GATE Institute Big Data for Smart Society –
Sofia University "St Kliment Ohridski",
Sofia, Bulgaria

Evaluating VLOP and VLOSE Implementation of the Strengthened EU Code of Practice on Disinformation in Bulgaria

White Paper

September 2023

Type of document:	White paper
Dissemination level:	Working Document
Authors:	Keith Peter Kiely
	Ruslana Margova
	Silvia Gargova
	Milena Dobreva
	Teodora Gandova
	Tsvetelina Stefanova
	Veneta Kireva
	Kalina Bontcheva
Acknowledgements:	Bulgarian-Romanian Observatory of Digital Media This assessment is based on research conducted by the GATE Institute and is partially supported by BROD, The Bulgarian-Romanian Observatory of Digital Media, which received funding from the European Union under contract number: 101083730. This document reflects the views only of the independent Consortium, and the Commission cannot be held responsible for any use which may be made of the information contained herein.
Licence:	This work is licensed under the Creative Commons Attribution 4.0 (CC-BY 4.0). If cited or quoted, reference should be made to the full name of the author(s), the title, the year and the URL of the source.
Version:	1.0
Delivery Date of document:	11.09.2023

Executive summary

In June of 2022, Google, Meta, Microsoft, TikTok, Twitter (rebranded as X) and a selection of advertising industry companies all signed up to the strengthened Code of Practice on Disinformation (European Commission, 2022). One of the goals of this strengthened version of the code was to empower the industry to adhere to self-regulatory standards in order to combat disinformation. The strengthened code also claims to set a more ambitious set of commitments and measures aimed at combating disinformation online.

Our aim here is to offer an assessment or evaluation of the implementation of the 2022 Code of Practice on Disinformation (CoP) by these companies in Bulgaria. Very little information exists on the implementation of the strengthened Code of Practice when it comes to Bulgaria by Very Large Online Platforms and Very Large Online Search Engines (VLOP and VLOSE) and this is a country which is particularly vulnerable to disinformation narratives.

This detailed analysis of VLOP and VLOSE compliance reports offers a general overview of the responses of Google, Meta, Microsoft, TikTok, and Twitter in line with three major pillars of (i) Advertising and Political advertising; (ii) Integrity of services; and (iii) Empowering Users, the research community and Fact-Checkers. We focus specifically on the data provided for Bulgaria in order to identify any important gaps or issues which should be addressed in future VLOP and VLOSE responses to disinformation and their compliance reports.

To effectively assess how each of the platforms has performed, we have applied an assessment scale to rate each response, along with a Table under each measure which applied the scale to the response. We also want to acknowledge that this scale is based on a methodology utilised by a first of its kind report led by the EDMO Ireland hub and German-Austrian Digital Media Observatory (GADMO) hub to assess the responses of VLOP and VLOSE to the strengthened Code of Practice (Park & Mündges, 2023).

Score	Interpretation
1	**Poor**: The response significantly falls short of meeting the requirements of the measure. For example, 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. Here is how we rated 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 rating represents 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 we believe this assessment to be correct e.g. the measure relates to displaying information alongside political advertising and the signatory's product does not allow political advertising.

Additionally, when assessing the responses of each platform to the measures, we have applied elements from the Kantar Public and Visionary Analytics Methodology (*A Monitoring Framework for the Code of Practice on Disinformation*, July 2023). This methodology uses the following indicators: Compliance, Clarity, Relevance, Usefulness, and Verifiability. We utilise these indicators when structuring our findings.

Generally speaking, this analysis found that VLOP and VLOSE responses under the **Advertising and Political Advertising pillar** regularly fell short in terms of compliance due to the responses either being too brief, lacking detail or relevancy. Partial responses could be indicative of the timeframe within which the reports were put together. There was indeed a sense of insufficient time to collect this data on the side of the online platforms but also promises that missing data will be appearing in future reports. Indeed, the area with the highest number of cases where the platforms declared the metric is not relevant to them is the advertising pillar.

In the **Integrity of Services pillar,** we can observe more detailed explanations, specifically from Google. Generally, responses here, while detailed, highlight another pertinent issue, the lack of verifiable or testable information and data. This is true for all platforms, but specifically in reports by Twitter, Microsoft, TikTok and Meta. Responses in this pillar also highlight and deliver specific data related to fake accounts removed, fake likes, fake followers, and accounts banned in Bulgaria provided by TikTok and LinkedIn. We can also see that generally Microsoft, Tik Tok and specifically Twitter view more measures as non-applicable while Meta offers commentary on all measures; Google seems to be handling more measures than others with a sufficient level of detail. However, at times that detail is not appropriate or relevant to the response or measure.

Full compliance and measures under the **Empowering Users, the research community and Fact-Checkers pillar** (areas 5-7 of the Strengthened code) are essential, given the challenges faced by citizens, researchers, and fact-checkers in places like Bulgaria. These challenges include political polarisation, media ownership concentration, strong Russian influence, and limited financial resources. These vulnerabilities make Bulgaria particularly susceptible to disinformation and misinformation and require both strong stakeholder cooperation, and effective VLOP and VLOSE strategies that involve empowering users, researchers, and fact-checkers at the local level. VLOP and VLOSE reports currently mention several measures that address this issue, including search interventions, media literacy campaigns, partnerships with fact-checkers and researchers, content removal, fact-checking, and labelling. However, there are gaps in the information provided, with some platforms failing to provide the required metrics. The VLOP and VLOSE reports also demonstrate that more data was provided by Google, Meta, Microsoft, and TikTok regarding Bulgaria compared to the first two pillars. It is unclear how this data relates to the Association of European Journalists – Bulgaria and the website factcheck.bg, as well as the Bulgarian section of AFP proveri, the only certified IFCN fact-checking members in Bulgaria. Overall, it is not surprising to discover such differences in the levels of reporting by the different VLOP and VLOSE but we have optimism that over time the reports will move in the direction of providing more details across all pillars.

In conclusion, it must be stressed that a key compliance failure under the third pillar is the ongoing highly limited, free access to data for researchers, which severely hampers not only the verifiability of the VLOP and VLOSE self-reported measures, but also threatens to undermine research-led policy making going forward due to the current highly limited ability of researchers to monitor independently disinformation and VLOP and VLOSE enforcement actions. The issue is severely exacerbated by Twitter's withdrawal from the Code of Practice and its introduction of unaffordable fees for data access. Unless some urgent regulatory action is taken under the DSA, representative, large-scale research into prevalence and impact of disinformation during the forthcoming European elections will not be possible.

There is also a key technological issue with all reports, which needs to be resolved urgently. This concerns the need to harmonise the structure of the CSV and JSON reports submitted by the different VLOPs and VLOSE, as currently their automatic analysis is practically impossible, also due to lack of documentation. Moreover, one of the CSV files contains a warning and refers to the full PDF report, implying that the uploaded CSV or JSON files are not sufficiently comprehensive. We gained further evidence of this, when attempting to count the occurrences of a keyword in CSV files and getting a count which differed from the one obtained from the PDF files. All those issues necessitated manual analysis of the PDF versions of VLOP and VLOSE reports, which is extremely time-consuming, slow, and error prone. Therefore we argue strongly for the commission taking a lead in providing all VLOP and VLOSE with standardised reporting templates in CSV and JSON, which are designed in consultation with EDMO and Hub researchers. Other than formatting issues, we also need to standardise on reporting periods (ideally weekly or bi-weekly), units of reporting (e.g. ad spend bands harmonisation across all platforms), and required detail of reporting to enable transparency and accountability.

Another key recommendation arising from this research is that data sharing by VLOP and VLOSE is not sufficient. To enable effective Code monitoring and independent research, the European Commission needs to fund and establish a shared, EU-wide large-scale data processing infrastructure, as well as a mechanism for researchers to share know-how.

At the same time, researchers at each EDMO Hub (including BROD) and those involved in other EU-funded research projects aimed at combating disinformation are incurring very high and unnecessary overheads in terms of time, skills, and effort being invested into data cleaning, harmonisation, storage, and access. We recommend that the EU invests in the development of shared, free, and comprehensive open-source data cleaning, harmonisation, storage, and analysis tools, which enable researchers from less resourced EU countries (such as those from Central and especially Eastern Europe) to carry out effective monitoring of the Code and VLOP and VLOSE actions on disinformation, as platforms are currently fairing the worst there in terms of effective enforcement of their policies against online abuse and disinformation.

We conclude by making just under 20 recommendations for actions that need to be undertaken urgently by the European Commission on one hand and VLOP and VLOSE on the other. These are centred on improving VLOP and VLOSE report quality, compliance, verifiability, transparency, and data provision for independent research and compliance monitoring. Further details are provided in Section 3.3.

