

CoP Monitor: Baseline Reports

Assessment of VLOP and VLOSE Signatory
Reports for the Strengthened Code of
Practice on Disinformation



CoP Monitor

Baseline Reports

Assessment of VLOP and VLOSE Signatory Reports for the Strengthened Code of Practice on Disinformation

Kirsty Park
Stephan Mündges



About EDMO Ireland Hub

EDMO Ireland combats disinformation by bringing together fact-checkers, media literacy experts, technologists, and academic researchers. In addition to fact-checking, EDMO Ireland advances an evidence-based understanding of disinformation and works closely with relevant stakeholders to inform policy and advance media literacy. Coordinated by Dublin City University (DCU), the EDMO Ireland consortium includes the DCU Institute for Future Media, Democracy and Society (DCU FuJo), TheJournal FactCheck, NewsWhip, and the University of Sheffield. EDMO Ireland is co-funded by the European Union and is one of 14 hubs of the European Digital Media Observatory's (EDMO) network.

About GADMO

The German-Austrian Digital Media Observatory (GADMO) is an alliance of fact-checkers, media literacy experts and scientists taking a coordinated approach to combating disinformation and misinformation. To this end, the leading fact-checking organisations in the German-speaking world are collaborating for the first time: Deutsche Presse-Agentur (dpa), Agence France-Presse (AFP), Austria Presse Agentur (APA) and the non-profit independent newsroom CORRECTIV. They cooperate with communication and data scientists from the Institute of Journalism and the Faculty of Statistics at TU Dortmund University and the AIT Austrian Institute Of Technology. The Institute of Journalism is leading GADMO's work on monitoring platforms' policies against disinformation and the implementation of the Code of Practice on Disinformation in Germany and Austria. GADMO is co-funded by the European Union and is one of 14 hubs of the European Digital Media Observatory's (EDMO) network.

About the authors

Dr Kirsty Park is the Policy Lead at EDMO Ireland and a postdoctoral researcher at DCU's Institute of Future Media, Democracy and Society. Her work focuses on media regulation, specialising in the DSA and CoP on Disinformation, and she recently led a European Parliament study assessing the risks of gendered harms for women and girls on social media.

Stephan Mündges is the manager of the Institute of Journalism at TU Dortmund University and one of the coordinators of GADMO. Within the research team, Stephan oversees the monitoring of platforms and their countermeasures against disinformation. A seasoned journalist by training, Stephan worked for the German public broadcaster ZDF for 12 years, where he specialised in reporting on digital topics, platform regulation, and disinformation campaigns.

List of Contributors

Mato Brautovic, University of Dubrovnik / ADMO
Eileen Culloty, Dublin City University / EDMO Ireland Hub
Dawn Holford, University of Bristol
Anastasia Kozyreva, Max Planck Institute for Human Development
Stephan Lewandowsky, University of Bristol¹
Trisha Meyer, Vrije Universiteit Brussel / EDMO BELUX
Susanne Wegner, TU Dortmund University / GADMO



¹ Stephan Lewandowsky has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 101020961 (PRODEMINFO).

Executive Summary

After a revision of the 2018 Code, the “Strengthened Code of Practice on Disinformation” was adopted in 2022 by 34 (now 44) Signatories including major platforms. The updated Code introduces stronger monitoring requirements and reporting through quantitative Service Level Indicators (SLIs) and Qualitative Reporting Elements (QREs), increased support for fact-checking, the establishment of a Transparency Centre for Signatory reporting and commits Signatories to collaborate and continuously update the Code through a Permanent Task-force.

This analysis examines the baseline reports published by the largest Signatories of the Strengthened Code in February 2023: Google, Meta, Microsoft, TikTok and Twitter/X. So far, no comprehensive, in-depth analysis of these reports has been published. Through a robust assessment scheme and three-stage coding process, we evaluated both qualitative and quantitative information provided by Signatories and assigned an overall score of 1 (poor), 2 (adequate) or 3 (good) for each Measure.

Our findings show that the overall quality of reporting is less than adequate. More than half of the applicable Measures exhibited incomplete or missing qualitative information. Notably, quantitative data was absent in 64% of relevant instances, and when provided, the methodology was sometimes questionable with data often imprecise or missing on a Member State-level. Furthermore, Signatories frequently included information unrelated to the specific Measure in question. In the "Empowering Research" section, Signatories' performance was particularly poor, indicating a need for substantial improvement in this area.

Our analysis highlights a discrepancy between a formal, report-centred approach to monitoring and a more substantive, investigative approach. The former focuses on verifying the fulfilment of reporting obligations, while the latter delves into the veracity and comprehensiveness of the information provided by Signatories. The substantive approach demands greater resources, expertise, and data access but is essential for meaningful monitoring.

Based on our findings, we offer several recommendations for future reporting:

For Signatories and the Permanent Task-force

- **Relevance:** Signatories should ensure each response directly addresses the specific requirements of the Code, providing all requested information.
- **Quantitative Data:** Reporting of quantitative data must be more comprehensive and thorough. Methodologies should be precise, and Member State-level data should be provided.
- **Transparency of Decision-making:** Permanent Task-force decisions on actions or workgroups relevant to Signatory Commitments should be communicated publicly,

ideally through the Transparency Centre, along with a version of the Code that incorporates any agreed upon updates.

- **Subscription Exemptions:** If Signatories are subscribed to Measures which are not currently relevant to them they should either unsubscribe from these Measures or receive an exemption at a Permanent Task-force level after a thorough review of the Signatories' claims.
- **Code Review:** The Code should be regularly reviewed to ensure that the phrasing and wording used is clear and understood by Signatories. Throughout the report we have noted multiple sections of the Code where we would advise clarification or review.

Sections of the Code

- **Scrutiny of Ad Placements:** There were multiple SLIs completely missing in this section with others missing Member State level breakdowns, while Signatories overused external links in describing policies. Based on Signatory responses, there is a need to clarify the requirements and wording of the three Commitments in this section.
- **Political Advertising:** Future monitoring should ensure that Signatories are scrupulously enforcing any policies relating to the prohibition of political or issue advertising.
- **Integrity of Services:** Quality and methodological rigour of the quantitative data regarding actions taken against manipulative tactics, techniques and procedures must improve.
- **Empowering Users:** Signatories need to improve reporting on media literacy and the incorporation of scientific evidence and user needs into the design of labelling and warning systems. Microsoft should also re-examine the relevance of the features offered by Bing Search in regard to the potential of search engines to spread disinformation.
- **Empowering Researchers:** This section achieved the lowest score across the analysis due to a high rate of incomplete or missing SLIs (100% for all except Google's services) as well as irrelevant and incomplete QRE responses. Signatories must do more in this area, particularly considering their upcoming obligations under Article 40 of the DSA. The Permanent Task-force should explore how reporting can capture the sustainability of research efforts and Signatory changes regarding resourcing.
- **Empowering the Fact-Checking Community:** Signatories must ensure better fact-checking coverage for the entire EU and disclose (and likely increase) funding allocated to fact-checkers.

For all stakeholders

- **Structural Indicators:** All parties involved with the development and implementation of Structural Indicators should intensify their efforts, ensuring that they are precisely defined and cover all areas of the Code.
- **Systematic monitoring:** To be effective in countering disinformation, the Code needs to be monitored appropriately. This should include both monitoring of transparency reports and verifying the accuracy of Signatory reporting. The European Commission should ensure that such monitoring is sufficiently funded.
- **Member State level monitoring:** EDMO, ERGA and other stakeholders should assess how Member State level monitoring can take place in a useful and efficient manner.

In conclusion, while the baseline reports represent a good first step in addressing the obligations set out in the Code, the overall quality of reporting needs significant improvement. Addressing the identified gaps will contribute to a more robust, comprehensive, and effective implementation of the Code of Practice on Disinformation.

Table of Contents

List of abbreviations	9
List of tables	9
Foreword by Ireland Hub	10
Foreword by GADMO	11
Foreword by Stephan Lewandowsky	12
1. Introduction	14
1.1. Scope of this report	15
1.2. Structure of this Report	16
2. Methodology	17
2.1. Assessment unit and assessment scheme	17
3. Overall results	20
3.1. Overall results by Section	21
4. Results by Section of the Code	23
4.1. Scrutiny of ad placements	23
4.1.1. Scope of the Section in the Code	23
4.1.2. Overall assessment	23
4.1.3. Code specific remarks	25
4.2. Political advertising	25
4.2.1. Scope of the Section in the Code	25
4.2.2. Overall assessment	26
4.2.3. Signatory specific remarks	29
4.2.4. Code specific remarks	30
4.3. Integrity of Services	31
4.3.1. Scope of the Section in the Code	31
4.3.2. Overall assessment	31
4.3.3. Signatory specific remarks	33
4.3.4. Code specific remarks	34
4.4. Empowering Users	34
4.4.1. Scope of the Section in the Code	34
4.4.2. Overall assessment	34
4.4.3. Signatory specific remarks	36
4.4.4. Code specific remarks	37
4.5. Empowering Researchers	38
4.5.1. Scope of the Section in the Code	38
4.5.2. Overall assessment	38

4.5.3. Signatory specific remarks	40
4.5.4. Code specific remarks	40
4.6. Empowering the Fact-Checking Community	41
4.6.1. Scope of the Section in the Code	41
4.6.2. Overall assessment	41
4.6.3. Signatory specific remarks	43
4.6.4. Code specific remarks	44
4.7. Transparency Centre	45
4.7.1. Scope of the Section in the Code	45
4.7.2. Overall assessment	45
4.7.3. Code specific remarks	46
4.8. Permanent Task-force	46
4.8.1. Scope of the Section in the Code	46
4.8.2. Overall assessment	46
4.8.3. Code specific remarks	46
4.9. Monitoring of the Code	47
4.9.1. Scope of the Section in the Code	47
4.9.2. Overall assessment	47
4.10. Crisis Response	48
4.10.1. Scope of the Section in the Code	48
4.10.2. Overall assessment	48
5. Conclusions on Future Monitoring	49
6. Recommendations	51

List of abbreviations

CoP	Strengthened Code of Practice on Disinformation
DSA	Digital Services Act
EDMO	European Digital Media Observatory
ERGA	European Regulators Group for Audiovisual Media Services
NATO StratCom CoE	NATO Strategic Communications Centre of Excellence
QRE	Qualitative Reporting Element
SLI	Service Level Indicator
UGC	User-generated content
VLOP	Very Large Online Platform
VLOSE	Very Large Online Search Engine

List of tables

Table 1: Assessment scheme	18
Table 2: Overall results by Signatory	20
Table 3: Overall results by Section	21
Table 4: Results Scrutiny of ad placements	23
Table 5: Measure-specific results Scrutiny of ad placements	24
Table 6: Results Political advertising	26
Table 7: Measure-specific results Political advertising	27
Table 8: Results Integrity of Services	31
Table 9: Measure-specific results Integrity of Services	32
Table 10: Results Empowering Users	34
Table 11: Measure-specific results Empowering Users	35
Table 12: Results Empowering Researchers	38
Table 13: Measure-specific results Empowering Researchers	39
Table 14: Results Empowering the Fact-Checking Community	41
Table 15: Measure-specific results Empowering the Fact-Checking Community	42
Table 16: Results Crisis Response	48

Foreword by Ireland Hub

As we approach a critical year of European elections, there are major concerns about the potential for disinformation and manipulation to disrupt the democratic process. At the same time, there are great expectations for the Strengthened Code of Practice on Disinformation and the Digital Services Act (DSA) to protect European citizens and safeguard European democracies. Ireland has a particular obligation to exercise oversight in this area as many technology companies maintain their European headquarters in Dublin. Indeed, thirteen of the nineteen companies that have been designated as Very Large Online Platforms and Very Large Online Search Engines are based in Ireland.

Between 2019 and 2021, DCU FuJo published three reports on the implementation of the original Code. Each report was funded by the Irish media regulator, Coimisiún na Meán, to assess progress and identify recommendations for development. These reports highlighted many critical shortcomings, both in terms of the Code itself and, crucially, the approach to it by many Signatories. Unveiled in 2022, the Strengthened Code represented an important development towards more systematic and detailed reporting of relevant information. However, it remains up to the Signatories, and especially the major platform Signatories, to act in good faith by providing relevant information.

