering consumers

a. Diluting the visibility of disinformation by improving the findability of trustworthy content and consider that users should be empowered with tools enabling a customised and interactive online experience so as to facilitate content discovery and access to different news sources representing alternative viewpoints, and should be provided with easily-accessible tools to report Disinformation.

b. Invest in technological means to prioritise relevant, authentic, and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels.

c. Transparency about political and issue-based advertising should reflect the importance of facilitating the assessment of content through indicators of the trustworthiness of content sources, media ownership and verified identity.

V. Empowering the research community

a. Important to "take the necessary measures to enable privacy-compliant access to data for fact-checking and research activities" and to "cooperate by providing relevant data on the functioning of their services, including data for independent investigation by academic researchers and general information on algorithms."

Commitments of the Code of Practice

Flowing from the above points, the Code of Practice then lists a set of 15 specific commitments, within the pillar structure, that the signatories agreed to implement.

I. Scrutiny of Ad Placements:

1. Relevant Signatories commit to deploy policies and processes to disrupt advertising and monetization incentives for relevant behaviours, such as misrepresenting material information about oneself or the purpose of one’s properties. These policies and processes can include, for example, the restriction of advertising services or limiting paid placements, and could potentially take place in partnership with fact-checking organizations. Such policies and processes may, as appropriate:

   a) Promote and/or include the use of brand safety and verification tools;

   b) Enable engagement with third party verification companies;

   c) Assist and/or allow advertisers to assess media buying strategies and online reputational risks;
d) Provide advertisers with necessary access to client-specific accounts to help enable them to monitor the placement of ads and make choices regarding where ads are placed.

II. Political and Issue-based Advertising

2. Signatories commit to keep complying with the requirement set by EU and national laws, and outlined in self-regulatory Codes,[10] that all advertisements should be clearly distinguishable from editorial content, including news, whatever their form and whatever the medium used. When an advertisement appears in a medium containing news or editorial matter, it should be presented in such a way as to be readily recognisable as a paid-for communication or labelled as such.

3. Relevant Signatories commit 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 elections), which could include actual sponsor identity and amounts spent.

4. Relevant Signatories commit to use reasonable efforts towards devising approaches to publicly disclose "issue-based advertising". Such efforts will include the development of a working definition of "issue-based advertising" which does not limit reporting on political discussion and the publishing of political opinion and excludes commercial advertising. Given the implications related to freedom of expression, Signatories encourage engagement with expert stakeholders to explore approaches that both achieve transparency but also uphold fundamental rights. The work to develop this definition shall not interfere with the areas covered by advertising self-regulatory organisations.

III. Integrity of Services

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. Such measures could include some of the measures in the Annex 2 to this Code.

6. Relevant Signatories commit to put in place policies on what constitutes impermissible use of automated systems and to make this policy publicly available on the platform and accessible to EU users.

IV. Empowering consumers

7. Relevant Signatories commit to invest in products, technologies and programs such as those referred to in Annex 2 to help people make informed decisions when they encounter online news that may be false, including by
supporting efforts to develop and implement effective indicators of trustworthiness in collaboration with the news ecosystem.

8. Relevant Signatories commit to invest in technological means to prioritize relevant, authentic and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels.

9. Relevant Signatories commit to invest in features and tools that make it easier for people to find diverse perspectives about topics of public interest.

10. Signatories commit to partner with civil society, governments, educational institutions, and other stakeholders to support efforts aimed at improving critical thinking and digital media literacy.

11. Signatories commit to encourage market uptake of tools that help consumers understand why they are seeing particular advertisements.

V. Empowering the research community

12. Relevant Signatories commit to support good faith independent efforts to track Disinformation and understand its impact, including the independent network of fact-checkers facilitated by the European Commission upon its establishment. This will include sharing privacy protected datasets, undertaking joint research, or otherwise partnering with academics and civil society organizations if relevant and possible.

13. Relevant Signatories commit not to prohibit or discourage good faith research into Disinformation and political advertising on their platforms.

14. Relevant Signatories commit to encourage research into Disinformation and political advertising.

15. Relevant Signatories commit to convene an annual event to foster discussions within academia, the fact-checking community and members of the value chain.
CHAPTER 5: STUDY FINDINGS

This section presents the findings based on the analysis carried out by the research team following the methodology established in the evaluation plan annexed to this report. Firstly, some overall study findings are presented, followed by an assessment of the effectiveness of each pillar and lastly, the findings for the other evaluation criteria.

This section provides a general overview of stakeholders’ opinions regarding the Code of Practice on disinformation and its implementation. With the exception of the monthly and annual reporting published by the Signatories, there are very few written sources assessing the implementation of the Code. In line with the evaluation plan, the information presented below is based on semi-structured interviews with stakeholders, on the two online surveys deployed for this study, and on evidence and material gathered for the four case studies.

Overall findings

The fact that the Code exists (as a first initiative to combat disinformation) was identified as an achievement in its own right by all stakeholders, except for two outliers – one civil society and one business association. However, the views on the extent of implementation of the Code and its effectiveness are diverse.

Some of the Signatories and non-Signatory stakeholders credit the Code with pushing them to be more proactive in the area of disinformation as well as in monitoring and reporting on this problem more closely. From the Signatory associations perspective, the Code has contributed to a greater awareness of the issue among their members and more initiatives to combat the phenomenon.

Nonetheless, at times it was pointed out that while the behaviour of platforms is slowly changing, this change is not necessarily the sole result of the implementation of the Code but rather the result of a number of initiatives, including the Code but also including the increased dialogue and communication on disinformation in general on the national, European and global levels. While the Code might not have been the initiator of these debates, it undeniably contributed and continues to contribute to them.

Several respondents from across the various stakeholder groups mentioned that there is a need for more/better definitions in the area of disinformation. Guiding principles, for example, on what should constitute an inauthentic behaviour or a trusted information source, should be developed. Here, the stakeholders, both from Signatory and non-Signatory parties, point out that the European Commission should take the leading role and facilitate the discussion and coordination among the wide group of stakeholders on this topic.

Similarly, the Signatories as well stakeholders in general indicated that it would be useful if a consensus across the Member States on what disinformation is could be reached. Currently there are several varying definitions/descriptions that need to be followed and that may be contradictory to each other. This hampers a unified approach
to the issue across the EU. Such initiative should, again, be led by the Commission by providing a common discussion forum for the MS.

When evaluating the Code, a first complicating factor when carrying out a detailed assessment is that the platforms’ self-assessment reports are not harmonised and further perceived by non-Signatory stakeholders as not user-friendly. This makes it difficult to compare the pre-Code and post-Code environments. However, going back to the fact that coordinated disinformation campaigns are still largely prevalent, it is not possible to say if consumers are more or less exposed to disinformation now than before the Code was established.65

Regarding the reporting commitment, it should be highlighted that the Signatories, particularly the platforms, went above and beyond of what was outlined in the Code. To complement the roadmaps of the Signatories on how they intend to implement the Code, a baseline report was produced by them in January 2019 to allow for comparison of the pre-Code and post-Code results and observe the development of effectiveness of the Signatories’ policies. On top of these, and in light of the then upcoming elections to the European Parliament, the European Commission asked the Signatory platforms to provide monthly reports on their effort to combat disinformation across the EU in the period leading up to the elections (i.e. January to May 2019).

Overall, these reporting requirements were met by the Signatories in terms of submission of the reports. While many highlighted the additional burden created by all this reporting and the monitoring it required, a few also mentioned that it helped them with taking stock of their policies and to see where further improvements are needed.

From the non-Signatory stakeholders’ point of view, the annual reports from the platforms are regarded as lacking enough factual details to be used to their full potential. They do not, for example, provide any indication of the quality of initiatives undertaken since signing the Code, or provide enough indicators regarding their effectiveness, or the platforms’ motivations for any policy changes. Besides having a common reporting template and an agreed set of indicators to be provided, no additional suggestions on how to improve these reports and their readability were made.

To this end, the already ongoing dialogue between the Signatories, Commission, and Sounding Board should continue. As mentioned by Signatory trade associations and Sounding Board members, the dialogue should be opened up to even more stakeholders from different areas, A careful consideration should be given to the way this dialogue would be structured to ensure a constructive cooperation.

65 For instance: In the run-up to the 2019 European elections (March-May), Avaaz reported a total of almost 700 suspect pages and groups to Facebook, which were followed by over 35 million people and generated over 76 million "interactions" (comments, likes, shares). Facebook has taken down 132 of the pages and groups reported, accounting for almost 30% of all interactions across the reported networks. Together the pages taken down reached 762 million estimated views over 3 months. Source: https://avaazimages.avaaz.org/Avaaz%20Report%20Network%20Deception%2020190522.pdf?slideshow.
With regards to **areas for improvement**, the following suggestions were raised:

- The Code should improve its common approach to terminology, interpretation of commitments, monitoring and reporting of the implementation; have clearer commitments with few of them being more targeted towards individual Signatories and; develop measurable KPIs as the current shortcomings make it difficult to monitor and evaluate the process.

- There is also no mechanism for action in case of non-compliance, besides the annual review by the Signatories and the independent assessment by a third party. Therefore, a mechanism for action in case of non-compliance of the Code’s Pillars could be considered.

- The Code is an approach “a la carte”, as the Signatories can choose to sign up for some commitments and not for the others. Because of this, the number of commitments could be expanded to include more areas and to provide the possibility of more targeted commitments or innovative approaches. These should also be discussed with other platforms and associations that have not signed the Code yet to provide them with more incentives to join the Code as they could suggest commitments aligned with their business models.

- A regular communication (such as a newsletter) could be set up, sending short updates to all Signatories on the progress of platforms to implement commitments of the Code or on the work done around the Code as well as distributed to a larger audience connected to the issue to increase the transparency and collaboration within the Code.

- The Commission should evaluate what more could be done to attract further players among the Code’s target groups. For example, it should be considered whether big advertisers should be invited to sign the Code to strengthen its legitimacy, even if they are more likely to already have internal guidelines to avoid disinformation.

The main draw for consumers of social media to a platform is the range of voices and the authenticity that comes with a platform designed to encourage free expression, however this leads to a dilemma. Namely, any attempt to clamp down on disinformation on social media is likely to restrict the so-much appreciated breadth and vibrancy of debate potentially leading to conflicts with freedom of expression. It should be noted that the Code of Practice already acknowledges this: ‘The aim of the Code of Practice is for the Signatories to actively contribute to tackling the spread of disinformation while at the same time maintaining the right to freedom of expression and open Internet (2.4 Aim and Scope of the Code)’ and that freedom of expression is not an absolute right. However, the study team notes that it is still crucial for social platforms to find the correct balance between freedom of expression on the one hand and fighting disinformation on the other, if they want to maintain a sustainable business model for the future.

In order to fully comprehend the information supplied in the two surveys, (National Regulatory Authorities and one with the expert public), it would be useful to have a view of what the general public, in their role as consumers, thinks of the implementation of the
Code of Practice. However, this exercise was outside the scope of this study; future studies may wish to consider the feasibility of collecting data from the general public on this point.

Lastly, regarding a mechanism to address any potential issues of non-compliance, the Commission should consider proposals for co-regulation within which appropriate enforcement mechanisms, sanctions and redress mechanisms should be established.
Effectiveness

This section provides the synthesised findings concerning the Effectiveness of the Code. Study findings are presented according to each of the Pillars.

Pillar I: Scrutiny of ad placements

Under the Code, the platforms recognised the need to “significantly improve the scrutiny of advertisement placements, notably in order to reduce revenues of the purveyors of Disinformation.”

Considering the commitments for this Pillar (see 4.3), the evaluation question that this section will address is:

- To what extent did the signatories deploy policies and processes to disrupt advertising and monetisation incentives for relevant behaviours?

In the annual self-assessment reports, Facebook outlined its general ad review procedure, although the report itself does not note any specific new activities or policy updates that have been implemented in response to the Code. Google also outlined its various existing ad policies: Misrepresentation policy, AdSense Misrepresentative content policy, Insufficient original content policy, AdSense Valuable inventory policy, Inappropriate content policy. It does not go further in stating efficacy as the only figures it provides with regards to ads suspended or removed are abstract and not couched in any meaningful context. The same can be noted for Microsoft, which has various ads policies already in place before the Code was signed and did not point to any new developments in this area since signing the Code. All platforms, therefore, generally stated in their reporting that they already had policies in place to ensure appropriate placement of ads, and that when they find an ad has broken their content guidelines, they have processes in place to remove it. While this is of course positive, it cannot be said that the Code itself has had a significant impact on the platforms in this regard. It does not appear to have inspired them to adapt their ad scrutiny policies any differently than before the Code was signed.

One weakness of this pillar is that it has not resulted in the platforms sharing any meaningful data with regards to disinformation and scrutiny of ad placements, something that the Commission repeatedly noticed during the monthly reporting and that continues to lack progress. Sharing these data would have allowed the Commission and national regulatory agencies to contextualise how their existing policies were affecting the overall levels of disinformation and what the platforms were reporting. For example, in the case of Google’s annual report, if more granular detail was provided on the total number of ads which violated the misrepresentation policies in each Member State, including the

---

67 Facebook, Facebook report on the implementation of the Code of Practice for Disinformation, September 2019.
68 Microsoft, Microsoft self-assessment and report on compliance with the EU code of practice on disinformation, October 2019.
actions Google took after discovering the violation and whether the ads were taken down or remained live, it would allow for a better evaluation of the Code’s effectiveness in this area. Without the provision of this data, any in-depth assessment of the effectiveness of the Code in this area will be limited.

Having said this, the National Regulatory Agencies (NRA) themselves perceived the Code to be quite effective in this pillar, with 85% of respondents saying it had either been ‘very’ or ‘somewhat’ effective. For the Expert Public the picture was more mixed, with 64% of respondents rating it as ‘very’ or ‘somewhat’ effective (Figure 3). It should be noted that both the NRAs and expert public often conflated the ‘scrutiny of ad placements’ pillar with the ‘political and issue-based advert’ pillar in written submissions.

Figure 2: NRA Responses to question on Pillar I effectiveness

![NRAs: How Effective do you Consider the Code in Pillar I - Scrutiny of Ad Placements?](image)

Source: Study team based on survey data (n=13)

One crucial difference between the two surveys is that 9% of expert public respondents see the Code as being highly ineffective in this area, whereas none of the NRAs saw this being the case.

---

69 Google, EC EU Code of Practice on Disinformation: GOOGLE ANNUAL REPORT, p7.
The interview feedback regarding the implementation of commitments under this pillar has also been diverse. Paid content advertising is generally seen as less of a problem, given that disinformation is perceived to be spread more widely through non-advertising and organic content than advertising content, although there is an admittedly grey area between these two forms with regards to boosted content. In addition to this, interviewees mentioned that this pillar is focused too narrowly on commercial actors and monetisation, while malicious non-commercial (e.g. state) actors, including third country actors, who are not covered by the scope, may in fact have a more significant impact on disinformation. However, these elements of disinformation do fall within the realms of pillars II, III and IV, which shows that, even within an engaged expert public, there exists an inconsistent understanding regarding how the Code itself is structured, which in turn limits the extent to which it can be deemed effective. It also suggests that the Code’s voluntary nature has resulted in less overall interest from key stakeholders and the general public, which may have affected resulted in less public pressure on the platforms and limited the extent to which the signatories felt compelled to be proactive in following the commitments, despite the knowledge that the Commission itself may then take a regulatory approach in the future.

While the platforms’ advertising policies are seen by some of the interviewees to be very effective when it comes to scrutiny of general advertisements, malicious click baiting is regarded as a key weakness of this pillar. There is indeed a consensus that clickbait is not equal to the dissemination of disinformation, however it is clear that articles which are promoted via click baiting are also more likely to contain disinformation and it is therefore relevant to consider the pillar’s interaction with clickbait.\textsuperscript{70} Since click baiting can indeed exist simply as a method of traffic generation, without contributing to the spread of disinformation, the platforms frequently state that they are not able to provide data that disaggregates click bait along these lines. The capturing of this data however

\textsuperscript{70} Chris Marsden and Trisha Meyer, Regulating Disinformation with artificial intelligence,
remains a key part of stopping the spread of fake news, and the Code lacks any mention of it.

One issue that was raised by stakeholders was that of monetisation algorithms, ad placement and ‘filter bubbles’. The term was coined by Pariser (2011) and research on the topic has been furthered by the work of Sunstein (2017), and several others. Filter Bubble Theory argues that users are rarely confronted with opinions differing from their own, which discourages balanced debate and is an ideal environment for fake news to flourish. There is strong evidence to support this theory, for example the fact that 70% of the YouTube videos users watch are suggested to them by the ‘recommended video’ algorithm, powered by Google.71

It should be noted that there has been research refuting the ‘filter bubble’ theory, such as work by Axel Bruns72 however further research continues to reveal a broad range of facets which suggest some sort of filter bubble exists. For example, a recent study by computer scientists at North-eastern University and the University of Southern California sought to force Facebook to show posts to users not already aligned with the ideology of the advertising; when they did, the cost of the advertising rose.73 This suggests that the algorithms themselves are not just open to abuse by purveyors of disinformation, but that they also provide financial incentives for advertising polarised views when these views match with the users.

To summarise the key findings under this pillar:

• This pillar of the Code has been effective in drawing further attention to scrutiny of ad placements as a fundamental tenet of disinformation practices and provides a strong baseline on which to build
• The Code has not effectively incentivised the platforms to provide data that is detailed enough to be of use in assessing the effectiveness of their existing policies with regards to scrutiny of ad placements
• There is no consistent understanding from stakeholders as to the details and implications of the Code under this pillar.
• The Code does not have a high enough public profile in this area to put sufficient pressure for change on platforms in this area.
• Future iterations of the Code could ideally refer to click baiting as a tool used in disinformation and specifically ad placements.
• The Code should also contain clearer definitions, which should be developed under guidance from the Commission. For example, a clearer definition of what “commercially reasonable efforts” mean with regards to implementing policies and processes

• With the monetisation of advertising being a central tenet of the platforms’ business models, this pillar should carry more weight in the way the Code is structured and presented.

• Minimum data reporting requirements should be mandatory for platforms in this area.

In view of the above, this area of the Code leaves room for improvement and assessing its effectiveness in this area will perhaps require more time; however it is important to emphasise that the provision of these commitments in the first place is a huge step forward in terms of preventing the spread of disinformation online. Despite some discussion that this the least clear-cut area of the Code in terms of the areas it covers, it is the most crucial on the most basic level: interrupting revenue flows for purveyors of disinformation who are motivated by financial gain. The largest success of the Code in this area at this stage therefore is strengthening the policy link between scrutiny of ad placements and disinformation, on a national and EU level, and setting up a knowledge base with which regulators can make balanced and meaningful obligations for Platforms with regards to data reporting, where data must be sufficiently contextualised and not presented merely in its abstract form.

**Pillar II: Political and issue-based advertising**

This section evaluates the extent to which the Code has been effective in increasing transparency of political and issue-based advertisements. This includes providing users with tools to help them understand why they have been targeted by a specific advertisement and combatting disinformation campaigns in the context of elections.74

In line with the commitments outlined in Chapter 4, the evaluation questions for this pillar focus on are:

• Did the Signatories ensure that all advertisements were clearly distinguishable from editorial content, including news, and labelled as a paid-for communication?

• Did the Signatories publicly disclose political advertising, including sponsor identity and amounts spent? Did the Signatories deploy reasonable efforts to develop approaches to publicly disclose "issue-based advertising", including a working definition of "issue-based advertising"?

After signing the Code, the platforms all brought forward measures to increase the transparency of political advertisements to varying degrees. Facebook, Google and Twitter all released searchable ad libraries in the months following the signing of the Code. Facebook launched their ‘ads transparency tools’ in late March 2019, which were expanded in November. Google launched the ‘Transparency Report’, which includes a searchable library, in May 2019 and Twitter expanded their Ads Transparency Centre, which had been available in the United States since 2018, to Europe in March 2019.

---

Twitter also updated and expanded its political advert policy in March 2019 for the European elections. This new policy required registration proof, EU Government ID verification, verification of EU VAT or company number (if applicable) then a letter with an access code was posted to the applicant’s address to confirm location within the EU. The platforms also updated their political advertising policies. Twitter, for example, banned advertising by state media, which has since been expanded to a platform-wide ban on political advertising in November 2019. It also introduced an authorisation procedure for what it calls ‘caused-based advertising’, that seeks to “educate, raise awareness, and/or call for people to take action in connection with civic engagement, economic growth, environmental stewardship, or social equity causes”.75

Google implemented a new EU political content ads policy in time for the elections to the European Parliament, which required verification and mandated in-ad ‘paid for by’ disclosures.

In April 2019, Microsoft updated its advertising policies to prohibit political ads globally, including on LinkedIn. This also covers issue-based advertising or using political ‘hot topics’, regardless of whether the ad has a political agenda.76 Facebook implemented mandatory disclaimers for political and issue-based adverts and went further on the commitment of issue-based advertising than any of the other platforms by listing specific topics that it requires verification for. These include: Civil and social rights, economy, crime, environmental politics, immigration, political values and governance and security and foreign policy.77

On the topic of verification, Facebook came into conflict with the EU institutions over its policies for the 2019 European Elections. Given the cross-border nature of the elections, local verification of campaigns was deemed to be an inappropriate verification method that was not suited to the cross-border nature of campaigns.78 It is therefore important to note that there are pros and cons to the actions taken in response to the Code’s commitments and any future iterations of the Code, or regulatory system, must take into account the unique nature of the EU’s political environment.

The platforms have all ensured that, where applicable, political ads are also labelled as ‘sponsored by’ and clicking the label reveals information regarding who has funded them. Critics have complained that these labels, however, are not always complete, and the amount of information contained in them has largely to do with how much the sponsor has shared with the platform.79 Specific details regarding the extent of the information that should be available to users however was not an explicit commitment of the Code, in this regard the Code only describes how sponsor identity and amounts spent ‘could’

76 Microsoft, Microsoft Self-assessment and Report on Compliance with the EU Code of Practice on Disinformation, 2019, p6.
77 https://www.facebook.com/business/help/1838453822893854
79 Gaia Giombelli and Erica Melloni, Country Case Study: Italy, Virtual Insanity, P8.
be included. This is a major flaw in implementation and transparency, but also in the framing of the Code itself.

One area where all the platforms noted difficulty in progressing was in developing a working definition of ‘issue-based advertising’. This is a key weakness when assessing the Code’s effectiveness and is looked at in one of the case studies. It was an explicit commitment for the Signatory platforms and one that has not been achieved on the terms laid out in the Code’s commitments. Facebook was the only platform that laid out defined parameters for issue-based ads. However, some have argued that the ‘paid for by’ disclaimer for political and issue-based advertising still lacks a clear outline of how Facebook understands the principles of issue-based advertising beyond outlining topics it includes within it.80

Having said this, Facebook went much further than the other platform on the question of issue-based ads and was the only platform to include them in its ad library. Google and Microsoft explicitly state that ‘issue-based ads’ as a separate category of advertisement are difficult to define, and Microsoft goes further by outlining that this prevents the formation of a workable policy in labelling or prohibiting them.81 Google does refer to ‘political issue advocacy’ while describing its political ads policy, however, it does not provide examples nor a clear definition of what this term actually means.82

Figure 4: NRA Responses to question on Pillar II effectiveness

With regards to the effectiveness of the pillar from the perspective of the National Regulatory Agencies and Expert Public, there is a distinct difference between the two.