Content

List of abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
AS/RS	Auto-Suggest and Related Search Suggestions Functionalities
BROD	Bulgarian Romanian Observatory of Digital Media
C2PA	Coalition for Content Provenance and Authenticity
CEE	Central and Eastern Europe
CGs	Community Guidelines
CIB	Coordinated Inauthentic Behaviour
CIO	Covert Influence Operations
CoP	Code of Practice on Disinformation
CSV	Comma-Separated Values File
DSA	Digital Services Act
DTAC	Digital Threat Analysis Centre
EDMO	European Digital Media Observatory
EEA	European Economic Area
EFCSN	European Fact-Checking Standards Network
EMIF	European Media and Information Fund
EU	European Union
FCI	Fact Check Impressions
GDI	Global Disinformation Index
I&A policies	Integrity and Authenticity Policies
IFCN	International Fact-Checking Network
IO	Influence Operations
IRA	Internet Research Agency
IRIE	Institute for Research on the Information Environment
JSON	JavaScript Object Notation File
KC	Knowledge Cards
ML	Machine Learning
MRC	Media Rating Council
NA	Not Applicable
NGED	NewsGuard Extension Downloads

NGI	NewsGuard Impressions
NGO	Non-Governmental Organisation
PDF	Portable Document Format File
PSA	Public Service Announcement
QRE	Qualitative Reporting Elements
SIEP ads	Social Issue, Electronic or Policies Adds
SLI	Service Level Indicators
SSA	Self-Serve Appeals
TH	Transparency Hub Viewership
ToS	Terms of Services
TTPs	Tactics, Techniques and Procedures
URL	Uniform Resource Locators
VLOP	Very Large Online Platform
VLOSE	Very Large Online Search Engines

List of tables

List of figures

Evaluating the Implementation of the 2022 EU Code of Practice on Disinformation in Bulgaria

1. Introduction

In June 2022, Meta, Google, Microsoft, TikTok, Twitter (rebranded as X, hereafter referred to as Twitter) and a selection of advertising industry companies signed up to the European Commission's Strengthened Code of Practice on Disinformation (CoP). The goal of the strengthened Code of Practice is to provide a set of voluntary and co-regulatory standards and commitments intended to reduce the spread of disinformation.

Disinformation is defined by the Code as: "misinformation, disinformation, information influence operations and foreign interference in the information space" (CoP, 2022). The CoP is structured into nine different sections. This white Paper focuses on responses to the first six sections of the CoP; Scrutiny of Ad Placements, Political Advertising, Integrity of Services (these sections look at ways in which disinformation may be monetized and the various tactics and techniques used), Empowering Users, Empowering the Research Community and Empowering the fact-checking community (These sections aim to move signatories towards cooperating with and ensuring adequate support is in place for various stakeholder groups). We focus on these sections as a starting point for analysing the baseline reports as they represent the most pressing challenges to the development of detailed and transparent research into the implementation of CoP and any subsequent measures which may follow implementation of the Digital Services Act (DSA).

It is also worth noting that each section of the CoP contains three parts or levels, the first contains the commitments that signatories have agreed to. These commitments are explained in detail in Measures i.e. actions and steps to be completed signatories. There are also detailed quantitative reporting requirements, the Qualitative Reporting Elements. hereafter referred to as QRE or Service Level Indicators, hereafter referred to as SLI.

As part of the Bulgarian Romanian Observatory on Digital Media (BROD), GATE alongside its 11 partners in BROD, is dedicated to studying and investigating the enforcement of the CoP and the developing Digital Services Act (DSA). Our research team at GATE has manually reviewed the responses of each of the major signatories: Google, Meta, Microsoft, TikTok, and Twitter specifically as they concern Bulgaria and have assessed these responses and their adherences to the relevant commitments in the revised CoP. We have evaluated the responses of signatory companies and divided the six sections into three pillars:

1. Advertising and Political advertising
2. Integrity of services
3. Empowering Users, the research community and Fact-Checkers.

GATE's research efforts include but are note focused on assessing companies' advertising placement businesses or the information curation and prioritisation systems addressed in the CoP. Consequently, GATE has partially assessed these commitments as part of its evaluation. Nonetheless, relevant research on the placement of ads by some of the CoP signatories has been conducted by organisations like the Global Disinformation Index. Unfortunately, restricted data access has limited the research community's ability to assess changes to information curation and prioritisation systems at scale. Therefore, GATE's recommendations for transparency outlined in this paper seeks to address this gap in order to enable more comprehensive assessment of information sorting algorithms in the future.

As a starting point we can point out that the data provided by each company do not cover consistent time periods:

Google's data covers the period of Q3'2022 (1 July 2022 - 30 September 2022).

TikTok's data covers 16 June - 16 December 2022.

Microsoft's data covers December 2022.

Meta's data covers 1 October 2022 - 31 December 2022.

Twitter's data covers the period of H2, 2022 (June - December 2022).

This analysis found that the responses from the **Advertising and Political Advertising** pillars regularly fell short in terms of compliance. Responses were determined to be too brief, lacking detail, or irrelevant to the report. This is likely a result of the short timeframe in which the report was compiled. While these results show that there was insufficient time for the companies to collect and report the data, it is possible that missing data will appear in future reports. Notably, the area with the highest number of responses from platforms indicating that the metric was not relevant to them was the advertising pillar.

In the **Integrity of Services pillar,** we can observe more detailed explanations, specifically from Google. In general responses here, while answers are detailed, they also highlight another pertinent issue, the lack of verifiable information. This is true to some degree for all platforms, but specifically in reports by Microsoft, TikTok, Twitter and Meta. Responses under this pillar also provided specific data related to fake accounts removed, fake likes, fake followers, and accounts banned in **Bulgaria** provided by TikTok and LinkedIn. We can observe that generally Microsoft and Tik Tok see more measures as non-applicable while Meta offers commentary on all measures; Google seems to be handling more measures than others with a sufficient level of detail. However, at times that detail is not appropriate or relevant to the response or measure.

Full compliance and measures under the **Empowering Users, the research community and Fact-Checkers pillar** (areas 5-7 of the Strengthened code) are essential, given the challenges faced by citizens, researchers, and fact-checkers in Bulgaria, which include political polarisation, media ownership concentration, strong Russian influence, and limited financial resources. These vulnerabilities make Bulgaria particularly susceptible to disinformation and misinformation and require strong stakeholder cooperation, and effective VLOP and VLOSE strategies that involve empowering users, researchers, and fact-checkers at the local level. VLOP and VLOSE reports currently mention several measures that address this issue, including search interventions, media literacy campaigns, partnerships with fact-checkers and researchers, content removal, fact-checking, and labelling. However, there are gaps in the information provided, with some platforms failing to provide the required metrics. The VLOP and VLOSE reports also demonstrate that more data was provided by Google, Meta, Microsoft, and TikTok regarding Bulgaria compared to the first two pillars, it is not entirely clear the relationship between the data for Bulgaria and the certified IFCN member in Bulgaria – the website factcheck.bg of the Association of European Journalists – Bulgaria and the Bulgarian fact-check section of AFP proveri.

In general, there is a large knowledge gap in how the figures provided were calculated for Bulgaria and who was involved on the local level in this process. Having direct access to that data would help the efforts of the research and fact-checking community. As stated above, a combination of facts make Bulgaria and Bulgarian research into disinformation vulnerable, greater efforts should be made by all signatories to provide the appropriate information and metrics and to work closely with researchers on the ground in Bulgaria and other high risk member states.

2. How effective were technology companies at enacting their CoP commitments in Bulgaria?

2.1. Advertising and Political Advertising

This pillar (areas 2 and 3 of the Strengthened code) is one of the most central in the Code of Practice and requires a review and improvement of practices around such possible disinformation activities with the aim of improving the transparency of their political advertising online. In Europe, political adverts are strongly regulated in a variety of ways, with some countries limiting political advertising completely in the months leading to an election. Big Technology companies claim that revenue from political advertising is negligible compared to other kinds of advertising. The strengthened Code aims to ensure that purveyors of disinformation do not benefit from advertising revenues. The strengthened code of practice realises the importance of political advertising and how a strengthened code will require clearer labelling and transparency measures. This includes providing information about the sponsor, ad spend, and display period of political ads. Additionally, signatories will create searchable ad libraries for political advertising. Signatories commit to stronger measures avoiding the placement of advertising next to disinformation, as well as the dissemination of advertising containing disinformation. The Code also sets up a more effective cooperation among the players of the advertising sector, allowing stronger joint action.