The present report assesses the first baseline reports submitted by the five biggest platform Signatories earlier this year. While we recognise that these baseline reports are the start of a process, it is also clear that Signatories have much more work to do to deliver meaningful data.

We were delighted to partner with GADMO to lead this investigation into the Strengthened Code and to work with colleagues from other hubs with the EDMO network and colleagues from universities outside the network. We hope this will be the beginning of a constructive process that helps to inform the development of the Code for the benefit of European citizens.

Eileen Culloty
EDMO Ireland Coordinator
Dublin City University

Foreword by GADMO

A pro-Russian disinformation campaign² promoting fake government and news websites as well as content by far-right politicians is one of the more recent examples. Mis- and disinformation is a serious problem in Germany and Austria. In this, digital platforms play an important role – both in hosting and spreading disinformation but also in countering and combating it.

That is why GADMO engages in monitoring platforms' policies and enforcement actions in this field. And it is the reason why we, together with our partners, monitor the reporting on and implementation of the Code of Practice on Disinformation by the largest platforms.

Proper monitoring is crucial for the effective implementation of the Code. This report was made possible through a voluntary collaboration, facilitated by the presence of the European Digital Media Observatory (EDMO) network of hubs. To work with the EDMO Ireland Hub and the other contributors, who lend us their time and expertise, was both inspiring and constructive. However, a more formalised and institutionalized approach is needed for sustained, effective monitoring of the Code.

GADMO is heavily invested in fact-checking. Four out of seven partners in our hub are the leading fact-checking organisations in Germany and Austria, every month we publish around 100 new fact-checks. We do so because fact-checking is one of the most valuable tools in the fight against disinformation. Unfortunately, as this analysis reveals, most platforms fall short in their efforts to support fact-checking. Coverage across all Member States is inadequate, and there is often a lack of transparency regarding funding for fact-checking organisations. When details are disclosed, the analysis shows: it is not enough.

While all platforms that are subject of this analysis engage in countering disinformation at least to some extent, mis- and disinformation in German roams quite freely on one platform that is not yet a Signatory of the Code: Telegram. We would welcome it if the relevant stakeholders would engage with Telegram and encourage the company to sign the Code of Practice.

We value the Code as an important (self-)regulatory framework to counter disinformation in the entire EU. As our analysis shows in detail, there are still several shortcomings in the baseline reports as well as in platforms' policies and, in particular, enforcement of such policies. We hope that this analysis might support all stakeholders in further developing and implementing the Code.

Stephan Mündges
GADMO Coordinator
TU Dortmund University

² <https://gadmo.eu/en/facebook-silent-as-pro-kremlin-disinformation-campaign-rages-on/>

Foreword by Stephan Lewandowsky

There are multiple indicators that democracy is in retreat worldwide. Even former beacons of democracy such as the UK and US have encountered events—the unlawful shutting down of Parliament in the UK and the violent insurrection on Capitol Hill in the US—that would have seemed unthinkable just a few years ago and that are difficult to reconcile with a healthy democracy.

There are numerous complex and intertwined reasons for democratic backsliding that cannot be readily disentangled. However, many scholars identify the advent of social media and the deluge of misinformation and disinformation during the last 10-20 years as contributors to the decline of democracy. It is difficult to explain how a violent mob would storm the seat of American democracy unless incited to do so by what was clearly misinformation, namely the baseless claim by Donald Trump that the 2020 election had been “stolen” from him.

Although the scientific debate about the exact role of misinformation and social media in shaping societies continues to be hotly debated, several studies have shown that social media can causally contribute to problematic behaviours offline, such as hate crimes against refugees or against ethnic minorities. It is now also firmly established that political segregation into opposing polarised communities on Facebook is extensive and in part driven by algorithms.

The European Union is leading the global response to those problems through its recent regulatory initiatives, which range from the recent Digital Services Act (DSA) through forthcoming legislation on AI and political advertising to the Strengthened Code of Practice (CoP) on disinformation which most large platforms have voluntarily signed up for.

This document analyses the first baseline reports submitted by the five biggest platform Signatories in early 2023 on the implementation of the commitments they signed up to under the Code. Signatories had six months after signing the Code to put in place actions to fulfil their commitments, and these reports provide a first set of qualitative and quantitative indicators of how platforms have lived up to their promises.

The baseline reports follow a common harmonised template consisting of 152 distinct indicators of performance. Although these indicators are based on self-report, they are sufficiently diagnostic to provide at least a rough sketch of platform performance. For example, the inescapable conclusion of this document is that Elon Musk’s Twitter failed every single indicator and gave every impression of blatant non-compliance (e.g., by self-plagiarising much of their report from previously published boilerplate text).

Our analysis also showed that the remaining large platforms (Google, Meta, Microsoft, and TikTok) provided good-faith reports into their activities, which, on average yield a glass that, depending on one’s expectations, is either half full or half empty.

The first baseline reports filed under the Strengthened Code of Practice present a step in the right direction. Our analysis also identified ways in which the process could be improved, which range from streamlining reporting to enhancing audit capabilities.

Those recommendations and Twitter's non-compliance suggest that a further necessary step forward is for the Code of Practice to become recognised as a 'Code of Conduct' under the DSA, in which case compliance with the Code would be subject to independent audits. Current timelines suggest that this could happen in 2024, and we hope that the detailed analyses reported in this document underscore the need for this transformation.

Stephan Lewandowsky
Chair in Cognitive Psychology
University of Bristol

1. Introduction

Disinformation, broadly defined, is a serious threat to free and democratic societies. False or misleading information is deliberately spread to undermine trust in public institutions, to incite violence or poison public debates. The perception of public opinion is distorted through manipulative techniques employed by malign domestic and/or foreign actors using digital communication platforms. In short, disinformation is a multi-faceted challenge for the European Union and its Member States. The EU institutions have adopted several measures to counter disinformation. One of them is the Code of Practice on Disinformation (hereinafter referred to as the Code or the CoP), a self-regulatory framework which was voluntarily signed by stakeholders in the digital information space. Published first in 2018, the Code has been further developed and has been re-published as the “Strengthened Code of Practice on Disinformation” in 2022. In this report, we always refer to this Strengthened Code not the former version.

The CoP defines disinformation as including “misinformation, disinformation, information influence operations and foreign interference in the information space”. (CoP, I.(a)) Its nine different sections cover a wide range of issues which can be broadly categorised into three classes: reaction, empowerment and monitoring. The first three sections (Scrutiny of Ad Placements, Political Advertising, Integrity of Services) address possible ways in which disinformation may be monetised and spread or manipulative tactics and techniques may be applied. The next three sections (Empowering Users, Empowering the Research Community, Empowering the Fact-Checking Community) oblige Signatories to foster cooperation with and ensure support for various stakeholder groups. The last three sections (Transparency Centre, Permanent Task-force, Monitoring of the Code) deal with various issues how the Code should be monitored, maintained and its results publicised. Each section is clearly structured into three levels: The first level consists of Commitments, which outline overall aims and actions that Signatories commit to. These are then elaborated upon in more detail in Measures - concrete steps the Signatories are obliged to take. In order to hold Signatories accountable to the Commitments and Measures they signed up to, there are detailed reporting elements which have to be fulfilled. They are either so-called “Qualitative Reporting Elements” (QRE) or “Service Level Indicators” (SLI). The latter can also be described as quantitative data points.

The Code currently has 44 Signatories, some of which are the world’s biggest technology companies. As several Commitments and Measures are very much tailored to be applied to large digital platforms, like social networks or platforms hosting and distributing user-generated content (UGC) it is fair to argue that holding Big Tech to account in the fight against disinformation is the prime objective of the CoP.

While participation in the Code remains voluntary, the Digital Services Act (DSA) which will be fully applied by early 2024, introduces a number of obligations relating to mitigating systemic risks for Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) with over 45 million users. Article 45 of the DSA encourages the development of voluntary codes of conduct to help develop risk mitigation measures and reporting frameworks and it cites the CoP as an example of a potential code of conduct. If the CoP becomes a code of conduct, then non-compliance or a refusal to participate may be

considered as an infringement of obligations potentially resulting in significant financial penalties. Considering that the CoP will likely be one of the first codes of conduct under the DSA, it is particularly important that the CoP is phrased without ambiguities, adhered to accurately by Signatories and monitored continuously and systematically.

1.1. Scope of this report

This report analyses the baseline reports submitted by the five biggest platform Signatories: Google, Meta, Microsoft, TikTok, and Twitter. Twitter³ was included in the analysis, as the work began before the company announced that it would leave the CoP.

Except in the case of Twitter, we are aware that this is the first iteration of the reports and that the reporting procedures as well as the contents reported will be further developed in the future. We believe that a critical, yet constructive analysis of the baseline reports can aid all stakeholders in compiling more concise and actionable reports in the future.

This analysis focuses mostly on the shortcomings of Signatories' reports and in some instances on the lack of clarity of the CoP. Before diving into this criticism, all stakeholders involved, including the European Commission, the Permanent Task-force and the Signatories are to be commended on their work towards the Strengthened Code of Practice on Disinformation.

As also pointed out in the European Regulators Group for Audiovisual Media Services' (ERGA) "report on the one year since the publishing of the Strengthened Code of Practice on Disinformation"⁴, the Strengthened Code implements several recommendations made by stakeholders and monitoring bodies in the past. Most importantly, the Strengthened Code obliges Signatories to report more granular and country-specific data. In comparison to its first iteration, the Strengthened Code is better structured, more detailed and clearer in what information has to be provided by Signatories.

Signatories have also launched a Transparency Centre online⁵, which provides an accessible, public resource for viewing Signatory reports and actions taken as part of the Code, a commendable and significant step forward.

In comparison to reporting under the original code, the harmonised reporting template of the baseline reports, which was developed with the support of ERGA, has led to more structure and clarity and the overall quality of information provided is much higher. The improved accessibility to Signatories' reports through the Transparency Center is an improvement as well.

We see this first round of reports as a positive step forward in the process of the Code but we hope to see further improvement as the Strengthened Code matures. Our findings may offer some guidance towards that.

³ The platform is now called X, but as it both committed to the Code of Practice as "Twitter" and withdrew as "Twitter" we continue to call it that for the purpose of this report.

⁴ https://rpms.sk/sites/default/files/2023-07/SG3_ERGA%20report%20CoP_final_for%20adoption.pdf

⁵ <https://disinfocode.eu/>

1.2. Structure of this Report

This report is structured and written so that relevant information for the different stakeholders is easy to find and quick to look up. In the first step we outline our methods applied and explain the reasoning behind our methodological choices. The bulk of this report consists of the Results section. We start by presenting overall results, followed by subsections corresponding to the different sections of the CoP.

The first six subsections are structured the same way:

- Scope of the Section in the Code
- Overall Assessment
- Signatory Specific Remarks (if applicable)
- Code Specific Remarks

For the last three subsections, which are all focussed on different monitoring procedures of the Code, we provide some general commentary but no numeric results as most Measures did not require individual reporting elements and some lacked Measures altogether. This is followed by a section evaluating the two submissions made by each Signatory on their response during a period of crisis, one covering the COVID-19 pandemic and the other covering the War of aggression by Russia on Ukraine.

The Results section is followed by Conclusions on Future Monitoring and Recommendations.

In every case where individual Signatories are discussed we address the Signatories in alphabetical order of the company names (i.e. Google, Meta, Microsoft, TikTok, Twitter).

2. Methodology

The focus of this analysis is on platform Signatories, designated as VLOPs / VLOSEs under the DSA, as these are known to be used by bad actors and unwitting users to spread dis- and misinformation. Furthermore, these platforms deserve the most scrutiny as they have the largest user bases. Comparing the list of Signatories of the Code of Practice⁶ with the list of VLOPs and VLOSEs⁷ results in a list of five companies: Google, Meta, Microsoft, TikTok and Twitter.