Source: Study team based on survey data (n=16)

80 Facebook, Facebook report on the implementation of the Code of Practice for Disinformation, September 2019, p3.
81 Microsoft, Microsoft Self-assessment and Report on Compliance with the EU Code of Practice on Disinformation, 2019, p5 and Google, EC EU Code of Practice on Disinformation, GOOGLE ANNUAL REPORT, 2019 p15.
82 Gaia Giombelli and Erica Melloni, Country Case Study: Italy, Virtual Insanity, P30.
About 75% of the NRAs (Figure 4) consider it to be ‘very’ or ‘somewhat’ effective, however this figure drops to 56% for the Expert Public, with a notable decline in those rating it as ‘very’ effective. Furthermore, almost a quarter of the Expert Public viewed the Code’s effectiveness under this pillar to be ‘very ineffective’. While a larger sample size for the NRAs may have produced a different outcome, 16 regulatory authorities can still give a reliable indication.

**Figure 5: Expert Public Responses to question on Pillar II effectiveness**

![Figure 5](source: Study team based on survey data (n=32))

Despite the policies outlined by the platforms, some interviewees believed that the transparency of political and issue-based advertising took a step backwards since the Code came into force. The main evidence cited is Facebook’s recent decision not to fact check political ads, which, although not an explicit commitment of the Code, could be seen to represent a step in the wrong direction in terms of transparency, as well as monitoring and preventing the spread of disinformation more generally. One interviewee provided a recent example of this policy in action from the United States, where a political advertisement which had been shown by Facebook’s own fact checkers (PolitiFact and Factcheck.org) as spreading disinformation, was permitted to remain live and has now been watched over 5 million times.

With regard to ‘ad labels’, which all signatory platforms are currently using, it has been pointed out that Facebook’s ‘paid for by’ tool, introduced after the Code and intended to prevent disinformation and election manipulation, failed multiple tests by a variety of organisations. For example, when a paid post is shared by a user, the labelling disappears. This makes paid content look genuine and using this loophole does not require specific technological knowledge or competencies. When looking at the Code itself, this labelling issue is not in direct contravention of the commitments. However, it does weaken the overall effectiveness of the ‘paid for by’ policy as a consequence this weakens the Code itself. While none of those interviewed were able to suggest a remedy

---

83 https://www.politifact.com
84 https://m.facebook.com/pg/DonaldTrump/posts/?ref=page_internal&mt_nav=0
to this issue, it is clear that transparency would be improved if some sort of labelling remained visible after the original advert has been shared, to indicate that its original source was not organic.

Regarding the ad libraries, this is one of the initiatives implemented in response to the Code that has attracted a large amount of negative attention from the research community and regulators. A multitude of stakeholders consulted for this evaluation raised issues with functionality that do not appear to have improved in the past 12 months. Overall, it was felt that ad archives did not provide a comprehensive picture of the actual situation in political advertising. For example, one major concern is that Facebook’s library did not provide data at a large enough scale to be usable and the information cannot be downloaded in a machine-readable format or exported as a full dataset. The archives themselves also need to be more complete in order to provide accurate information and allow an assessment of what is included and what not. For example, studies have shown that Google and Twitter’s ad libraries are incomplete, with political parties being unable to find their own adverts in some cases. Facebook’s library also showed different results depending on the geographic location of the search, with the general conclusion being that all ad libraries have been developed as a ‘symbolic gesture’ with little thought being put into functionality.

Overall, there is also the perception from stakeholders that the Code, as it stands, places a disproportionate emphasis on political or issue-based advertising and should be updated to recognise that disinformation is also highly prevalent in terms of “organic” content by individual users, not just via advertising. In this sense, a stronger link between this pillar and other pillars which directly affect ‘organic content’, such as pillars III and IV, could be considered as the Code moves forward.

As is the case across this analysis, under this pillar there should also be a stronger international definition of disinformation and issue-based advertising agreed across the platforms and with oversight at the European level. Implicit within this is a need to maintain a balance between freedom of speech and other fundamental rights, such as non-discrimination.

It is important to reiterate that disinformation is not inherently political. The toolset of spreading disinformation is frequently applied to politics but the differences between the two must be made clear. In the fields of health, for example, this has been seen often. The grouping of ‘political and issue-based’ advertising under a single pillar and often conflation between the two may not be the most productive format for the Code or any future iterations.

In summary:


Ibid, P3.

Article 21 of the European Charter of Fundamental rights.
• Measures regarding political advertising have seen stronger development than issue-based advertising since the Code was signed.

• The results of interviews and survey revealed that, while efforts have been made by platforms in the area of political and issue-based advertising, stakeholders still see significant room for improvement.

• It became clear that even the experts consulted for this study have significantly different opinions on what falls within the scope of ‘issue-based’ advertising, seeing national culture as a crucial factor in deciding the scope.

In conclusion, the principles of this pillar could be further strengthened but overall good progress has been made. Some transparency measures, such as the ad libraries, could be made more usable and certain key technical issues improved. This includes dealing with the issue of ad label disappearance following organic reposting, as this is open to significant abuse by coordinated disinformation actors. Overall, the platforms should invest more in the measures they have implemented, to avoid stakeholders seeing them as simply ‘token’ gestures.

Pillar III: Integrity of services

In this section, the analysis is looking into how the Code of Practice was implemented regarding Pillar III. Integrity of Services. In line with the commitments outlined in Chapter 4, the evaluation questions for this pillar focus on are:

• To what extent have the signatories put in place clear policies regarding identity and the misuse of automated bots on their services?

• To what extent have the policies put in place by signatories been effective in identifying and deactivating automated bots?

• To what extent have the signatories put in place policies on what constitutes impermissible use of automated systems?

As detailed in their monthly and annual reporting, all the Signatory platforms have measures in place to counter manipulative and inauthentic behaviour on their services. The reports also outline how the policies implemented under this pillar continue to be re-evaluated and modified to ensure they address the latest tactics and techniques applied by purveyors of inauthentic and malicious behaviours.

While the platforms provide numbers on, for example, how many accounts were closed/removed, taken action against etc., oftentimes, these are provided on a global scale rather than focusing on the EU results. With regard to fake accounts, one justification could be that these may often be detected and taken action against within moments of creation and, therefore, their intended targets and/or aims are unknown.

While commendable that Facebook, in their annual report, lists all the coordinated inauthentic behaviour cases it detected and removed, it would be beneficial for the reports to outline how these cases were processed and whether Facebook intends to
investigate further, either internally or with the support of external researchers, to further analyse their impact, influences, or the actors involved.

**Figure 6: NRA Responses to question on Pillar III effectiveness**

![NRAs: How effective do you consider the Code in Pillar III - Integrity of services?](image)

Source: Study team based on survey data (n = 11)

With regard to the effectiveness of the pillar from the perspective of the National Regulatory Agencies and Expert Public, the two target groups are in agreement. About 46% of the NRAs (Figure 6) would say that regarding this pillar the Code has been ‘somewhat ineffective’. Regarding the expert public, this number stands at 47% (Figure 7). While the sample size may not be entirely representative of the two target groups, it does still appear that the scope of the Code is currently insufficient.

**Figure 7: Expert Public Responses to question on Pillar III effectiveness**

![Expert Public: How effective do you consider the Code in Pillar III - PolitiIntegrity of Services?](image)

Source: Study team based on survey data (n=30)

One shortcoming highlighted by the respondents is the fact that there is lack of data regarding this pillar that would allow independent verification of the information provided by the platforms. Another is that while the platforms are more transparent about computational propaganda and coordinated inauthentic behaviour, they are less effective when it comes to individual users and smaller networks.
To this end, several interviewees mentioned that platforms should provide clearer statistics on which accounts/bots/users/fake accounts were deleted and why. Some mentioned examples of granularity include: What proportion of accounts were active before they were cancelled; the reach of the deactivated accounts (e.g. how much of the fake content was seen/commented on/shared by genuine users); languages fake accounts are in, etc. This information should be provided in an aggregated manner.

To non-Signatory stakeholders, it currently seems that it has not become harder for malicious actors to spread disinformation. On the contrary, some interviewees stated that it has become easier in recent months. When asked for data/sources to indicate the perceived increase in activity of malicious actors, stakeholders were unable to provide concrete examples. However, the perceived increase in malicious activity could be attributed to the fact that now, because of the Code, stakeholders and users are now more aware of such behaviour due to increased transparency regarding the misuse of platforms by such actors rather than an actual increase of inauthentic behaviour stemming from the implementation of the Code.

Society is now in a second stage of disinformation, where purveyors of disinformation are becoming more aware of the efforts to stop them and are adjusting their behaviour accordingly, e.g. deep fakes, mirroring IP ranges, randomising specific patterns to provide a few examples. This does not result in less disinformation online, but more organised and complex forms of disinformation that are harder to combat. This problem is also recognised by the Signatories who are adjusting their policies on an ongoing basis to be able to keep up with such developments. One such adjustment includes extending the abilities of their detection systems by feeding in the information connected to identified fake and/or malicious accounts and users. The platforms even acknowledge that at times, rather than blocking fake accounts, they might leave them active. This is to monitor their activity, see the impact they might have and learn from the observed behaviour.89

Regarding the monetisation and impact of inauthentic behaviour, a recent NATO StratCom publication90 outlines an experiment where StratCom, in order to test the ability of four social media platforms (Facebook, Instagram, Twitter and Youtube) to identify and remove purchased services (i.e. comments, views, likes) from social media manipulation service providers. While the numbers produced by the experiment were quite alarming, it needs to be highlighted that the experiment consisted of purchasing engagement on posts older than six months and that had a neutral message. Therefore, they were selected on purpose to avoid impacting genuine engagement. This was intended to counter that fact that the added advantage of social media is often that it allows immediate interaction. It is therefore challenging to correlate the impact of the experiment with more “current and/or controversial posts”.

89 Information provided during interviews with Signatory platforms.
90 NATO StratCom Centre of Excellence (2019). Falling Behind: How Social Media Companies are Failing to Combat Inauthentic Behaviour Online. Available at: https://www.stratcomcoe.org/how-social-media-companies-are-failing-combat-inauthentic-behaviour-online
This is also highlighted in the position of platforms when assessing the report’s results. Since the majority of the bought engagement that was reported was on posts that had low activity or were dormant and, therefore, had no impact on spreading disinformation. While this should not diminish the report’s results and the issue it brings an attention to, it also needs to be put into perspective given that at least 10 billion items are shared across social media platforms on a daily basis. At the same time, the data that are reported by the Signatories under this Pillar are not proportionate to these numbers.

To this end, there is a need for continuous discussion on metrics across the various stakeholders. For example, it makes no sense to provide a ratio of how many fake accounts were closed in relation to the total number of fake accounts on an individual platform as at any point in time as the platforms cannot know the exact number of fake accounts. A possibility would be to monitor this retro-actively and to monitor when each detected fake account was created, for how long it was active and what was its reach of genuine and/or fake accounts/users. Another possibility would be to establish a list of common criteria focusing on the difficulty of creating a fake account that could be measurable across the social media platforms. In this regard, the methodology applied by the StratCom report could serve as a guide to creating such an approach.91

To summarise the main findings under this pillar:

- The non-Signatory stakeholders seems to be less informed about the developments and efforts put forward by the Signatories;
- The platforms have many tools and policies in place to combat inauthentic behaviour. Due to the nature of these malicious actors, oftentimes these tools and policies are reactive rather than pro-active as the platforms cannot always predict the new developments in this area;
- For the same reason, it is hard to directly assign the development and improvement of tools and policies to combat inauthentic behaviour to the implementation of the Code;
- Due to the reactive nature of this pillar, policy makers (both at European and Member State level) might do well to proceed with caution when proposing regulation and related initiatives as the regulatory process is unlikely to keep up with technological developments. To this end, any regulation in this area should take a form of principles or general objectives to be achieved giving the platforms the possibility to implement these via the most appropriate means in line with their business models. Nonetheless, further cooperation and exchange of information among the impacted actors should be further promoted and supported whether there is a regulatory action or not.

91 In order to perform their experiment across the selected platforms, the StratCom team devised seven criteria against which the platforms’ abilities to counter the malicious use of their services would be compared. The criteria included, for example, the success in blocking the creation of inauthentic accounts, ability to remove traces of inauthentic accounts or responsiveness to reports of inauthentic activity.
While the platforms do have tools and policies in place to reduce inauthentic behaviour, the focus is predominantly on preventing the creation of fake accounts/users and on detecting and deleting them. The status of such accounts/users (active, dormant, inactive etc.) and the reach of their activities is less known and reported on. In this regard, the platforms might already monitor this type of information and make use of it when developing their tools. Instead they may prefer to keep this information private so as to not give an advantage to malicious actors. Nonetheless, to provide a better picture of the actual situation regarding the inauthentic behaviour and malicious actors, the platforms should provide data on the proportion of active accounts from the total number of removed fake accounts/users, the reach and influence of such accounts on genuine and other fake users as well as the number of fake accounts/users each platform is monitoring.

While the StratCom report might have focused on low activity content, similar studies and experiments should be promoted and supported by the Signatories as well as by policymakers (both at European and national levels) and other independent entities. The partnerships with researchers could be extended to also include this area. The findings for these types of studies would help understand the behaviour of malicious actors better and would contribute to the improvements of platforms’ algorithms and tools designed to fight such actors.

**Pillar IV: Empowering consumers**

In this section, the central question is to what extent the implementation of the Code of Practice on Disinformation has been effective in the area of its fourth pillar (empowering consumers).

In line with the commitments outlined in Chapter 4, the evaluation questions for this pillar focus on are:

- To what extent are the Signatories investing in products, technologies and programmes to help people make informed decisions when they encounter online news that may be false?
- To what extent are the Signatories investing in technological means to prioritise relevant, authentic and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels?
- To what extent are the Signatories investing in features and tools that make it easier for people to find diverse perspectives about topics of public interest?
- To what extent are the Signatories partnering with civil society, governments, educational institutions, and other stakeholders to support efforts aimed at improving critical thinking and digital media literacy?
- To what extent are the Signatories encouraging market uptake of tools that help consumers understand why they are seeing certain advertisements?
Together with the fifth Pillar (empowering the research community, see below), this pillar is where the perceptions of Signatory and non-Signatory stakeholders differ the most. To start with the position of the National Regulatory Authorities, who belong to neither category and can be seen as the most independent category of stakeholders, a clear majority thinks that the Code is either very or at least somewhat effective for this Pillar, see Figure 8.

**Figure 8: Results from the survey for National Regulatory Authorities**

![Pie chart showing results](source: Study team based on survey data (n=14))

On the other hand, the expert public is much less happy with the effectiveness of the Code for this Pillar, as only one quarter of the respondents states that they think that the Code is either very or at least somewhat effective for this Pillar, see below. When considering the relevance of the Code for this Pillar, only 17% of the experts expressed that they feel that the policies put in place by the Signatories to empower consumers are taken as a result of the implementation of the Code of Practice and 35% states they feel this is not the case (48% expressed that they do not know, N=23).

**Figure 9: Results from the survey for the expert public**

![Pie chart showing results](source: Study team based on survey data (n=30))

To provide an insight on the opinion of the general public, who are the users of the platforms, a report from the Digital News Project has shown that social media is less
trusted by citizens than other news media in its ability to separate fact from fiction; 24% mentioned that social media is able to do so versus 40% for news media in general. On the one hand, the substantial minority that does trust social media mentions this is because of the broad range of views that have a place on it and the authenticity of the platforms.

On the other hand, the majority that does not trust social media mentions that this is the case because of the platforms’ biases and agendas (i.e. there is a sense from respondents that feeds are becoming polluted with inaccurate information, extreme agendas, and strong opinions, perhaps encouraged by social media algorithms). They also point towards other social media users for fuelling this phenomenon by sharing content without critically assessing it.92

To implement the commitments under Pillar IV, most platforms currently have some consumer empowerment features in place:

- Facebook has a context button which appears alongside links shared on its News Feeds and which is designed to provide more background and information on the publishers of the content with the aim of letting people decide for themselves what to read, trust and share. Moreover, since November 2019, Facebook is labelling state-controlled media as such based on its own definition and standards with input from more than 40 experts around the world. Lastly, they participate in the organisation of several campaigns and workshops to enhance the media literacy of citizens.93

- Google has features such as Breaking News and Top News to ensure the prominence of authoritative content. YouTube only surfaces content from authoritative sources on its Breaking News and Top News shelves and it further provides context via information panels in the search results for certain historical and scientific topics that have often been subject to disinformation. Lastly, they participate in the organisation of several campaigns and workshops to enhance the media literacy of citizens (most noteworthy, as of September 2019, they claim to have trained over 1 million people in person across the European Union through their Be Internet Citizens and Be Internet Awesome programmes) and they hosted the Global Media Literacy Summit, a full day event with over 180 delegates from 32 countries in London.94

- Twitter has the ‘verified accounts’ function which means that celebrities, journalists, news organisations and politicians have verified accounts on Twitter (signalled by a blue badge with a white ‘v’ next to their name), which authenticates that the information coming from those accounts is in fact coming from those individuals. Twitter further has a feature called “Top Tweets” but it is unclear how content is ranked in the timelines of users choosing “Top Tweets,” instead of

93 Facebook Monthly Self-Assessment Reports, Code of Practice on Disinformation, 2019.
chronologically ordered Tweets. As Twitter has not subscribed to Commitment 8 (on prioritisation of authoritative content), it is difficult to assess the platform’s performance in this area. One recent update is the introduction of “Topics” which allows users to follow conversations about a topic like how one would follow an account with one single tap. Like with Top Tweets, Topic suggestions will appear in the timeline and in search results based on what users tend to look for and already follow on Twitter. This helps to expose users to more views and conversations around topics, rather than single voices from one account. However, it should be noted that it would be better if users would actually be exposed to alternative viewpoints as well rather than ones based on what they follow already to prevent thinking in bubbles too much.95

- Mozilla included in its Milestones for the Implementation of the Code of Practice on Disinformation several tools linked to Pillar IV, most importantly a rollout of enhanced security features in the default setting of Firefox which highlights the quality of websites and provides other information about the website relevant to empower consumers.96

- Microsoft has a partnership with NewsGuard, which reviews online news sites across a series of nine journalistic integrity criteria. As of the end of August 2019, NewsGuard had rated the 2,805 websites responsible for 96.01% of the news and information consumed and shared online in the United States. Of those sites (many of which are also popular in the EU), 834, or 29.7%, were given an overall Red rating.97

The most important general conclusion of this pillar is that the reports are generally unclear on the extent to which the tools are available across EU Member States in local languages. The assessment of the study team based on desk research looking into these tools is that the tools are mostly not available across all EU Member States.98 Moreover, the platforms’ reporting is not detailed enough to assess the relevance and impact of the consumer empowerment tools in place. Specifically, information on the uptake and actual use of these tools is lacking.

Most interviewees, coming from all the stakeholder groups, did see a limited increase in consumer awareness on the topic of disinformation. However, often they did not see a (causal) relationship with the Code of Practice as the basis of this development and they rather put it down to the increased (media) attention on the topic of disinformation, often referred to as fake news, during recent months. However, a counterargument could be made pointing towards the fact that the phenomenon of disinformation might be in the spotlight precisely because of the Code.

The assessment of the study team is that Pillar IV should be considered as an area where policy makers (both at European and Member State level) have the most

95 Twitter Monthly Self-Assessment Reports, Code of Practice on Disinformation, 2019.
96 Mozilla Milestones for the Implementation of the Code of Practice on Disinformation.
98 It differs from tool to tool where, but most of these are only available in around five Member States, usually the only the larger and more digitally advanced ones.
opportunities to step in and propose regulation and related initiatives. The main reason for this is that there are very few economic incentives for the platforms to do so themselves without being pushed. The ultimate goal, admittedly an ambitious one, would be to ensure that consumers firstly, get access to better quality information (limiting their exposure to disinformation), and secondly, when they do get exposed to disinformation that they know this is the case (both through flagging the information as such and the better awareness of the consumer itself through training and education).

A key objective seems to be to ensure that users get to see the highest quality and most relevant content first. In order to do so and to improve the implementation of the commitments under Pillar IV, platforms need to work with publishers, fact-checkers, and other content creators to better label the trustworthiness of different kinds of content. For instance, as suggested by stakeholders connected to the Sounding Board in interviews, platforms could engage more with traditional media to develop transparency and trustworthiness indicators for information sources (which falls under Commitment no. 7), which can then be used to feed content ranking algorithms, eventually providing users with access to a plurality of credible information sources. A good practice example that can be mentioned in this regard is the Trust Project which is a consortium of top news companies led by an award-winning journalist which is developing transparency standards that help consumers easily assess the quality and credibility of journalism. Several platforms (i.e. Google, Facebook and Bing) are involved in the project.99

One specific solution suggested by interviewees from traditional media organisations to achieve this is to “push up” content from so-called “trusted information providers” which relates to Commitment Number 8 of the Code. However, the platforms pointed towards the difficulties in distinguishing such outlets as well as the fact that not all content published by these outlets is necessarily trustworthy (e.g. even trusted information providers can publish click bait content). According to the platforms, a definition is not the only thing needed of these trusted information providers, but also a body that defines which outlets comply with these criteria. Indeed, the interviewees from the traditional media organisations did not manage to define trusted information providers and their preference for this solution might be motivated by their own presumed status as such.

An alternative would be to design a set of guidelines to which content would be compared in order to be “pushed up.” Although this solution is already more precise than the one mentioned above, there are also several practical problems linked to this relating first of all to its higher costs and secondly to questions relating to who will be responsible for designing these guidelines as well as enforcing them. A good practice that could be mentioned to establish which content could merit to be pushed up are the standards developed by Reporters sans Frontieres under the Journalism Trust Initiative.100

To summarise the main findings under this Pillar:

99 For more information, see: https://thetrustproject.org/.
100 For more information, see: https://jti-rsf.org/en/.
• Together with the fifth pillar (empowering the research community, see below), this pillar is where the perceptions of Signatory and non-Signatory stakeholders differ the most;

• Where platforms and national regulatory authorities seem to be rather positive about the improvements made over the last year, academics, experts and Sounding Board members are much more critical;

• Most of the platforms have many tools in place to work on empowering consumers. What could really be improved is the consistent roll-out of these tools across all Member States and consistent reporting on the impact of these tools;

• Most stakeholders did see an increased awareness of consumers on the topic of disinformation. However, it is difficult to establish the degree to which this is a direct effect of the Code of Practice;

• Several good practises (e.g. the Trust Project and the Journalism Trust Initiative) should be further developed and more widely implemented. This can then act as a minimum standard for all platforms to live up to further down the line; and

• Pillar IV should be considered as an area where policy makers (both at European and Member State level) have the most opportunities to step in and propose regulation and related initiatives as there is not always a clear economic incentive from platforms to do this themselves.

The study’s finding is that platforms do have the consumer perspective in mind when designing their platforms and its tools as their entire business model is aimed at creating a service that is as attractive as possible for consumers. However, this does not automatically lead to consumer empowerment as the possibilities for this are not always know, and sometimes not even desired, by the (majority of the) consumers. There is thus a gap where policy makers could step in to ensure that consumer empowerment is secured to the largest extent possible. It is especially important to streamline the various tools across all Member States and where possible, also across platforms, to avoid differentiated levels of consumer empowerment depending on the economic viability to create this.