Measure 1

QRE 1.1.1 *- Signatories will disclose and outline the policies they develop, deploy, and enforce to meet the goals of Measure 1.1 and will link to relevant public pages in their help centres.*

SLI 1.1.1 *- Signatories will report, quantitatively, on actions they took to enforce each of the policies mentioned in the qualitative part of this service level indicator, at the Member State or language level. This could include, for instance, actions to remove, to block, or to otherwise restrict advertising on pages and/or domains that disseminate harmful Disinformation.*

Google

Google AdSense has rolled out a number of policies and processes geared towards disrupting the monetisation incentives of malicious and misrepresentative actors. Examples of AdSense policies are named such as Unreliable and Harmful Claims, Replicated Content, Manipulated Media, Dangerous or Derogatory Content. In terms of the SLI, it is noted that the number of Actioned AdSense Pages for Bulgaria was 56,529. The number of Actioned AdSense Domains for Bulgaria was 12. Additionally the estimated cost of Blocked Requests on pages for Bulgaria was €95,500.94, while the estimated costs of Blocked Requests on Domains for Bulgaria was €119,496.31 (Google, 2023, pp. 4-9). However, **without sufficient context, it is hard to evaluate whether these measures are sufficiently comprehensive or effective.**

Meta

In Meta's report Measure 1.1 outlines the policies in place for both Facebook and Instagram. These include monetisation policies for partners, content, page and professional mode demonetization rules, Instagram Partner Monetization Policies, Instagram Content Monetization Policies, as well as terms of service, commercial terms and community guidelines or standards. Meta also noted that for SLI 1.1.1, SLI 1.1.2, & SLI 1.2.1, **they were not able to deliver this SLI** in the time provided for the baseline report and that they would work to improve their SLIs across chapters in their next report in January-June 2023 (Meta, 2023, pp. 11-12).

Microsoft

Notes that LinkedIn does not allow misinformation and disinformation on that platform, either organic content or in the form of advertising content. The response references LinkedIn's Professional Community Policies. LinkedIn also provides additional specific examples of false and misleading content that violates its policy via a Help Centre article on Fake or Misleading Content. LinkedIn's advertising policies incorporate the above provision and prohibit disinformation and misinformation, and fraudulent and deceptive ads. Claims in ads must have factual support. For Microsoft advertising it is noted that in December 2022, it rolled out revised network wide policies to avoid publishing and carriage of harmful Disinformation and the placement of advertising next to disinformation content.

On the issue of the number of ads which LinkedIn and Instagram have restricted under the Misinformation policies in QRE 2.1.1, it is claimed that no ads were restricted. The same can be said for SLI 1.2.1 (Microsoft, 2023, pp. 11-19, 26). **The response is brief and uninformative as to the goals of the measure**.

TikTok

Offers a comprehensive list of policies related to advertising, political adverts and paid ads or landing pages. TikTok claims not to allow paid ads or political actors to place adverts. It is also noted that advertisers are required to meet a number of requirements with respect to their landing page. TikTok also provides verification in the context of ads and they claim they are currently trailing mandatory verification for accounts belonging to a government, politician or political party in the US. According to the data provided there were no ads removed under the Covid-19 misinformation policy, no ads removed under the political content ad policy and no ads removed under the misleading and inauthentic or deceptive behaviours policy for Bulgaria (TikTok, 2023, pp. 2-6). However, **due to data access issues we were not able to verify whether this really means that no such ads existed on TikTok or that the platform was unable to identify and moderate them**.

Twitter

Twitter offer no specific response to to measures 1.1 - 1.6. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, pp.1-3).

QRE 1.2.1 - *Signatories will outline their processes for reviewing, assessing, and augmenting their monetisation policies in order to scrutinise and bar participation by actors that systematically provide harmful Disinformation.*

SLI 1.2.1 *Signatories will report on the number of policy reviews and/or updates to policies relevant to Measure 1.2 throughout the reporting period. In addition, Signatories will report on the numbers of accounts or domains barred from participation to advertising or monetisation as a result of these policies at the Member State level, if not already covered by metrics shared under Measure 1.1 above.*

Google

In response, Google advertising again points to Google Ads and AdSense. They note that updating monetisation policies on climate change and product policy updates in response to COVID-19 misinformation. In addition, changes made due to the invasion of Ukraine where Google Advertising has adapted and enforced policies to protect users. **Bulgaria-specific data for the SLI is missing** (Google, 2023, pp. 9-10).

Meta

Addresses change to community standards and are not specific to demonetization for both Instagram and Facebook. **Does not address the commitment adequately** (Meta, 2023, p. 12).

Microsoft

It is noted in this response that LinkedIn does not offer an ad revenue share program and does not allow third-parties to monetise content they post to LinkedIn by running ads against it. For Microsoft Advertising it is noted that they work with selected, trustworthy publishing partners and require these partners to abide by strict brand safety-oriented policies to avoid providing revenue streams to websites engaging in misleading, deceptive, harmful, or insensitive behaviours. **SLI is not relevant** as LinkedIn does not allow third parties to monetise content they post to LinkedIn by running ads against it and Microsoft Advertising block domains globally, not at the Member State level (Microsoft, 2023, pp. 18-19).

TikTok

TikTok does not offer ad revenue sharing in Europe. Instead TikTok provides creator monetisation opportunities such as the TikTok Creator Fund. **No SLI's provided** (TikTok, 2023, pp. 6-7).

Twitter

Twitter offer no specific response to to measures 1.1 - 1.6. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, pp.1-3).

QRE 1.3.1 - Signatories will report on the controls and transparency they provide to advertising buyers with regards to the placement of their ads as it relates to Measure 1.3.

Google

In addition to QRE 1.2.1, Google Ads states that it provides advertisers with additional controls and helps them exclude types of content that, while in compliance with AdSense policies, may not fit their brand or business (Google, 2023, p. 10). These controls are said to let advertisers apply content filters or exclude certain types of content. Advertisers can exclude content such as politics, news, sports, beauty, fashion and many other categories listed here.

Meta

Offers very similar, but not identical responses for both Instagram and Facebook. The responses mention brand safety controls which aim to prevent ads from running alongside certain types of content on Facebook or Instagram. This response notes that advertisers can see and update brand safety settings directly and offers additional details on this aspect of transparency (Meta, 2023, pp. 12-13).

Microsoft

Offers an overview of the range of information and tools which LinkedIn gives advertisers, transparency and control regarding the placement of their advertising. LinkedIn also publishes a Feed Brand Safety score for advertisers and the public. This score measures the number of ad impressions on the LinkedIn platform that appeared adjacent to content removed for violating LinkedIn's Professional Community Policies, including disinformation. On the LinkedIn Audience Network, LinkedIn also provides tools to assist advertisers in controlling where their ads appear within the network. Microsoft Advertising provides its customers with campaign reporting and functionalities to monitor and control ad placement across the Microsoft Advertising network (Microsoft, 2023, p. 20).

TikTok

Reports that TikTok partners with a number of industry leaders to provide a number of controls and transparency tools to advertising buyers with regard to the placement of ads (TikTok, 2023, pp. 7-8). These include the TikTok Inventory Filter, The TikTok Brand safety Verified by DoubleVerify, the TikTok Brand safety by Integral Ad Science. TikTok have also partnered with third parties to offer post-campaign solutions that enable advertisers to assess the suitability of user content that ran immediately adjacent to their ad in the For You feed: Zefr, IAS and DoubleVerify.

Twitter

Twitter offer no specific response to measures 1.1 - 1.6. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, pp.1-3).

***QRE 1.5.1** - Signatories that produce first party reporting will report on the access provided to independent third-party auditors as outlined in Measure 1.5 and will link to public reports and results from such auditors, such as MRC Content Level Brand Safety Accreditation, TAG Brand Safety certifications, or other similarly recognised industry accepted certifications*

Google

The response discusses the Trustworthy Accountability Group and notes that Google is currently enrolled in the Verified by Trustworthy Accountability Group and its Trustworthy Accountability Group ID status is active (Google, 2023, pp. 11-12). It is also noted that Google partakes in Audits including those conducted by independent accreditation organisations such as the Media rating Council (MRC). YouTube is also stated to be a founding Platform member of the Global Alliance for Responsible Media.

Meta

This answer notes that in November 2022, Facebook received accreditation from the Media Rating Council (MRC) for content-level brand safety on Facebook covering Meta's Partner Monetization policies, Content Monetization policies, and associated content-level brand safety and suitability controls applied to Facebook in-Stream Video and Instant Articles on various devices (Meta, 2023, p. 13). Instagram is also said to be in scope of accreditation from the MRC in 2023.

Microsoft

LinkedIn does not offer a content monetisation or an ad revenue share program. Microsoft Advertising undergoes yearly Media Rating Council (MRC) accreditations via Third-Party audit (Microsoft, 2023, p. 21). The MRC accreditation certifies Microsoft Advertising's click measurement system adheres to the industry standards for counting ad clicks and the processes supporting this technology is accurate.

TikTok

Note that TikTok has achieved the TAG Brand Safety Certified seal by the Trustworthy Accountability Group in Europe and globally. TikTok has been certified by the Internet Advertising Bureau (IAB) for the IAB Gold Standard 2.1. **They also claim they will be complying with their independent audit obligations under the DSA**.

Twitter

Twitter offers no specific response to measures 1.1 - 1.6. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, pp.1-3).