2.1. Assessment unit and assessment scheme

The Code of Practice is highly detailed, setting many requirements for the Signatories in terms of information and data provision. However, the level of detail varies considerably among the different Measures, QREs and SLIs. Similarly, the specificity of the information provided by the Signatories also varies significantly.

Assessing the quality of the information provided by the Signatories often requires extensive background knowledge. The analysis of the baseline reports frequently necessitates normative judgments. To simultaneously develop a structured, comprehensive, and as objective as possible analysis process, which delivers actionable results that do justice to the detail level of the Code and the baseline reports, we devised a customised assessment procedure.

We decided to focus on the Measure level as a unit of assessment. The Commitments were deemed too broad for our purposes, while reporting elements were considered too specific. Given that reporting elements are operationalizations of Measures, and this being the first round of reports, analysing at the measure level seems the most suitable initial step. It offers an objectively analysable level of detail that is fair and not excessively detailed.

One issue that arose is that some Measures share QREs and SLIs (for example, Measure 31.1 and 31.2). This is a factor to consider in future iterations of the analysis. In subsequent rounds, it might be valuable to consider coding at the reporting elements level to provide more detailed feedback and gain deeper insights into the effectiveness of the Code.

The second step involved developing an assessment scheme. For this, three quantitative and two qualitative variables were identified. Details can be found in the table below.

⁶<https://digital-strategy.ec.europa.eu/en/library/signatories-2022-strengthened-code-practice-disinformation>

⁷https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413

Table 1: Assessment scheme

Variable	Expression	Explanation
Missing qualitative responses	yes	Information that is requested in a Measure or QRE is not provided comprehensively and/or the information provided by the Signatory does not meet the obligations set out in the Measure or QRE. As some QREs cover several aspects of an issue, this variable should also be coded as “yes” when the Signatory fails to provide information of one or more relevant aspects.
	no	The qualitative responses provided by the Signatory are comprehensive, concise and to the point.
	irrelevant	If a Measure does not contain QREs, this variable might be coded irrelevant.
Missing quantitative data	yes	SLIs are ignored or relevant data that is requested as part of a Measure or SLI is not provided by the Signatory. This might also be the case if Member State specific data is requested but not provided. The variable is also to be coded “yes” if data provided is unsuitable or the methodology is ill-suited to provide reliable data.
	no	Quantitative data provided by the Signatory is comprehensive and the methodology to collect or calculate the data is appropriate.
	irrelevant	Measure does not contain SLI or there is reasonable ground why an SLI cannot be provided at this point in time.
Overall Score	1	Poor: The response significantly falls short of meeting the requirements of the Measure. This is the case for responses that lack major details, are incomplete or irrelevant, or fail to address the specific information requests outlined in the measure.
	2	Adequate: The response shows effort towards meeting the requirements of the Measure but there are notable issues or areas that require improvement. This is the case for responses that partially address the question, but may lack important details, evidence, or context.
	3	Good: The response fully meets the requirements of the Measure. This is the case for responses that are complete, relevant, and provide clear and comprehensive information that directly addresses the specific information requests outlined in the measure.
	n/a	Not Applicable: If a Signatory claims a Measure they subscribed to is not relevant to their services and the Assessor believes this claim to be correct e.g. the

		Measure relates to displaying information alongside political advertising and the Signatory's product does not allow political advertising.
Signatory level comments	free text	Qualitative comments explaining the Assessor's results for the previous three variables, e.g. 'Lacked detail on x'. Assessors should include any observations or feedback on the response.
Code level comments	free text	Qualitative comments regarding the CoP; this might include comments on wording, structure, needed clarification etc.

This scheme provides quantitative feedback, assessing the overall worth of the baseline reports, along with those by each Signatory included in this analysis. Concurrently, it offers qualitative insights into deficiencies and outlines practical steps to enhance future reporting.

To ensure that results are as objective as possible, we implemented a three-stage coding process. Initially, two assessors independently coded specific sections. Each assessor also had the opportunity to flag their product-specific coding for individual Measures as “Needs a second opinion”. After the completion of these preliminary two coding rounds, a final third round was conducted to confirm the final results. In this concluding stage, the authors of this report evaluated every instance of disagreement between the first two assessors taking into account assessor comments and the content of the reports to reach a final result. Additionally, the third assessor checked coding overall for inconsistencies.

3. Overall results

Table 2: Overall results by Signatory

Signatories	% of Measures missing qualitative information	% of Measures Missing quantitative data	Average Overall Score
Google	44%	53%	2.1
Meta	52%	65%	2.0
Microsoft	54%	59%	1.9
TikTok	43%	50%	2.0
TOTAL w/o Twitter	49%	58%	2.0
Twitter	100%	100%	1.0
TOTAL	55%	64%	1.9

Note: Scores correspond to grades of 1 = Poor, 2 = Adequate and 3 = Good.

While Twitter is no longer a Signatory to the Code, the analysis shows that it was already not engaging with the process well before its withdrawal. Twitter's baseline report achieved the lowest possible overall score of 1 (Poor) and a 100% rate of missing or incomplete QREs/SLIs across all measures. In their report, Twitter left many sections blank and provided only general comments on some Commitments. In cases where information was supplied, it often appeared to be a simple copy-and-paste from their own website. It is fair to infer from the report that even before withdrawing as a Signatory, Twitter did not intend to comply with the Code when submitting their report. For this reason, we will not discuss Twitter's results in following analysis sections, focusing instead on remarks directed to the remaining Signatories.

With Twitter excluded, the average score among Signatories is 1.9, less than adequate, and no Signatory achieved more than an overall grade of 2 (Adequate) except for Google with 2.1. In Measures containing QREs, 49% were considered to have either incomplete or missing qualitative information while the rate was 58% for quantitative elements in Measures containing SLIs, showing much room for improvement.

Relevance is still an issue, with Signatories often providing information which was not directly relevant to the QRE/SLI in question. For example, responses asked to focus on processes for implementing policies may instead focus on details of policy. There was also a tendency for Signatories to redirect to either outside links such as policy pages on a website or responses to other QREs. Assessors felt that many redirected answers were only partially relevant to the QRE/SLI in question, which left the burden of extracting the relevant information to assessors. Signatories must also bear in mind that future monitoring methodologies may employ strategies such as assigning different assessors to various sections of the Code and so where possible, each QRE or SLI response should be complete and self-contained.

The lack of quantitative data is considerable and not acceptable. In some instances the methodology used was questionable while in others the data was not precise. For several SLIs no data on Member State level was provided. In certain cases, quantitative data was missing entirely. Signatories providing reliable, comprehensive data on an ongoing basis is

key for monitoring the Code and evaluating the Signatories' actions to counter dis- and misinformation.

3.1. Overall results by Section

Table 3: Overall results by Section

Section	% of Measures missing qualitative information	% of Measures Missing quantitative data	Average Overall Score
Scrutiny of ads placements	58%	64%	1.9
Political Advertising	56%	0%	2.2
Integrity of services	56%	94%	1.9
Empowering Users	37%	52%	1.9
Empowering Researchers	83%	89%	1.6
Empowering the Fact-Checking Community	63%	72%	1.7

Note: Scores correspond to grades of 1 = Poor, 2 = Adequate and 3 = Good.

The table shows that there are considerable differences in how well Signatories meet their obligations in the first six sub-sections, which represent the major pillars of the Code and contain most QREs and SLIs. While qualitative information is missing in all sections, only 37% of Measures in the “Empowering Users” section is lacking at least some qualitative information compared to 83% for the “Empowering Researchers” section. “Empowering Researchers” is also the worst rated section at 1.6, with QRE responses that generally failed to address all the requested information while SLIs were incomplete for all Signatories except for Google. With the exception of Political Advertising, all sections scored just below adequate with a score of less than 2.

The “Empowering the fact-checking Community” section was the second lowest at 1.7 with qualitative information as well as quantitative data lacking in a significant number of Measures. Also notable is the lack of quantitative data (94%) in relevant Measures in the “Integrity of Services” section. This can be attributed to missing data on a Member State level, a poor methodology when computing data or, in some instances, a complete absence of data.

Overall, Signatories performed best in the “Political Advertising” section, with a score of 2.2. Quantitative data requirements have been met by those Signatories that subscribed to the Commitments of this section. However, three out of five Signatories opted out of most Measures which they had subscribed to in this section, claiming that they were not relevant to their services.

The process of analysis stressed the difference between a formal, report-centred approach to monitoring and a more substantive, investigative approach. While the former checks whether all reporting obligations are met by Signatories and whether information and data provided are relevant and concise, the latter is concerned with investigating the veracity and comprehensiveness of information and data reported by Signatories. Of course, the substantive, investigative approach is by far more laborious, requires more expertise in all

the different fields covered by the Code and needs appropriate resources as well as data access.

Having conducted a first comprehensive formal and report-centred monitoring, it is clear to us that a more investigative approach is vital to effectively monitor the Code. Though we did not dive deep into most issues, we identified several instances where the assertions made by Signatories in their reports might not align with the actual situation. For example, the purported enforcement of TikTok's prohibition on political advertising appears to be inadequate (refer to p. 28). There's also a noticeable disparity between what Signatories assert they do to combat Fake Engagement and research findings that challenge the effectiveness of these counter-measures (see p. 49). Furthermore, fact-checkers have already critiqued the statements made by Signatories concerning the commitments in the "Empowering the Fact-Checking Community" section (see p. 43).

With regards to the monitoring process, our analysis also found that there is room for improving the institutional framework in which monitoring is carried out. A clear delineation of competencies and responsibilities in the monitoring process is needed. The involvement of multiple entities—including the EU Commission, ERGA, EDMO, the Permanent Task-force, and private contractors—adds layers of complexity and dilutes responsibilities and resources. Should the CoP transition into a code of conduct under the DSA, this situation might even further be complicated. Thus, it is important that both formal and investigative monitoring are conducted on a regular basis and that responsibilities for such monitoring are clearly defined. In addition, comprehensive Structural Indicators should be swiftly designed and implemented, as they could provide a foundation for more substantive and effective monitoring (for details, see p. 47).

4. Results by Section of the Code

4.1. Scrutiny of ad placements

4.1.1. Scope of the Section in the Code

Disinformation can be spread via or monetized through advertising. In order to limit both, Signatories actively involved in the purchasing, selling, and placement of digital advertising are committed to enhancing their policies and systems. This improvement pertains to determining the eligibility of content for monetisation, the controls for monetisation and ad placement, and the data reporting on the accuracy and effectiveness of these controls and services around ad placements.

Signatories also agree to take the necessary steps to provide independent third-party auditors with access to their services and data. Furthermore, Signatories pledge to advance the development, improve the availability, and take practical steps towards the use of brand safety tools and partnerships, as well as adapt their current ad verification and review systems if necessary.

4.1.2. Overall assessment

Table 4: Results Scrutiny of ad placements

Signatories' services	% of Measures missing qualitative information	% of Measures Missing quantitative data	Average Overall Score
Google Advertising	42%	20%	2.3
Facebook	67%	100%	1.7
Instagram	67%	100%	1.6
LinkedIn	20%	0%	2.6
Microsoft Advertising	67%	50%	1.9
TikTok	38%	40%	1.9
Twitter	100%	100%	1.0
TOTAL	58%	64%	1.9

Both of Meta's services performed badly with an average score of 1.7 for Facebook and 1.6 for Instagram, indicating that overall, their responses in this section fell below 2, a grade of adequate. This is in part attributable to a lack of quantitative data with both services reaching a 100% failure rate across measures for SLIs while 67% of their Measures in this section did not completely satisfy the qualitative requirements within QREs.

Both Microsoft Advertising and TikTok scored an overall of 1.9 in this section. The best average scores came from Google Advertising with 2.3 and LinkedIn with 2.6. In the case of LinkedIn, this score is reflective of the fact that it missed no quantitative responses and had the lowest incomplete rate across qualitative responses at 20% across the Measures.