**Pillar V: Empowering the research community**

In this section, the aim is to establish whether the Code has been effective in empowering the research community and fact-checkers. In line with the commitments outlined in Chapter 4 the evaluation questions for this pillar focus on are:

• To what extent are the Signatories supporting good faith independent efforts to track disinformation and understand its impact?

• To what extent are the Signatories not prohibiting or discouraging good faith research into Disinformation and political advertising on their platforms?

• To what extent are the Signatories encouraging research into disinformation and political advertising?
• To what extent are the Signatories convening annual events to foster discussions within academia, the factchecking community and members of the value chain?

From the Signatories’ point of view, the relationship between platforms and researchers is perceived as good, even though they note that there is still room for improvement. Signatories interviewed agreed that independent analysis led by researchers is a key step toward promoting shared understanding of the threats related to foreign information operations.

According to the annual self-assessment reports drafted by Signatories of the Code, the platforms all implemented several policies and tools intended to provide researchers and the fact-checking community with access to platform data. This includes, in particular, access to the repositories of political ads, a resource that did not exist in the EU in 2018.

To facilitate the access to data for researchers, new datasets have been made available by platforms. Twitter for example disclosed a significant archive of state-backed information operations on Twitter in October 2018. Other datasets were made available in January, June, August and September 2019, providing access to more than 30 million Tweets. It has been reported that researchers in 15 EU countries accessed these datasets over 20 thousand times.101

The platforms also launched partnering programmes to combat disinformation. It is the case, for example, of the Mozilla Foundation which launched joint campaigns on transparency involving 71 researchers and 37 civil society organisations.102 Facebook reported that in April 2018 it launched a partnership with Social Science One (SS1)103, a group of 83 academic researchers, to share data with the academic research community while maintaining stringent privacy protections.104

Microsoft also implemented partnering programmes with researchers (TAP), research institutions (Princeton University, Oxford Internet Institute) and with industry, including the participation of Bing News in the Trust Project, a consortium of top news and digital companies that aims to make it easier for the public to identify quality news.105 In collaboration with the International Fact Checking Network, Google News Lab launched FactCheck EU106 in March 2019 to provide fact checks from 19 organizations from 10 countries in 11 languages. Google also worked with First Draft107 to provide training boot camps for journalists in Frankfurt, Brussels, Milan, and Madrid. Furthermore, Google also introduced new tools for researchers and the fact checking community: a ‘Fact Check Explorer’,108 which allows for exploration of Fact Checking journalism, and the ‘Fact

101 Twitter Self-Assessment Report, Code of Practice on Disinformation, September 2019
102 Mozilla Self-Assessment Report, Code of Practice on Disinformation, September 2019
103 https://socialscience.one/
104 Facebook report on the implementation of the Code of Practice for Disinformation Annual Report, September 2019
105 Microsoft Self-Assessment Report, Code of Practice on Disinformation, September 2019
106 https://factcheckeu.info/en/
107 https://firstdraftnews.org/
108 https://toolbox.google.com/factcheck/explorer
Check Markup Tool’ ¹⁰⁹, which allows fact checkers to easily mark their own articles as fact-checks in a way that is machine-readable via the ‘ClaimReview’ mark-up.

Events with the research community have also been organised by the platforms to foster discussions on disinformation. For example, it was reported that to help newsrooms and journalists prepare for the European Elections, Facebook’s News Partnership team conducted 10 training events across the EU from March through May 2019. The team trained over 400 journalists on how to tell digital stories on their platforms, how to spot false news and how to maintain integrity in digital reporting.¹¹⁰ Another example of events sponsored by one of the Signatories is the News Impact Summit, which took place in Lyon in November 2019, and which was powered by the Google News Initiative. Experts presented how to make the most of digital tools to cover elections, by sharing experiences that ranged from creating new apps and delivering election results in real time to smartphones, to experimenting with platforms that flag and debunk misinformation during elections.¹¹¹

Several survey respondents stated that, thanks to the Code of Practice, researchers have access to more databases and tools such as “Crowdtangle”, launched by Facebook¹¹²,¹¹³. From the results of the survey, it appears that, even though the level of cooperation between platforms and researchers is perceived as not sufficient, it is still greater than before and the quality of the data accessible has risen. For example, Twitter’s API is an important data source for academics. Several researchers interviewed agreed that Twitter in particular made efforts regarding access to data by researchers and fact-checkers. Fact-checkers and researchers have increasingly used the few data available to conduct investigations, or report on abusive behaviour, in particular from political parties, activist groups and political candidates, as reported by a survey respondent.

It was agreed by most researchers surveyed that it is difficult to assess whether the policies put in place by platforms on the collaboration with researchers result from the Code of Practice or from a wider public pressure. They, however, also stated that the Code must have contributed to materialising the demand from fact-checkers, researchers and users.

Despite the initiatives implemented by Signatories and the increase of research noted by researchers, the vast majority of stakeholders agreed that the **pillar of the Code aimed at empowering the research community is the one which has proven to be the least advanced.** Many stakeholders interviewed or surveyed showed disappointment regarding the achievements linked to the collaboration between platforms and researchers on disinformation.

¹⁰⁹ https://developers.google.com/fact-check/tools/api/
¹¹⁰ Facebook report on the implementation of the Code of Practice for Disinformation Annual Report, September 2019
¹¹² https://www.crowdtangle.com/features
¹¹³ Crowdtangle is a platform used by media companies around the world, allowing analysts to track the popularity of news items and other public postings across multiple platforms.
Most of the survey respondents estimate that the cooperation between platform and fact-checkers/researchers is not effective (84%). They noted that there is limited engagement with the research community and that the tools set up by platforms are still too weak. Even though some stakeholders agreed that Crowdtangle is an example of a good research tool, they noted that the tool is also owned by Facebook, and they hence see a conflict of interest regarding this tool. The lack of transparency regarding the access to data is a common concern raised by the researchers interviewed.

The Ad Archives of Google and Facebook in particular are not seen as fit for purpose by some stakeholders. The data that could be extracted were deemed unreliable and it was noted that the archives were damaged with bugs, which ultimately made these tools effectively useless as a transparency tool for researchers, journalists or stakeholder for whom this data were intended. Many stakeholders hence denounced the lack of user friendliness of the platforms’ databases. To improve the user friendliness of the large ads databases, some researchers suggested that these are made downloadable, to avoid manual checking of ads.

A report from the European Commission from October 2019 summarising and analysing the self-assessments produced by Signatories confirmed the main opinion of interviewees regarding the access to data from platforms. The report states 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”. It was also noted that only a very limited community of researchers has been offered access to platforms’ data sets. In the case of some platforms, the access to data is mainly focused on the US research community. This latter concern has also been raised by several researchers interviewed.

It has been reported by many researchers that the access to data of platforms has not really improved after the creation of the Code. A vast majority of the respondents to the survey of the expert public estimated that they do not have more and easier access to privacy-compliant platforms’ data since the platforms signed the Code (74%). Some academics also raised the issue that the choice of platforms to grant access to researchers sometimes seemed arbitrary, and that this has increased distrust between platforms and the research community. In its report of October 2019 analysing the self-assessments of Signatories, the European Commission confirms that, while the platforms provide APIs to facilitate the running of queries by researchers and fact-checkers, concerns remain regarding the limited functionalities of the APIs and the searchability of the repositories given that platforms have the ability to alter or restrict access on a unilateral basis.

115 Ibid.
Regarding the partnership between Facebook and SS1, in December 2019 the Co-Chairs and European Advisory Committee of SS1 issued a public statement presenting that the initiative “has made important progress over the last 18 months, but Facebook has still not provided academics with anything approaching adequate data access.” The Co-Chairs and Committee noted that even though Facebook is the “first company to explore a significant platform-academic partnership model”, SS1’s work with the platform has been continuously delayed. The representatives of SS1 call on Facebook and other platforms to “make accurate and representative data available for scientific research into the most pressing issues of public concern”.117

Many researchers stressed that academic access to APIs should be possible, with verification steps if need be, to allow the monitoring of both real-time and past disinformation spreading. Some researchers asked that access to data is allowed with no limits nor vetoes from the side of platforms. According to these stakeholders, such an open approach would reinforce the trust between platforms and researchers. Another reason for the difficulty of accessing data is the Cambridge Analytica scandal, as platforms become more wary about accommodating requests for personal data. For a platform to be able to provide the personal data of users, they need consent from users. Several academics interviewed noted that the implementation of GDPR has been a setback for access to data as it made it more difficult for platforms to deal with.118 Few platforms interviewed agreed themselves that GDPR made it more difficult to share data. The summary report published by the European Commission in October 2019 to analyse the self-assessment of Signatories confirmed that “although views apparently differ among the platforms, the platforms generally cite alleged risks of data protection violations as inhibiting cooperation with the research community”.119 Some non-Signatory interviewees also mentioned that they had the feeling that GDPR is sometimes used as an excuse not to engage in any information exchange and that there are ways around it.

However, GDPR should not hamper platforms from sharing data with the researchers. The GDPR aims at strengthening the rights of data subjects and specifically how their personal data is used and how it is protected. As long as the principles established by the GDPR are respected by platforms and researchers, such as lawfulness, fairness, transparency of the data shared, data collected used only for the purposes explicitly specified or limiting data collection to what is necessary to serve the purposes at the time of the collection, the data can still be shared with researchers.

The platforms also sometimes find it challenging to partner with researchers. Some platforms interviewed counterargued that the requests for data from researchers are rarely unified and consistent, and that the researchers are sometimes unclear about what they expect to get out of the data. It is often stated by Signatories that the researchers

117 https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one
118 Regulation on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=FR
underestimate what can be done with available data e.g. by monitoring Search or YouTube results as many have done to assess the behaviours of platforms’ algorithms. Similarly, often when researchers ask for raw data, they do not want raw data but populated datasets which are not always available.

The case study on the relationship between academics and platforms, particularly pertaining to data requests, has argued that some oversight should be introduced on the relationship between academics and platforms and the topic of data requests. This oversight must be both regulatory, from independent regulatory authorities (convened in the European ERGA organisation) to ensure that the data requests are useful to make studies to inform policy, and academic, to ensure that all requests serve sound academic purposes. This will also improve the cooperation between academics and regulators leading to more efficient and evidence-based policy intervention. There are certain best practices, most notably SS1[^120], in this area which still need to be improved but can act as a blueprint for the relationship between academics and platforms.

The case study on the partnerships between fact-checkers and platforms concluded that further work should be done to strengthen the existing partnership and to enable the creation of more alliances combating disinformation. This should be done mainly through an increase of media literacy programmes, enabling the platforms and fact-checkers to communicate further the purposes and benefits of such partnerships to the public. An increased communication on the topic would also enable the public to better trust the fact-checking community and to increase the number of users of fact-checking tools. Democratizing simple and user-friendly fact-checking tools among all Signatory platforms could contribute to enhancing the impact of the fact-checking partnerships.

In summary:

- The relationship between platforms and researchers is perceived as good by the Signatories.
- Several survey respondents stated that, thanks to the Code of Practice, researchers have access to more database and tools.
- However, many researchers interviewed showed disappointment regarding their collaboration with the platforms.
- Researchers noted that there is limited engagement with the research community and that the tools set up by platforms are still too weak, not transparent enough and not really user friendly.
- A vast majority of the respondents to the survey of the expert public estimated that they do not have more and easier access to privacy-compliant platforms’ data since the platforms signed the Code (74%).

[^120]: [https://socialscience.one/](https://socialscience.one/)
Some academics also raised the issue that the choice of platforms to grant access to researchers sometimes seemed arbitrary, and that this has increased distrust between platforms and the research community.

Several academics interviewed noted that the implementation of GDPR has been a setback for access to data as it made it more difficult for platforms to deal with.

According to platforms, the requests from researchers may sometimes be unclear or unrealistic, which makes the collaboration more difficult. Regarding the data requests from academics to platforms, the case study has argued that more independent oversight is needed to guide these relationships, both from regulators and from senior academics.

Regarding the partnerships between fact-checkers and platforms, the case study suggested that, to improve their effectiveness, there should be more media literacy programmes, enabling the platforms and fact-checkers to communicate further the purposes and benefits of such partnerships to the public.

It is clear that efforts have been made by the Signatories to support and encourage good faith research into disinformation. The annual self-assessment reports drafted by Signatories demonstrated that all Signatories implemented several policies and tools to increase the collaboration with researchers and fact-checkers, and notably to provide them with access to platform data. The Signatories made available new datasets, launched partnering programmes with researchers, research institutions or industry to combat disinformation. They also organised events with the research community to foster discussions on disinformation.

However, the initiatives implemented by platforms to encourage the research into disinformation should be further developed to be more effective. The majority of researchers showed that access to data is still limited, or that the databases are not user-friendly. Many also noted that a lack of trust remains between researchers and platforms.

From the evaluation of the activities implemented by Signatories, combined with this study's analysis of the opinion of stakeholders interviewed, the commitments of Pillar V of the Code appear to be the least well implemented by Signatories. Even though the criteria of the Evaluation Framework are fulfilled for Pillar V (e.g. annual events are organised, research into disinformation is supported, good faith research is not prohibited by Signatories), the initiatives remain too weak and sporadic. Efforts should be done from both sides to restore the trust between platforms and researchers. On the one hand, platforms should facilitate further the access to data and improve the transparency of the access. The platforms should also do further improvements on the user-friendliness of the databases. On the other hand, researchers should ensure that their requests for data to platforms are clear and transparent.

**Efficiency**

The assessment of efficiency of the Code predominantly focused on the Signatories, since the burden of implementing the Code largely falls with them. This section will address the following questions:
what level of resources did the signatories devote to combating disinformation? what benefits did the signatories observe as a result of compliance with the code? have there been any added costs or administrative burdens placed on the signatories as a result of the code?

with regards to the above, the platforms were not able to provide specific figures and as such quantitative analysis is difficult. the overall message received from the signatory platforms is that they are investing significant resources: financial, technical, working hours, and human resources potential, into meeting requirements of the code, be it the ads transparency centre (ATC), policies, or product changes, such as reporting features. administratively, the signatories suggest that the administrative burden has been high in terms of reporting requirements and in responding to requests for feedback.

A quantitative financial investment estimation connected to the implementation of the code is not feasible to estimate as these investments are broadly counted as part of standard product or policy development or fall within a more significant human resources allocation. This also raises the question to what extent signatory investments specific to the code can be distinguished and to what extent these resources are overheads that would have been spent otherwise. In terms of the policy formation process, platforms did note that signing the code placed a greater emphasis on updating and improving certain policies, which made the process of executing those changes more efficient. Other benefits that were noted during the interviews were the consolidation of best-practice and industry standards, which will increase the efficient allocation of resources in the future and make interactions with government bodies smoother.

There are however some examples of quantified costs. For Mozilla specifically it was reported that, not including the financial investments on their Electronic Tracking Protection (ETP) tool, between 2018-2019, they estimate their costs for 2018-2019 to fall between USD 500,000 and USD 1 million (EUR 463,140 – EUR 926,280).

For the other platforms, various grants would also feature as a cost of the code. For example, Google is participating in and providing financial support to the Trust Project and Microsoft provides administrative and financial support for the Technology | Academics | Policy (TAP) forum, the Oxford Technology & Elections Commission and a two-part project at the Oxford Internet Institute. Specific figures for the level of investment in initiatives such as these were however not provided during the data collection.

The most obvious conclusion with regards to efficiency would, therefore, be that if the platforms collaborated on grant allocation and began to delineate disinformation investment from core budgets, they may increase their efficiency and bring down the costs of implementing the code or its future iterations.

One comment on efficiency should also be noted with regards to the National Regulatory Authorities (NRA), where the amount being invested to combat disinformation in general varies to a large extent. While most NRA respondents are not investing any specific
financial resources to combat disinformation, one respondent, an EU candidate country, is reportedly investing 100,000 EUR or more.

Figure 10: Overview of investment into fighting disinformation

![Survey Results](image)

Source: Study team based on survey data

Although platforms report that administrative burden has been high in terms of reporting requirements and in responding to requests for feedback, which has in turn increased the costs to them of report, if they were to begin separating their budget lines and human resources for anti-disinformation activities, these costs could be optimised and therefore reduced. This would also result in higher quality administrative information from the platforms and more precise costings for anti-disinformation activities, which could in turn help the Commission and national regulatory agencies make reporting requirements more tailored, reducing human resource and financial burdens further for the platforms.

It is clear that the issue of disinformation is not going to reduce in the coming years, platforms will therefore need to come to terms with increasing financial burden, not only from policymakers but also from consumers who will increasingly demand greater control over the social media they consume, the platforms could therefore leverage financial advantage by moving forward with these efficiency savings as quickly as possible. Options would also include collaborating on best-practice and industry standards within the sector.

In summary:

- Collaboration with other platforms to define industry standards on levels of financial investments should be a way for platforms to reduce administrative burden in the long-term, although it increases costs in the short term.
- Platforms will need to find ways of delineating more accurate figures on the financial resources being invested in adherence to future responsibilities vis a vis combatting disinformation.
- National Regulatory agencies should also agree on levels of financial investment in combatting the spread of disinformation.
Relevance

This section aims to establish whether the Code of Practice on Disinformation has been relevant to combat disinformation by replying to the following evaluation questions:

- To what extent do industry, regulatory and other relevant stakeholders estimate that the focus of the Code of Practice is relevant?
- To what extent do industry, regulatory and other stakeholders estimate that the focus of the Code of Practice is still relevant?
- Should the Code be abandoned?
- Should there be any new commitment added to the Code?

There is a consensus that the Code is relevant to combat disinformation. Representatives of an association of advertisers highlighted that the Code has been a trigger for many initiatives on disinformation. Many stakeholders, among which experts and Signatories, agreed that initiatives to combat disinformation already existed before the Code, but that the existence of the Code raised awareness of the topic on a larger scale and allowed them to promote measures to combat the phenomenon more easily.

Several Signatories interviewed agreed that the Code provides an impetus for them to define and abide by best practices. The Code also represents an added pressure for them for to re-assess those best practices over time in a continuous dialogue with the European Commission. It was agreed that the Code provides them with guidance on points of importance for the European Commission and that it has prompted specific launches such as EU Political Ads Transparency report by Google and related verification policies and processes. These stakeholders also noted that nowhere else in the world does such an evolutive instrument exist. By comparison, other regulatory instruments are bound to prove less effective when it comes to a complex, multi-faceted, fast-changing, not well-defined set of issues. Academics interviewed also agreed that the Code enabled them put pressure on platforms and pushed them to take more concrete actions.

Overall, stakeholders agree that the Code should not be abandoned. Most of the interviewees, including Signatory platforms, indicate that abandoning the Code would send a wrong message. It was unanimously agreed that the Code should continue to evolve, based on the conclusions of the self-evaluations of the Code’s implementation by Signatories, as well as on the latest research developments. This evolution is particularly needed because the field of research and policymaking on disinformation is still relatively new and developing.

As regards the survey respondents, 61% of the expert public and 86% of national regulators estimated that the Code had a positive effect on fighting disinformation in the EU and its Member States and confirmed that it has incentivised stakeholders do take

---

121 https://transparencyreport.google.com/political-ads/region/EU
more actions to combat the phenomenon. A vast majority of national regulators (85%) who replied to the survey estimated that this positive effect was a direct result of the Code.

Despite the unanimous opinion that the Code should not be abandoned, many stakeholders pointed out that the instrument could be better operated and its scope better tailored.

Some Signatory platforms and associations argued that the scope of the Code is too limited as it focuses mainly on several companies and does not encompass a broader approach including all types of stakeholders. They suggested that a more inclusive approach should be taken to build cross-sectoral cooperation in order to build trust and address gaps in the Code. They also stressed the need for more Signatories to the Code. A Signatory also proposed to create a specific Code for platforms only, which would be more restrictive.

Several Signatory associations confirmed these points of views as they estimate that greater cooperation needs to be built across the news ecosystem, to strengthen a societal approach to tackling the disinformation issue. As seen in other areas, such as countering violent extremism, collaborative and cross-sectoral approaches are required to tackle evolving societal challenges. In their view, continuing to have disinformation framed as a competitive space between sectors will hinder any efforts to address it going forward. A regulator also agreed that there is disparity on how the Code is implemented among platforms and stakeholders.

Several survey respondents noted that the lack of enforcement mechanisms and the flexibility that was given to tech platforms in implementing the Code – e.g. having the flexibility to decide what constitutes a ‘political’ ad, and what did not – led to a rather fragmented approach taken by platforms. Most survey respondents (89% of expert public and 53.8% of national regulators) agreed that to improve the Code and facilitate the combat against disinformation, the Code should become mandatory. Few stakeholders interviewed also agreed with this view.

To the question “how could the Code be improved to facilitate the combat against disinformation?”, the survey respondents also replied in majority that the commitments should be more precise and more concrete (73.7% of expert public and 76.9% of national regulators chose this option) and that more platforms should be incentivised to sign the Code (63.2% of expert public and 69.2% of national regulators). The survey respondents suggested that concepts in the Code should be better defined (52.6% of expert public and 53.8% of national regulators) and that more commitments should be added (42.1% of expert public and 46.2% of national regulators).

In summary, the key findings for the Relevance criterion cover:

- Many stakeholders agreed that the existence of the Code enabled them to raise awareness on the topic of disinformation at a larger scale and to promote more easily measures to combat the phenomenon.
• Signatories also mentioned that the Code provides them with guidance on points of importance for the European Commission and that it has prompted specific verification policies and processes.

• Overall, stakeholders agree that the Code should not be abandoned and that it should continue to evolve based on the conclusions of the evaluation of the Code’s implementation by Signatories as well as on the latest research developments.

• However, some Signatory platforms and associations argued that the scope of the Code is too limited as it focuses mainly on several companies and does not encompass a broader approach including all types of stakeholders.

• Several Signatory associations estimated that greater cooperation needs to be built across the news ecosystem, to strengthen a societal approach to tackling the disinformation issue.

• Most survey respondents (89% of expert public and 53.8% of national regulators) agreed that to improve the Code to facilitate the combat against disinformation; the Code should become mandatory.

Overall, the Code is an appropriate tool to combat the phenomenon. The responses to the evaluation questions corresponding to the Relevance criterion in the Evaluation Framework enable to conclude that the Code incentivises stakeholders to implement policies and processes to limit the spread of disinformation. It also enables stakeholders to raise awareness on the topic and to promote more easily the measures aiming at fighting disinformation. From the analysis of stakeholders’ opinions, it appears clear that the Code should not be abandoned and that it should adapt to future developments. However, the Code would benefit from more Signatories. The potential new Signatories could represent all type of stakeholders, to favour a cross-sectoral approach in tackling the issue.

**Coherence**

This section aims to establish to what extent the commitments set up in the Code of Practice are overlapping or in synergy with other EU interventions which have similar objectives.

The analysis focuses on the following evaluation questions:

• To what extent are the commitments of the Code covered by other EU interventions?

• To what extent are the commitments of the Code in synergy with other EU interventions with similar objectives?

• To what extent are the commitments of the Code overlapping with other EU interventions with similar objectives?

Not all stakeholders were able to comment, however those who could suggest that there are synergies between the Code of Practice on Disinformation and other codes of
**conduct or policies.** An academic interviewed noted that she supported the idea of the Code of Practice on Disinformation mainly because there were other EU instruments in parallel to it, which formed part of a converging European framework.