***QRE 1.6.1** - Signatories that place ads will report on the options they provide for integration of information, indicators and analysis from source raters, services that provide indicators of trustworthiness, fact-checkers, researchers, or other relevant stakeholders providing information e.g. on the sources of Disinformation campaigns to help inform decisions on ad placement by buyers.*

Google

Reports that Google Ads also provides its advertising partners with functionalities that empower them to retain command over the placement of their advertisements, the manner in which their ads are displayed, and their targeted audience (Google, 2023, p. 13).

Meta

This response talks about the Brand safety controls in place at both Instagram and Facebook (Meta, 2023, pp. 13-14). **This response is not addressing the commitment. Similarly, responses for QRE 1.6.2 – QRE 1.6.4 are as unresponsive.**

Microsoft

Reports that LinkedIn has integrated a number of brand safety tools and services to help advertisers understand and control the placement of their ads, and to help avoid placing ads next to disinformation content and/or in places or sources that repeatedly publish disinformation. It is also noted that Microsoft Advertising partners with Information Integrity experts, such as NewsGuard and GDI, as source and references of Disinformation domains. Microsoft Advertising is actively blocking domains that these sources deem as Disinformation (Microsoft, 2023, pp. 21-22).

TikTok

Zefr is said to be working on providing TikTok with additional measurements related to misinformation. Zefr recently acquired the AI misinformation company Adverif.ai, which is powered by fact-checking data from more than 50 organisations globally (TikTok, 2023, p. 9). **No SLI data provided**.

Twitter

Twitter offers no specific response to measures 1.1 - 1.6. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, pp.1-3).

Table 1 - *Ratings for Measure 1*

Commitment 1							
	Google	**Meta**		**Microsoft**		**Tik Tok**	**Twitter**
		Facebook	*Instagram*	*LinkedIn*	*Microsoft Advertising*		
Measure 1.1	3	2	2	3	1	3	1
Measure 1.2	2	2	2	3	3	2	1
Measure 1.3	2	3	3	3	3	3	1
Measure 1.5	3	2	2	2	2	3	1
Measure 1.6	2	3	3	3	3	2	1

Measure 2

QRE 2.1.1 *– Signatories will disclose and outline the policies they develop, deploy, and enforce to meet the goals of Measure 2.1 and will link to relevant public pages in their help centres.*

SLI 2.1.1 - *Signatories will report, quantitatively, on actions they took to enforce each of the policies mentioned in the qualitative part of this service level indicator, at the Member State or language level. This could include, for instance, actions to remove, to block, or to otherwise restrict harmful Disinformation in advertising messages and in the promotion of content.*

Google

In general, the response here seems pretty detailed with specific and relevant policies covering inappropriate content, including dangerous or derogatory content & hacked political materials. In addition, misrepresentation policies cover issues such as unreliable claims, misleading representation and clickbait adverts. Google has offered the following details for Bulgaria which claim that there were 312,683 creatives actioned for destination requirements. The number of creatives actioned for inappropriate content in Bulgaria was 1,994, while the number of creatives actioned for misrepresentation was 110,307 (Google, 2023, p. 18).

Meta

Similar responses were given for both Facebook and Instagram, which details how advertisers on Meta must follow the Terms of Service, Community Standards and Advertising Standards with links included. It is noted that

Misinformation is considered as unacceptable content under Meta's Advertising Standards. For SLI, **member state/language and different types of moderation (e.g. restriction not only blocking) data is missing,** EU level data claims that for the number of ads removed in line with these policies for Facebook were: Misinformation Policy: Over 8,800, overall ads removed 2.9million. The same stats are used for Instagram (Meta, 2023, pp. 17-18).

Microsoft

As noted in response to QRE 1.1.1, LinkedIn prohibits misinformation and disinformation on its platform, whether in the form of organic content or in the form of advertising content. LinkedIn's Professional Community Policies, which apply to all content on the platform, expressly prohibit false and misleading content, including misinformation and disinformation: In December 2022, Microsoft Advertising rolled out revised network-wide policies to avoid the publishing and carriage of harmful Disinformation and the placement of advertising next to Disinformation content. Such policies prohibit ads or sites that contain or lead to Disinformation. To enforce this policy, "We may use a combination of internal signals and trusted third-party data or information sources to reject, block, or take down ads or sites that contain disinformation or send traffic to pages containing disinformation. We may block at the domain level landing pages or sites that violate this policy." (Microsoft, 2023, p. 25). They claim that on Bulgarian level they have blocked 29 individual advertisements, 61 impressions and 1 unique domain (Microsoft, 2023, p. 26). **Given the significantly higher numbers reported by Google, this begs the question of whether Microsoft is finding and detecting all violating content and how their numbers can be verified independently by researchers**.

TikTok

TikTok's answers here are noticeably brief and do not address in detail issues of tools or methods. They note that they are in the process of supplementing the existing approach by producing dedicated advertising-focused misinformation policies and claim they will provide more information in the next report. According to the data provided there were no ads removed under the Covid-19 misinformation policy, no ads removed under the political content ad policy and no ads removed under the misleading and inauthentic or deceptive behaviours policy for Bulgaria (TikTok, 2023, p. 10).

Twitter

Twitter offers no specific response to measures 2.1 - 2.4. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, p. 6).

QRE 2.2.1 *- Signatories will describe the tools, methods, or partnerships they use to identify content and sources that contravene policies mentioned in Measure 2.1 - while being mindful of not disclosing information that'd make it easier for malicious actors to circumvent these tools, methods, or partnerships. Signatories will specify the independent information sources involved in these tools, methods, or partnerships.*

Google

All newly created ads or ads that are edited by users are said to undergo a review for policy violations. The review of new ads is carried out using either automated mechanisms, human manual reviews, or a combination of both. A link is provided with more information on how the ad review process works, please see the about the ad review process page (Google, 2023, pp. 19-20).

Meta

Meta offers a similar response to QRE 2.1.1, where the details of Meta's advertising standards are laid out, and notes that particular types of content are ineligible to monetize, this includes content debunked by fact-checkers. Repeated attempts to post false information by advertisers will result in. Adverts that include misinformation that violate community standards or misleading medical information are also ineligible for monetization (Meta, 2023, p. 17).

Microsoft

LinkedIn works with numerous partners to facilitate the flow of information to tackle purveyors of disinformation, including disinformation spread by state-sponsored and institutional actors. LinkedIn maintains an internal Threat Prevention and Defence team composed of threat investigators and intelligence personnel to address disinformation. This team works with various other internal teams, including an Artificial Intelligence modelling team, to develop leads into threat actor campaigns. **Microsoft Advertising** employs dedicated operational support and engineering resources to enforce its advertising policies detailed below, combining automated and manual enforcement methods to prevent or take down advertisements that violate its policies. Every ad loaded into the Microsoft Advertising system is subject to these enforcement methods, which leverage machine-learning techniques, automated screening, the expertise of its operations team, and dedicated user safety experts (Microsoft, 2023, p. 28).

TikTok

Report that in order to identify content and sources that breach their COVID-19 advertising policy, all ads go through moderation prior to going live on the platform. It also refers to the user report button and how this will initiate a review and action if necessary. It is noted that TikTok operates a "recall" process whereby ads already on TikTok will go through an additional stage of review if certain conditions are met, including reaching certain impression thresholds. Additional reviews are also conducted on random samples of ads (TikTok, 2023, p. 12). They will explore additional tools, methods or partnerships as they work on enhancing disinformation advertising policies. **The response seems slightly short. It also does not say anything about tools, partnerships in Bulgaria or language-specific methods directly.**

Twitter

Twitter offer no specific response to measures 2.1 - 2.4. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, p. 6).

QRE 2.3.1 *- Signatories will describe the systems and procedures they use to ensure that ads placed through their services comply with their advertising policies as described in Measure 2.1.*

SLI 2.3.1 *- Signatories will report quantitatively, at the Member State level, on the ads removed or prohibited from their services using procedures outlined in Measure 2.3. In the event of ads successfully removed, parties should report on the reach of violatory content and advertising*.

Google

Same response as for QRE 2.2.1. Google ads claims it will explore opportunities **to provide more granular information for future reports** (Google, 2023, p. 20).

Meta

The responses for both Facebook and Instagram are similar yet not identical. They offer details on the ad review system at Meta. The review process can include the specific components of an ad such as images, video, text and targeting information, as well as an ad's associated landing page and other information (Meta, 2023, pp. 17-18). **Again there is a lack of data provided on breakdown by member states for the SLI.** General figures repeat those detailed above for EU adds removed for misinformation policy and total ad removal overall.

Microsoft

All advertising that runs on **LinkedIn**'s platform is subject to LinkedIn's Advertising Policies. LinkedIn has implemented both automated and manual systems to help ensure that advertising on the platform complies with its Advertising Policies, and that ads that do not comply with its policies are removed. Please see QRE 2.2.1. **Microsoft Advertising** blocks sites or domains that our Information Integrity expert partners deem as spreading Disinformation. Microsoft Advertising also rejects all ads associated to such domains and instructs its publishing partners to block ads from showing on such domains. We did not receive any request to remove content during the reporting period, which may be due to the proactive nature of our actions. In terms of the SLI **there were no restricted ads or impressions for LinkedIn with respect to Bulgaria**, however for Microsoft Advertiser 39 ads were prohibited in Bulgaria (Microsoft, 2023, pp. 29-30).