Table 5: Measure-specific results Scrutiny of ad placements

Measure	Google	Meta		Microsoft		TikTok	Twitter
	Google Advertising	Facebook	Instagram	LinkedIn	Microsoft Advertising	TikTok	Twitter
1.1	3	1	1	3	1	3	1
1.2	2	1	1	n/a	2	1	1
1.3	2	3	3	2	2	3	1
1.4	not subscr.	not subscr.	not subscr.	n/a	not subscr.	3	not subscr.
1.5	3	2	2	n/a	2	2	1
1.6	3	2	2	3	3	2	1
2.1	3	2	2	3	2	2	1
2.2	1	1	1	3	3	1	1
2.3	1	2	2	3	1	2	1
2.4	3	2	2	2	2	1	1
3.1	2	1	1	2	2	2	1
3.2	2	1	1	2	2	2	1
3.3	2	2	1	3	1	1	1

All the Signatories shared information about their policy and procedures for demonetizing disinformation. However, there were frequent instances of insufficient detail or a lack of relevance to the QRE requests, and a heavy reliance on external links rather than directly providing the relevant information.

Additionally, 64% of Measures were missing quantitative information, with this figure reaching 100% for Facebook and Instagram. Some SLIs were fully incomplete while others lacked a Member State breakdown, for instance, SLI 2.3.1 which details the number of policy violating ads removed and their reach at the point of removal.

Commitment 1 largely addresses advertising from the buyer's perspective focusing on factors such as eligibility requirements for advertisers or controls for ad placements while Commitment 2 focuses on the content of advertising and ensuring that such content does not propagate disinformation. In practice, assessors found that this distinction was not clearly understood by Signatories and answers across Commitment 2 tended to be brief and redirect to responses under Commitment 1.

The responses to all Measures across Commitment 3 were particularly weak with Signatories providing little detail about their cooperation with relevant players. Signatories also exhibited a tendency to repeat answers instead of customising their responses to the specific requirements of each measure, although this may in part relate to the wording of the measures in Commitment 3 as covered in the Code specific remarks section.

4.1.3. Code specific remarks

The distinction between Commitment 1 and 2 could be more clearly delineated. In particular QRE 1.1.1 and QRE 2.1.1 appear to overlap. In Measure 1.1 the line “avoiding the publishing and carriage of harmful Disinformation to protect the integrity of advertising supported businesses” led to responses which focused on prohibited content and related policies. It may be useful to remove references to content in QRE 1.1.1 and make Commitment 1 more clearly focused on processes, verification and ad placement with Commitment 2 covering advertising messages and policies relating to prohibited content.

The text of Measure 1.1. describes the development of an additional SLI to report on “demonetisation efforts, including data related to the volume of advertising that support disinformation sources” for the baseline reports via a working group set up by relevant signatories. It appears that this has been fulfilled with the inclusion of SLI 1.1.2, however it would be helpful for any agreed upon updates to the Code to be publicly communicated through the Transparency Centre.

While the focus of Measure 1.2 appears to be on the implementation of policies e.g. “content review processes”, the QRE for the measure seems to focus on policy reviews and frequency of updates. It may be useful to include a QRE which outlines the process for implementing the policies described in Measure 1.1.

Assessors noted that the wording of the three measures (3.1, 3.2, 3.3) within Commitment 3 are very similar and 3.3. seems redundant. Considering that signatories did repeat answers across these measures, we recommend clearly delineating the requirements of these measures and ensuring that they are not overlapping.

4.2. Political advertising

4.2.1. Scope of the Section in the Code

This section of the Code attempts to establish certain safeguards for political and issue advertising, while simultaneously respecting the principle of free speech. A significant challenge, however, is defining the scope of issue advertising.

The section is extensive, containing 10 Commitments and 28 Measures and possibly subject to changes in the future. At several points in the Section’s text the European Commission’s proposal for a Regulation on the transparency and targeting of political advertising is referenced. As this regulation is still negotiated by the relevant legislative bodies, our analysis is not informed by the proposal but solely by the Code’s text.

Signatories are obliged to a number of Measures to ensure transparency and integrity in political and issue advertising, concerning e.g. clear labelling, providing identity verification systems for sponsors, and disclosing information about political and issue advertising on the service and in ad libraries.

4.2.2. Overall assessment

Table 6: Results Political advertising

Signatories' services	% of Measures missing qualitative information	% of Measures Missing quantitative data	Average Overall Score
Google Advertising	63%	0.00%	2.2
Facebook	45%	0.00%	2.4
Instagram	50%	0.00%	2.4
Messenger	0%	i	3.0
LinkedIn	67%	i	2.3
Microsoft Advertising	67%	i	2.0
TikTok	100%	i	1.5
Twitter	100%	i	1.0
TOTAL	56%	0.00%	2.2

On average Signatories performance in this section was adequate. Google and Meta provided the necessary quantitative data which contributed to an overall score above 2. There is still a considerable lack of qualitative information. For TikTok and Twitter most Measures were coded as not applicable as these Signatories claimed that most Measures in this section were irrelevant for their services and Assessors agreed with that claim. Meta's Messenger is only subscribed to one Measure of this section.

Table 7: Measure-specific results Political advertising

Measure	Google	Meta		Microsoft		TikTok	Twitter
	Google Advertising	Facebook	Instagram	LinkedIn	Microsoft Advertising	TikTok	Twitter
4.1	1	3	3	3	3	2	1
4.2	n/a	n/a	n/a	n/a	n/a	n/a	n/a
5.1	2	3	3	2	2	1	1
6.1	1	3	3	n. subscr.	n. subscr.	n/a	n/a
6.2	2	1	1	n. subscr.	n. subscr.	n/a	n/a
6.3	2	1	1	n. subscr.	n. subscr.	n/a	n/a
6.4	n/a	3	3	n. subscr.	n. subscr.	n/a	n/a
7.1	3	3	3	n. subscr.	n. subscr.	n/a	n/a
7.2	3	2	2	n. subscr.	n. subscr.	n/a	n/a
7.3	3	3	3	2	1	n/a	n/a
7.4	1	1	1	n. subscr.	n. subscr.	n/a	n/a
8.1	3	3	3	n. subscr.	n. subscr.	n/a	n/a
8.2	3	3	3	n. subscr.	n. subscr.	n/a	n/a
9.1	3	2	2	n. subscr.	n. subscr.	n/a	n/a
9.2	3	2	2	n. subscr.	n. subscr.	n/a	n/a
10.1	1	2	2	n. subscr.	n. subscr.	n/a	n/a
10.2	3	3	2	n. subscr.	n. subscr.	n/a	n/a
11.1	1	2	2	n. subscr.	n. subscr.	n/a	n/a
11.2	2	3	3	n. subscr.	n. subscr.	n/a	n/a
11.3	2	3	3	n. subscr.	n. subscr.	n/a	n/a
11.4	2	2	2	n. subscr.	n. subscr.	n/a	n/a
13.1	n/a	n/a	n/a	n. subscr.	n. subscr.	n/a	n/a
13.2	n/a	n/a	n/a	n. subscr.	n. subscr.	n/a	n/a
13.3	n/a	n/a	n/a	n. subscr.	n. subscr.	n/a	n/a

Measure	Meta
	Messenger
6.5	3

As apparent from

Table 7 Microsoft did not subscribe to most Measures and TikTok as well as Twitter were rated n/a for most Measures as they claim these Measures are irrelevant for their services. The reasoning behind both approaches is the same: that these services do not allow the publication of political and issue advertising. How well these bans are enforced in practice is beyond the scope of this analysis. Nonetheless, this question warrants some attention.

First, even though TikTok claims that it does not allow political and issue advertising, the service does allow “cause-based advertising”. The term is not defined in TikTok’s baseline report, nor could we find any reference to it in TikTok’s Advertising Policy pages. The term needs to be defined precisely and the distinction from issue advertising must be drawn clearly. Second, a brief exploratory query of TikTok’s new ad library has surfaced numerous instances in which political and/or issue advertising was spread on the platform.⁸ Previous research has also shown that political advertising has escaped TikTok’s moderation systems and been published on the platform by partisan influencers.⁹ Similar research on LinkedIn and Microsoft’s Advertising systems has not yet been conducted to our knowledge. If a Signatory decides to ban political or issue advertising on its services, it must ensure that this policy is enforced thoroughly.

The ensuing remarks mainly pertain to Google and Meta. Overall, both Signatories have adequately fulfilled their reporting obligations. One major caveat is that Google’s responses only concern political advertising, not issue advertising. More on that under “Signatory specific remarks”. Additionally, there are shortcomings concerning certain Measures:

The effective labelling of political or issue ads is pivotal for enhancing user transparency, a goal codified in Commitment 6, which is not sufficiently met by Google and Meta. For instance, both Signatories failed to provide comprehensive information on research conducted and published to boost users’ recognition and understanding of labels. Further details on the deficiencies related to Commitment 6 are included in the “Signatory specific remarks” subsection.

Both Meta and Google performed well on Measures under Commitment 7, except for 7.4, which mandates “reporting on research and publishing data on the effectiveness of measures to verify the identity of political or issue ad sponsors”. In response to this Measure, Meta redirects to their response to QRE and SLI 7.1.1, where they report on policy and rejection numbers but omit any research or data concerning the effectiveness of their policy. Google neglects to address the Measure altogether.

As for ad libraries, covered under Commitments 10 and 11, both Signatories have adequately fulfilled their reporting obligations, although quantitative data on API usage is lacking. Meta discusses a pilot program offering additional access during the U.S. elections. However, the report lacks clarity on whether and how this program will be extended in the future, but it does show that providing extra access is feasible.

⁸ Research about this issue by five EDMO hubs is currently ongoing. A report about the findings is expected to be published soon.

⁹ “These Are “Not” Political Ads: How Partisan Influencers Are Evading TikTok’s Weak Political Ad Policies”: <https://foundation.mozilla.org/en/campaigns/tiktok-political-ads>

It must be noted that this analysis only encompasses the reporting by Signatories and does not delve into the veracity of the Signatories' claims. We recognize that ad libraries have faced criticism in the past.¹⁰ Whether they meet the obligations delineated in the Code in reality necessitates further comprehensive research.

Commitment 13 was rated “n/a” as it requires Signatories to "engage in ongoing monitoring and research to understand and respond to risks related to Disinformation in political or issue advertising", an activity to occur at least annually; as it has not been a year since this Measure was implemented, they affirm their intention to collaborate with the Task-force on it.

4.2.3. Signatory specific remarks

Google

Google's ad policies as set out in response to Measure 5.1, only apply to political ads and not issue ads. In addition to political/election ads, which cover political parties, politicians, and referendums, Google maintains a 'Sensitive Events' policy, though its scope is significantly limited and does not cover established definitions of issue ads. We consider this problematic as the latter clearly falls within the ambit of this section.

On several Measures the information provided is insufficient to meet obligations, e.g.: Regarding Commitment 4, which involves adopting a common definition of 'political and issue advertising', Google's response falls short. The Signatory does not define the scope or provide relevant policies, but instead offers a minimal response, indicating an intention to comply with the EU Political Ads Regulation once finalised.

Commitment 6, focused on the accurate labelling of political and issue ads, is operationalized into several measures, some of which Google has not provided sufficient information for. For instance, Google has not detailed how they "develop a set of common best practices and examples for marks and labels on political or issue ads and integrate those learnings" (Measure 6.1).

In response to Measure 6.2, Google declares that some ad formats oblige advertisers themselves to include a 'Paid for by' disclosure directly in the ad. However, it remains unclear why this is a requirement, since other formats automatically generate a 'Paid for by' disclosure. Additionally, it is unclear whether this process necessitates additional checks by the platform and, if so, whether such checks are being conducted.

Google does not fully answer Measure 6.3, which requires investment and participation in research to improve users' identification and comprehension of labels. While their reply does address research, there are no specifics about the publication of such research, suggesting it may only be internal.

Meta

The quantitative data provided by the Signatory could be more detailed. For instance, Signatories are required to offer data regarding the volume of ads labelled as political or

¹⁰ E.g.: “A security analysis of the facebook ad library” <https://nyuscholars.nyu.edu/en/publications/a-security-analysis-of-the-facebook-ad-library>

issue ads (SLI 6.2.1) and on the volume of ads rejected due to non-compliance with the necessary verification processes (SLI 7.1.1). The provided figures, such as 'over 8,000', lack precision, leading to questions about rounding methods and rationale. Moreover, the data is aggregated for both Facebook and Instagram, which deviates from the typical approach of reporting these platforms separately.