There are 6 relevant EU instruments established in parallel to the Code of Practice on Disinformation, which were all mentioned by various stakeholders as presenting potential synergies with the Code:

- E-Commerce Directive
- Audiovisual Media Services Directive
- Copyright Directive
- General Data Protection Regulation
- Directive on security of network and information systems
- Code of conduct on hate speech

The **E-Commerce Directive** is the legal framework for online services in the Internal Market. The purpose of the Directive is to remove obstacles to cross-border online services in the EU and provide legal certainty to business and citizens. The Directive sets out basic requirements on mandatory consumer information, steps to follow in online contracting and rules on commercial communications (e.g. online advertisement and unsolicited commercial communications). It also sets rules on limitations of liability of intermediary service providers.

The **Audiovisual Media Services Directive (AVMSD)** governs EU-wide coordination of national legislation on all audiovisual media, both traditional TV broadcasts and on-demand services. It notably sets up an obligation of loyalty and an obligation for platforms to give visibility to the media as well as protecting minors from harmful content and promoting media literacy.

The **Copyright Directive** lays down rules which aim to harmonise further Union law applicable to copyright and related rights in the framework of the internal market, taking into account, in particular, digital and cross-border uses of protected content. It lays down rules on exceptions and limitations to copyright and related rights, on the facilitation of licences, as well as rules which aim to ensure a well-functioning marketplace for the exploitation of works and other subject matter. It also sets up the identification of illegal content.

The **General Data Protection Regulation (GDPR)** is an essential step to strengthen individuals' fundamental rights in the digital age and facilitate business by clarifying rules for companies and public bodies in the digital single market. It lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data. The GDPR sets up requirement of transparency and privacy even if they are made available in open access.
The Directive on security of network and information systems (NIS) is the first piece of EU-wide legislation on cybersecurity. It provides legal measures to boost the overall level of cybersecurity in the EU. It lays down obligations for all Member States to adopt a national strategy on the security of network and information systems; creates a Cooperation Group in order to support and facilitate strategic cooperation and the exchange of information among Member States and to develop trust and confidence amongst them; creates a computer security incident response teams network (‘CSIRTs network’) in order to contribute to the development of trust and confidence between Member States and to promote swift and effective operational cooperation;

establishes security and notification requirements for operators of essential services and for digital service providers; lays down obligations for Member States to designate national competent authorities, single points of contact and CSIRTs with tasks related to the security of network and information systems. It also sets up transparency requirements on cyber threats, regarding hybrid threats.

There are also synergies between the Code of Practice on Disinformation and the Code of conduct on hate speech. To prevent and counter the spread of illegal hate speech online, in May 2016, the Commission agreed with Facebook, Microsoft, Twitter and YouTube a “Code of conduct on countering illegal hate speech online”. In the course of 2018, Instagram, Google+, Snapchat and Dailymotion joined the Code of Conduct. Jeuxvideo.com joined in January 2019. The implementation of the Code of Conduct is evaluated through a regular monitoring exercise set up in collaboration with a network of organisations located in the different EU countries. Using a commonly agreed methodology, these organisations test how the IT companies are implementing the commitments in the Code. The last evaluation shows that the Code delivered successful results: the companies are now assessing 89% of flagged content within 24 hours and 72% of the content deemed illegal hate speech is removed.

There are some areas where parallels can be drawn which are useful for future regulation of the platforms. For example, on consumer rights, Amazon or eBay adopted a similar approach to the social media platforms in that they take no responsibility if something ‘goes wrong with a product’, however they now have responsibility to provide information on pricing algorithms and obligations towards ensuring quality.

Several stakeholders interviewed, notably a consumer organisation, academics and regulators however agreed that there are certain overlaps which may be problematic between the Code of Practice on Disinformation and other EU policies with similar objectives. It was stated by many stakeholders (researchers, signatory associations, consumer organisations) that there are certain overlaps between the Code and the GDPR. The GDPR provides an enhanced protection, which might however enter in

---

conflict with the Code. Some argued that platforms use GDPR and its protection as an excuse to not implement the Code/provide data.

Several stakeholders also stated that there is an overlap between the Code of Practice on Disinformation and the Code of Conduct on Hate Speech. According to some interviewees, both Codes addresses the issue of viral rhetoric and flooding of information, which destroys the quality of information. However, other interviewees noted that what is contained in the Code of Conduct on Hate Speech is different than what is entailed in the scope of the Code of Practice on Disinformation. It was stressed that the scope of both instruments should not be confused. Several signatory associations agreed to say that disinformation is not necessarily illegal. According to them, there is therefore a high risk that a regulatory approach infringes upon other freedoms, such as the freedom of expression/speech. This view was confirmed by several other stakeholders interviewed who stressed the importance to distinguish between legal and illegal content and the need to use different approaches to deal with this difference. This nuance is one that has been a foundation of the European Commission’s approach so far, including with the Code.

It was suggested that independent technology experts (not in research collaboration with any of the platforms) lacked both in the High-Level Expert Group, as well as the Sounding Board of the Forum. According to some stakeholders, without these experts, most knowledge about how the platforms work technically was in the hands of the technology companies.

All stakeholders have a level of responsibility, therefore, overarching principles to set up an online responsibility framework encompassing all areas should be developed. It was mentioned by an academic that a coordination centre could be set up as well. 35% of the expert public who replied to the survey indicated that a body composed of national government through regular authorities could be the one overseeing the platforms. 31% of the expert public estimated stressed that it should be an independent body consisting of platform representatives, policymakers and experts who should oversee the platforms.

In summary, the Code of Practice compliments and augments the effectiveness of other related EU initiatives and serves to create and overlapping but clearly defined blanket of protection to consumers of social media and other related media. The main exception to this is in GDPR which, considering its extent, does provide the platforms with competing obligations vis-à-vis data protection and the sharing of data. These will need to be clarified in future iterations of the Code of Practice.

In summary:

- All stakeholders need to continue to work in collaboration and, in some areas, deepen collaboration, with the creation of an online responsibility framework.
- This framework should explicitly map and appreciate the interrelation of various initiatives, with the aim of improving effectiveness and eliminate any overlaps.
Independent technology experts must feature more heavily in the evaluation and monitoring processes of future iterations of the Code, and should feature in future High Level Expert Groups and various advisory bodies, to compliment feedback from the platforms with regards to any competing data provision commitments.

**EU added value**

This section aims to establish the added value of having a Code of Practice on Disinformation at the EU-level, in addition to the various Member State initiatives that exist in the area.

The analysis focuses on the following evaluation questions:

- Would have the same results in relation to the strategic objectives of the Code have been possible without the EU intervention?
- What would be the most likely consequences of withdrawing the Code?

To reply to these questions, a combination of several data collection tools was used. For this evaluation criteria, the analysis mainly relied on interviews with relevant stakeholders such as the platforms themselves, independent experts and stakeholders involved in the Sounding Board.

Many actions that have been taken at the Member State level to fight disinformation (legal or policy measures such as awareness campaigns) have been initiated in the years 2018-2019. For instance, France, Germany, Hungary and Lithuania have passed, or are developing, laws that regulate social media. The most far-reaching has probably been introduced in France in December 2018125 and the most debated has probably been the one introduced in Hungary in March 2020126 (more details on these developments are described in Section 6).

The global and European dialogue on disinformation has also become more important during these years, notably due to several severe interferences from foreign states in election processes. Against this backdrop, this period also saw the creation of the Code. There thus seems to be a clear correlation between the enhanced European and international policy response to disinformation phenomenon, the enhanced media awareness for the phenomenon and the design and implementation of the Code of Practice. It is difficult to say whether the Member States’ laws have been passed as a reaction to the initiatives at the EU-level, whether this worked the other way around or whether there has been interactions where both have acted as catalysts for each other. No instances have been found where Member State initiatives were stopped or delayed.

125 [https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=FE4B2D41B8095626DE8FF07129A9EFA.tpsfr41s_2?cidTexte=JORFTEXT000037847556&dateTexte=&oldAction=rechJO&categorieLien=id&idJO=JORFCONT000037847553](https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=FE4B2D41B8095626DE8FF07129A9EFA.tpsfr41s_2?cidTexte=JORFTEXT000037847556&dateTexte=&oldAction=rechJO&categorieLien=id&idJO=JORFCONT000037847553)

126 [https://foreignpolicy.com/2020/03/31/hungarys-orban-given-power-to-rule-by-decree-with-no-end-date/](https://foreignpolicy.com/2020/03/31/hungarys-orban-given-power-to-rule-by-decree-with-no-end-date/)
because of the Code of Practice. On the contrary, the Code very often triggered more initiatives, for a more detailed discussion on Member State initiatives, see Section 6.

However, piecemeal regulation by Member States is unlikely to force online platforms to comply with dozens of different and potentially contradictory standards, so an EU approach is considered by most of the interviewed stakeholders to be the best solution. The EU is considered to have more bargaining power, which is needed to hold the large international platforms accountable. At the same time, Member State initiatives are seen as short-term and usually politically motivated (e.g. initiated because of upcoming elections), while an overall systematic approach should come from the EU in addition to this. Especially since interviewees feel that the most important threat is coming from state-actors outside of the EU.

Moreover, comparability of data is another important reason to favour an EU-led approach. This argument was mostly used by interviewees from platforms as well as by interviewed academics. There were also some specific concerns from non-Signatory interviewees that any action other than at the EU, or even international, level would simply result in “forum shopping” by the platforms and make any regulation ineffective. Interviewed Signatory platforms considered that an EU-approach is desirable to create more clarity for the platforms and save costs with regards to compliance on their side. An EU led approach, therefore, seems to be favoured by most, if not all, interviewed stakeholders.

Regarding the interaction between EU and Member State initiatives it appears that, for some more detailed policy interventions, many interviewed stakeholders from several categories expressed the opinion that the EU could provide a more broader framework of minimum standards to live up to (e.g. in a Directive on Disinformation). This could be further implemented with specific measures at the Member State level. In this way, the EU could provide the basis on which Member State legislators could take a more detailed approach with specific actions. The reason for this is that the possibility should be created for local and regional alliances of several Member States to fight disinformation more profoundly in addition to any action at the EU level in the spirit of a ‘multi-speed Europe.’ In any case, the most important condition is to ensure sufficient synergies between the EU and Member State initiatives.

Lastly, several stakeholders (both Signatories and non-Signatories) expressed the opinion that the EU level approach is a good starting point but that this is only a first step and that, it should be expanded at a later stage and adopt a global approach. In this way, Europe can act as an example to the rest of the world. One stakeholder mentioned that this might happen in a similar way to the GDPR which had an impact on the rest of the world where many countries have followed the EU approach, even without an ex-ante global coordination in any organised way.

To summarise the main findings on this topic:

- Fighting disinformation is an area where the EU has a clear added value considering the cross-border dimension of the issue and the fact that a lot of
political weight is needed to counterbalance the big influence and interests of multinational platforms.

- In addition to public action towards the platforms, this could be done by providing a set of minimum standards (e.g. in the form of a Directive) which Member States can further implement with more specific actions considering each Member State’s national circumstances.

A key finding is that disinformation is really a topic where the EU can lead the way and provide a global standard; the Code of Practice can act as a starting point for this. This should be done using a mixture of raising public awareness of the topic, agreeing on broad guidelines at the EU level in the form of legislation and encouraging Member States to take specific measures in addition to that.

**Sustainability**

The research questions pertaining to sustainability were predominantly addressed as part of the interview programme. However, the study’s two online surveys that targeted academic and other experts and National Regulatory Authorities respectively also contributed to addressing issues raised under the sustainability criterion. The literature review was less effective in addressing sustainability questions. On the topic of sustainability of the Code’s impact, the research questions focused on the extent to which direct and indirect impacts generated by the Code can be expected to last over time. With regards to the interviews, the discussions also focused on understanding whether interviewees considered the Code’s impacts to be sustainable and the reasons behind this.

There was no stark division in opinion between stakeholder type, although experts who are focused on disinformation as a global phenomenon and who tend to take a wider perspective on the topic, also tended to consider the Code as one of many initiatives that aim to combat disinformation. In contrast, stakeholders who considered themselves very well informed about the Code and its activities tended to give it more prominence.

The Signatories to the Code were generally positive towards its sustainability – however, this response is also to be expected considering their public commitment to it.

Although a minority of non-Signatory interviewees (two stakeholder organisations) claimed the impact of the Code would not be sustainable – since the Code was too weak or had had no impact to date – most interviewees suggested the Code had generated some lasting change.

The same interviewees took the view that the Code’s sustainability was very weak since “all it had achieved” was to start a conversation about combatting disinformation. The fact that several big platforms or media companies had not signed up to the Code was also cited as a reason to why sustainability of the Code was seen as problematic.
Having said this, most of the discussions on sustainability of the Code were cautious but positive. Many of the responses to the sustainability question were in line with the following interviewee quote:

“Yes, [the Code is a] step in [the] right direction. There is a need for continuous assessment of its relevance and continuous improvement, but [the Code] is a good base.”

There seems to be a wider agreement that the Code has laid a practical foundation for the foreseeable future. Although it was implemented quickly in order to help safeguard the integrity of the 2019 European Elections and/or put pressure on the big platforms to act quickly during the election campaign, the commitments in the Code will remain valid since they provide an agreed framework under which cooperation and individual measures (by the Signatories) can be further strengthened.

The national regulatory authorities survey results also indicate that this practical foundation also extends to the national level, which indicates sustainability in this regard too. When asked the question “Do you estimate that the Code provided additional value to the national policy framework of your country to combat disinformation?” the respondents were largely positive – 75% compared to those who responded with a ‘No’ (see Figure 11).

**Figure 11: Additional value of the Code**

<table>
<thead>
<tr>
<th>Do you estimate that the Code provided additional value to the national policy framework of your country to combat disinformation?</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="chart.png" alt="Pie chart" /></td>
</tr>
</tbody>
</table>

- 75% Yes (please explain)
- 25% No (please explain)

*Source: Study team based on survey data*

When asked to justify their response, the respondents suggested that the Code activities stemming from EU level action were also becoming visible at national level in the Member States. One respondent suggested that the Code had served as a source of inspiration for some national initiatives.

Indeed, the national regulatory authority representatives who also responded to the survey were inclined to say that their organisation’s activities had also changed as a result of the Code (see Figure 12). According to free text responses, these activities included national monitoring and evaluation activities of disinformation as well as an increased political will to push for a national policy on disinformation.
With this in mind, to a certain extent the Code’s actions seem to have triggered actions at Member State level which in turn could strengthen the likelihood of the Code achieving sustained impacts. These findings also strengthen the EU added value of the Code, which is discussed in a preceding section.

The sustainability of impacts achieved was deemed as positive by the Signatories. Although it is to be expected that stronger sustainability could be seen from those Signatories who put in more effort implementing measures to combat disinformation. However, if the Code continues to be implemented, non-Signatories may become more engaged over time if they see the benefits derived by the current Signatories. A driver in this regard is that the disinformation phenomenon is increasingly being called out by growing public opinion and concern.

The fact that the Code is based on the principle of self-regulation and requires different measures by different Signatories does however hamper its potential sustainability. Survey and interview respondents were divided on this question, each group providing a different argument.

Those who considered that the industry-driven approach hampered the sustainability of impacts argued that, due to the self-regulatory nature of the Code, the effectiveness of measures risked being compromised (e.g. lack of oversight and monitoring of implementation), which in turn negatively affected the substantiality of impacts achieved.

However, many of those consulted put forward a counter argument, highlighting that, thanks to its design, which is the result of Signatory engagement, the Code’s activities can be implemented by a wide range of organisations. In this sense, the self-regulatory approach is adaptable. It was acknowledged by interviewees that since the Signatories of the Code operate under different business models, a la carte measures also need to be in place. One interviewee highlighted that a flexible or “agile” approach to self-regulation also provides the opportunity to change or adjust the course of action, thereby maximising impact and sustainability of impact.

Among those stakeholders positive to the Code, there was also a consensus that the Code needs to be given time to prove whether it is dealing with the issue in an effective
and sustainable way. At the same time, it needs to keep evolving to stay relevant considering political and technological developments.

The main findings for the Sustainability criterion are summarised below:

- Overall, the Code’s outcomes to date can be considered to have fairly good sustainability, predominantly thanks to the Code’s high relevance and concrete actions undertaken. As a very minimum, there is an agreement that the Code has laid a foundation to a common framework for combatting disinformation, which can be further strengthen in future iterations of the Code.

- However, since sustainability is in effect measuring the longevity of (positive) outcomes produced, a strengthening of the Code’s commitments and improved implementation would also lead to better sustainability.

- EU action on disinformation has prompted debates and at times action at Member State level. Supporting and complementary action at national level is highly likely to further strengthen the sustainability of current activities led by the EU.
CHAPTER 6: INITIATIVES AT MEMBER STATE LEVEL

In parallel with action at EU level, legislative and policy measures to combat disinformation are also being implemented at Member State level. Most of these legislation and policy measures aimed at tackling disinformation were developed during the years 2018-2019.

This time period saw an increase of communication and debate around the phenomenon of disinformation, and it is thus safe to say that the enhanced awareness on the matter has led the EU Member States to take initiative.

Implementation of non-legislative measures to combat disinformation

Most EU Member States rely mainly on policy measures to combat disinformation and have not taken any legislative measures to date. This is the case notably for the Nordic countries. Most of the policy measures taken by these Member States consist of media literacy programmes implemented in schools and high schools. Denmark, for example, adopted an Action plan to strengthen safeguards against any interference on Danish democracy and society in May 2019. The Danish authorities bolstered their efforts to get ahead of disinformation problems by re-purposing some media literacy material from Sweden. In order to avoid falling for misinformation, the government is distributing brochures with specific tips for recognising such information.

Similarly, in Finland no specific legislative act has been adopted to counter disinformation either, however, the Finnish government also invested in strengthening media literacy using partnerships between schools and fact-checkers. Some Finnish schools, for example, partnered with Finnish fact-checking agency Faktabaari (FactBar) to develop a digital literacy “toolkit” for elementary to high school students teaching about the EU elections.

Latvia is another Member State where the focus of the government has been placed on the implementation of policy measures and media literacy. The government in Latvia has been working hard to build media literacy, particularly within its population. School workshops that teach Latvian teachers and students how to differentiate fact from fiction have risen in recent years.

Cyprus also adopted a similar path. The independent regulatory authority responsible for monitoring the audio-visual media services in Cyprus, the Cyprus Radio Television Authority, designed and implemented media literacy programmes to educate high school students to recognise disinformation. One of the workshops designed for these students concentrates on the critical evaluation of the content disseminated through the various forms of mass media, and mainly through social media and online sites. Various

---

127 https://www.disinfo.eu/2019/06/19/eu-g7-regulations/
examples of media content (articles) are provided to students who are asked to evaluate the truthfulness, the veracity, and the credibility of the content.\textsuperscript{129}

Campaigns to raise awareness on the phenomenon of disinformation have also been launched, for example, in the \textit{Netherlands}. In February 2019, the country launched a public awareness campaign aimed at informing people about the spread of misinformation online. The campaign, which came months ahead of the EU Parliamentary elections, was predominantly waged on social media. The aim of the campaign was to make Dutch voters more aware of the possible presence of disinformation and help people to recognise it. Prior to this campaign, the Dutch authorities have also emphasised the importance of education, with national broadcaster NOS airing a 90-minute television special in March 2018 called \textit{News or Nonsense (Nieuws of Nonsense)}, which gave examples of what people should look out for.

In the framework of a joint initiative of the Ministry of the Economy, the Ministry of the Family, Integration and the Greater Region and the Ministry of National Education, Children and Youth in \textit{Luxembourg}, the BEE SECURE initiative was launched.\textsuperscript{130} The initiative includes a “Share Respect - Stop Online Hate Speech” campaign, within which a file about false information on the Internet was published.\textsuperscript{131} The file presents information on the different types of false information circulating on the Internet, on the motivation of authors to disseminate false information as well as on the means to verify if the information can be true. It also gives specific advice to parents on coping with their children’s media consumption.

The \textit{Swedish} Civil Contingencies Agency, which was tasked with spotting potential influence operations by foreign actors, devised an awareness campaign for journalists and the public about the spread of misleading information and propaganda. In collaboration with researchers from Lund University, the Agency published a handbook describing the principles and methods of identifying, understanding and countering information influence activities. A section of this handbook is dedicated to disinformation and it presents the different techniques used by malicious actors to spread false information.\textsuperscript{132}

Awareness campaigns combating disinformation can also be found in \textit{France}, where the civil initiative “What The Fake” fights against hate speech, extremism and online manipulation.\textsuperscript{133} The initiative produces and disseminates campaigns to combat extremist discourses and manipulation processes that proliferate on the internet. It intervenes in three main directions: the fight against hate speech, extremism, and manipulation, especially in the face of the proliferation of “fake news” and the vitality of

\textsuperscript{129} https://cmpf.eui.eu/the-role-of-national-regulatory-authorities-in-tackling-disinformation/
\textsuperscript{130} https://www.bee-secure.lu/fr/a-propos
\textsuperscript{131} https://www.bee-secure.lu/fr/campagnes/share-respect/quest-ce-que-le-hate-speech/fausses-informations-sur-internet
\textsuperscript{132} https://www.msb.se/RibData/Filer/pdf/28698.pdf
\textsuperscript{133} https://what-the-fake.com/qui-sommes-nous/
conspiracy theories. It publishes articles and guidelines on how to recognise and cope with false information.\textsuperscript{134}

Other type of non-legislative measures taken by Member States included the development of collaboration and alliances between experts, which led to the creation of innovative tools to fight disinformation. In \textbf{the Netherlands}, “Drog” is a platform, created by a multidisciplinary team of academics, journalists, and media experts, that conduct research, give talks, offer workshops and educational programmes, and create innovative tools that help build resistance to disinformation. “Drog” runs a sophisticated simulation game that aims to shock participants into understanding that they should be critical of the information they get online and through social media feeds. It runs experiments in schools, government services and military headquarters by letting participants create their own fake news and spread it in simulated news environments including social media.\textsuperscript{135}

Platforms combating disinformation have been also implemented in \textbf{Greece}, where since 2013, the website Ellinica Hoaxes is debunking disinformation and fake news.\textsuperscript{136} Since its establishment, the website has debunked around 3,000 disinformation items. They range from articles on websites, TV, newspapers, to fake memes circulating via social media.

In \textbf{Croatia}, the website Factographer is specialising in fact checking, that is, verifying the factual accuracy of claims in public space.\textsuperscript{137} Factographer is a member of International Fact Checking Network (IFCN), an international network of fact-checking organisations. It was launched in 2015 as a joint co-publishing project of the Croatian Journalist Society and GONG. The purpose of the Factographer is to use the fact-checking method, using reliable sources, to evaluate the accuracy and validity of statements found in media, and to prevent an unfounded position in public discourse from becoming an inalienable fact. Factographer also deals with suppressing misinformation in the public space, and since April 2019 has been part of Facebook’s Third-Party Fact Checking Program.\textsuperscript{138}

In \textbf{the UK}, the “Online Harms White Paper”, published in April 2019, does not foresee an implementation of specific legislation to combat disinformation but it presents several initiatives to combat the phenomenon, including specific tools and awareness campaigns.\textsuperscript{139} For example, the new ‘RESIST’ counter-disinformation toolkit equips the government, public services and partner country communicators with the knowledge and skills they need to identify, assess and respond to disinformation.\textsuperscript{140} A pilot public disinformation communications campaign has also been launched. It provides the public

\begin{footnotesize}
\textsuperscript{134} https://what-the-fake.com/2018/03/14/sy-retrouver-entre-faux-vrai-internet/
\textsuperscript{135} https://aboutbadnews.com/
\textsuperscript{136} https://www.ellinikahoaxes.gr/about-us/
\textsuperscript{137} https://faktograf.hr/
\textsuperscript{138} https://www.facebook.com/help/publisher/182222309320722
\textsuperscript{139} https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper
\textsuperscript{140} https://gcs.civilservice.gov.uk/guidance/resist-counter-disinformation-toolkit/
\end{footnotesize}
with the skills they need to recognise and respond to disinformation, showing people how it can affect them and what they can do about it.\(^{141}\)

**Implementation of combined non-legislative and legislative measures to combat disinformation**

Some EU Member States also combined legislative and policy measures to fight disinformation. It is the case of **France** which now has a rather influential legislative framework to fight disinformation. In 2018, the country notably passed a law considered to be western Europe’s first attempt to officially ban false material.\(^{142},^{143}\) The law triggered intense debates and it was rejected twice by the senate before being passed by the parliament. Critics were notably arguing that the controversial law could jeopardise democracy and censor the press.