TikTok

Same answer as QRE 2.2.1 (TikTok, 2023, p. 12).

Twitter

Twitter offers no specific response to measures 2.1 - 2.4. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, p. 6).

QRE 2.4.1 - *Signatories will describe how they provide information to advertisers about advertising policies they have violated and how advertisers can appeal these policies*

SLI 2.4.1 *- Signatories will provide relevant information to advertisers about which advertising policies have been violated when they reject or remove ads violating policies described in Measure 2.1 above or disable advertising accounts in application of these policies and clarify their procedures for appeal.*

Google

This response was supported by detailed and relevant qualitative and quantitative data. Google outlines the notification policy for adverts that do not follow their policies as well as the appeals process, Self-serve appeals (SSA) process and details, and then goes on to provide the details for the number of ads and appeals for member states. Bulgaria had 452 Ad appeals, of which 106 were successful. 108 appeals were partially successful and 238 appeals failed (Google, 2023, pp. 20-21).

Meta

The responses for both Facebook and Instagram are both similar yet not identical. The text again details the ad review system which **Meta** employs. **No SLI data provided.** "We were not able to deliver this SLI in the time provided for this baseline report. We are working to improve our SLIs across chapters in our next report in January-June 2023." (Meta, 2023, pp. 19-20).

Microsoft

Once an ad is restricted or rejected for violation of policies, LinkedIn sends the advertiser an email notification outlining the rejection reason and policy violated. The email also provides information on how advertisers can address the violation. For Microsoft advertising, advertisers are also notified through Email, Prompts in the campaign user interface, and notifications from the assigned account representatives (Microsoft, 2023, pp. 31-32).

TikTok

TikTok notes that where an advertiser has violated an advertising policy, they are informed by way of a notification visible on their TikTok Ads Manager account or a representation. Then advised of any violations. **No data available for the SLI** (TikTok, 2023, p. 13).

Twitter

Twitter offers no specific response to measures 2.1 - 2.4. All that is offered is an overview of Twitter's existing advertising policies (Twitter, 2023, p. 6).

Table 2 - ***Ratings for Measure 2***

Commitment 2							
	Google	**Meta**		**Microsoft**		**Tik Tok**	**Twitter**
		Facebook	*Instagram*	*LinkedIn*	*Microsoft Advertising*		
Measure 2.1	3	2	2	3	2	2	1
Measure 2.2	1	1	1	3	3	1	1
Measure 2.3	1	1	2	2	1	1	1
Measure 2.4	3	2	2	2	2	1	1

Measure 6

QRE 6.1.1 - *Relevant Signatories will publicise the best practises and examples developed as part of Measure 2.2.1 and describe how they relate to their relevant services.*

Google

Notes that Google Advertising's additional compliance on these measures will be based on the EU Political Ads legislation (Google, 2023, p. 30).

Meta

Extensive explanation provided around the use of disclaimers for ads about social issues, elections or politics. There are same responses for Facebook and Instagram (Meta, 2023, p. 29)

Microsoft

Note that commitment 6 is not relevant or pertinent for LinkedIn and Microsoft advertising as they do not allow political or issue-based advertising as set out in more detail under measure 5.1 (Microsoft, 2023, p. 41).

TikTok

TikTok argues that this measure does not apply to them as they prohibit political advertising, they do not allow political actors to place advertising nor do they allow ads and landing pages. They say they allow cause based advertising and public services advertising from government agencies, non-profits and other entities if they are not driven by partisan political motives (TikTok, 2023, p. 15).

Twitter

Twitter respond that Commitments 4-13 are not relevant to their current approach to political and issue advertising in Europe at the time of writing. They claim this may change and that Twitter will relaunch its Advertising Transparency Centre in line with the DSA (Twitter, 2023, p.10).

SLI 6.2.1 - *Relevant Signatories will publish examples of how sponsor identities and other relevant information are attached to ads or otherwise made easily accessible to users from the label.*

Google

The response was detailed and offered information on how election adverts in regions where verification is required must contain a disclosure which identifies who paid for the ad. For various ad formats, advertisers are responsible for including a "Paid for by" disclosure directly in the ad, followed by the name of the organisation or individual paying for it. These formats include Third Party ad serving on Google Display Network and YouTube, Audio creates and Native creatives on DV360 and Video creatives on DV360 (except for creatives served on YouTube). Apart from the disclosures present within the ads, ads from verified advertisers also include 'About This Ad' and 'Why this Ad' features, which give users access to additional information related to the advertiser identity and on why particular ads are being shown on Google Search, YouTube and other Google services. This response covers the response QRE 6.2.2. SLI 6.2.1 then offers the breakdown across markets. In Bulgaria there were 1,727 creatives from verified advertisers labelled for EU Election Ads. The amount spent by verified advertisers on Creatives labelled for EU election ads in Bulgaria was €138,959.92 (Google, 2023, pp. 30-32).

Meta

The response of Meta notes that ads about social issues, elections, or politics require authorizations and a 'paid for by" disclaimer if the ad content includes specific things including advocacy for a politician, candidate or party etc. Data is provided for the number of ads accepted and labelled with social issues, elections or politics (SIEP) disclaimers on Facebook and Instagram combined. For **Bulgaria** the number was over 1,100 (Meta, 2023, pp. 30-31).

Microsoft

Note that commitment 6 is not relevant or pertinent for LinkedIn and Microsoft advertising as they do not allow political or issue-based advertising as set out in more detail under measure 5.1 (Microsoft, 2023, p. 41).

TikTok

TikTok argues that this measure does not apply to them as they prohibit political advertising, they do not allow political actors to place advertising nor do they allow ads and landing pages. They say they allow cause based advertising and public services advertising from government agencies, non-profits and other entities if they are not driven by partisan political motives (TikTok, 2023, pp. 15, 17).

Twitter

Twitter responded that Commitments 4-13 are not relevant to their current approach to political and issue advertising in Europe at the time of writing. They claim this may change and that Twitter will relaunch its Advertising Transparency Centre in line with the DSA (Twitter, 2023, p.10).

Table 3 - *Ratings for Measure 6*

Commitment 6								
	Google	**Meta**			**Microsoft**		**Tik Tok**	**Twitter**
		Facebook	*Instagram*	*Messenger*	*LinkedIn*	*Microsoft Advertising*		
Measure 6.1	1	2	2	N/A	N/A	N/A	N/A	N/A
Measure 6.2	2	1	1	N/A	N/A	N/A	N/A	N/A
Measure 6.3	1	1	2	N/A	N/A	N/A	N/A	N/A
Measure 6.4	N/A	3	3	N/A	N/A	N/A	N/A	N/A

Measure 6.5		N/A	N/A	3				

Measure 7

***QRE 7.1.1** - Relevant Signatories will report on the tools and processes in place to collect and verify the information outlined in Measure 7.1.1, including information on the timeliness and proportionality of said tools and processes*

***SLI 7.1.1** - Relevant Signatories will publish meaningful metrics on the volume of ads rejected for failure to fulfil the relevant verification processes, comparable to metrics for SLI 6.2.1, where relevant per service and at Member State level.*

Google

QRE 7.1.1 reiterates that all election ads run by verified election advertisers in regions where election ads verification is required must contain a disclosure that identifies who paid for the ad. To provide transparency for users, these ads are included in the Political Advertising Transparency Report. The verification process is also outlined. In Bulgaria 232 adverts were rejected due to unverified advertisers attempting to run EU election ads in Q3'2022 (Google, 2023, pp. 34-37).

An interesting note for Google, they are not signed up to Commitment 12, which commits signatories to increase oversight of political and issues advertising and assist in the creation, implementation and improvement of political or issue advertising policies and practices.

Meta

Meta's response includes notes on the identity confirmation requirements. Mentions that local representatives in a country can complete authorisations for that country. In terms of SLI response, the number of unique ads rejected for not complying with the policy on SIEP ads on both Facebook and Instagram from October 1 to December 31 2022 in Bulgaria was over 3,600 (Meta, 2023, pp. 35-37). **Again very imprecise numbers and same numbers for both platforms seems strange**.

Twitter

Twitter responded that Commitments 4-13 are not relevant to their current approach to political and issue advertising in Europe at the time of writing. They claim this may change and that Twitter will relaunch its Advertising Transparency Centre in line with the DSA (Twitter, 2023, p.11).