Finally, the reporting periods are oddly defined; for SLI 6.2.1, the data covers the period "from November 15 to December 31, 2022 in EU member states in Q4 2022 in EU member states [sic]", while for SLI 7.1.1, it spans "from October 1 to December 31, 2022 in EU member states." The rationale behind these distinct periods is unclear, adding to the overall ambiguity of the reporting. In particular, as Meta deviates from the reporting period set out in the guidelines of the reporting template (16 December 2022 to 16 January 2023).

4.2.4. Code specific remarks

TikTok and Twitter are subscribed to most Measures in this section but de facto opted out of most of those. This put us in a difficult position. We have been accommodating by evaluating relevance. In subsequent monitoring iterations, assessors should not bear the responsibility of determining whether certain Measures apply to the Signatories. The Signatories should either opt out from non-applicable Measures or an exemption system should be established within the Permanent Task-force.

Assessors did find that this section was more challenging to evaluate due to QREs doubling over multiple Measures. Where possible, we would encourage the development of individual QREs and SLIs for each Measure.

There is a direct crossover between QRE 4.1.1 and QRE 5.1.1, both of which ask Signatories to share their policies relating to their definition of political advertising. Ideally, duplication should not be necessary. We would recommend that Commitment 5 focuses on the policies in place regarding political and issue-based advertising, including among those who prohibit it. This should provide exact definitions of what is and is not allowed, including any customised terminology e.g. "cause-based advertising" outside the scope of the agreed upon definition in Measure 4. Additionally to stating policies, Commitment 5 should also include a QRE regarding how Signatories enforce policies relating to prohibited advertisements, particularly among those who do not permit political advertising.

There is also some crossover between 7.2.1 and 7.3.2, both of which report on the actions taken against actors evading verification processes, although 7.3.2 specifically refers to political and issue ad sponsors.

There is ambiguity in the wording of Measure 6.2 regarding what the term "examples" refers to. Does it denote sample images, illustrations of scenarios where labelling would be applicable, or descriptions of their display and interaction functionality? A clarification would be worth considering.

Furthermore, both QRE 10.2.1 and 11.1.1 demand quantitative information on monthly usage of advertising repositories and APIs, respectively. These quantitative elements are better suited to a distinct SLI.

4.3. Integrity of Services

4.3.1. Scope of the Section in the Code

The "Integrity of Services" section holds substantial importance in the Code as it addresses the threats of coordinated inauthentic behaviour, influence operations and deepfakes among others. It emphasises the need to implement and promote safeguards by platforms against impermissible manipulative behaviours and practices. This section addresses a wide range of tactics, techniques, and procedures (TTPs) that are commonly used in disinformation campaigns and foreign information manipulation and interference.

Commitment 15 stresses the need to ensure the trustworthiness of algorithms used for detecting, moderating, and sanctioning impermissible conduct and content. It also obliges Signatories to combat deepfakes and AI-generated manipulative content, calling for robust measures to counteract these sophisticated forms of disinformation. Finally, Commitment 16 underscores the importance of Signatories working collaboratively to combat disinformation and influence operations that might be active across platforms.

4.3.2. Overall assessment

Table 8: Results Integrity of Services

Signatories' services	% of Measures missing qualitative information	% of Measures Missing quantitative data	Average Overall Score
Google Search	33%	100%	2.2
YouTube	43%	100%	2.3
Facebook	43%	100%	2.1
Instagram	43%	100%	2.1
Bing Search	50%	50%	1.8
LinkedIn	71%	100%	1.4
TikTok	57%	100%	1.7
Twitter	100%	100%	1.1
TOTAL	56%	94%	1.9

The average overall score of this section is 1.9, which is less than adequate. Though Google and Meta performed slightly better, achieving a score above 2. The high percentage of Measures missing quantitative data at least in part is alarming. So is LinkedIn's performance in this section. It is not just the absence of robust quantitative data that is concerning, but the qualitative information provided is also in part insufficient. The results show that all Signatories have room for improvement in each Commitment of this section.

Table 9: Measure-specific results Integrity of Services

Measure	Google		Meta		Microsoft		TikTok	Twitter
	Google Search	YouTube	Facebook	Instagram	LinkedIn	Bing Search	TikTok	Twitter
14.1	3	3	3	3	1	2	2	2
14.2	1	1	1	1	1	1	1	1
14.3	3	3	3	3	3	3	3	1
15.1	2	2	2	2	1	3	2	1
15.2	3	2	3	3	1	1	2	1
16.1	1	2	2	2	2	1	1	1
16.2	n. subscr.	3	1	1	1	n. subscr.	1	1

Assessors determined that every Signatory performed poorly on Measure 14.2. This particular Measure is complex, requesting Signatories to maintain and update a list of public policies outlining prohibited behaviours, detailing how they combat threats and harms, as well as obligations to publicise and improve relevant metrics. In QRE 14.2.1 Signatories are asked i.a.: "Relevant Signatories will also develop further metrics to estimate the penetration and impact that Fake/Inauthentic accounts have on genuine users and report at the Member State level (including trends on audiences targeted; narratives used etc.)" (CoP, p.16) Signatories failed to respond to this request. SLIs 14.2.1-4 ask for specific data on TTPs at Member State level. These SLIs are mostly missing or incomplete.

The data provided by Signatories in response to several SLIs has to be assessed thoroughly. Though it is tempting to believe in provided data as objective and unambiguous indicators, a detailed analysis shows that this is not always the case. See Signatory specific remarks for LinkedIn and TikTok for details.

All Signatories (except Twitter) performed well on Measure 14.3, which obliges them to cooperate on defining TTPs in the Permanent Task-force.

All Signatories have room for improvement in their responses to Commitment 16, which asks Signatories to "share relevant information about cross-platform information manipulation, foreign interference in information space and incidents that emerge on their respective services for instance via a dedicated sub-group of the Permanent Task-force or via existing fora for exchanging such information" (Measure 16.1). The sub-group mentioned had not been set up at the time of the reporting. However, the information provided about exchange through existing fora is lacking detail in many cases or is not provided at all. No data is provided in response to SLI 16.1.1 which asks to quantify "actions taken as a result of the collaboration and information sharing", as some Signatories state due to lack of time. We look forward to meaningful data in the next reports.

Measure 16.2 obliges Signatories to "share information on the tactical migration of known actors of misinformation, disinformation and information manipulation across different

platforms". Assessors found that, except for YouTube, responses provided are insufficient and lack substance.

4.3.3. Signatory specific remarks

Microsoft

In the report for Bing Search, Microsoft argues convincingly regarding Measure 14.1 that many TTPs are irrelevant for Search Engines. However, the information provided should be in part more detailed, e.g. Microsoft's information regarding the "quality and credibility (QC) assessment of websites is superficial and should be laid out in more detail. In QRE 14.1.2 Microsoft states that "Bing dedicates meaningful resources to maintaining the integrity of the platform"; this statement needs further explanation and data in order to assess its veracity.

The report for LinkedIn fell short in multiple ways. Many qualitative responses lacked substance and detail. The proper quantitative data was missing. In detail:

QRE 14.1.1 is not answered in detail and missing information regarding seven out of twelve TTPs, which are relevant to the LinkedIn platform. Quantitative data provided is incomplete and not robust. For SLI 14.2.1 data on only four TTPs is presented. Microsoft argues that data for others cannot be computed. This may be a legitimate reason, however, other Signatories found ways to compute data for other TTPs as well. Additionally, how some of the data is computed is methodologically not robust: The reported numbers for three TTPs are a subset of inauthentic accounts reported for TTP1. All data based on this subset is likely to be underrepresenting the TTPs in question. For example, for "Use of fake followers or subscribers" Microsoft reports "a subset of the fake accounts reported in TTP 1 SLI 14.2.1 that followed a LinkedIn profile or page". This ignores the fact that one fake account may follow more than one profile or page. Therefore, Microsoft's numbers would be probably much higher for this TTP if reported properly.

Microsoft essentially states that Commitment 15 is "not applicable" to LinkedIn. With this claim, Microsoft contradicts itself. The first QRE of this Commitment demands Signatories to report on "policies in place for countering prohibited manipulative practices for AI systems that generate or manipulate content". However, Microsoft stated in their report regarding QRE 14.2.1 that "LinkedIn also acts vigilantly to maintain the integrity of all accounts and to ward off bot and false account activity (including "deep fakes")." Thus, the QRE does apply to LinkedIn. QRE 15.2.1 asks Signatories to "report on their policies and actions to ensure that the algorithms used for detection, moderation and sanctioning of impermissible conduct and content on their services are trustworthy". This QRE is also applicable to LinkedIn considering they stated in for QRE 14.2.1 that "We are investing in artificial intelligence technologies such as advanced network algorithms that detect communities of fake accounts". If they do so, they should also be able to report on policies and actions to ensure that these systems are trustworthy.

TikTok

TikTok reports that the number of Fake accounts as a percentage of MAUs is 0.0067 % - which is suspiciously low. In comparison, Meta estimates the number on their platforms to be circa 5 %. TikTok's claim warrants thorough investigation which is beyond the scope of this

analysis. Additionally, the quantitative data provided is insufficient: For SLI 14.2.4 there is no data on Member State level provided, for some TTPs it is missing entirely.

The Signatory fails to disclose information regarding two TTPs ("Inauthentic coordination of content creation or amplification, including attempts to deceive/manipulate platforms algorithms" and "Coordinated mass reporting of non-violative opposing content or accounts") and provides only superficial information regarding implementation of policies against TTPs. TikTok could, for example, provide additional information about the number of people who work in CIO investigations and/or Content Moderation. Furthermore, the statement, "The implementation of these policies is also ensured through enforcement measures applied in all Member States" (QRE 14.2.1) is not backed up by any details.

4.3.4. Code specific remarks

The wording of commitment 15 could be improved. The word "and" between the two clauses makes it seem that it is only relevant to those who distribute content AND operate AI systems. Additionally, the phrasing 'disseminate AI generated and manipulated content through their services (e.g., deepfakes)' may be wrongly interpreted as the Signatories are disseminating this content rather than being used to disseminate such content.

4.4. Empowering Users

4.4.1. Scope of the Section in the Code

This section of the Code focuses on protecting users from disinformation by empowering them to detect and report misleading content. This is achieved through enhancing media literacy efforts, ensuring recommender systems and services are designed with safety and transparency, and providing functionality to flag content and appeal enforcement actions. Signatories also commit to the development of tools relating to media literacy, authenticity of content, providing easy access to fact-checks and warning labels, and making informed decisions as well as engaging with research in both the development of tools and in assessing their effectiveness.

4.4.2. Overall assessment

Table 10: Results Empowering Users

Signatories' services	% of Measures missing qualitative information	% of Measures Missing quantitative data	Average Overall Score
Google Search	36%	33%	2.1
Youtube	36%	50%	1.7
Facebook	31%	43%	1.9
Instagram	31%	43%	1.9
Messenger	50%	100%	1.0
WhatsApp	0%	100%	2.0
Bing Search	27%	33%	2.1
LinkedIn	24%	50%	2.3

TikTok	14%	25%	2.5
Twitter	100%	100%	1.0
Total	37%	52%	1.9

The total average score across this section of the Code is 1.9, indicating a below adequate performance. TikTok scored highest (2.5), followed by LinkedIn (2.3) with Bing Search (2.1) and Google Search (2.1) achieving an adequate score, followed closely by Facebook and Instagram (1.9) and Youtube (1.7). Both Messenger and Whatsapp only subscribed to one measure, but Messenger in particular received a poor grade overall.