It includes chapters on the following topics:

- The implementation of an accessible and visible reporting mechanism;
- The transparency of algorithms;
- The promotion of content from press and news agencies and from audio-visual communication services;
- The fight against accounts disseminating false information on a massive scale;
- The information of users on the nature, origin and modalities for dissemination of content;
- The identity of individuals providing remuneration in return for the promotion of information content; and
- The promotion of media literacy.

Candidates and political parties are now able to seek accelerated court action against "false information" online during the three months before an election. The law also allows the CSA, the French National Broadcasting Agency, to render the authority to suspend television channels "controlled by a foreign state or under the influence" of that state if they "deliberately disseminate false information likely to affect the sincerity of the ballot.” This means that, in the context of elections, France has the power to take on any foreign TV station suspected of spreading “false news.” In parallel to the legislative measures passed to fight disinformation and secure the online environment (e.g. Law against hate speech, 2019), France has also taken some policy measures to combat the phenomenon. As mentioned previously, the French government supported the creation of “What the Fake”, a civil initiative developed in the aftermath of the terrorist attacks of 2016, which fights against hate speech, extremism and online manipulation.

\(^{141}\) [https://sharechecklist.gov.uk/]
Lithuania has also implemented both legislative and policy measures to fight disinformation. On the legislative side, it implemented the Law on the Provision of Information to the Public which states in its Article 19, paragraph 2: “It shall be prohibited to disseminate disinformation and information which is slanderous and offensive to a person or degrades human dignity and honour”. The law enables the Radio and Television Commission to block media that spread disinformation. The decision on temporarily blocking a channel is made through the courts following an application by the Commission. As said by the President of Lithuania in a press release: “Lithuania was among the first Member States to launch active measures against disinformation at the national level. We have adopted laws allowing to suspend the broadcasting of TV channels by a court decision for the incitement of hatred or dissemination of fake news”. On the policy side, initiatives such as “Debunk.eu” have been launched in the country. “Debunk.eu” is an initiative that unites the media, the society, and the state to fight against disinformation.

On the legislative side, Germany does not have a general law that prohibits the creation and dissemination of fake news. However, depending on the facts of the case, there are a number of civil and criminal law provisions that may be applicable to safeguard individuals or the public from false news on social networks. A requirement is that the information is capable of defaming a person or negatively affecting public opinion of the person. The crime of defamation is punishable with imprisonment not exceeding one year or a fine and, if committed publicly or through the dissemination of written materials, with imprisonment not exceeding two years or a fine.

In 2017, the Network Enforcement Act (so-called Facebook Act) was passed and it explicitly aims to combat hate speech and fake news on social networks. The explanatory memorandum stated that: “fighting fake news on social networks [is] a priority. To do so requires improvements in law enforcement on social networks in order to promptly remove objectively criminal content, such as incitement to hatred, abuse, defamation or content that could lead to a breach of the peace by misleading authorities into thinking a crime has been committed”.

---

146 https://de debunk.eu/
149 Ibid. § 186
The law in its current form does not create any new duties for social media platforms, with the exception of reporting requirement, but imposes high fines for non-compliance with existing legal obligations. On the non-legislative side, the German Press Agency (DPA) is carrying out fact checks via its “DPA fact check” programme. The Agency’s journalists choose topics of nationwide importance and verify the accuracy of the related content according to three areas: “statement”, “assessment” and “facts”. These fact checks are intended to enable readers to form their own opinion based on facts, even with complex topics. The aim is that the readers question and ultimately disprove unfounded prejudices or misleading manipulations.

Other Member States tried to or are in the process of adopting laws to fight disinformation, but these pieces of legislation are being highly debated. It is the case in Hungary, where, citing the urgency of the coronavirus pandemic, Hungary’s parliament has approved in March 2020 a bill to allow the Prime Minister to rule by a decree. The bill includes harsh penalties of up to five years in prison for spreading what the government deems “fake news”.

In Ireland, the government presented a bill on political bots and advertising in 2017 which renders the use of a bot to create 25 or more personas on social media punishable by up to five years in prison or fines of up to EUR 10,000. However, this bill has lapsed with the dissolution of the Dáil, the principal chamber of the Oireachtas (Irish legislature).

Italy has also seen several attempts to introduce legislation to tackle the spread of disinformation, none of which have yet been adopted. The first bill, tabled in February 2017 and heavily criticised by experts, proposed introducing fines and prison sentences for those spreading ‘false, exaggerated or biased news on data or facts that are manifestly false or unproven’. Under another draft law, put forward by the centre-right Forza Italia party in May 2019, users would only be allowed to create new social media accounts if they provided their social security number.

Other measures in some Member States

At least one EU Member State has implemented other measures for the online sphere focussing on matters such as cyber-security and terrorist threats, with no specific
mention to disinformation. In the Czech Republic, the Center Against Terrorism and Hybrid Threats was created in 2017, with a view to national security objectives related to the new threats arising from radicalization and disinformation campaigns.\(^{159}\) With regard to disinformation, the Center's scope of action is restricted to content that may stimulate "radical or terrorist forces", not having "the power to censor or remove content that it deems inappropriate, but to serve as a warning and communication".

**Conclusions on Member States' measures to combat disinformation**

From the mapping and analysis of the Member States' initiatives to combat disinformation, it appears clear that only a few Member States have implemented legislative measures to combat the phenomenon. Most countries have favoured non-legislative measures such as media literacy programmes or awareness campaigns. These "soft" measures aim to explain in a clear and, often, interactive manner what is false information, how to recognise it, and how to deal with it. They are directed at the general public, as well as at communication professionals. Some of the media literacy programmes are also implemented in school, high schools and universities.

The majority of the national legislative measures that were implemented specifically to deal with disinformation were subject to intense debate as they were seen as too draconian by many experts. It is the case, for example, in Hungary, France and Ireland. In some Member States, such as Germany, a number of civil and criminal law provisions may be applicable to safeguard individuals or the public from fake news in social networks.

The table below presents a list of the non-legislative and legislative measures put in place by Member States to combat disinformation:

### Table 3: Legislative and non-legislative measures at national level

<table>
<thead>
<tr>
<th>Member State</th>
<th>Legislative measures</th>
<th>Non-legislative measures</th>
</tr>
</thead>
</table>
| Austria      | N/A                  | • AIT, the OVE initiative ScienceClip.at and the Austrian Ministry of Defence teach media skills to effectively deal with disinformation on the Internet<sup>160</sup>  
• Workshop on the use of AI technologies against fake news<sup>161</sup> |
| Belgium      | N/A                  | • In 2018, an Expert group and citizen consultation platform to tackle misinformation was launched. The expert group composed of academics, media representatives and NGOs elaborated several recommendations on the subject adopted by the Belgian authorities  
• Fact-checking website: www.stopfakenews.be |
| Bulgaria     | N/A                  | • Disinformation Unit within the Council of Ministers as part of the Action Plan against Disinformation for coordinated measures by the European Commission and the European Parliament. |
| Croatia      | N/A                  | • Fact-checking website: <https://faktograf.hr/>  
• Media Literacy Days to focus on respect in media, recognising fake news<sup>162</sup> |
| Cyprus       | N/A                  | • Media literacy programmes: The independent regulatory authority responsible for monitoring the audiovisual media services in Cyprus, namely the Cyprus Radio Television Authority, designed and implemented media literacy programmes in an attempt to educate high school students to recognise disinformation and fake news. It considers the development of a number of different workshops, one of which focuses on the negative consequences of disinformation and emphasises the importance of evaluating the information and of assessing media content that the public consumes.<sup>163</sup> |
| Czech Republic | N/A                  | • The Center Against Terrorism and Hybrid Threats was created in 2017, with a view to national security objectives related to the new threats arising from radicalisation and disinformation campaigns.<sup>164</sup> |
| Denmark      | N/A                  | • In 2019, the Action plan to strengthen safeguards against any interference on Danish democracy and society was adopted. Moreover, the Danish authorities bolstered their efforts to get ahead of misinformation problems by re-purposing some media literacy material from Sweden. In order to avoid falling for misinformation, the government is distributing brochures with important tips.<sup>165</sup> |
| Estonia      | N/A                  | • In 2006, Estonia was one of the first countries to create a Computer Emergency Response Team to manage security incidents. “Baltic elves” – volunteers who monitor the internet for Russian threats. |

---

<sup>161</sup> [https://www.eurosint.eu/articles/workshop-vienna-explores-use-ai-technologies-against-fake-news](https://www.eurosint.eu/articles/workshop-vienna-explores-use-ai-technologies-against-fake-news)
<sup>162</sup> [https://mzo.gov.hr/news/media-literacy-days-to-focus-on-respect-in-media-recognising-fake-news/1746](https://mzo.gov.hr/news/media-literacy-days-to-focus-on-respect-in-media-recognising-fake-news/1746)
<sup>165</sup> [https://um.dk/en/news/newsdisplaypage/?newsid=1df5adb-d1df-402b-b9ac-57fd4485f4a4](https://um.dk/en/news/newsdisplaypage/?newsid=1df5adb-d1df-402b-b9ac-57fd4485f4a4)
### Legislative measures

<table>
<thead>
<tr>
<th>Member State</th>
<th>Legislative measures</th>
<th>Non-legislative measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>N/A</td>
<td>Partnership between schools and fact checkers. As an example, some Finnish schools partnered with Finnish fact-checking agency Faktabaari (FactBar) to develop a digital literacy “toolkit” for elementary to high school students learning about the EU elections. ¹⁶⁸</td>
</tr>
<tr>
<td>France</td>
<td>• Law n° 2018-1201 of 22 December 2018, relating to the combat against manipulation of information¹⁶⁹</td>
<td>• What The Fake: French civil initiative that fights against hate speech, extremism and online manipulation. The initiative produces and disseminates campaigns to combat the extremist discourses and manipulation processes that proliferate on the internet.¹⁷⁰ • Fact-checking programmes: Les décodeurs du Monde,¹⁷¹ Check News de Libération¹⁷², Observateurs de France 24.¹⁷³</td>
</tr>
<tr>
<td>Germany</td>
<td>• Act to Improve the Enforcement of Rights on Social Networks, Network Enforcement Act, NetzDG, Sept. 1, 2017</td>
<td>• German Press Agency (DPA) fact-checking programme: “DPA fact check” programme.¹⁷⁴ • Munich Municipal Library –has been holding social community courses for pupils from the 5th grade upwards since 2012. The topics they focus on are privacy, data protection and copyright laws – and, since 2016, also fake news.¹⁷⁵</td>
</tr>
<tr>
<td>Greece</td>
<td>N/A</td>
<td>• Since 2013, the website Ellinika Hoaxes is debunking disinformation and fake news. Since then they have debunked, roughly, 3,000 fake news. They range from articles in websites, TV, newspapers, to fake memes circulating via social media.¹⁷⁶</td>
</tr>
<tr>
<td>Hungary</td>
<td>• Bill on disinformation, March 2020¹⁷⁷</td>
<td>• “Misinformation and Propaganda Through the Eyes of Hungarian Students” programme, including a workshop for Hungarian students with international experts, a public discussion with Hungarian and international experts, projects related to misinformation prepared by participating students.¹⁷⁸</td>
</tr>
</tbody>
</table>

¹⁶⁶ https://www.disinfo.eu/resources/estonia-2
¹⁶⁷ https://www.kaitseliit.ee/en/cyber-unit
¹⁶⁹ https://www.disinfo.eu/resources/france-2
¹⁷⁰ https://what-the-fake.com/
¹⁷¹ https://www.lemonde.fr/les-decodeurs/
¹⁷² https://www.liberation.fr/checknews,100893
¹⁷³ https://observers.france24.com/fr/
¹⁷⁴ https://www.dpa.com/de/unternehmen/faktenchek/#faktenchek-regeln
¹⁷⁵ https://www.goethe.de/en/kul/bib/21001176.html
¹⁷⁶ https://www.ellinikahoaxes.gr/
¹⁷⁸ https://cmds.ceu.edu/misinformation-and-propaganda-through-eyes-hungarian-students
### Assessment of the implementation of the Code of Practice on Disinformation

<table>
<thead>
<tr>
<th>Member State</th>
<th>Legislative measures</th>
<th>Non-legislative measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ireland</td>
<td>On 05 November 2019, the Irish Government approved a proposal to Regulate Transparency of Online Political Advertising. The detailed Proposal is outlined in the Progress Report of the Interdepartmental Group on the Security of Ireland’s Electoral Process and Disinformation (IDG). It follows from a public consultation and open policy forum on the issue.</td>
<td>Fact-checking organisation: <a href="https://www.thejournal.ie/">https://www.thejournal.ie/</a>. The Government established an Interdepartmental Group (IDG) on Security of the Electoral Process and Disinformation in December 2017 to consider the issues arising from recent experiences in other democratic countries with regard to the use, and misuse, of social media by external, anonymous or hidden third parties.</td>
</tr>
<tr>
<td>Italy</td>
<td>N/A</td>
<td>Fact-checking organisation: Pagella Politica. ‘Enough-With-the-Hoaxes’ Media Literacy Campaign. ‘Red Button’ Portal: Two months before the general election in 2018, a special cyber police unit under the Italian state police launched an initiative to protect the election. This included a so-called ‘red button service’ through which citizens could report fake news stories directly to the police.</td>
</tr>
<tr>
<td>Latvia</td>
<td>N/A</td>
<td>Media literacy programmes: School workshops that teach Latvian teachers and students how to differentiate fact from fiction have risen in recent years.</td>
</tr>
<tr>
<td>Lithuania</td>
<td>Law on the Provision of Information to the Public</td>
<td>“Debunk.eu” is an initiative that unites the media, the society, and the state to fight against disinformation.</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>N/A</td>
<td>The BEE SECURE initiative includes actions to raise awareness of the more secure use of new information and communication technologies. In the framework of the Share Respect - Stop Online Hate Speech campaign, BEE SECURE publishes a file on false information on the Internet and gives tips on how to identify it.</td>
</tr>
</tbody>
</table>

---

180 [https://assets.gov.ie/39188/8c7b6b8c1d0d046ebe915963abfe427e90.pdf](https://assets.gov.ie/39188/8c7b6b8c1d0d046ebe915963abfe427e90.pdf)
181 [https://assets.gov.ie/39188/8c7b6b8c1d0d046ebe915963abfe427e90.pdf](https://assets.gov.ie/39188/8c7b6b8c1d0d046ebe915963abfe427e90.pdf)
182 [https://pagellapolitica.it/](https://pagellapolitica.it/)
183 [https://www.bastabufale.it/](https://www.bastabufale.it/)
187 [https://debunk.eu/](https://debunk.eu/)
188 [https://www.bee-secure.lu/fr/a-propos](https://www.bee-secure.lu/fr/a-propos)
<table>
<thead>
<tr>
<th>Member State</th>
<th>Legislative measures</th>
<th>Non-legislative measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malta</td>
<td>N/A</td>
<td>• Information Literacy Workshop&lt;sup&gt;189&lt;/sup&gt;</td>
</tr>
<tr>
<td>Netherlands</td>
<td></td>
<td>• &quot;Drog&quot; is a platform, created by a multidisciplinary team of academics, journalists, and media-experts, that conduct research, give talks, offer workshops and educational programmes, and create innovative tools that help build resistance to disinformation,&lt;sup&gt;190&lt;/sup&gt;</td>
</tr>
<tr>
<td>Poland</td>
<td>N/A</td>
<td>• The ‘Cybersecurity Doctrine of the Republic of Poland’ was adopted in 2015 by the National Security Bureau as a response to the increase in hybrid threats, propaganda, disinformation, and psychological influence operation by foreign states and non-state actors. The doctrine maps out tasks for state institutions, notably security agencies and armed forces, the private sector, as well as NGOs.&lt;sup&gt;191&lt;/sup&gt; • Fact-checking organisation: <a href="https://demagog.org.pl/">https://demagog.org.pl/</a></td>
</tr>
<tr>
<td>Portugal</td>
<td>N/A</td>
<td>• In March 2019, ahead of the European parliamentary elections, the Portuguese parliament discussed fake news and misinformation, with the Socialist Party (PS) demanding cybersecurity measures from the government to guarantee safe elections. This was the first time Portugal’s MPs addressed this issue, since the disinformation phenomenon sparked growing awareness during the UK’s Brexit referendum, allegedly influenced by disinformation.&lt;sup&gt;192&lt;/sup&gt; • Fact-checking organisation: Poligrafo.&lt;sup&gt;193&lt;/sup&gt;</td>
</tr>
<tr>
<td>Romania</td>
<td>N/A</td>
<td>• Rubrika is the first fully automatic news analyser in Romania, allowing users to research all news websites and Facebook pages from Romania, and analyse the content they publish. Also, Rubrika is a news aggregator that displays news only from trustworthy news sources.&lt;sup&gt;194,195&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>189</sup> https://www.um.edu.mt/newspoint/notices/um/2017/11/informationliteracyworkshopbeyondfakenews<br>
<sup>190</sup> https://www.aboutbadnews.com/<br>
<sup>191</sup> https://en.bbn.gov.pl/ftp/dok/01/DCB.pdf<br>
<sup>192</sup> https://www.disinfo.eu/resources/portugal-2<br>
<sup>193</sup> https://poligrafo.sapo.pt/<br>
<sup>194</sup> https://rubrika.ro/<br>
<sup>195</sup> The Romanian Government has launched the online platform Covid 19 Știri Oficiale, which gathers all the news and official announcements from the authorities related to the evolution of the new coronavirus epidemic in Romania and the measures taken against it, in an attempt to counter disinformation and fake news about the coronavirus situation. Available at: https://stirioficiale.ro/informatii
## Assessment of the Implementation of the Code of Practice on Disinformation

<table>
<thead>
<tr>
<th>Member State</th>
<th>Legislative measures</th>
<th>Non-legislative measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slovakia</td>
<td>N/A</td>
<td>Factual is a fact-checking website, with true or false indicators for each piece of news, by volunteer contributors of national and international news. The project was founded by a group of NGOs, corporations, entrepreneurs and policy consultants interested in good governance.(^{196})</td>
</tr>
<tr>
<td>Slovenia</td>
<td>N/A</td>
<td>Fact-checking organisation: AFP Slovakia</td>
</tr>
<tr>
<td>Spain</td>
<td>N/A</td>
<td>Fighting Fake News Media Literacy Workshop for Students(^{197})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Workshop on fake news, targeting youngsters, youth workers, educators, students and others interested in the disinformation topic.(^{198})</td>
</tr>
<tr>
<td>Sweden</td>
<td>N/A</td>
<td>The Civil Contingency Agency (MSB), which was tasked with spotting potential influence operations by foreign actors, devised an awareness campaign for journalists and the public about the spread of misleading information and propaganda.(^{200})</td>
</tr>
<tr>
<td>UK</td>
<td>N/A</td>
<td>UK Online Harms White Paper presents several initiatives to combat disinformation including specific tools and awareness campaigns.(^{201})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘RESIST’ counter-disinformation toolkit equips the government, public services and partner country communicators with the knowledge and skills they need to identify, assess and respond to disinformation.(^{202})</td>
</tr>
</tbody>
</table>

\(^{196}\) [https://www.factual.ro/](https://www.factual.ro/)
\(^{197}\) [https://business.facebook.com/events/387148322083792/](https://business.facebook.com/events/387148322083792/)
\(^{199}\) [https://www.disinfo.eu/resources/spain-2](https://www.disinfo.eu/resources/spain-2)
\(^{200}\) [https://www.thelocal.se/20180115/sweden-to-create-new-authority-tasked-with-countering-disinformation](https://www.thelocal.se/20180115/sweden-to-create-new-authority-tasked-with-countering-disinformation)
\(^{201}\) [https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper](https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper)
Assessment of the implementation of the Code of Practice on Disinformation

<table>
<thead>
<tr>
<th>Member State</th>
<th>Legislative measures</th>
<th>Non-legislative measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>• A pilot public disinformation communications campaign has been launched. It provides the public with the skills they need to recognise and respond to disinformation, showing people how it can affect them and what they can do about it.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The Digital, Culture, Media and Sport Committee has published its final report on Disinformation and ‘fake news’.</td>
</tr>
</tbody>
</table>

203 [https://sharechecklist.gov.uk](https://sharechecklist.gov.uk/)
204 [https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179102.htm](https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179102.htm)
From the analysis of the interviews carried out for this Assessment of the Code of Practice on Disinformation, a clear correlation was found between the enhanced European and international awareness around disinformation, the enhanced communication about the phenomenon and the fact that new important initiatives were taken at Member State level to fight disinformation. Seventy-five percent of the national authorities who replied to our survey estimated that the Code provided additional value to the national policy framework of their country to combat disinformation. Most of them also estimated that it enabled their country to better define disinformation in national measures and to put in place more national policies.

To this, it can be concluded that a pan-European instrument and initiatives at Member State level are complementary to combat disinformation. National measures, either legislative or non-legislative are useful complement to the Code of Practice on Disinformation, as they enable to further raise awareness on disinformation at the national level and, in some cases, on the Code itself as it is directly mentioned in some of the media literacy programmes. A part of these national initiatives existed before the Code entered into force, but many of them were facilitated by the existence of the Code.
CHAPTER 7: METRICS AND KEY PERFORMANCE INDICATORS

This chapter discusses metrics and key performance indicators proposed by the research team as suitable for measuring change in the phenomenon of disinformation. As such, this chapter attempts to address one of the key criticisms of the Code of Practice, which is that in its current form, it is not possible to measure objectively the progress of the Code’s implementation by each of its signatories.

At the time of the initial implementation of the Code, a set of draft monitoring indicators were initially proposed and adopted, however, these were not designed to detect medium- or longer-term results or outcomes with regards to the Code’s commitments. Rather, the purpose of these indicators was to guide the annual self-assessment reports completed by the signatories.

Several steps can be taken to ensure future assessments of the Code and disinformation in general. These can be based on further quantitative data that was not collected during 2019, the Code’s first year of implementation.

Firstly, the logic model and evaluation question framework, developed and validated as part of this study’s Evaluation Plan, constitute two basic methodological tools for future research, evaluation and monitoring. Naturally as the Code might evolve and bring on-board new Signatories, these tools will need to be re-tested and – if found in need of an update – adapted before being re-deployed. In this regard, the logic model and evaluation framework should be considered as ‘living documents’ which should document and reflect the aim, activities and outcomes of the Code.

Secondly, deploying a robust set of Key Performance Indicators (KPIs) would also help address the somewhat piecemeal and differentiated implementation (i.e. across the Pillars, across platforms and across Member States), which is one of the key findings of this report.