***QRE 7.3.1** - Relevant Signatories will report on the tools and processes in place to request a declaration on whether the advertising service requested constitutes political or issue advertising.*

***QRE 7.3.2** - Relevant Signatories will report on policies in place against political or issue ad sponsors who demonstrably evade verification and transparency requirements on-platform*

Microsoft

Notes that, as mentioned in QRE 5.1.1, LinkedIn's advertising Policies prohibit political advertising. It is also noted that LinkedIn has implemented both automated and manual systems to help ensure that advertising on the platform complies with these policies. Ads that do not comply are removed. LinkedIn members can also report adverts for what they believe might be breaches and a team will review. Microsoft Advertising's policies also prohibit certain types of ads that might be considered issue-based.

For QRE7.3.2 it is noted that Microsoft advertising employs dedicated operational support and engineering resources to enforce restrictions on political advertising using a combination of proactive and reactive mechanisms (Microsoft, 2023, p. 43). Proactively blocking political ads. Reactively removing adverts that violate their policies. For Microsoft advertising the response for QRE7.3.1 could be said to be off topic (on definition of political/issue ads, not on tools and processes for determining ads). The response to QRE 7.3.1 is under QRE7.3.2, no detail is provided on QRE7.3.2 (policies in place on evasion of verification/transparency requirements).

TikTok

TikTok argues that this measure does not apply to them as they prohibit political advertising, they do not allow political actors to place advertising nor do they allow ads and landing pages. They say they allow cause based advertising and public services advertising from government agencies, non-profits and other entities if they are not driven by partisan political motives (TikTok, 2023, p. 15).

Twitter

Twitter responded that Commitments 4-13 are not relevant to their current approach to political and issue advertising in Europe at the time of writing. They claim this may change and that Twitter will relaunch its Advertising Transparency Centre in line with the DSA (Twitter, 2023, p.11).

Table 4 - *Ratings for Measure 7*

Commitment 7							
	Google	Meta		Microsoft		TikTok	Twitter
		Facebook	*Instagram*	*LinkedIn*	*Microsoft Advertising*		
Measure 7.1	3	2	2	N/A	N/A	N/A	N/A
Measure 7.2	3	2	2	N/A	N/A	N/A	N/A
Measure 7.3	3	2	2	2	1	1	1
Measure 7.4	1	1	1	N/A	N/A	1	1

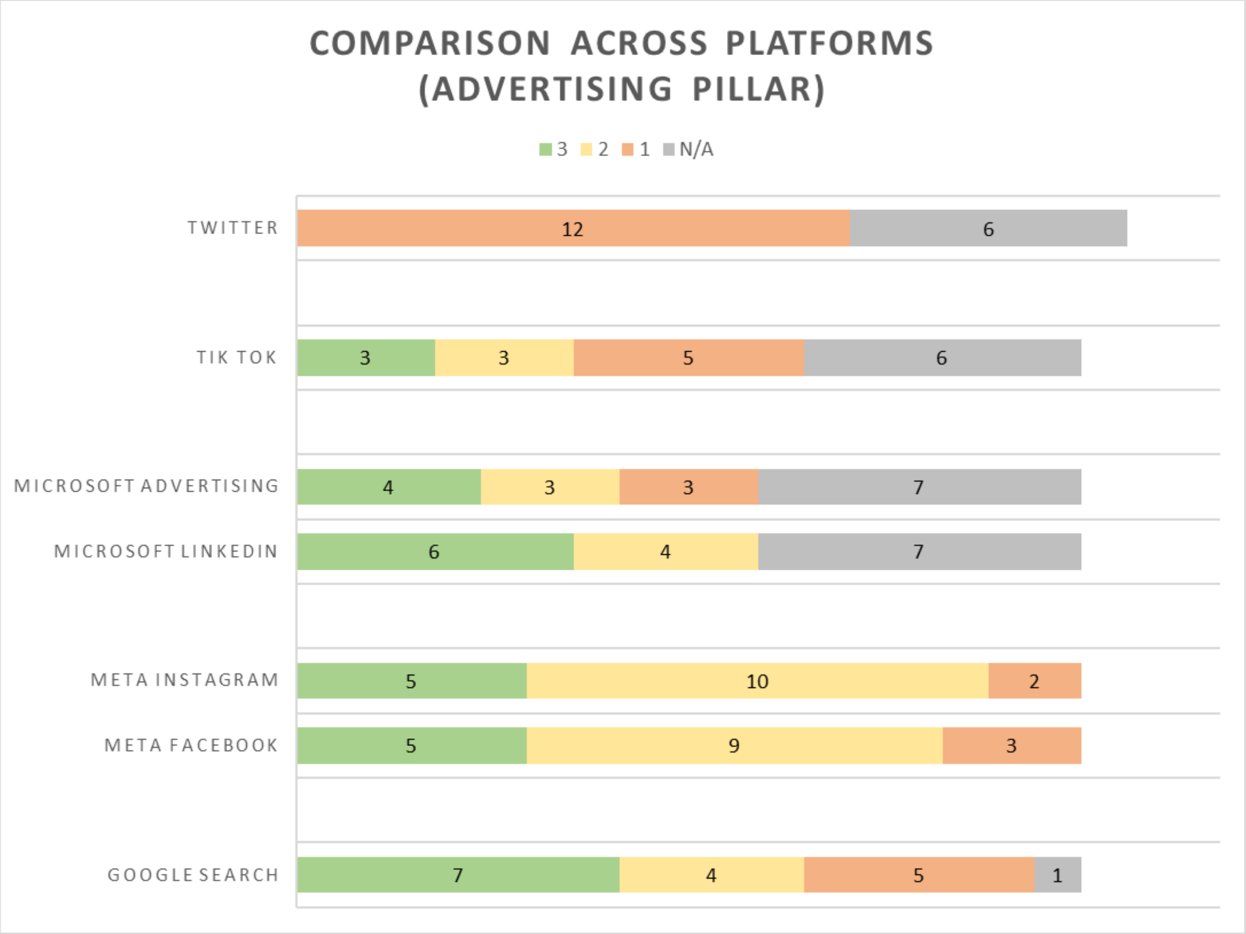
2.1.1. Advertising Pillar Conclusion

On **Measure 1** there was a fairly even degree of compliance from measures 1.1 to 1.6. One of the key questions comes around Microsoft advertising's response in QRE1.1.1 which was brief and did not offer very much detail in order to satisfy QRE1.1.1 or SLI 1.1.1. For **Measure 2** TikTok, Google and Facebook were rated low in compliance with respect to measures 2.2 and 2.3. These low ratings were generally reflective of a lack of information being provided in the response, such as a lack of member state level information. This lack of data was generally accompanied by an indication that more data will be presented in forthcoming reports. With respect to **Measure 6,** neither Microsoft nor TikTok participated in this response, claiming it is not applicable to them. To measure 6.1. Google provided an only partial response which included no description of current operations and procedures. For measure 6.3 Google offered an unclear response which left questions around where the referred to research materials could be found. For **Measure 7**, only Google and Meta were full participants and both scored low for compliance with Measure 7.4 due to the lack of data and response that Google will make efforts to provide in future reports. Meta simply refers to QRE 7.1.1 and SLI 7.1.1 and this seems to indicate they find QRE 7.4.1 irrelevant. TikTok also scored low for compliance on 7.4 through no response. They refer to measure QRE 5.1.1 but measure 7.4 is on (effectiveness) of declaring political/issue ads // verifying the identity of political or issue ad sponsors.

There is a degree of confusion on Measure 7.4 Meta simply refers to QRE 7.1.1 and SLI 7.1.1 and this seems to indicate they find QRE 7.4.1 irrelevant. It is also fair to say that the reported actions taken or measures were relevant overall in terms of what was outlined in the measures, commitment and pillar overall. Just generally, there was a trend of responses which lacked the requested data. **Very often data on the state level was missing. Where data is provided, it is useful, however there are conflicts in terms of time period reported which hinders comparability. While links and data are provided there are questions on the verifiability of some of the information provided.**