Table 11: Measure-specific results Empowering Users

Measure	Meta		Google		Microsoft		TikTok	Twitter
	Facebook	Instagram	Google Search	YouTube	LinkedIn	Bing Search	TikTok	Twitter
17.1	2	1	3	2	2	3	3	1
17.2	2	2	2	3	1	1	2	1
17.3	2	2	2	2	2	2	2	1
18.1	2	2	n. subscr.	1	2	n. subscr.	3	1
18.2	2	2	2	2	2	2	3	1
18.3	2	2	2	2	1	2	3	1
19.1	3	3	2	1	3	1	3	1
19.2	1	1	2	1	3	2	3	1
20.1	n. subscr.	n. subscr.	n. subscr.	n. subscr.	3	2	n. subscr.	n. subscr.
20.2	n. subscr.	n. subscr.	n. subscr.	n. subscr.	3	3	n. subscr.	n. subscr.
21.1	2	2	2	1	2	3	3	1
21.2	1	1	n. subscr.	1	n/a	n. subscr.	1	1
21.3	1	1	1	2	n/a	2	2	1
22.1	n. subscr.	n. subscr.	n. subscr.	n. subscr.	3	n. subscr.	n. subscr.	n. subscr.
22.2	n. subscr.	n. subscr.	n. subscr.	n. subscr.	n/a	3	n. subscr.	n. subscr.
22.3	n. subscr.	n. subscr.	n. subscr.	n. subscr.	2	1	n. subscr.	n. subscr.
22.7	n. subscr.	n. subscr.	2	2	1	3	2	1
23.1	3	3	3	3	3	2	3	1
23.2	2	2	2	1	3	1	2	1
24.1	2	2	n. subscr.	2	3	n. subscr.	3	1

	Meta	
	Messenger	Whatsapp
25.1	1	2
25.2	1	2

Commitment 17 deals with media literacy and assessors found that there was a tendency among Signatories to focus only on media literacy initiatives and not on the development of media literacy tools, particularly in Measure 17.3 which asked for details of partnerships with media literacy experts relating to both. Signatories should also be reminded that the scope of media literacy and critical thinking goes beyond simple instruction about how to use a platform or to adjust settings on a platform.

On the safe design of recommender systems in Commitment 18, only TikTok achieved a good grade across all Measures. Assessors felt that this section of the Code was too vague and that Signatories in general needed to provide more detailed data to meet SLI requirements.

Commitment 21, dealing with tools for allowing users to access fact-checks and warning labels saw some of the lowest scores, particularly on Measures relating to research and incorporating scientific evidence and user needs into the design of labelling and warning systems.

Commitment 25 deals with features or initiatives to help users critically assess content on messaging apps and as such only features two Meta services, Messenger and Whatsapp, who both subscribed to no other Measures in this section. Messenger scored a poor grade (1) overall, providing minimal detail in QREs and a complete lack of SLI data. Whatsapp scored adequate (2), providing better detail in QREs about its work with partner organisations and tools and features to limit the virality of disinformation, but the SLI information provided was incomplete and irrelevant.

4.4.3. Signatory specific remarks

Meta

QRE 17.1.1 asks Signatories to describe tools they develop or maintain to help users improve their media literacy and critical thinking skills. Instagram described an on-platform campaign for youth to raise awareness about the online safety features contained within Instagram such as reviewing privacy settings or managing screen time. This information is irrelevant to the aims of the Commitment.

In Measure 21.2, both Meta and Instagram state “We know this program is working” in regards to their labelling and warning systems but provide little evidence to back up such a claim. Similarly, on 21.3 both platforms provide a vague description of working with experts but it needs a much clearer description of how they incorporate scientific evidence and user needs in their implementation of these programs.

Microsoft

On Commitment 18, Microsoft states in relation to Bing Search that, “Unlike social media services, for which this Commitment appears primarily oriented, search engines do not typically cause or facilitate the viral propagation of disinformation”. Consequently, its answers to Measure 18.2 relating to the development and enforcement of publicly documented, proportionate policies to limit the spread of harmful false or misleading information are limited with the SLI entirely missing. However, in the next Measure, 18.3, Microsoft highlights its role in facilitating research on data voids, in which media manipulators can use obscure search queries, gaming of auto-complete search recommendations and search optimisation techniques to spread disinformation, which would appear to contradict the Signatory’s claim that 18.2 is irrelevant for Bing Search.

Similarly in Commitment 19 on transparency about recommender systems, Microsoft made the claim that their service under Bing Search does not function as a recommender system. While this did not affect the grading, assessors did question whether personalised search or features like providing personalised image and video results alongside search results should be considered within the scope of a recommender system. This also applies to Commitment 22 and we did consider that Microsoft’s linking to its Webmaster Guidelines for Measure 22.3 was a poor quality answer.

While our analysis focused on the more objective fulfilment of reporting obligations, we would question whether the arguments Microsoft made regarding the relevance of the above Measures to Bing Search would stand up to scrutiny from experts in this area. The role of search engines and the various features available to search users such as personalised search or auto-complete suggestions should be given greater consideration particularly when considering how disinformation may contribute to systemic risks under the DSA.

On Commitment 21, Microsoft explained that LinkedIn does not use labelling, which leads to the question of whether LinkedIn should continue to commit to Measures relating to labelling if they do not plan to add labels.

4.4.4. Code specific remarks

Assessors noted that the wording of the Code in Commitment 18 is too broad. For instance, in 18.1 Signatories are asked to report on the “risk mitigation systems, tools, procedures, or features” relating to the design of both their recommender systems and “systemic approaches in the design of their products, policies, or processes, such as pretesting”, which made assessment difficult due to such a large scope. Additionally, 18.2 appears broadly focused on general disinformation policies and their enforcement which overlaps with previous sections of the code. Signatory responses would likely be more useful and task simplified for Signatories if the QREs and SLIs were reviewed and their scope more clearly defined.

Measure 19.2 includes one SLI relating to providing users with options for selecting their preferred options with recommender systems. However, some Signatories provided qualitative responses within their SLI answers. There is no QRE within the Code so assessors were unclear as to whether this QRE was added by Signatories to provide context or something agreed at a Task-force level, and as such the Measure was graded only on the

basis of the SLI. It may be worth providing a QRE to understand what options Signatories have made available for users in addition to the metrics requested under the SLI or providing an updated version of the Code if this Measure was already agreed upon at a Task-force level.

We would also recommend that the Permanent Task-force agree to a definition of recommender systems and ask that search engines consider any aspects of their services which would meet such a definition. These decisions should also be publicly available and incorporated into future monitoring work for commitments which deal with recommender systems.

4.5. Empowering Researchers

4.5.1. Scope of the Section in the Code

This section of the Code focuses on empowering researchers and encouraging disinformation research. Signatories commit to the provision of automated access to non-personal data and public data, working with an independent, third-party body to provide data access to vetted researchers, and supporting good faith research into disinformation.

4.5.2. Overall assessment

Table 12: Results Empowering Researchers

Signatories' services	% of Measures missing qualitative information	% of Measures missing quantitative data	Average Overall Score
Google Search	60%	0%	1.8
Youtube	50%	50%	2.2
Facebook	83%	100%	1.8
Instagram	83%	100%	1.8
Bing Search	83%	100%	1.8
LinkedIn	100%	100%	1.5
TikTok	100%	100%	1.2
Twitter	100%	100%	1.0
Total	81%	89%	1.6

The performance across this Section was particularly poor with all Signatories except YouTube achieving a score of less than adequate. Assessors noted a clear issue across SLIs, with all receiving a 100% missing rate except for Google's services.

In addition to missing quantitative data, the issue of relevance, redirecting to other QREs and incomplete answers were a large factor in the overall low scoring in this section. For example, QRE 27.4.1 asks Signatories to describe pilot programs for sharing data with vetted researchers including details about the nature of the programs, number of research teams engaged and research topics or findings where possible. Despite the straightforward nature of this request, no Signatory provided all elements of the requested information with an appropriate level of detail.

It should be noted that some Signatories did indicate that they plan to improve for future reports. Meta and Microsoft stated their intention to provide better metrics and TikTok has made significant progress in this area since the baseline reports were released, but the poor results in the overall assessment table reflects the fact that Signatories did not make enough effort to provide clear and structured answers within this section of the report.

Table 13: Measure-specific results Empowering Researchers

Measure	Meta		Google		Microsoft		TikTok	Twitter
	Facebook	Instagram	Google Search	YouTube	LinkedIn	Bing Search	TikTok	Twitter
26.1	2	2	2	2	1	2	1	1
26.2	2	2	n. subscr.	3	1	2	1	1
26.3	1	1	2	3	2	2	1	1
27.1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
27.2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
27.3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
27.4	2	2	1	1	1	1	1	1
28.1	2	2	2	2	2	2	2	1
28.2	2	2	2	2	2	2	1	1
28.3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
28.4	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

A number of Measures in this section were graded as not applicable. Under Commitment 27, we determined, based on Signatory responses, that the work on the first three measures relating to the establishment of an independent body for providing access to data for vetted researchers will likely take place at a Permanent Task-force level. This work had not begun at the time of reporting. Similarly, Measures 28.3 and 28.4 involve collaborative work with other bodies that were planned but had not yet proceeded to a level that facilitates updates. Once work does begin on these Measures, Signatory answers should provide adequate detail beyond just stating the fact that they are participating in Task-force work.

The Measures that were assigned numerical grades include those relating to providing public access to data (Commitment 26), engaging in pilot programs towards sharing data with vetted researchers (Measure 27.4), ensuring adequate human resources to engage with the research community (Measure 28.1) and providing transparency about what data is currently available (Measure 28.2). Considering the fact that these Signatories will have obligations regarding researcher access to data under Article 40 of the DSA when it is fully applied in early 2024, we consider the lack of effort, clarity and progress from Signatories on these Measures to be particularly concerning.

4.5.3. Signatory specific remarks

Google

For YouTube, Google only lists Google Trends as a public, real-time data source. Google Trends, while useful for tracking general search interests over time, may not be the most pertinent tool when it comes to dissecting the nuances of disinformation. Relevant data would ideally encompass specifics on flagged content, videos that were removed due to policy violations, the nature of videos that are recommended to users, and more. Furthermore, YouTube does not provide metrics related to user engagement or the number of views videos receive.

Microsoft

With Bing Search, the available data listed under Measure 26.1 is also being claimed as applicable to Measure 26.2. However, it appears that all these sources do not offer 'real-time or near real-time' access to data as specified for Measure 26.2. Additionally, this Measure explicitly refers to data which is 'subject to an application process', which is not the case for any of the listed data sources. We would question whether Measure 26.2 is currently relevant to Bing Search and recommend that it clearly delineates between its answers for the two Measures in future reporting.

TikTok

Assessors found TikTok's performance on Commitment 26 to be poor. At the time of reporting the platform had yet to implement any tools that allow for the provision of continuous, real-time, or near real-time data for research purposes. While they indicate that they are working on this, they fail to offer any specific details or timelines. The SLIs for this commitment are coded as lacking quantitative data. While it is understandable that TikTok cannot provide usage data for non-existent tools, the concern remains that such tools should already be in place according to the Commitment's requirements.

We would like to acknowledge that since the reports were published TikTok has published their commercial content library, a tool with public data access that is useful to researchers, and introduced a Research API to Europe.¹¹

4.5.4. Code specific remarks

The responses to Measures 27.1-27.3 offer a reminder that Permanent Task-force decisions should be communicated publicly, ideally through the Transparency Centre. Additionally, it would be useful to provide timelines for planned actions, workgroups or collaborations at a Task-force level.

Twitter highlighted its 'industry-leading API program' in its baseline report, but very shortly after announced it would end the program. Additionally, there have been continued concerns in the research community relating to Meta's commitment towards providing its CrowdTangle program with the resources necessary to ensure its sustainability and effectiveness¹². In a wider context, Google and Microsoft have also made reductions to responsible AI research

¹¹ <https://library.tiktok.com/ads>, <https://developers.tiktok.com/products/research-api/>

¹² <https://algorithmwatch.org/en/crowdtangle-platform-transparency/>

teams as part of cost-cutting efforts¹³, which highlights how essential continued resourcing is for research partnerships and efforts. While Measure 28.1 deals with resourcing it may be worth considering if additional QREs or SLIs could focus on sustainability by capturing changes to the levels of resources dedicated to research as well as planned changes in terms of resourcing.