Regular assessment of the KPI data – in addition to regular monitoring – could allow for the Signatories to review their performance against the KPIs in real time and to – when necessary – adapt their activities. Thus, the KPIs would constitute a tool for gauging progress on an ongoing basis while also providing evidence towards more comprehensive evaluations of the Code’s effectiveness. These KPIs would also allow monitoring of the development and spread of disinformation as such.

Generally, when policymakers or evaluators design KPIs, they should adhere to RACER criteria. This means that KPIs should be Relevant, Accepted, Credible, Easy and Robust in order to be effective:

- **Relevant**: the data collected should be relevant to the objectives set (i.e. encompass all five pillars of the Code);
• **Accepted:** the data collected (evidence) should be accepted among all stakeholders, e.g. it should be politically neutral;

• **Credible:** similarly, the evidence collected should be understandable to all levels of experts as well as laymen;

• **Easy:** the KPIs should be easy to monitor (and on a regular basis); and

• **Robust:** against manipulation and poor interpretation of the data.\(^{205}\)

The indicators that are proposed here have been developed by the study team and are based on the desk research undertaken and extensive stakeholder consultations, including the pre-consultation interviews. The study also developed a case study dedicated to KPIs (see Annex 1). Thus, this proposal draws on extensive input from experts, industry and policymakers.

This study’s proposal is to include 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. Given the heterogeneity across and within each of the platforms, the first set of KPIs needs to be more generalist than the second set.

With this approach, the study also suggests to move from a focus on the “production” of disinformation to include also its “distribution”. This enables assessment of progress made and the actual (attributable) impact of disinformation. Indeed, the real impact of the phenomenon cannot be grasped only by measuring how much disinformation is out there, but it needs to consider also behavioural aspects, such as human reactions to disinformation and why/how consumers and individuals share disinformation that they come across. Building on the “Disinformation ABC” by Camille François, which stipulates that an efficient set of measures should aim at taking into account the multifaceted nature of disinformation (Actors, Behaviour and Content), this study adds a D for Distribution of content.\(^{206}\)

The two sets of KPIs are as follows:

- Structural indicators;
- Service-level indicators.

A detailed description of each set of indicators is provided below together with a suggestion regarding the data collection and assessment process.

---

\(^{205}\) European Commission Better Regulation Guidelines Tool #41. Monitoring Arrangements and Indicators

ARRANGEMENTS AND INDICATORS

\(^{206}\) Camille François (Transatlantic Working Group), Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses
**Structural indicators**

A set of **structural indicators**, focusing on outcomes, should be designed which 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.

In line with the ReutersOxford method[^207], the structural 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. Secondly, the structural KPIs could also collect data on – for example – a monthly basis on newly created content by purveyors of disinformation and the development of users’ engagement with such content. Obtaining regular information on these aspects would help to provide indications on prevalence (i.e. existing cases) and incidence (i.e. new sources) on disinformation.

In order to be able to monitor the prevalence of disinformation (e.g. how many purveyors of disinformation there are) as well as its incidence (e.g. how many new purveyors are there in a given period of time) within the EU, there is a need to set up sampling groups in each Member State to be able to monitor the spread of disinformation as well as to consider the language particularities in each of the MS. These groups should consist of representative samples of the population in each MS where the participating individuals would agree for their online interaction with news and disinformation content to be monitored in line with GDPR provisions.

With the use of the sample groups, the study team suggests the following structural indicators to be collected in each MS:

### Table 4: Proposed Structural Key Performance Indicators

<table>
<thead>
<tr>
<th>Proposed KPIs</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence of authoritative and disinformation sources in accessing news</strong></td>
<td>Number of authoritative and disinformation sources through which the sample groups access news and in what language</td>
</tr>
<tr>
<td><strong>Direct versus indirect access to authoritative and disinformation source</strong></td>
<td>Whether the sample accesses authoritative/disinformation content directly (e.g. via browser) or indirectly through social media and in what language</td>
</tr>
<tr>
<td><strong>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)</strong></td>
<td>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</td>
</tr>
</tbody>
</table>

[^207]: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/Measuring%20the%20reach%20of%20fake%20news%20and%20online%20distribution%20in%20Europe%20CORRECT%20FLAG.pdf
Such data should be collected and processed by an independent body, for example, the European Digital Media Observatory (EDMO)\(^{209}\), with support provided by platforms and ERGA. In order to set up the sampling groups, the independent body should collaborate with ERGA and fact-checking and research networks in order to set up the representative samples in each country. The body would then, in cooperation with the platforms, set up the mechanism in which the sample data would be collected ensuring GDPR compliance. The data created by the sample groups would then be regularly collected (i.e. monthly) and assessed by the body in cooperation with selected researchers.

The sample groups could also be used to collect specific service-level indicators (described below) and would provide a foundation against which the service-level indicators could be measured/compared with.

**Service-level indicators**

*Secondly*, a set of *service-level indicators*, focusing on results, should be developed to measure the progress of each platform in its combat against disinformation.

In this regard, what would be interesting to find out is not just the quantity of disinformation that is present on the platforms, but to focus on the actual impact it has had. The first step to knowing this would be to increase the understanding with regards to how disinformation spreads and how it reaches users including organic disinformation and disinformation in closed groups. The goal is to understand how disinformation enters and flows through society, how it impacts users and why. When understanding the entire lifecycle of disinformation and the impacts it has, one can understand the weaknesses of it and consequently develop strategies to interrupt its spread and limit its impact, in the process creating a more resilient and information literate society. To this end, there is a need to have service-level KPIs not only focusing on each of the pillars of the Code but also on the general level of disinformation. As mentioned previously, the service-level KPIs could also make use of the sample groups created for the structural level KPIs to collect some control data.

Examples of service level KPIs that could be considered are presented in the table below. Whenever possible, the data should be provided per MS for comparative purposes.

---

208 This number would be proportionate to the population of each MS, i.e. smaller MS could consider top 30/50 websites.

209 The Observatory is being established by the European Commission.
### Table 5: Proposed Service-level KPIs per Pillar

<table>
<thead>
<tr>
<th>Code Pillar</th>
<th>Proposed KPIs</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General level</strong></td>
<td>Direct investment spent on identifying/deactivating disinformation content as percentage of turnover</td>
<td>Quantify investments of platforms in fighting disinformation</td>
</tr>
<tr>
<td></td>
<td>Direct investment spent on identifying/deactivating disinformation content per user as percentage of total turnover per user</td>
<td>Quantify investment of platforms in fighting disinformation in terms of human resources</td>
</tr>
<tr>
<td></td>
<td>Number and types of tools available per MS</td>
<td>Qualitative assessment of whether all tools platforms developed in connection to combating disinformation are available in all MS</td>
</tr>
<tr>
<td></td>
<td>Ratio of total donations made to NGOs, civil society, etc to support initiatives combating disinformation (e.g. media literacy campaigns) against total turnover</td>
<td>Quantify investment of platforms into independent initiatives to combat disinformation</td>
</tr>
<tr>
<td></td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td><strong>Pillar I: Scrutiny of ad placements</strong></td>
<td>Total turnover of advertising operators from advertisement placement</td>
<td>Revenue created during a certain time period</td>
</tr>
<tr>
<td></td>
<td>Total of foregone (lost) revenue due to certain accounts being closed due to being purveyors of disinformation</td>
<td>Estimate of revenue lost due to closure of certain accounts due to their link with disinformation</td>
</tr>
<tr>
<td></td>
<td>Total advertisement revenue from top 100 websites identified as purveyors of disinformation</td>
<td>The amount of money that flows to the most prominent purveyors of disinformation due to advertisement placement</td>
</tr>
<tr>
<td></td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td><strong>Pillar II: Political and issue-based advertising</strong></td>
<td>Number of mislabelled political and issue-based advertising</td>
<td>The prevalence of false positives and negatives in the ad libraries</td>
</tr>
<tr>
<td></td>
<td>Ratio of total turnover of issue-based advertising with revenue lost due to accounts closed down due to breach of issue-based advertising policies</td>
<td>The prominence of issue-based labelling and its impact on revenues</td>
</tr>
<tr>
<td></td>
<td>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</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>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</td>
<td></td>
</tr>
<tr>
<td></td>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

### Assessment of the implementation of the Code of Practice on Disinformation

<table>
<thead>
<tr>
<th>Pillar III: Integrity of Services</th>
<th>Ratio (estimate) of inauthentic accounts/users that remained alive/active after creation against all active accounts</th>
<th>What proportion of all accounts are considered to be fake/inauthentic that are not caught by the platforms policies before their creation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>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</td>
<td>Measure of the actual impact of inauthentic accounts/users. Can be used to monitor the role of inauthentic accounts/users in promoting disinformation</td>
</tr>
<tr>
<td></td>
<td>Ratio of directly contracted employees (or FTE) tasked with identifying/deactivating disinformation content as percentage of total number (or FTE) of staff</td>
<td>Quantify investments of platforms in fighting disinformation</td>
</tr>
<tr>
<td></td>
<td>Number of fake accounts/fake users deactivated following their report by a genuine user</td>
<td>Measure magnitude of the problem of fake accounts/fake users</td>
</tr>
<tr>
<td></td>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pillar IV: Empowering Consumers</th>
<th>Ratio of complaints submitted by users about disinformation content against number of complaints followed up</th>
<th>Effectiveness of platforms to follow up with their users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>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</td>
<td>Approximation of the “safety” of content on platforms</td>
</tr>
<tr>
<td></td>
<td>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</td>
<td>The effectiveness of tool designed to empower consumers</td>
</tr>
<tr>
<td></td>
<td>Number of users that interacted with disinformation content produced by inauthentic accounts/users that were notified when such content was removed</td>
<td>How effective the platforms are on informing users that they have interacted with a disinformation content</td>
</tr>
<tr>
<td></td>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pillar V: Empowering the Research Community</th>
<th>Ratio of number of academic/research organisations that enter into relevant arrangements with platforms against number of data requests received</th>
<th>Effectiveness of cooperation of platforms and the research community</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ratio of donations made to academic/research organisations for research projects against total investment into combating disinformation</td>
<td>Quantify investment of platforms into research supporting fight against disinformation</td>
</tr>
<tr>
<td></td>
<td>With these indicators, the aim is to investigate how effective the cooperation with the research community is.</td>
<td></td>
</tr>
</tbody>
</table>

Directorate-General for Communications Networks, Content and Technology
The majority of data input for these indicators would be provided by the platforms, however, it should be an independent body, for example, the EDMO who then further assesses and analyses the data provided, including the data’s reliability, with ERGA providing further support/oversight if needed.

The data collected as part of the structural KPIs sampling group could serve as a control measure to compare the results of these KPIs of the overall platform users and the representative samples in each Member State.

**Data provision and assessment**

There are a number of considerations that apply to both sets of indicators. Many of these indicators are based on data provided by the platforms, it is crucial to independently verify this data. As mentioned above, the European Digital Media Observatory which is currently being set up will create a data hub which can be used to audit data delivered by the platforms. It may also help to implement the pillar which is currently lagging in implementation (empowerment of researchers).

In its work, the EDMO should cooperate with research and fact-checking networks, particularly when it comes to interpreting data at Member State level. With this regard, EDMO will need to ensure that the data provided by platforms are securely stored and that access to them is strictly monitored and in line with GDPR provisions.

Another issue is whether any measures should be included to enforce the commitments to combat disinformation (measured by these KPIs) in the Code itself. In other words, whether any enforcement measures should be included in the Code (see, for example, a suggestion in section 5.1). The independent regulatory authorities (convened in the ERGA organisation) have a role to play here at the Member State level as well. Moreover, it could be considered to give ERGA a more important role as an overarching body in the oversight of monitoring and assessment of data generated by the KPIs.

It is important to point out that most of the structural and service level indicators have been designed to be collected on a monthly basis in order to analyse the development of the phenomenon of disinformation over time. This also adds the possibility of making a qualitative assessment on top of the quantitative assessment based on the general indicators.
CHAPTER 8: CONCLUSIONS

This section presents the overall study conclusions. These conclusions have been developed based on, evidence collected from the various data collection tools throughout the study as well as the study team’s analysis based on this data. The conclusions are presented per evaluation criterion together with suggestions for further improvements.

Relevance

The findings of this study show that the implementation of the Code of Practice has produced overall positive results. There is a consensus amongst most stakeholders that the Code of Practice is a highly welcomed and needed initiative. Since disinformation continues to be a widespread problem, the Code remains highly relevant and its aims and activities still address current needs. The Code has established a common framework under which to agree and implement activities to tackle disinformation. Specifically, the Code, and the preparatory activities carried out before its establishment, has contributed to the debate on disinformation, raised awareness, and provided guidance among wider stakeholders (e.g. civil society, policymakers, the media and publishing sector) and among the Signatories. Discussions facilitated by the work of the Code has also contributed towards wider action and measures to combat the phenomenon.

Therefore, the Code should not be abandoned but its implementation should continue. However, several improvements to the agreement and its implementation would contribute to further (positive) impact. In tandem with its continuation, debate, research and other suitable activities should also follow to firstly, promote further improvement of the Code’s effectiveness and sustainability so as to maximise its impacts and secondly, to continue to assess the scope of the Code and the specific mechanism for implementation.

Effectiveness

With regards to its overall effectiveness, an initial achievement of the Code is that is has set a foundation on which further activities can be continued.

This study's research shows that the Code has had a positive effect with regard to improved cooperation between policymakers and the Signatories of the Code. Indeed, the Code has established a platform for negotiation that concretely has produced results in the form of regular monitoring of Signatory activities and action to combat disinformation challenges. The monitoring processes produce reports on change over time, which allows for better transparency of social media platforms during elections and other political campaigns. To further support the evaluation and monitoring of the Code, this study also provides methodological recommendations that can be used to further improving this process, especially in the form of a set of KPIs.

However, certain challenges are connected to the Code and its implementation. Firstly, it is a voluntary document and as such there are no means to enforce the commitments promised by the Signatories. This has led to at times fragmented implementation (e.g. across the various Pillars, across the various platforms and across Member States). Further
time investment and continued efforts will be required to ensure that all five Pillars reach higher levels of effectiveness over time.

Overall, the Pillars 1 to 3 (scrutiny of ad placements, political and issue-based advertising and integrity of services respectively) have produced a more positive change compared to Pillars 4 and 5 (empowering consumers and empowering the research community).

With regards to Pillar 4, the Code has contributed towards improved awareness of consumers on the topic of disinformation. Wider awareness of the Code among the general public (in their role as consumers) may also attract more Signatories to the Code so as to ensure their reputation and social corporate responsibility commitments. Especially better reporting on the results of the different tools Signatory platforms have put in place and a more consistent roll-out across all EU Member States is needed.

For Pillar 5, there is evidence that the Signatory social media platforms have implemented several policies and tools to increase the collaboration with researchers and fact-checkers, and notably to provide them with access to platform data. Nevertheless, difficulties persist. These may, at least partly, stem from a lack of sufficient dialogue between the research community and the Signatories since the former group is overall unsatisfied with the response from the latter to data requests. Out of the five Pillars, Pillar 5 will require the most substantial efforts to improve effectiveness. These efforts should be done from both sides to restore the trust between platforms and researchers.

Indeed, questions regarding the effectiveness of the individual Pillars of the Code illustrate that disinformation is a complex topic; it is a nebulous concept and at times subjective. The more concrete the action agreed upon, the easier it can be (measured to be) effective. For example, it may be the case that measures regarding political advertising have seen stronger development than issue-based advertising because stakeholders find the remit of issue-based advertising unclear as it lacks a common definition. Indeed, other terminology used in the Code also needs to be agreed among stakeholders in the near future to ensure not just a common language but also to improve future evaluations.

With regards to the term “disinformation” itself, it may be useful to further try to deconstruct this concept within the remit of the Code. As outlined in this Report, understanding intent is crucial when exploring behaviour as well as the impact of this phenomenon. A refined definition that differentiates between misinformation (sharing false information but with no intent to harm), disinformation (sharing false information with the intent to harm) and malinformation (sharing genuine information with the intent to harm) in line with Wardle & Derakshan could, for example, be considered. The reasoning here is that a better understanding of intent may also improve the effectiveness of the specific actions to combat specific behaviours.

Efficiency

This study’s assessment of the efficiency of the Code predominantly focused on the Signatories, since the burden of implementing the Code largely falls with them. The Signatory platforms were not able to provide specific figures and as such quantitative analysis was not possible to undertake as part of the study investigations.
The overall message received from the Signatory platforms is that they are investing significant resources (financial, technical, human resources) to meet the requirements of the Code. They also suggest that the administrative burden has been high in terms of reporting requirements of the Code and in responding to requests for feedback for data.

There is no reason to fundamentally question the assessment of the Signatories on this particular point, since their claims of costs can partially be verified (e.g. by an independent assessment of the work required to compile the monitoring reports, respond to request for data, etc.).

However, since the Code is a self-regulatory initiative it is also logical that the costs should fall on the industry involved in the exercise.

With regards to improving efficiency, one approach would be for the Signatories to work more closely together in implementing the commitments under the Code. The platforms could for example collaborate on grant allocation and begin to delineate disinformation investment from core budgets.

**Coherence**

This study's investigations conclude that the Code of Practice is coherent with other EU initiatives in the same area.

By and large, there are no overlaps with activities implemented through other mechanisms, including the E-Commerce Directive, the Audiovisual Media Services Directive, the Copyright Directive, the General Data Protection Regulation, the Directive on security of network and information systems, and the Code of conduct on hate speech.

**EU added value**

Although there is a divergence in opinion among stakeholders with regards to the effectiveness of the principle of self-regulation, there is wide acknowledgement that the Commission is right in pursuing a dialogue with the social media platforms. There is also acknowledgement that the Code constitutes a first and crucial step in the global fight against disinformation. In this sense, the Code shows European leadership on an issue that is international in nature.

There seems to be a clear correlation between the enhanced European and international awareness around disinformation, the enhanced communication about the phenomenon and the fact that new important initiatives are taken at Member State level to fight disinformation.

The Code has led to increased reflection among Member States with regard to activities to understand and combat disinformation. Some Member State authorities are planning/undertaking activities relevant to the Code (e.g. planning disinformation strategies, preparing (better) monitoring of the phenomenon, etc.).

There is also a consensus that disinformation is a topic where the EU has an added value and where it should continue to lead and coordinate action. The EU may also wish to
consider providing a minimum set of standards for Member States to adhere. Mostly to ensure consistency, to create more bargaining power towards the platforms and because a large part of the threat comes from state actors outside of Europe.

**Sustainability**

The study’s findings on sustainability (i.e. considering the longer-term outcomes of the Code) are cautious albeit positive.

Although the Code has met with some criticism, especially in the beginning and in particular with regards to it “lacking teeth”, this study has shown the Code to be overall effective, yet still more action is required (e.g. concerning a common understanding of key concepts and concerning the implementation of Pillars 4 and 5).

It should also be recognised that the Code has been implemented during a time period when the concept of disinformation has been regularly publicly discussed. This correlation has probably helped to give the Code a higher profile, however the Code has also contributed towards the debate, thus reinforcing the need for action.211

The high profile of the Code also appears to have helped to prompt action at national level, which could contribute to stronger and more granular action at Member State level. In turn, action by policymakers at national or regional level could strengthen the sustainability of the results and outcomes stemming from the Code’s activities.

**Priorities for improving the Code**

Although the overall assessment of the Code of Practice of the study is positive, the evidence also suggests that the Code could benefit from certain key improvements.

Firstly, more consistent reporting adhering to certain minimum information standards could allow for an even better assessment of the effectiveness of the implementation of the Code, especially allowing comparisons between the various platforms. This reporting should focus on actual impacts and results of measures rather than mere statistics; a proposal of Key Performances Indicators to be used is suggested by this study.

Independent auditing of the data delivered by the platforms in their reports could eliminate the debate on whether this data is correct and representative. This would ensure a level playing field in discussing the effectiveness of the Code and eliminate the information asymmetry that is currently there. In turn, this might get the more critical stakeholders on board with the project of the Code or at least allow them to discuss with the platforms on an equal basis.

The Code would also benefit from wider agreement on the terminology it uses as well as the definition of key concepts. Specific ideas for this have been made under the

---

211 The study authors would like to reiterate that this assignment was carried out before the Covid-19 crisis and that the events that followed the Covid outbreak were not part of the study remit. However, the fact that the pandemic has become the topic of a new wave of disinformation has reinforced the need for action. See for example information from the UK Parliamentary Office of Science and Technology. [https://post.parliament.uk/analysis/covid-19-misinformation/](https://post.parliament.uk/analysis/covid-19-misinformation/)
effectiveness criterion in this study. Equally, the Code would benefit from getting a better understanding of the concept of disinformation itself as well as its characteristics and its impacts.

To improve the consistency of the implementation of the Code across Member State, it could be considered to set minimum standards for the platforms to adhere to in each Member State (for example, if a tool is not implemented in all MS at the same time, a provisional calendar should be communicated when users in each of the MS can expect to be able to use these tools). Minimum standards would ensure consistency in the implementation of the Code across the Union and help create more bargaining power towards the platforms. This is important given that a large part of the threat comes from state actors outside of Europe.

Good practices related to the implementation of the commitments under the various pillars of the Code, either implemented by one of the Signatory platforms or by other relevant stakeholders, could be exchanged more regularly. In the long term, these could form part of the Code and act as minimum standards for all signatories to adhere to.

The Code could also guarantee a wider positive impact if it had a larger number of Signatories signed up to it commitments. Therefore, all efforts should be made to widen the base of platforms signing up to it.

Lastly, although the Code of Practice is a self-regulatory instrument – and the first of its kind – introducing a mechanism for action in case of non-compliance of the (insufficient) implementation of the commitments that platforms signed up to, could be considered to enhance the credibility of the agreement. To that effect, the Commission should consider proposals for co-regulation within which appropriate enforcement mechanisms, sanctions and redress mechanisms should be established.
ANNEX 1: CASE STUDIES

Each case study was developed using desk research and several interviews. The findings were then written up as 2-3-page mini reports. The most important findings have been integrated in the main analysis of this study.

The table below provides an overview of the selected topics, their reasoning and the type of stakeholders targeted for interviews.

Table 6: Case study selection

<table>
<thead>
<tr>
<th>Topic</th>
<th>Reasoning</th>
<th>Targeted stakeholders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue-based advertising</td>
<td>While there is no problem in defining what is political-based advertising, with issue-based one it is not so easy as there are often national, even regional nuances that need to be considered. This case study explores what would be the minimum measures defining issue-based advertising.</td>
<td>• National regulators&lt;br&gt;• Academics&lt;br&gt;• Civil society&lt;br&gt;• Signatories</td>
</tr>
<tr>
<td>KPIs/Metrics</td>
<td>Looking at the current reporting requirements, how could these be improved and harmonised across the reporting organisations. What constitutes a strong KPI in this area? Should these be set on a European or national level. What is the best way to monitor specific harmful activities, e.g. is it better to delete fake accounts or monitor their behaviour to see how they develop?</td>
<td>• Signatories&lt;br&gt;• Academics&lt;br&gt;• Researchers&lt;br&gt;• Civil society</td>
</tr>
<tr>
<td>Data requests</td>
<td>What are the main barriers for researchers to get access to the data they request, and whether a common template for such requests could be established?</td>
<td>• Researchers&lt;br&gt;• Academics&lt;br&gt;• Signatories</td>
</tr>
<tr>
<td>Partnerships with fact-checkers</td>
<td>How effective are these partnerships in verifying and promoting trusted content?</td>
<td>• Fact-checkers&lt;br&gt;• Signatories</td>
</tr>
</tbody>
</table>
Case study on Issue-Based Advertising

This case study goes deeper into the topic of ‘issue-based advertising’ as part of Pillar II of the Code of Practice.