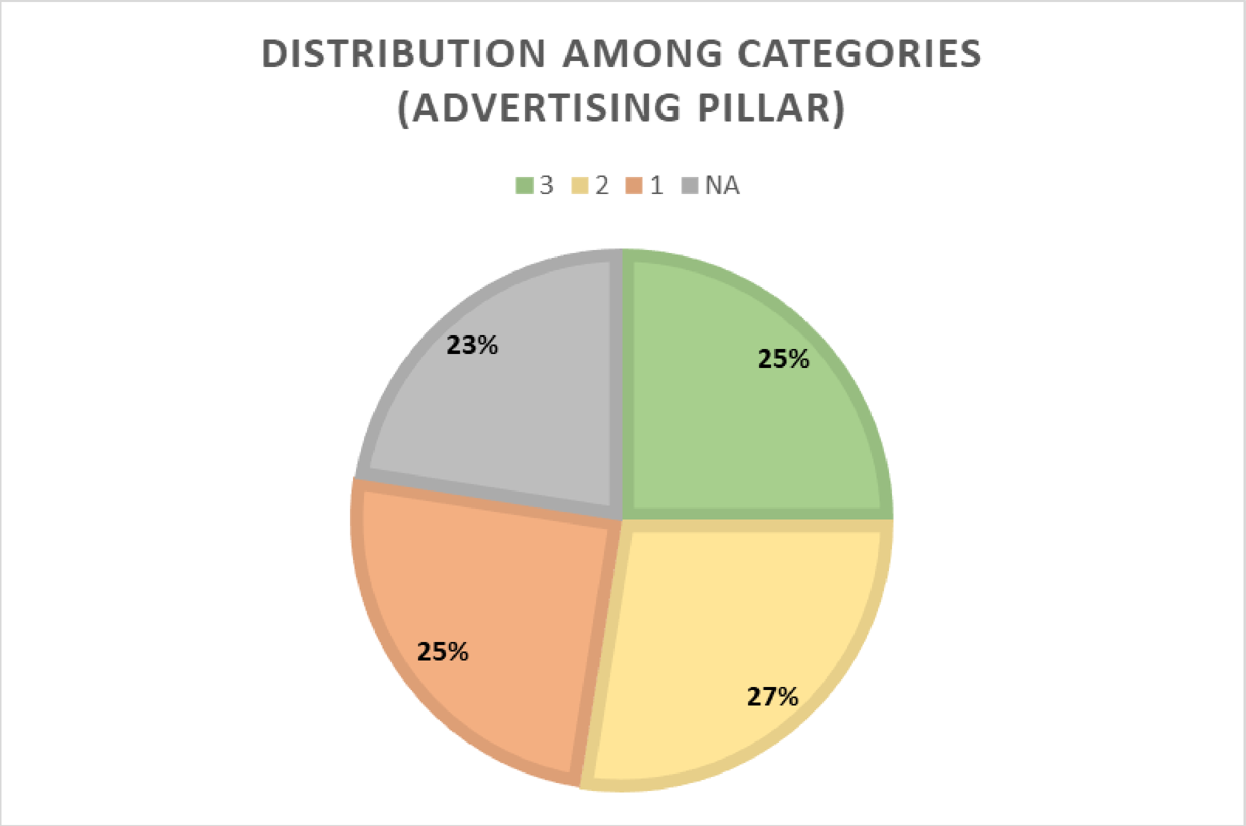
In terms of **Bulgaria**, On SLI 1.1.1, Google noted that the number of Actioned AdSense Pages for Bulgaria was 56,529. The number of Actioned AdSense Domains for Bulgaria was 12. Additionally, the estimated cost of Blocked Requests on pages for Bulgaria was €95,500.94, while the estimated costs of Blocked Requests on Domains for Bulgaria was €119,496.31. Again for SLI 2.1.1, Google reported that there were 312,683 creatives actioned for destination requirements in Bulgaria. The number of creatives actioned for inappropriate content in Bulgaria was 1,994, while the number of creatives actioned for misrepresentation was 110,307. Microsoft reported under SLI 2.3.1 that there were no restricted ads or impressions for LinkedIn with respect to Bulgaria, however for Microsoft Advertiser 39 ads were prohibited in Bulgaria. For SLI 6.2.1, Google again reported that there were 1,727 creatives from verified advertisers labelled for EU Election Ads. The amount spent by verified advertisers on Creatives labelled for EU election ads in Bulgaria was €138,959.92. And in response to SLI 7.1.1 Google reported that 232 adverts were rejected in Bulgaria due to unverified advertisers attempting to run EU election ads in Q3'2022. **This leaves us with some interesting data and also a finding that it is largely Google who are providing member level details and specifically details on Bulgaria in this pillar. Microsoft also appeared once, but TikTok and Meta have provided no country level data for Bulgaria in this Pillar**.

Figure 1 - *Advertising Pillar Rating Comparison*



In Figure 1 (above), the green colour stands for responses rated as 'Good', the yellow colour represents a rating of 'Adequate', the red colour covers 'Poor' responses and the grey colour represents N/A. As Figure 1 notes **Google** scored the highest number of 'Good' responses within this pillar with 7, while simultaneously scoring the joint highest number of 'Poor' responses with 5. The other company with the same number of 'Poor' responses was **TikTok. Microsoft** scored the highest number of N/A responses under the advertising pillar, with 7 of these responses rated for both Microsoft Advertising and LinkedIn. LinkedIn, however also scored the second highest number of 'Good' responses with 6 rated as such. **Twitter** offered the poorest set of answers with no specific responses to measures 1.1 - 1.6, only an overview of Twitter's existing advertising policies. Twitter further responded Commitments 4-13 are not relevant to their current approach to political and issue advertising in Europe. They claim this may change and that Twitter will relaunch its Advertising Transparency Centre in line with the DSA (Twitter, 2023, p.10). In Figure 2 (below), the distribution of the various ratings of the responses under this pillar can be seen. 21% of Measures were not responded to, 29% of responses overall were rated as 'Good', 32% were rated as 'Adequate' and 18% were rated as 'Poor' responses.

Figure 2 - *Distribution of Rating Categories*



2.2 Integrity of Services

The Code aims to strengthen the measures to reduce manipulative behaviour used to spread disinformation (e.g. fake accounts, bot-driven amplification, impersonation, malicious deep fakes), and establish stronger cooperation among signatories to fight the challenges related to such techniques. A cross-service understanding of unpermitted manipulative behaviours and practices to spread disinformation should be agreed upon among signatories. They are also required to periodically review the list of tactics, techniques and procedures (TTPs) employed by malicious actors, and will implement clear policies, covering the range of behaviours and practices identified. Our findings about research on disinformation in the EU note that most visible examples of disinformation are not simply false content. Misrepresentation of sources, communities, and popularity lies at the heart of disinformation tactics and narratives. Commitments designed to deal with inauthentic and misrepresentative behaviour are therefore crucial to tackling the reality of contemporary disinformation dissemination.

Measure 14

QRE 14.1.1 *- Relevant Signatories will list relevant policies and clarify how they relate to the threats mentioned above as well as to other Disinformation threats.*

Google

Responds that Google's search systems are designed to elevate authoritative information and combat the threats listed in Commitment 14 at scale. It is stated that many of those TTPs are not relevant to search engines (e.g. TTPs 1-5, TTP 11) the answer goes on to outline how Google Search's ranking systems directly tackle threats like TTP4-TTP10. The response also links to Google Search's Overall Content Policies and Google Search Webmaster Guidelines. YouTube's systems are also said to prioritise authoritative sources. YouTube also has various policies which set out what is not allowed on YouTube, accessed via the landing page in YouTube's health Centre which addresses TTPs. TTPs 1, 2, 3, 5, 9, 10, and 11 are covered, in whole or in part, by YouTube's Spam & Deceptive Practices Policies (Google, 2023, pp. 55-56). **The responses need more details about proactive efforts to detect TTPs**. With respect to the Threat Analysis Group and the Trust & Safety teams more details would be helpful in terms of clarification and demonstration of effectiveness. **The answer to QRE 14.1.2 redirects to QRE 16.1.1 which does not supply sufficient information.**

Microsoft

LinkedIn's User Agreement (in particular section 8 LinkedIn "Dos and Don'ts") and Professional Community Policies detail the impermissible manipulative behaviours and practices that are prohibited on the platform. Fake accounts, misinformation, and inauthentic content are not allowed, and active steps to remove it from the platform are taken. **Bing Search** does not have a news feed for users, allow users to post and share content, or otherwise enable content to go "viral." Therefore, addressing misinformation in organic search results often requires a different approach than may be appropriate for other types of online services. Many of the TTPs are more pertinent to social media (e.g., those relating to user accounts, subscribers/followers, influencers, or targeting users of a service) and do not apply to search engines. Instead, Bing's policies primarily address TTP Number 10, which concerns the use of deceptive practices to attempt to deceive or manipulate search ranking algorithms, such as by exploiting data voids, spam tactics, or keyword stuffing (Microsoft, 2023, pp. 50-51).

Meta

Meta outlines their approach to Coordinated Inauthentic Behaviour (CIB) as being grounded in behaviour based enforcement. They focus on violating behaviours exhibited by violating actors, rather than on content. When CIB networks are taken down it is behaviour and not content based. The response for both Facebook and Instagram is similar, focusing on additional efforts to tackle inauthentic behaviour by fake accounts at scale by blocking accounts from being created, removing accounts when they sign up or removing existing accounts (Meta, 2023, pp. 54-56). **This is a very comprehensive overview of the process which could be expanded upon by adding information about Fake Engagement.**

TikTok

TikTok specially-designed tools to implement policies are instruments that toggle to disclose branded content - see QRE 14.1.1 or human investigations to detect deceptive behaviours (for covert influence activities - see QRE 14.1.2). TikTok notes that it wishes to implement the following measures in the next 6 months: Increased transparency into the work on CIO and published (here) insights into the networks they identify and remove from their platform globally; expanded the rollout of their branded content toggle to all EEA users. In **SLI 14.2.1-14.2.4** it is noted that the number of fake accounts removed by TikTok in Bulgaria was 7602, these fake accounts had 267,899 followers at the time of removal. Likewise, the number of fake 'Likes' removed for Bulgaria was 75,224, with the number of fake 'Likes' prevented estimated to be 4.6 million. The number of fake followers removed for Bulgaria is said to be 829,090, while the number of fake follows prevented is said to be 120,064. The number of accounts banned under the impersonation policy in Bulgaria is said to be 118. The number of times the branded content toggle has been used to disclose the existence of a commercial relationship was 7,239 in Bulgaria (TikTok, 2023, pp. 30-47).