In Measure 28.2, Signatories are asked to 'be transparent on the data types they currently make available to researchers across Europe'. Responses to this Measure varied in approach, potentially indicating a lack of understanding of the purpose behind the QRE and Measure, while it appears that the types of data available have already been covered in Commitment 26. We recommend clarification or rewording.

4.6. Empowering the Fact-Checking Community

4.6.1. Scope of the Section in the Code

Fact-checking and debunking are key components in the fight against dis- and misinformation. The Code acknowledges that platforms have to cooperate with fact-checkers. To this end, platforms are required to enter into agreements with fact-checkers to ensure comprehensive coverage across all Member States. These agreements should also guarantee a fair compensation structure. Furthermore, platforms must seamlessly integrate, display, or otherwise consistently utilise the work of fact-checkers within their services. Additionally, to perform their role effectively and accurately, fact-checkers should be granted robust access to data from the platforms, which will aid them in identifying, analysing, and debunking misinformation or disinformation swiftly and efficiently.

4.6.2. Overall assessment

Table 14: Results Empowering the Fact-Checking Community

Signatories' services	% of Measures missing qualitative information	% of Measures Missing quantitative data	Average Overall Score
Google Search	29%	100%	2.1
Google YouTube	29%	100%	2.0
Facebook	71%	33%	1.9
Instagram	71%	33%	1.9
Microsoft Bing	71%	67%	1.4
Microsoft LinkedIn	100%	100%	1.1
TikTok	29%	33%	1.9
Twitter	100%	100%	1.0
TOTAL	63%	72%	1.7

The results in this section cover a broad range with fairly big differences between Signatories' performance in the different categories. While assessors found that for Google less than a third of Measures were missing qualitative information, all Measures containing SLIs are lacking quantitative data. Nonetheless, Google scored highest overall. It must be

¹³ <https://www.ft.com/content/26372287-6fb3-457b-9e9c-f722027f36b3>

noted that for LinkedIn all relevant Measures are missing qualitative information and quantitative data, resulting in an overall score of only 1.1 (Poor).

Table 15: Measure-specific results Empowering the Fact-Checking Community

Measure	Google		Meta		Microsoft		TikTok	Twitter
	Google Search	YouTube	Facebook	Instagram	LinkedIn	Bing Search	TikTok	Twitter
30.1	2	2	2	2	1	2	3	1
30.2	3	3	2	2	2	2	3	1
30.3	3	3	1	1	1	1	2	1
30.4	1	1	1	1	1	1	1	1
31.1	n/a	n/a	n/a	n/a	n/a	n. subscr.	n/a	n/a
31.2	2	1	3	3	1	2	2	1
31.3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
31.4	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
32.1	1	1	1	1	1	1	1	1
32.2	n. subscr.	n/a	n/a	n/a	n/a	n. subscr.	n/a	n/a
32.3	3	3	3	3	1	1	1	1
33.1	n. subscr.	n. subscr.	n. subscr.	n. subscr.	n. subscr.	n. subscr.	n. subscr.	n. subscr.

The overarching goal of this section is to ensure comprehensive fact-checking coverage across all Member States. This goal has not been achieved. According to the information provided, most Signatories apart from Meta have only a limited number of agreements with fact-checkers.

Looking at the formal reporting obligations, TikTok and Google provided relevant and concise qualitative answers for most of the QREs. However, the latter failed to provide quantitative data. Most Signatories failed to provide sufficient information regarding the funding of fact-checkers as codified in Measure 30.2.

All observed Signatories failed to provide qualitative information regarding Measure 30.4. The QRE of this Measure obliges Signatories to “report, ex ante on plans to involve, and ex post on actions taken to involve, EDMO and the elected body representative of the independent European fact-checking organisations [...]”. However, observed Signatories apparently did not formulate any plans to involve EDMO or the EFCSN. Though they look “forward to engaging with EDMO on our fact-checking efforts” (Meta) or stand “ready to cooperate in accordance with this QRE at the appropriate time” (LinkedIn), the QRE clearly asks for more specific information.

Three out of four Measures in Commitment 31 have been rated n/a. Measure 31.1 is operationalized in reporting elements under 31.2, which was rated. Measures 31.3 and 31.4

deal with setting up, funding, governing and applying a repository of fact checks. As the Code allows more time for this effort, Signatories were not yet obligated to report on it.

All Signatories performed poorly on Measure 32.1 and 32.2, which oblige them to “provide fact-checkers with information to help them quantify the impact of fact-checked content over time” (Measure 32.1) and to “provide appropriate interfaces, automated wherever possible, for fact-checking organisations to be able to access information on the impact of contents on their platforms and to ensure consistency in the way said Signatories use, credit and provide feedback on the work of fact-checkers” (Measure 32.2). The request for quantitative data is mostly ignored and information provided as part of QREs is often irrelevant.

As stated previously, our analysis is mainly focussed on the Signatories’ compliance with the reporting obligations set out in the CoP. We want to acknowledge that a more in-depth analysis of commitments made by platforms in this section has been published by the European Fact Checking Standards Network (EFCSN). Regarding this section the EFCSN issues a harsh verdict: “Some of the VLOPs are not just ignoring their commitments, but one could think that they are actively trying to re-write what was agreed, misrepresenting their current policies to portray them as genuine efforts to comply with the Code of Practice.” Taking into accounts the EFCSN’s analysis as well as our own, overall the commitments made by the Signatories have not been fulfilled. The reporting as well as the substance of Signatories actions in this field have to improve significantly in order to meet the standards set out in the CoP.

4.6.3. Signatory specific remarks

Google

Google (reporting for both Search and YouTube) stresses in their report their financial commitments to the International Fact Checking Network (IFCN) as well as their financial contributions through the European Media and Information Fund (EMIF) in response to Measure 30.1. There are good reasons to install third-party bodies as intermediaries between platforms and fact-checking organisations to maintain the independence of the latter. Most importantly, this construct ensures that allocation of funding is independent and not influenced by corporate interests. However, if they are to be compliant with the Code, Google should state the number of agreements with fact-checking organisations through EMIF and IFCN per member state and language. It is problematic that Google does not disclose the fact that EMIF’s support for fact-checking activities is only one of four funding lines within the fund’s portfolio.

In relation to Measure 32.1, it is surprising to see Google describing tools that are universally accessible to webmasters as specialised tools for fact-checkers.

As for YouTube, there is a lack of data for SLI 31.1.1. The Signatory claims that it cannot provide metrics on impact as “YouTube’s use of fact-checks does not involve taking specific actions with regards to content that is fact-checked”, but for QRE 21.1.1 they state: “YouTube’s fact check information panels provide additional context by highlighting relevant, third-party fact-checked articles above search results for relevant queries.” So at least for this feature there should be metrics available.

Meta

Meta is one of only two Signatories that provide sufficient quantitative data for Commitment 30, demonstrating an almost complete coverage of all Member States except for Malta. However, some relevant qualitative information is missing, e.g. the response to QRE 30.1.1 emphasises IFCN certification but overlooks the specifics of the relationship between the Signatory and the fact-checker. In particular, Meta should provide more information on the conditions for the agreements in questions, so that the conditions fact-checkers are working under can be assessed. Similarly, QRE 30.1.3 offers a list of miscellaneous ways the Signatory supports fact-checkers, but doesn't delve into the primary resource allocation.

For Measure 30.3, the response provided does not adequately illustrate the signatory's cooperative efforts.

Microsoft

Assessors found that LinkedIn does not meet the obligations set out in this section in several ways. In addition to the remarks made in the "Overall assessment section", LinkedIn does not adhere to Commitment 30, which stipulates cooperation between fact-checkers and platforms.

Regarding Measure 30.1, the current reporting lacks detailed information. For instance, there is an undefined "pilot arrangement" and there's no clear indication whether the two listed fact-checking organisations focus on the EU or globally. Moreover, there's no insight provided on how LinkedIn intends to achieve fact-checking coverage in each Member State and support fact-checkers' work at the Member State level. In terms of SLI 30.1.1, it should clarify whether the fact-checking organisations are located in EU Member States, focus their work on these states, and specify the languages they cover. As for Measure 30.2, further detail is necessary on the financial aspect of supporting fact-checking organisations. In response to Measure 30.3, LinkedIn merely asserts reliance on independent global news agencies, which doesn't convey a spirit of cooperation. They could, however, illustrate adherence to this Measure by showcasing the cross-border work conducted by these agencies.

For Measure 31.2, qualitative information is noticeably absent. Also, SLI lacks data at the Member State level. The reported number of content pieces reviewed by fact-checkers, which stands at 252, is markedly low compared to Facebook's reported figure of 150,000 articles written by third-party fact-checkers.

Similar to Google Search, Microsoft categorises tools and features that are universally accessible to all webmasters as being specifically for "webmasters of fact-checking organisations" (32.1).

4.6.4. Code specific remarks

QRE 30.1.1 asks for qualitative and quantitative data. The latter is not defined but examples are given: "for instance: contents fact-checked, increased coverage, changes in integration of fact-checking as depends on the agreements and to be further discussed within the Task-force". It would be worthwhile to define quantitative data more precisely and make this a separate SLI.

Regarding Measure 31.2, Google and Bing report that they are unable to provide metrics concerning the impact of actions taken in response to fact-checks. However, it seems probable that there exists a methodology to assess these measures' impact on search results, for instance. In cases like these, Signatories could be required to secure an exemption from the Permanent Task-force. Otherwise Signatories can easily avoid providing valuable information without clarifying the actual issues or obstacles, or offering alternative data. This undermines effective external monitoring.

4.7. Transparency Centre

4.7.1. Scope of the Section in the Code

This section discusses the establishment and funding of a website that houses all pertinent details related to the execution of the Code's Commitments and Measures. The section's scope also includes the site's maintenance and development.

This and the following two sections all concern in various ways monitoring procedures. As the majority of the Measures were marked "i" (irrelevant), we did not assign an Overall Score. Several Measures in these sections lack individual reporting elements, and some Commitments even lack Measures entirely. However, we do offer some general observations on how Signatories fulfil the responsibilities outlined in these three sections.

4.7.2. Overall assessment

The Transparency Centre website, under the domain disinfocode.eu, was launched in February 2023. It's worth noting that the deadline set out in Measure 34.1 (which required the website to be established within six months of the Code being signed) was exceeded by nearly two months. However, the website was made available in time for the baseline reports.

Details regarding the funding of the website, as per Measure 34.2, have not been disclosed. Regardless, the existence and functionality of the website indicate that the required funding has been allocated.

As stated in their reports, the Signatories have hired an unnamed contractor for the creation and maintenance of the website. There is a lack of clarity about who is responsible for running the website and maintaining its contents. The only contact information available is an obscure email address, which leaves room for improvements in terms of transparency.

On a positive note, the website is user-friendly, contains relevant information, and is searchable. As of the time of reporting, the Signatories were unable to provide any statistics concerning the website's usage as it was just launched. Overall, the objectives set out in this section were achieved.

4.7.3. Code specific remarks

While no QREs or SLIs are mandated for most Commitments, clarity around the expectations for Commitments/Measures without QREs or SLIs in the transparency reports would be beneficial. It's unclear whether Signatories need to take any action or provide any specifics about how they fulfilled such requirements, or if it suffices to simply acknowledge their adherence to these Measures. Given that these Commitments relate to transparency, and the general public lacks access to the internal processes of the Permanent Task-force and its working groups, it would be helpful if Signatories included a brief summary of their efforts to meet their Commitments, even if this summary is only a couple of paragraphs long. Although this is addressed in Measure 36.3, it would be more effective if such a requirement were highlighted earlier in the report or structure to ensure it is not overlooked. This would further promote the overarching goal of transparency and clarity in the implementation of these commitments and measures.

4.8. Permanent Task-force

4.8.1. Scope of the Section in the Code

The objective of this section is to set up a Permanent Task-force which is aimed at developing and amending the code when necessary. It mandates that all Signatories actively participate in the Permanent Task-force. In addition to the Signatories, the Task-force also comprises representatives from EDMO, ERGA, and the European External Action Service (EEAS). It is chaired by the European Commission.