Focus

There are specific challenges related to this topic for platforms, EU institutions and Member States that require careful consideration in the context of the Code itself and any future iterations. The data sources for this case study are largely news articles, platform reports and the Code itself, complimented with further interviews of academics, regulatory agencies of member states and civil society organisations.

Current situation

At present there is no universal definition of ‘issue-based advertising.’ With this in mind, the below table draws together various defining characteristics of issue-based advertising, as distinct from political advertising. The Code itself placed the onus squarely on the platforms to coordinate an accepted definition, similarly, the High-Level Expert Group on Disinformation did not clearly define ‘issue-based’ as distinct from ‘political advertising’.

Table 7: 'How to define issue-based advertising'

The Cambridge Dictionary defines issue-based advertising as ‘advertising that is not intended to sell a product or service, but rather to change people’s opinions or behaviour.’

Of course, issue-based advertising has been targeting consumers long before the rise of social media, for example the 1986 ‘Don't Die of Ignorance’ Aids campaign in the UK.

The key distinction to make is that, in contrast to political advertising, issue-based adverts do not have to be linked to politics, politicians, or electoral processes and are not explicitly connected to an ideology.

Recent topics on social media which can be categorised as ‘issue-based’ include:

- Health, for example the anti-vaccine movement.
- Environment, for example palm oil and habitat loss of Orangutans.
- Conflict, for example the humanitarian crisis in Yemen.

The main challenge, therefore, with regards to issue-based ads, as opposed to political advertising, is that issues are constantly changing and gain momentum internationally while still largely being framed in, and defined by, national context.

The task of defining and identifying them is also made more challenging because they often lack the same telling and consistent indicators there are for political advertising, such as explicit reference to elections, candidates and political parties, which are universal across the EU.

One of the key challenges with the Code of Practice is that, although it is pan-European in nature, what is and is not ‘issue-based’ is highly dependent on national context. Indeed, the national regulators themselves have widely varying definitions of political and issue-based advertising. For example, during interviews it was noted that, since political advertising is not permitted in France, the national regulator does not use the notion of "political advertising" nor "issue based" advertising but instead talks about "information content linked to a debate of general interest."

In Ireland, the definition of 'issue-based' is not enshrined in legislation but comes from a High Court judgement - Colgan vs. the IRTC (1998), which regulates adverts in this context depending on whether they aim to procure changes in law or government policy (domestic or foreign), or intend to further the interests of a political party. This judgement also makes explicit mention of national context and its role in defining 'issue-based.' In Hungary an advert is categorised as either ‘political’ or ‘other’, with no explicit room for a separate definition for 'issue-based'.

Indeed, the stakeholders interviewed for this case study pointed out that 'issue-based' as a distinct concept is still relatively new in Europe, whereas it has been part of advertising lexicon in the United States for longer. This of course poses an issue for platforms when implementing the Code’s commitment of defining ‘issue-based advertising’.

One core question with regards to issue-based advertising, as outlined in Table above, is how to view it in relation to political advertising, and where the distinction between the two lies. In this debate it is clear context plays a large role. One example which illustrates this well is the 2005 ‘make poverty history’ campaign in the UK. This campaign was declared political by UK national regulator, Ofcom, and banned from advertising. This decision was made not solely based on the content of the ad, but by the context in which this campaign took place, the 31st G7 summit in Scotland. Indeed, definitions of political vs issue-based cannot be undertaken without appreciating the national context.

One further example which illustrates this is the provision of abortion information in Ireland. In other European countries, advertising on this issue would not be seen as political, however, in Ireland it is clearly a political question. This is emphasised most recently by the referendum in 2018 and Google and Facebook’s decision to block adverts related to the abortion referendum.

With this in mind, to explore this topic more deeply it is necessary to look at how the Code of Practice was originally designed and how it has been implemented by the platforms, as well as some specific peculiarities with regards to issue-based ads as a topic.

---

Looking at the table below, the Code is clearly focused on ensuring that issue-based ads are transparent in terms of who is promoting them, as well as committing the platforms to remain cognizant of specific political contexts in which an issue-based advert is placed (proximity to election, sensitivity of national debates etc). The chief commitment connected to issue-based ads was for the platforms to develop a working definition of them. Naturally, lacking a definition means it is hard to assess which ads should be affected, although the commitments themselves do suggest some criteria for definition, they do not go far enough in this regard.

Table 8: What the Code Says

<table>
<thead>
<tr>
<th>What the Code says</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Goal</td>
</tr>
<tr>
<td>“Ensure transparency about political and issue-based advertising.”</td>
</tr>
<tr>
<td>Conditions</td>
</tr>
<tr>
<td>“Approaches to issue-based advertising developed should be reflective of the European market for political and issue-based advertising, and take note of the European Commission Recommendation on election cooperation networks, online transparency, protection against cybersecurity incidents and fighting disinformation campaigns in the context of elections to the European Parliament.”</td>
</tr>
<tr>
<td>Specific Transparency Indicators</td>
</tr>
<tr>
<td>“Trustworthiness of content sources, media ownership and/or verified identity.”</td>
</tr>
<tr>
<td>Outputs</td>
</tr>
<tr>
<td>“Relevant Signatories commit to use reasonable efforts towards devising approaches to publicly disclose “issue-based advertising”.”</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>“Such efforts will include the development of a working definition of “issue-based advertising” which does not limit reporting on political discussion and the publishing of political opinion and excludes commercial advertising.”</td>
</tr>
</tbody>
</table>

When looking at the policies the platforms currently have, summarised in the Table above, the main point to be noted with regards to the platforms is that Facebook was the only one to outline a clear definition of which topics it sees as issue-based advertising. The other signatories consciously decided not to define it. This is clear evidenced by the fact that they did not define it, despite it being a commitment to the Code.

Twitter’s ban on political advertising is important to note here. While it may help overall with the issue of disinformation in Europe, its prohibition does not mean Twitter can avoid making a judgement on its definition, or the definition of what is ‘issue-based’. In a 2019 interview relating to the ban, Del Harvey, Twitter’s vice President of Trust and Safety, outlined that Twitter is intended to be an ‘open platform, where people are held accountable by other users.’ The ban on political advertising should, therefore, be seen as an attempt to avoid certain topics becoming ‘siloed’, and an attempt to allow for greater scrutiny by users themselves, rather than Twitter taking an active role in stopping the creation of disinformation content itself. With this in mind, it appears rational that Twitter avoids hard definitions. Having said this, their policy on ‘cause-based’ ads does outline some topics that match the definition of ‘issue-based’ of this case study. However, the confusion in terms here only serves to emphasise further that a baseline of pan-European terminology must be agreed upon for any Code to work in practice, this requires coordination of all stakeholders.

With regards to the specific commitment of the signatories to coordinate on definitions of issue-based advertising, the Code clearly did not provide the intended results. Indeed, more broadly, the failure of the platforms to effectively coordinate with regards to third-party evaluation of the Code’s implementation appears representative of a number of coordination difficulties. It is fair to assume that coordination on the definition of ‘issue-based’, as with third party evaluation, was attempted to a certain extent but the methods employed by the platforms were simply not robust enough.

Table 9: Current Platform Policies

<table>
<thead>
<tr>
<th>Platform</th>
<th>Policy on Issue-based advertising</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microsoft</strong></td>
<td>Advertising that exploits political agendas, sensitive political issues or uses “hot button” political issues or names of prominent politicians is not allowed regardless of whether the advertiser has a political agenda.218</td>
</tr>
<tr>
<td><strong>Twitter</strong></td>
<td>Cause-based: “Twitter restricts the promotion of and requires advertiser certification for ads that educate, raise awareness, and/or call for people to take action in connection with civic engagement, economic growth, environmental stewardship, or social equity causes.”219 Issue-based “Twitter's Political Content Policy includes issue advocacy, which applies to ads that refer to an election or a clearly identified candidate or ads that advocate for legislative issues of national importance.”220</td>
</tr>
</tbody>
</table>
| **Facebook** | The ad will be labelled, and “special authorisation or disclaimers may be needed if an ad focuses on the following social issues” within the European Union:221  
  - civil and social rights  
  - crime  
  - economy  
  - environmental politics |
### Addressing future needs

The major commitment of the Code of practice with regards to issue-based advertising, cooperation between platforms to define the term itself, was not met. The Code therefore has not yet succeeded to incentivise the platforms to work collaboratively in this regard and is consequently unfit for purpose. Therefore, any analysis must start from the fundamental point of view that much more needs to be done in this area if issue-based advertising is to be kept as a distinct and separate facet of disinformation within policymakers’ conceptualisation of the phenomenon. Within this challenge, policymakers must evaluate the extent to which the benefit of having ‘issue-based’ in addition to ‘political advertising’, which is simpler to define, outweigh the costs in terms of reaching a workable definition with all stakeholders.

Currently, political advertising and issue-based advertising are often treated as a pair in reporting by the platforms, NRAs and the Commission. This has led to fractured policies and definitions by all stakeholders.

Yet, if future iterations of the Code are to successfully approach the question of ‘issue-based advertising’, they must deal with the issue of national context as a defining factor. The European Union is an entity that is familiar with the challenge of developing supranational frameworks which are flexible enough to stretch across different national contexts, and a variety of policy areas exist which could be used as a basis (energy, research and innovation, infrastructure, environment). This framework must also go beyond the political and social landscape of member states and take into account the existing legal context; the member states themselves have a wide variety of different definitions and legal basis for issue-based and political advertising.

Nevertheless, the national regulators interviewed for this case study were optimistic that it is possible to develop some broad pan-European guidelines to support the creation of a definition of ‘issue-based’ advertising. ‘Issue-based’ advertising is an important part of

---


Directorate-General for Communications Networks, Content and Technology
disinformation and although more time and effort may be needed to understand its peculiarities, the Code should continue to refer to the need to develop a set of criteria for defining ‘issue-based’ at the national level including all stakeholders in its creation. The coordination required in this regard should be undertaken by the European Commission, with significant input from all stakeholders. An output of this could be an ‘issue-based advertising checklist’, with clear criteria for defining what is an ‘issue-based ad’. Participation in developing this checklist should be mandatory for platforms.

**List of stakeholders contacted**

<table>
<thead>
<tr>
<th>Type of stakeholder</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Regulatory Authority</td>
<td>Broadcasting Authority of Ireland</td>
</tr>
<tr>
<td>National Regulatory Authority</td>
<td>CAS (Conseil Supérieur de l’Audiovisuel)</td>
</tr>
<tr>
<td>National Regulatory Authority</td>
<td>National Media and Infocommunications Authority</td>
</tr>
</tbody>
</table>
**Case study on Key Performance Indicators**

**Focus**

This case study looks into the Key Performance Indicators (KPIs) and the reporting connected with these as set up by the Code of Practice on Disinformation. It focuses on the current situation, that is after the Code’s first year of implementation, as well as analyses what changes with regard to the KPIs could be implemented in the next iterations of the Code.

This case study used information from the signatories’ monitoring reports, interviews with Signatories and other stakeholders as well as additional literature research.

**Current situation**

In the Code of Practice, a whole chapter is dedicated to commitments by its Signatories to measuring and monitoring the effectiveness of the Code. Within this chapter, the Signatories commit to produce annual reports on their work done to combat disinformation and to cooperate with an independent third party selected by the Signatories to evaluate the annual reports produced against the commitments made.

Regarding the reporting commitment, it needs to be highlighted that the Signatories, particularly the platforms, went above and beyond of what was outlined in the Code. To complement the roadmaps of the Signatories on how they intend to implement the Code, a baseline report was produced by them in January 2019 to allow for comparison of the pre-Code and post-Code results and observe the development of effectiveness of the Signatories’ policies. On top of these, and in light of the then upcoming elections to the European Parliament, the European Commission asked the Signatory platforms to provide monthly reports on their effort to combat disinformation across the EU in the period leading up to the elections (i.e. January to May 2019).

Overall, all these reporting requirements were well met by the Signatories in terms of submission of the reports. While many highlighted the additional burden created by all this reporting and the monitoring it required, a few also mentioned that it helped them with taking stock of their policies and to see where further improvements are needed.

From the perspective of non-Signatories, there were a number of criticisms raised with regard to the content of these reports, including from the European Commission. Many different stakeholder groups (e.g. trade and business associations, researchers and academics etc.) point to the fact that the reports do not have a common structure and that each of the Signatory is interpreting the reporting needs in their own way to fit their business models. This makes it very hard to compare the achieved results across the Signatories and to evaluate the overall impact of the Code on tackling disinformation.

Further criticism is connected to the reports containing a lot of qualitative information and assessment while being very scarce on providing actual quantitative data, meaning that the Code’s KPIs are in majority of cases not reported on. Or in the instances where quantitative data is included, it is oftentimes provided in an aggregate form and it is unclear whether the numbers include only results from the EU or global ones.
This issue is connected to the discussion on the validity of the KPIs included in the Code. Many stakeholders point out that these KPIs are too vague and provide very little value to assessing the performance of the Signatories’ policies. Some stakeholders, many among research peers, point out that the indicators included in the Code should not be labelled as ‘Key Performance Indicators’ as they do not fulfil the RACER criteria\(^{223}\).

From the Signatories side, they concede that the KPIs might be quite broad but this was done on purpose to allow the Signatories to adjust them to their varied business models. Nonetheless, they are open to further discussion regarding the re-evaluation of the KPIs. However, they caution against using too rigid indicators so as for the comparison to not punish those platforms that are unable to provide data due to their business models and user bases being so different as to prevent them from collecting the data for the required indicators.

**Addressing future needs**

The reviewed literature, the stakeholders interviewed in the course of the study and the case study research all agree that the first step in designing strong KPIs is the need to define what exactly is a key performance indicator and what is its objectives, i.e. what it intends to measure. In connection to the Code, it means considering whether the KPIs should focus on measuring the outcomes of the various tools and policies implemented (e.g. total number of deactivated fake accounts) or on the outcomes of the actions taken as a result of the tools and policies in place (e.g. average number of disinformation content shared by fake accounts before being deactivated). Based on the feedback collected as part of the study, the KPIs should contain a balanced mix with some focusing on outcome while the others on the impacts.

Another point that needs to be adhered to when designing and selecting the KPIs is the clear description of their purpose, contextual information and data to be collected. This is to avoid vague indicators open to interpretation leading to incomparable results. As an example of this, the Signatories highlight that while the ‘number of policy infringements’ is an important metric, it is not necessarily always meaningful. This is because these numbers fluctuate based on external factors, and it is hard to interpret the results. For instance, an increase in registered infringements may mean that the platform has become better at spotting and taking action against bad actors, or it may mean the platform is as effective as it was in the prior year but there have been more attempts to place policy-infringing ads on its services.

Another observation that needs to be kept in mind is that all Signatories have different services, business models and user bases. This means that by selecting more prescriptive KPIs in order to facilitate comparison between the platforms, if selected hastily, it could lead to even bigger differences in the reported results rather than approximation. For instance, while a KPI to measure transparency around political advertising might be appropriate for a social media platform that generates revenue by displaying advertising alongside user-
generated content, it might be wholly inappropriate for other types of services, such as search, or for Signatories that prohibit all political advertising. In this instance, it would be better to focus on the prominence of trustworthy and authoritative sources of content, for example, through the use of algorithmic tools rather than on the disinformation content blocked.

**List of stakeholders contacted**

<table>
<thead>
<tr>
<th>Type of stakeholder</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signatory</td>
<td>Microsoft</td>
</tr>
<tr>
<td>Sounding Board member</td>
<td>ACT</td>
</tr>
</tbody>
</table>
Case study on partnerships with researchers - Data requests

This case study focuses on partnerships between the platforms and researchers and particularly aims at improving the data requests researchers could make towards platforms for data needed for their research.

Focus

This case study aims to explore the questions of what currently are the main barriers for researchers to get access to the data they request (current situation in Chapter 1) and how this could be improved going forward, especially exploring the questions whether a common template for such requests could be established (addressing future needs in Chapter 1). This is particularly relevant as the main analysis undertaken for this study has shown that Pillar V (empowering the research community) is the weakest in terms of implementation, while a good-functioning relationship between platforms and researchers is fundamental to enable independent assessment of the platforms’ performance for all Pillars. Moreover, interviews have shown that many researchers are unhappy about their current cooperation with the platforms.

Building on the work done during the main analysis of this study, for this case study additional desk research has been carried out focusing specifically on this topic and additional interviews have been carried out with three academics working on the disinformation topics and having made requests for data to platforms on a regular basis as well as one platform dealing with such requests.

Current situation

To further substantiate the unhappiness of research with the current cooperation with the platform, in the expert survey conducted for this study 4 out of 25 experts considered the implementation of Pillar 5 to be somewhat effective, while 10 out of 25 experts considered this somewhat ineffective and 11 out of 25 considered it very ineffective. Of the 15 experts who answered the question, all answered that that they do not feel that the cooperation between the platforms and fact-checkers/researchers is effective and 10 out of 15 considered that they feel that the quality of the accessible data has not improved since the introduction of the Code (only 5 out of 15 considered that the quality of the data did improve.

The extra interviews conducted for this case study show that researchers face challenges when trying to cooperate with platforms and making data requests. Firstly, the interviewees expressed the specific concern that commercially, a lot of the data they need is actually available, but that in their opinion the information should be available to all researchers for free and they shouldn’t have to pay for it. Further concerns stated were stated are that:

- There is no systematic access to this data and that it is based only on ad-hoc requests;
- The data is only outcome-oriented and not process-oriented (i.e. providing insights into how platforms address the issue of disinformation technically and how the algorithms they use work);
- The platforms can freely select the data they wish to share;
The data is only used by a small group of academics (i.e. social scientists, mostly from the political science area); and

The current system is too much centred on the United States and its academic institutions while largely ignoring academics from other parts of the world.

The interviewed researchers recommended that the general scientific standards for data sharing should be followed and that in this sense, lessons can be learned from other disciplines such as pharmaceutics, where companies are obliged by national regulators to provide detailed information based on public health concerns, although the concerns are not equally big for platforms, there is still a rather big concern in terms of democratic values.

On the other hand, platforms mentioned in interviews that researchers could also be clearer about what data they need exactly and for what kind of purposes. This goes back to the fact that it will probably not be easy to convince platforms to share all their data to everyone, but it might be more acceptable for them to share specific pieces of data for specific purposes. Right now, platforms perceive that requests for data from researchers are rarely unified and consistent, and that the researchers are sometimes unclear about what they expect to get out from the data. They also pointed to the different standards needed for sharing public data (for which the burden in term of efforts and privacy protection is less) and private data which requires much more consideration.

However, the main problem identified by the research team of this study is that the Code of Practice in its current set-up leaves much of these decisions up to the platforms. Looking at the first commitment for Pillar V, it is a very good idea to share privacy protected datasets, undertake joint research, or otherwise partner with academics and civil society organisations if relevant and possible. However, it is currently up to the platforms to decide what relevant and possible means. Specifically, it is up to the platforms to decide which cases constitute “good faith independent efforts,” the underlying condition for accepting data requests according to the Code.

Addressing future needs

Going forward, one of the solutions could be independent oversight to data requests made by researchers. In this way, it is not up to the platforms to decide which requests should be honoured. This responsibility could for instance be given to independent regulatory authorities (convened in the European Regulators Group for Audiovisual Media Services, ERGA) as part of their enhanced responsibilities as proposed by this study.

However, at the same time there always needs to be some academic oversight on the requests; considering this second part, a good practice to be replicated is Social Science One, a new type of partnership between academic researchers and the private sector to advance the goals of social science in understanding and solving society’s greatest challenges. A commission of senior academics, who have signed confidentiality agreements and forego the right to publication, acts as the trusted third party and as a kind of broker between the private companies (e.g. platforms) on the one hand and academics

224 https://socialscience.one/.
on the other hand. The commission and partnering companies firstly agree on the scope of a research project, after which the company provides the commission access to relevant company data and additional support by answering questions. The commission identifies relevant datasets and issues Requests for Proposals to which outside academics can respond. The academics to undertake the research are then selected by the commission based on academic and social merit. The selected academics in turn receive privacy-preserving access to the data and retain the freedom to publish research findings on agreed topics without pre-approval from the companies. If the company breaches this agreement in any way, such as withholding relevant information, the commission has the right and the obligation to publicly report the violation.

One particularly successful initiative to mention here is that in April 2018, a project on “The effects of social media on democracy and elections” was launched through the system together with Facebook.225 This process help academics to be clearer in communicating towards the platforms and organise or unify themselves to do so. This way, the collective bargaining negotiation process is done centrally rather than on an ad-hoc basis for each request as currently is the case. Even though the system Social Science One provides is not perfect and there are researchers who are critical of it, it provides a first step in the right direction in terms of its set-up at least.

The platforms pointed out that when considering introducing such oversight it is important to avoid slowing down the dynamic nature of partnerships between academics and platforms. They also pointed out that it runs a significant risk of making such relationships significantly more burdensome and less attractive to enter into in the European Union than in other jurisdictions.

In conclusion, some oversight should be introduced on the relationship between academics and platforms and the topic of data requests. This oversight must be both regulatory from independent regulatory authorities (convened in the European ERGA organisation) to ensure that the data requests are useful to make studies to inform policy and academic to ensure that all requests serve sound academic purposes. This will also improve the cooperation between academics and regulators leading to more efficient and evidence-based policy intervention.

List of interviews conducted

<table>
<thead>
<tr>
<th>Type of stakeholder</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>George Washington University</td>
</tr>
<tr>
<td>Academic</td>
<td>George Washington University</td>
</tr>
<tr>
<td>Academic</td>
<td>Copenhagen Business School</td>
</tr>
<tr>
<td>Platform</td>
<td>Microsoft</td>
</tr>
</tbody>
</table>

225 [https://socialscience.one/our-facebook-partnership](https://socialscience.one/our-facebook-partnership)
Case study on partnerships with fact-checkers

This case study investigates the topic of partnerships between the Signatories of the Code of Practice on Disinformation (platforms, associations) and fact-checking organisations. It aims at assessing to which extent such partnerships already exist, whether they are effective in verifying and promoting trusted content and what could be done to improve them.

Focus

This topic is closely linked to the commitments of the Code regarding scrutiny of ad-placements given that partnerships with fact-checkers is mentioned as a possible format within which the policies and processes used to disrupt advertising could take place (II.A.1. of the Code). This topic is also related to the commitment of the Code under Pillar V, which aims at empowering the research community, given that further partnerships would help “support good faith independent efforts to track Disinformation and understand its impact, including the independent network of factcheckers facilitated by the European Commission upon its establishment”. Finally, it also relates to the pillar IV of the Code, which aims at empowering the consumers, as partnerships may lead to the creation of tools to provide consumers with the information they need to help assess the trustworthiness of sources of news they see on the platform.

For this case study, we used mainly information from the signatories’ monitoring reports and from the interviews with platforms and fact-checkers.

Current situation

One of the challenges of this case study is the lack of definition of trusted content or trusted information providers. Several interviewees suggested that it would be useful to have common definitions on what is trusted content, to coordinate more effectively across Member States.

Several platforms, signatories of the Code, have put in place effective partnerships with fact-checking organisations to combat disinformation. It is notably the case of Microsoft, which is collaborating with News Guard Technology, a tool which reviews online news sites across a series of nine journalistic integrity criteria, such as whether the site regularly publishes false content, reveals conflicts of interest, discloses financing, and publicly corrects reporting errors. Microsoft is partnering with NewsGuard to provide a free plug-in for the Microsoft Edge web browser (also available for other browsers including Chrome and Firefox), as well as an opt-in news rating feature for the Edge mobile application on both iOS and Android. This tool empowers Edge users to benefit from the analysis done by NewsGuard and to better identify the most reliable news and information sites.

As of the end of August 2019, NewsGuard had rated the 2,805 websites responsible for 96.01% of the news and information consumed and shared online in the United States. Of
those sites (many of which are also popular in the EU), 834, or 29.7%, were given an overall Red reliability rating. As of the end of August 2019, NewsGuard has also rated, in the native language, 700 sites responsible for 90% of the news and information consumed and shared online in the United Kingdom, Germany, France, and Italy—countries where NewsGuard launched, with analysts and editors native to those countries, in May 2019.229 On top of this tool, and as reported by NewsGuard itself, Microsoft and NewsGuard have held common events on disinformation, including in Brussels, to train representatives of the Silicon Valley platforms on the importance of labelling news sources at the domain, or website, level so that all news from untrustworthy sites comes with warnings to news consumers to proceed with caution, as Microsoft mobile Edge does. Microsoft has also worked with NewsGuard to protect Microsoft’s brand safety so that its advertisements only appear on trustworthy news websites.

In November 2018, Gallup measured the impact on news consumers of having access to NewsGuard’s ratings and labels230 and it found for example that 91% of the survey respondents find the NewsGuard Nutrition Labels helpful; 90% generally agree with the ratings and respondents trusted the ratings more because NewsGuard ratings are done by “trained journalists with varied backgrounds and 69% would trust social media and search companies more if they took the simple step of including NewsGuard in their products. Both Microsoft and NewsGuard have indicated to very satisfied with this partnership, which is estimated to be very effective. Microsoft indicated to be working with NewsGuard to make the tool evolve continuously.

Facebook also started a third-party fact-checking programme in December 2016, within which it partnered with fact-checkers helping to identify false stories so they can be stopped from spreading on Facebook231. The programme in the EU includes Croatia, Czechia, Denmark, France, Germany, Greece, Ireland, Italy, Lithuania, Poland, Portugal, Slovakia, and Spain. According to an interview with Facebook, the platform is also launching cooperation with partners in Belgium, the Netherlands and in Latvia in March 2020.232 All fact-checkers partnering with Facebook are independent and certified through the International Fact-Checking Network (IFCN).233,234 Facebook’s fact-checking program uses a combination of technology and human review to detect and reduce the spread of false news stories. When fact-checkers rate an article as false, Facebook shows it lower in News Feed — reducing future views by over 80% on average.235

The platform also uses the information from fact-checkers to train its machine learning model, so that it can catch more potentially false news stories and do so faster. It was reported by Facebook that, in practical terms, Facebook’s day-to-day activities with fact-checking organisations have included for example: an annual regional two day meetings with its partners from Europe to discuss Europe specific partner questions, exchange ideas

---

229 https://www.newsguardtech.com/press/newsguards-first-year/
231 The full list of partners of Facebook is available here: https://www.facebook.com/help/publisher/182222309230722
233 https://www.poynter.org/ifcn/
234 International Fact-Checking Network focuses on training fact-checkers around the world, translating the IFCN’s Code of Principles into ten languages, and providing tools and training to the fact-checking community.
and educate partners with existing and new tools and an annual two day meeting with all fact-checking partners globally to build capacity and strengthen the fact-checking community.

Several studies from academic and journalistic institutions, using different methodologies, established that the overall volume of false news is trending downward since the 2016 US election, notably on Facebook. The researchers found that on Facebook, interactions with these false news sites declined by more than half after the 2016 election, suggesting that “efforts by Facebook following the 2016 election to limit the diffusion of misinformation may have had a meaningful impact.”

Pagella Politica, which is partner of Facebook in the so-called Third-Party Fact-Checking project, also noted positive results of the partnership with Facebook. This partnership consists of flagging and fact-checking material (articles, videos, images) suspected to convey false information. Pagella Politica writes articles and uses a tool provided by Facebook to collect the suspected material, evaluates it and links it to its relevant fact-checking articles. The partnership also involves a variety of means in order to coordinate communication, such as a mailing list, a Slack channel, periodic meetings (both online and offline). Pagella Politica notes that the Facebook tool has given them useful resources, privileged access to information and a closer connection with other fact-checking projects.

Google also partnered with the International Fact-Checking Network. In collaboration with this Network, Google News Lab launched FactCheck EU in March 2019 to provide fact checks from 19 organizations from 10 countries in 11 languages. Google also worked with First Draft to provide training boot camps for journalists in Frankfurt, Brussels, Milan, and Madrid. Google also introduced new tools for researchers and the fact checking community: a ‘Fact Check Explorer’, which allows for exploration of Fact Checking journalism, and the ‘Fact Check Markup Tool’, which allows fact-checkers to easily mark their own articles as fact-checks in a way that is machine-readable via the ‘ClaimReview’ mark-up. Pagella Politica also works with Google in using the Fact-Check Markup Tool. However, the fact-checking organisation noted that this is a rather loose partnership, without formal agreements, nor meetings. Pagella Politica also notes that for its partnership with Google, it is difficult to evaluate the objective effectiveness of the partnership (i.e. relatively to the general public), since the fact-checking organisation does not have access to the relevant data. It is thus difficult to say how Pagella Politica’s articles would spread without the use of the Markup Tool.

Despite the policies and processes put in place through partnerships between the Signatories and the fact-checkers, several interviewees estimate that further efforts remain to be done to empower researchers and consumers and to render the partnerships with fact-checkers more effective. Although researchers interviewed during our study on the Code of Practice on Disinformation agreed that fact-checkers and researchers have increasingly used the few data available to conduct investigations, or

---

237 Google self-assessment report
Assessment of the implementation of the Code of Practice on Disinformation

report on abusive behaviour, in particular from political parties, activist groups and political candidates, as reported by a survey respondent, the vast majority of stakeholders agreed that the pillar of the Code aiming to empower the research community is the one which has proven to be the least advanced. Most of the survey respondents (expert public) conducted in the context of the study on the Code of Practice on Disinformation estimate that the cooperation between platform and fact-checkers/researchers is not effective (84%).

A fact-checking organisation noted that many platforms still do not partner with fact-checkers, with concrete actions such as the implementation of tools and they only donate funds to these organisations. A platform interviewed also noted that there are difficulties which might hinder the development of effective partnerships with fact-checkers. First of all, fact-checkers do not exist in all countries, and different places have different standards of journalism as well as varying levels of press freedom. Even where fact-checking organisations do exist, they are not enough to review all potentially false claims online. It can take hours or even days to review a single claim. And most false claims are not limited to one article — they spread to other sites. There are other challenges, too, such as how to treat opinion and satire. The interviewee noted that there can be a fine line between misinformation and satire or opinion. For example, sometimes people try to call their sites “satire” as cover for their true motivation — to spread fake stories. This can make it more difficult for fact-checkers to assess whether an article should be rated “false” or left alone.

A fact-checker stressed that one of the factors hindering the development of partnerships between platforms and fact-checkers is notably the fact that there is a lack of data on the effectiveness, or lack thereof, of these partnerships, which then puts a brake on the knowledge and support for these partnerships by the general public. The fact-checking community is more generally willing to expand the collaboration to other platforms and stakeholders, compared to the platforms.

Addressing future needs

It was noted by a fact-checker that the EU could incentivize the creation of partnerships between the platforms and fact-checking organisations by enforcing the existing duties in the Code of Practice on Disinformation, notably the duties set in the commitment to empower consumers. To fulfil this commitment, and increase the effectiveness of the partnerships, journalistically based news reliability service could be used or created by platforms. Fact-checking tools should be made available to the users of all platforms that operate the largest social media. Moreover, it was suggested for the European Commission to use a unique reliability index to track the progress of the tools on each platform.

To improve the partnerships between platforms and fact-checkers, a fact-checker also suggested for more resources to be attributed tofact-checkers so that they can increase the number of articles written and investigations led. It was suggested for the EU to sponsor scientific research on the partnerships to show their (eventual) effectiveness. The side effect of such sponsoring, however, could be the lack of partiality produced as a result of the funded research. A platform agreed that financial support to the fact-checking network is important, notably during crisis situations, such as the one of the COVID-19 where accurate
information about the virus is essential. Several platforms agreed that boosting the collaboration and supporting a healthy ecosystem for fact-checking organisations is essential. Supporting and helping to build capacity for long-established fact-checking organisations and networks, like the International Fact-Checking Network could be a way to boost the ecosystem.

Another key element which needs to be strengthened to enable an effective collaboration between platforms and fact-checkers is the promotion of media literacy. A platform states that: “it is important for Europeans of all ages to acquire media literacy skills to enable them to meaningfully take part in society and to evaluate the credibility of information they encounter online and to access alternative points of view”. It was acknowledged that, in the long-term, media literacy initiatives are one of the best instruments to counter the malicious effects of online disinformation. Supporting quality journalism is also seen as part of building media literacy and supporting the fact-checking ecosystem.

A platform recommends that, to improve the partnerships, there should be more focus put on communicating and explaining the partnerships programmes from different angles, and on equipping fact-checking partners with the information they need to speak about the programme effectively. Another suggestion is to enable cross-fact checking partner coaching through a system of informal mentors. This could also include offering feedback to platforms on the onboarding processes of these fact-checkers and how the platforms can best support their work.

It was noted also that more alignment would be helpful on fact-checkers’ top priorities and their needs for tools to assist them with fact-checking.

Finally, it was raised that one of the factors that could improve the effectiveness of partnerships between platforms and fact-checkers is the trust that the public puts into fact-checkers. The fact-checkers selected by platforms are being approved by the International Fact-Checking Network, for example, in the case of Facebook and Google, and the platforms make sure that the fact-checkers have high standards of accuracy, fairness and transparency. However, some platforms noted that they receive accusation of bias despite measures taken to select the fact-checkers. A platform noted that it is important that consumers trust the fact-checkers so that their work and partnerships gain effectiveness.

It appears clear that despite existing partnerships between platforms and fact-checkers, further work should be done to strengthen the existing partnership and to enable the creation of more alliances combating disinformation. This should be done mainly through an increase of media literacy programmes, enabling the platforms and fact-checkers to communicate further the purposes and benefits of such partnerships to the public. An increased communication on the topic would also enable the public to better trust the fact-checking community and to increase the number of users of fact-checking tools.

---

238 In this context, WhatsApp announced a $1 million support to the International Fact-Checking Network to expand the battle against COVID-19 related misinformation. The grant will be used to support the journalism of the CoronaVirusFacts/DatosCoronaVirus alliance, the collaborative project that was launched by the IFCN in January and now includes more than 100 fact-checkers in 45 countries. Source available at: https://www.poynter.org/fact-checking/2020/ifcn-receives-1-million-from-whatsapp-to-support-fact-checkers-on-the-coronavirus-battlefront/
Democratizing simple and user-friendly fact-checking tools among all Signatory platforms could contribute to enhancing the impact of the fact-checking partnerships.

**List of stakeholders contacted**

<table>
<thead>
<tr>
<th>Type of stakeholder</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signatory platform</td>
<td>Facebook</td>
</tr>
<tr>
<td>Signatory platform</td>
<td>Microsoft</td>
</tr>
<tr>
<td>Fact-checking organisation</td>
<td>Newsguardtech</td>
</tr>
<tr>
<td>Fact-checking organisation</td>
<td>Pagella Politica</td>
</tr>
</tbody>
</table>
### ANNEX 2: LIST OF STAKEHOLDERS CONTACTED

The table below provides an overview of all the stakeholders that the study team contacted throughout the study. Some of the stakeholders did not wish to contribute to this study.

Table 10: List of stakeholders contacted

<table>
<thead>
<tr>
<th>Type of stakeholder</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signatory (online platform)</td>
<td>Facebook</td>
</tr>
<tr>
<td>Signatory (online platform)</td>
<td>Google</td>
</tr>
<tr>
<td>Signatory (online platform)</td>
<td>Mozilla</td>
</tr>
<tr>
<td>Signatory (online platform)</td>
<td>Twitter</td>
</tr>
<tr>
<td>Signatory (online platform)</td>
<td>Microsoft</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>European Association of Communication Agencies</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>European Digital Media Association</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>Interactive Advertising Bureau Europe</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>World Federation of Advertisers</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>SAR Marketing Communication Association</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>Association des agences conseils en communication</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>Union of Belgian Advertisers</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>Associste komunikacnich Agentur</td>
</tr>
<tr>
<td>Signatory (trade association)</td>
<td>DG CNECT</td>
</tr>
<tr>
<td>Academic</td>
<td>An academic from Cardiff University</td>
</tr>
<tr>
<td>Academic</td>
<td>An academic Oxford University</td>
</tr>
<tr>
<td>Media</td>
<td>News Media Europe</td>
</tr>
<tr>
<td>Academic</td>
<td>An academic from University of Utrecht</td>
</tr>
<tr>
<td>Academic</td>
<td>An academic from George Washington University</td>
</tr>
<tr>
<td>Academic</td>
<td>An academic from Aarhus University</td>
</tr>
<tr>
<td>Media</td>
<td>EBU</td>
</tr>
<tr>
<td>Media</td>
<td>ACT</td>
</tr>
<tr>
<td>Media</td>
<td>Euractiv</td>
</tr>
<tr>
<td>Media</td>
<td>A journalist contributing to La Stampa</td>
</tr>
<tr>
<td>Civil society / consumer organisations</td>
<td>Bureau Européen des Unions de Consommateurs</td>
</tr>
<tr>
<td>Sounding Board</td>
<td>International Fact-Checking Network</td>
</tr>
<tr>
<td>Sounding Board</td>
<td>European Federation of Journalists</td>
</tr>
<tr>
<td>Sounding Board</td>
<td>Copenhagen Business School</td>
</tr>
<tr>
<td>Regulators</td>
<td>AGCOM</td>
</tr>
<tr>
<td>Regulators</td>
<td>BAI</td>
</tr>
<tr>
<td>Fact-checkers</td>
<td>Fact-checkers from the Aristotle University of Thessaloniki</td>
</tr>
<tr>
<td>Non-signatory platform</td>
<td>Snapchat</td>
</tr>
<tr>
<td>Non-signatory platform</td>
<td>Reddit</td>
</tr>
<tr>
<td>Fact-checkers</td>
<td>Pagella Politica</td>
</tr>
<tr>
<td>Fact-checkers</td>
<td>Les Decodeurs</td>
</tr>
</tbody>
</table>
## Type of stakeholder

<table>
<thead>
<tr>
<th>Type of stakeholder</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Civil society organisations</td>
<td>Avaaz</td>
</tr>
<tr>
<td>Academic</td>
<td>An academic from the Bocconi University</td>
</tr>
<tr>
<td>Academic</td>
<td>An academic from the Sorbonne Nouvelle University</td>
</tr>
<tr>
<td>Media</td>
<td>AFP</td>
</tr>
<tr>
<td>Civil society organisations</td>
<td>European Digital Rights</td>
</tr>
<tr>
<td>Civil society organisations</td>
<td>Consumer Choice Center</td>
</tr>
<tr>
<td>Civil society organisations</td>
<td>Global Disinformation Index</td>
</tr>
<tr>
<td>Civil society organisations</td>
<td>Internet Society</td>
</tr>
<tr>
<td>Civil society organisations</td>
<td>Centre on Regulation in Europe</td>
</tr>
<tr>
<td>Regulators</td>
<td>CAS (Conseil Supérieur de l’Audiovisuel)</td>
</tr>
<tr>
<td>Regulators</td>
<td>National Media and Infocommunications Authority</td>
</tr>
<tr>
<td>Regulators</td>
<td>Rada pre vysielanie a retransmisiu</td>
</tr>
<tr>
<td>Regulators</td>
<td>Conseil supérieur de l’audiovisuel de la Communauté française de Belgique</td>
</tr>
<tr>
<td>Regulators</td>
<td>Krajowa Rady Radiofonii i Telewizji</td>
</tr>
</tbody>
</table>
ANNEX 3: BIBLIOGRAPHY


Camille François (Transatlantic Working Group), Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses

Chris Marsden and Trisha Meyer, Regulating Disinformation with artificial intelligence,

Code of practice on Disinformation Summary of the signatories’ first reporting – January 2019

COM (2018) 236 final

Constine, J., Facebook now has 2 billion monthly users… and responsibility, 27 June 2017

CSA, Recommendation no. 2019-03 of 15 May 2019 of the Conseil supérieur de l’audiovisuel to online platform operators in the context of the duty to cooperate to fight the dissemination of false information

DG Connect, 2018, Report of the independent High-level Group on fake news and online disinformation, A multi-dimensional approach to disinformation


European Commission Better Regulation Guidelines Tool #41. Monitoring Arrangements and Indicators ARRANGEMENTS AND INDICATORS


European Parliament - IPOL_STU(2019)608864_EN-Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States


Facebook Monthly Self-Assessment Reports, Code of Practice on Disinformation, 2019.

Facebook report on the implementation of the Code of Practice for Disinformation Annual Report, September 2019

Facebook, Facebook report on the implementation of the Code of Practice for Disinformation, September 2019

Gaia Giombelli and Erica Melloni, Country Case Study: Italy, Virtual Insanity, P30.


Google Annual Report, 2019

High level Group on fake news and online disinformation, A multi-dimensional approach to disinformation, 2018

Irish Government, First report of the interdepartmental group on security of Ireland’s electoral process and disinformation, June 2018


Microsoft Self-Assessment Report, Code of Practice on Disinformation, September 2019

Microsoft, Microsoft self-assessment and report on compliance with the EU code of practice on disinformation, October 2019.

Microsoft, Microsoft Self-assessment and Report on Compliance with the EU Code of Practice on Disinformation, 2019

Microsoft, Microsoft Self-assessment and Report on Compliance with the EU Code of Practice on Disinformation, 2019


Mozilla Self-Assessment Report, Code of Practice on Disinformation, September 2019

NATO StratCom Centre of Excellence (2019). Falling Behind: How Social Media Companies are Failing to Combat Inauthentic Behaviour Online. Available at: https://www.stratcomcoe.org/how-social-media-companies-are-failing-combat-inauthentic-behaviour-online


Ott 2016 The age of Twitter: Donald J. Trump and the politics of debasement http://www.tandfonline.com/doi/abs/10.1080/15295036.2016.1266686


Regulation on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Available at:


Solon, O. and Freedom house, Manipulating Social Media to Undermine Democracy, November 2017


Twitter Monthly Self-Assessment Reports, Code of Practice on Disinformation, 2019.

Twitter Self-Assessment Report, Code of Practice on Disinformation, September 2019


Weblinks


https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20Web_0.pdf


https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20Web_0.pdf


https://late.europa.eu

https://www.wired.com/2017/02/veles-macedonia-fake-news/


https://www.theguardian.com/world/2017/may/06/emmanuel-macron-targeted-by-hackers-on-eve-of-french-election
https://www.1843magazine.com/technology/rewind/the-true-history-of-fake-news
https://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/
https://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/

https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

Assessment of the implementation of the Code of Practice on Disinformation

https://techcrunch.com/2017/06/27/facebook-2-billion-users/
https://www.theguardian.com/news/series/cambridge-analytica-files
https://www.rusemb.org.uk/press/2029
https://www.siliconrepublic.com/companies/facebook-and-the-journal-team-up-on-fact-checking-project
https://www.facebook.com/business/help/1838453822893854
https://www.politifact.com
https://m.facebook.com/pg/DonaldTrump/posts/?ref=page_internal&mt_nav=0
https://socialscience.one/
https://factcheckeu.info/en/
https://firstdraftnews.org/
https://toolbox.google.com/factcheck/explorer
https://developers.google.com/fact-check/tools/api/
https://www.crowdtangle.com/features
https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one
Assessment of the implementation of the Code of Practice on Disinformation

https://transparencyreport.google.com/political-ads/region/EU
https://foreignpolicy.com/2020/03/31/hungarys-orban-given-power-to-rule-by-decree-with-no-end-date/
https://www.disinfo.eu/2019/06/19/eu-g7-regulations/
https://www.bee-secure.lu/fr/a-propos
https://www.msb.se/RibData/Filer/pdf/28698.pdf
https://what-thefake.com/qui-sommes-nous/
https://aboutbadnews.com/
https://www.ellinikahoaxes.gr/about-us/
https://faktograf.hr/
https://www.facebook.com/help/publisher/182222309230722
https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper
https://gcs.civilservice.gov.uk/guidance/resist-counter-disinformation-toolkit/
https://sharechecklist.gov.uk/
Loi organique n° 2018-1201 du 22 décembre 2018 relative à la lutte contre la manipulation de l’information (1). Available at:
https://www.legifrance.gouv.fr/eli/loi/2018/12/22/MICX1808387L/jo/texte
Assessment of the implementation of the Code of Practice on Disinformation


https://www.dpa.com/de/unternehmen/faktencheck/#faktencheck-regeln
https://foreignpolicy.com/2020/03/31/hungarys-orban-given-power-to-rule-by-decree-with-no-end-date/
https://www.eurosint.eu/articles/workshop-vienna-explores-use-ai-technologies-against-fake-news
https://mzo.gov.hr/news/media-literacy-days-to-focus-on-respect-in-media-recognising-fake-news/1746
https://um.dk/en/news/newsdisplaypage/?newsid=1df5adbb-d1df-402b-b9ac-57fd4485ffa4
https://www.disinfo.eu/resources/estonia-2
https://www.kaitseliit.ee/en/cyber-unit
https://www.disinfo.eu/resources/france-2
https://what-the-fake.com/
https://www.lemonde.fr/les-decodeurs/
https://www.liberation.fr/checknews,100893
https://observers.france24.com/fr/
https://www.dpa.com/de/unternehmen/faktencheck/#faktencheck-regeln
https://www.goethe.de/en/kul/bib/21001176.html
https://www.ellinikahoaxes.gr/
https://cmds.ceu.edu/misinformation-and-propaganda-through-eyes-hungarian-students
https://assets.gov.ie/39188/8c7b6bc1d0d046be915963abfe427e90.pdf
https://assets.gov.ie/39188/8c7b6bc1d0d046be915963abfe427e90.pdf
https://pagellapolitica.it/
https://www.bastabufale.it/
https://www.interno.gov.it/it/notizie/progetto-red-button-contro-fake-news
https://debunk.eu/
https://www.bee-secure.lu/fr/a-propos
https://www.aboutbadnews.com/
https://www.disinfo.eu/resources/portugal-2
https://poligrafo.sapo.pt/
https://stirioficiale.ro/informatii
https://www.disinfo.eu/resources/slovakia-2
https://business.facebook.com/events/387148322083792/
https://www.disinfo.eu/resources/spain-2
https://www.thelocal.se/20180115/sweden-to-create-new-authority-tasked-with-countering-disinformation
https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper
https://gcs.civilservice.gov.uk/guidance/resist-counter-disinformation-toolkit/
https://sharechecklist.gov.uk/
https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179102.htm
https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/Measuring%20the%20reach%20of%20fake%20news%20and%20online%20distribution%20in%20Europe%20CORRECT%20FLAG.pdf
GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.