4.8.2. Overall assessment

The Permanent Task-force has been established with Signatories actively participating in its proceedings. One testament to their active involvement is the creation of the harmonised reporting template. Additionally, the Task-force has structured itself with various subgroups, some of which are co-chaired by platform signatories.

From the outside, there is no possibility to evaluate the work contributed to the Task-force by individual Signatories.

4.8.3. Code specific remarks

If external assessors are to monitor the compliance of the Code, there is a need for greater transparency regarding the work of the Task-force. Assessors in various sections noted that they often had to infer the work of Task-force subgroups, timelines and decisions which were referenced by Signatories but had not been made public. We would recommend that the Permanent Task-force provide regular and comprehensive updates on their work, particularly as it relates to the reporting and work of Signatories and their Commitments. The Transparency Centre may be an appropriate place to provide such updates.

4.9. Monitoring of the Code

4.9.1. Scope of the Section in the Code

This section outlines the monitoring framework, including various reporting requirements. It is largely technical in nature, detailing measures that do not have individual reporting elements, and there are even some Commitments that lack Measures altogether.

This section also contains a Commitment to develop Structural Indicators. The aim of these Indicators is to assess the effectiveness of the Code in reducing the spread of online disinformation. As both authors are involved in the EDMO Expert Group on Structural Indicators for the Code of Practice on Disinformation, we provide some general comments on this Commitment, even though the respective Commitment was not assigned any numerical values in our analysis.

4.9.2. Overall assessment

Commitments 38 to 40 pertain to the reporting process, which includes the allocation of adequate resources, setting a deadline for baseline reports, and ensuring consistent reports thereafter. In general, these commitments appear to have been satisfactorily fulfilled on a formal level. However, there's one notable issue:

The responses from most Signatories, when questioned about the provision of sufficient financial and human resources, tend to emphasise their general resources. For instance, they often mention Trust and Safety teams and personnel responsible for content moderation in general. This approach fails to delineate resources specifically dedicated to the implementation of the Code.

While the detailed regulation of the Code's monitoring is commendable, there's a crucial missing element: a clearly defined protocol specifying which body is responsible for carrying out the monitoring work. The reports are extensive, and Signatories are required to provide a significant amount of information and data. Yet, it remains unclear who has the responsibility of verifying the comprehensiveness and correctness of this reporting.

Moreover, there appears to be minimal progress regarding Commitment 44, wherein Signatories commit to “to be audited at their own expense, for their compliance with the commitments undertaken pursuant to this Code”. This area could benefit from further attention and more robust implementation.

The initial set of Structural Indicators was anticipated to be released within nine months following the signing of the Code (Commitment 41). Thus, they were due after the baseline reports were published. However, it must be stressed that these indicators have not yet been released and to our knowledge neither did the Working Group responsible table a workable proposal for Structural Indicators in due time (as foreseen in Measure 41.3). This lapse is particularly concerning considering the paramount importance of Structural Indicators in critically evaluating the substance of the Signatories' reporting.

From the information available to us, there are two prevailing tracks being pursued: A short-term solution being offered by a private contractor to the EU Commission. And a long-term solution, for which EDMO has tabled suggestions through a working paper¹⁴.

Regrettably, at least publicly there has been minimal progress in this area. It's imperative that strides are made in this direction ahead of the next reporting cycle.

4.10. Crisis Response

4.10.1. Scope of the Section in the Code

Commitment 42 of the Code specifies that the European Commission can request specific information from signatories in the event of special situations like elections or crisis via a rapid response system developed by the Permanent Task-force.

The baseline reports included two sections for reporting on each Signatory's response during a period of crisis, one covering the COVID-19 pandemic and the other covering the War of aggression by Russia on Ukraine.

4.10.2. Overall assessment

Table 16: Results Crisis Response

	Google	Meta	Microsoft	TikTok	Twitter
COVID 19	3	2	3	3	1
Ukraine	3	2	3	3	1

We have provided an overall grade for each Signatory as there were no specific measures, QRE's or SLIs. We find that Google, Microsoft and TikTok provided comprehensive overviews of their crisis response work. Meta's responses were considered useful but assessors found that it was difficult to determine which services Meta was referring to with each action, making its report less clear. Characteristic of the rest of their reporting, Twitter's responses were largely incomplete and the information that was provided was poor.

The provided template for reporting on crisis actions worked well and we would recommend that a similar harmonised approach is used for future crisis response reporting.

¹⁴

https://cadmus.eui.eu/bitstream/handle/1814/75558/WP%202023_34.pdf?sequence=1&isAllowed=y

5. Conclusions on Future Monitoring

The Strengthened Code and the harmonised reporting template have introduced a much higher level of robustness and comparability to Signatory reports, but this also comes with a much higher requirement in terms of monitoring. The monitoring process is time consuming and requires expert knowledge in many areas.

Section I(p) of the Code acknowledges that, in addition to the European Commission, Signatories are expected to cooperate with both EDMO and ERGA for monitoring the Code, but there is no clear structure establishing the competencies and responsibilities for monitoring compliance. Additionally, if the CoP becomes a code of conduct under the DSA as envisioned, then compliance and monitoring efforts may also fall under the realm of the Commission's DSA enforcement team and the European Centre for Algorithmic Transparency (ECAT). Considering the scope of the work involved, the need for adequate planning and resourcing, and any impact from assessment of systemic risk assessments under the DSA, we would urgently advise the development of a clearly communicated plan from the Commission to ensure that the Code will be monitored appropriately.

There are also two additional issues to consider when it comes to Monitoring. First, The Commission would like to see the Code monitored at a Member State level, focusing on aspects of QREs or SLIs which provide information about a Member State. This could potentially take place either through ERGA representatives or EDMO hubs. In its current implementation, there is very little Member State level data in reports beyond figures contained in SLIs and as such it is hard to see how such work would be undertaken and how valuable it would be. We would discourage any duplication of work given the scope of work involved and the various resourcing constraints applicable to both EDMO and ERGA. We would recommend that relevant stakeholders, including EDMO and ERGA should assess how Member States can be involved in a useful and effective way, taking into account the capacity and resources available among stakeholders.

Second, there is a need to distinguish between the more objective monitoring of the reports e.g. did a Signatory provide the requested information for a QRE or SLI and monitoring the actual implementation of Measures e.g. whether the reported information is in fact true. This has already been established in previous monitoring of the Code, particularly the CovidCheck report¹⁵ which found a mismatch between what actions platforms claimed they had undertaken and the reality for users.

While such work requires domain expertise, it often also requires significant investigatory and research capacity, as well as access to data, to verify claims. For example, in Commitment 14, Signatories outline their policies against mentioned TTPs and describe how they enforce them, but it is unclear how well these enforcement systems actually work. In a series of yearly reports, the NATO StratCom CoE has demonstrated that platforms are unable to stop manipulative techniques from commercial actors and in the latest iteration of its Social Media Manipulation report, the authors find that 90 to 100 % of inauthentic activity remains on platforms four weeks after it was purchased¹⁶. Similar work would be extremely

¹⁵ <https://doras.dcu.ie/26472/>

¹⁶ <https://stratcomcoe.org/publications/download/Social-Media-Manipulation-2022-2023-DIGITAL.pdf>

valuable across the Code and should be considered an essential part of monitoring the Code. It may also be a suitable avenue through which Member State level monitoring can take place, particularly through EDMO hubs, but it is essential to stress that this work will require adequate resourcing and funding.

6. Recommendations

In the subsections of this report, we have provided detailed feedback and recommendations relating to Signatories and the Code. In addition to this detailed feedback, we provide the following overall recommendations.

General Recommendations

- Signatories should ensure that their responses are directly answering the specific requirements of a QRE or SLI, that outside links are not used as a replacement for providing information within the reporting format and that in the case where answers for QREs or SLIs are overlapping, it would be preferable to repeat the relevant information rather than redirect to responses for other QREs or SLIs. For future reporting, these standards would be a fair expectation of Signatories and they are also consistent with the recommended guidelines for the reporting template.
- Reporting of quantitative data must be more comprehensive and thorough. Methodologies should be precise, and Member State-level data should be provided.
- The reporting template should include the description for all Measures including QREs and SLIs as this would aid readability and likely make it easier for the platforms themselves to fill out the reports.
- The reporting period should be defined and publicised. In the case of baseline reports, a predefined reporting period was specified in the reporting template. However, this period was disregarded in several instances. To ensure comparability of quantitative data, Signatories should adhere to the same reporting period when compiling data in the future.
- To aid in transparency, Permanent Task-force decisions should be communicated publicly, ideally through the Transparency Centre. The current version of the Code including any updates should also be available on the Transparency Centre.
- Signatory descriptions of future plans should include an indicated timeline for when to expect such actions in the format of quarter/year. The Permanent Task-force and Signatories should also provide timelines for planned actions, workgroups or collaborations at a Task-force level.
- If Signatories are subscribed to Measures which are not currently relevant to them they should either unsubscribe from these Measures or receive an exemption at a Permanent Task-force level after a thorough review of the Signatories' claims.
- We urge all parties involved to work on Structural Indicators more intensely, ensuring that they are precisely defined and cover all areas of the Code.

- We urge the development of a clearly communicated and resourced plan to ensure that the Code will be monitored appropriately. This should include both monitoring of transparency reports and verifying the accuracy of Signatory reporting.
- EDMO, ERGA and other stakeholders should assess how Member State level monitoring can take place in a useful and efficient manner.
- The European Commission should continuously assess whether certain platforms should be encouraged to join the CoP, as there are major digital platforms that have not yet become Signatories.

Scrutiny of Ad Placements

- The Permanent Task-force should ensure the difference between Commitment 1 and 2 through either rewording or clarification to Signatories. Rewording or clarification should also apply to the three Measures of Commitment 3 to avoid overlapping responses.

Political Advertising

- All Signatories, including those who prohibit political or issue advertising, should provide an overview of their political advertising policies in Commitment 5. This should include clear parameters for what is and is not prohibited as well as definitions for any customised terminology e.g. TikTok should clearly define "cause-based advertising".
- Monitoring should ensure that Signatories are properly enforcing any policies relating to the prohibition of political or issue advertising. Additionally, Commitment 5 should include a QRE asking for information about the enforcement of such policies.
- Signatories should be encouraged to publish their internal research, e.g. on the effectiveness of labelling political and issue advertising.
- Google should be encouraged to broaden the definition of political advertising in their advertising policies to include issue advertising.
- The Permanent Task-force should review instances of duplication between QREs and where possible develop individual QREs and SLIs for each Measure.

Integrity of Services

- Quantitative data regarding all TTPs relevant to Signatories' services must be provided, where applicable on a Member State level.
- The Permanent Task-force should work towards a harmonisation of methodologies in computing the data for Measure 14.2.

Empowering Users

- Signatories should ensure that their media literacy actions incorporate tools as well as initiatives and that this work focuses on critical thinking and not just online safety or privacy settings.
- Signatories should increase their efforts and ensure that they provide concrete details on incorporating scientific evidence and user needs into the design of labelling and warning systems.
- Microsoft should be encouraged to re-examine the relevance of the features offered by Bing Search in regards to Commitment 18, 19 and 22, considering both any features which could function as recommender systems and the role of data voids in contributing to disinformation.
- The Permanent Task-force should agree upon a definition of recommender systems among Signatories and ask that search engines consider any aspects of their services which would meet such a definition.

Empowering Researchers

- Signatories must do more to ensure that they are empowering researchers and put more effort into providing relevant and clear information for Measures in this section.
- On Measure 28.1, the Permanent Task-force should consider additional QREs or SLIs which focus on the sustainability of research efforts and Signatory changes regarding resourcing.

Empowering the Fact-Checking Community

- Signatories (except for Meta) must work towards a more comprehensive fact-checking coverage of all Member States.
- Signatories should disclose (and likely increase) funding for fact-checkers.
- Signatories should provide fact-checkers with more information to help them quantify the impact of fact-checked content.