Twitter

In the second half of 2022, Twitter's Threat Disruption team reportedly collaborated with 3 stakeholders to counter coordinated attempts at manipulating the platform, often involving the spread of disinformation. The frequency of these efforts varied based on the prevalence of such activity on the platform. Here are some key investigations and actions taken by Twitter:

Investigation: 2022-09-06 Disruption: 2022-09-23 Actioned Assets: 149 accounts Meta (formerly Facebook) shared 1133 accounts engaged in coordinated inauthentic behavior, primarily linked to a Russian influence operation. Twitter investigated these accounts, which exhibited ties to Russian infrastructure and geopolitical interests (Twitter, 2023, pp.18-19).
Investigation: 2022-09-28 Disruption: 2022-09-30 Actioned Assets: 15 accounts Twitter investigated a Russian disinformation campaign using fake media outlets associated with Sputnik, targeting European audiences. The investigation revealed that these accounts were linked to Sputnik assets and displayed identity deception through location inconsistencies, but consistently aligned with Russian geopolitical interests (Twitter, 2023, pp.18-19).
Investigation: 2022-06-30 Disruption: 2022-07-01 Actioned Assets: 7 accounts An investigation into politically motivated inauthentic behavior attributed to the Internet Research Agency (IRA) by Google. The IRA-linked accounts were engaging in pro-Putin and anti-American content. Twitter identified and suspended a small number of these accounts based on the information provided by Google (Twitter, 2023, pp.18-19).

QRE 14.2.1 - *Relevant Signatories will report on actions taken to implement the policies they list in their reports and covering the range of TTPs identified/employed, at the Member State level*.

SLI 14.2.1 - *Number of instances of identified TTPs and actions taken at the Member State level under policies addressing each of the TTPs as well as information on the type of content.*

Google

The response notes that **Google Search** relies on a combination of people and technology to enforce its policies, Machine Learning (ML) is said to play a critical role in content moderation on Google Search. There are inbuilt systems which weight authoritativeness. Algorithms examine various factors and signals "to raise authoritative content and reduce low quality content" (Google, 2023, p. 59). The response also points to Google Search's publicly available website, How Search Works and notes that search is looking to continuously improve the quality and effectiveness of its automated systems to protect platforms and users from harmful content. The process for testing of algorithmic high standards is also discussed. As is the Google Search Quality Rater Guidelines. **YouTube** responds with links to its policies on Community Guidelines, Channel Monetisation Policies, including Advertiser Friendly Content Guidelines. In terms of implementation and enforcement it is described as a joint effort between people and technology as before. It discusses enforcement guidelines, differentiation between violative and non-violative material (Google, 2023, pp. 59-60). If this results in high levels of accuracy, then the team is expanded and the rest of the process follows. There was no information provided regarding the following: "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.)" (European Commission, 2022, p. 16) **SLIs are also missing or inappropriate for this specific commitment**.

Microsoft

Once again, limited implementation, in particular regarding the implementation of Bing's anti-abuse policies set out in Measure 14.1. Reference is made to LinkedIn's Professional Community Policies and how these processes manage disinformation and misinformation. QRE 14.2.1 provides only the SLI notes that the number of fake accounts LinkedIn prevented or restricted until December 1-31, 2022 in Bulgaria was: 65,426. The same number is used for the number of actions taken by type and the number of identified TTPs. The number of instances of fake accounts reported in TTP1 that engaged with a feed or post for Bulgaria was 216. The number of instances of fake accounts reported in TTP1 that followed a LinkedIn profile or page was 12,418 for Bulgaria. The number of LinkedIn pages or groups created in December 2022 by the fake accounts reported in TTP 1 SLI 14.1.1 was 1 for Bulgaria. SLI 14.2.2-4 is not applicable as stated by Microsoft (Microsoft, 2023, pp. 53-57).

Meta

Meta notes that they report on a quarterly basis on enforcement actions taken under the two more relevant policies to the commitment. Fake accounts policies is the first and it is reported that in Q3'2022 Facebook took action against 1.5 billion fake accounts (96% of which were found proactively). They provided details on estimates for the percentage of fake accounts vis-a-vis monthly active users globally. The second is the coordinated inauthentic behaviour policies and they offer lists of three networks which were taken down which targeted at least one country in Europe, originating in the United States, Russia and China (Meta, 2023, pp. 57-59). There was no information provided regarding the following: "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.)" (European Commission, 2022, p. 16) In terms of the SLI responses to 14.2.1-4 data on Member State Level missing and instead we are provided with data from the Q3'2022 quarterly report. It does offer some intriguing data on the networks and the countries they are targeting (Meta, 2023, p. 60) but is **not responsive to what the SLI and Commitment set out here. The response is also limited in terms of focus of their reporting to specific data around coordinated inauthentic behaviour,** when the TTPs mentioned in the Commitment can be applied outside of CIB operations (i.e. when no coordination is needed between different actors). So **the focus is slightly misaligned here**.

TikTok

TikTok notes that implementation of their policies is done by different means, including specifically designed tools or human investigations to detect deceptive behaviours (TikTok, 2023, p. 29). The implementation of these policies are said to be ensured through enforcement measures applied in all member states. TikTok goes on to describe the process when an account amounts to an impersonation, or is involved with networks removed in the past as part of a CIO. The data they provide is limited to Q3'2022. For Bulgaria during this period the number of fake accounts removed was 7,602. The number of followers of fake accounts identified at the time of removal for Bulgaria was 267,899. The number of fake likes removed for Bulgaria was 75,224 and the number of fake likes prevented were 4,600,356. The number of fake followers removed in Bulgaria was 829,090, while the number of fake follows prevented for Bulgaria was 120,064. 118 accounts were banned in Bulgaria under impersonation policy (TikTok, 2023, pp. 30-38). 7,239 was the number of times branded content toggle has been used to disclose the existence of a commercial relationship in Bulgaria.

Twitter

No response provided on this measure (Twitter, 2023, p. 23).

Table 5 - *Ratings for Measure 14*

Commitment 14								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 14.1	2	3	3	3	1	2	2	2
Measure 14.2	1	1	1	1	1	2	1	1
Measure 14.3	3	3	3	3	3	3	3	1

Measure 15

Signatories that develop or operate AI systems and disseminate AI-generated and manipulated content through their services commit to take into consideration the transparency obligations and the list of manipulative practices prohibited under the proposal for Artificial Intelligence Act.

Google

Google states that it has deployed AI principles setting out Google's commitment to develop technology responsibly including issues relevant to this commitment. Google search also deployed and enforced the Manipulated Media Policy. YouTube has deployed and enforced Misinformation Policies and Spam and Deceptive Practices Policies. In QRE 15.2.1 Google Search outlines its application of AI for Training and high quality data, Rigorous Evaluation, Responsible application design and Minding Google Search's footprint. YouTube's product, policy and enforcement decisions are guided by considerations such as Value openness and accessibility, respecting end-user rights, build for everyone. **YouTube also offers limited and superficial information on algorithms** (Google, 2023, pp. 79-83).

Microsoft

Not applicable - statement that if/when such features are launched that LinkedIn and Bing will determine any appropriate measures to implement (Microsoft, 2023, pp. 66-68).

Meta

Facebook states that it has dedicated AI models and systems to identify manipulated media, including deep fakes. They claim to remove such videos and for other manipulated media they treat it as disinformation. **They do not plan further enhancements to this** (Meta, 2023, pp. 64-67).

TikTok

Respond that they expect to be able to report on additional developments to help ensure their AI algorithms **comply with Measure 15.2 of the Code, in the next reporting period** (TikTok, 2023, pp. 49-50).

Twitter

Twitter's policy on synthetic and manipulated media aims to prevent the sharing of content that deceives or confuses people and leads to harm. They may label or remove tweets containing misleading media to provide context and authenticity. To be labeled or removed, content must feature significantly altered, manipulated, or fabricated media, shared deceptively or with false context, and potentially cause widespread confusion, public safety issues, or serious harm. Twitter considers factors like media editing, added visual or auditory elements, use of filters, and AI-generated content to determine if media are misleading. If they are unable to determine whether content is being shared with false context, they will not take action. Context matters too; if media are presented as factual when they're not, Twitter may take action. If content poses threats, promotes abuse, risks violence, or hinders public services, it may be removed. Media not causing immediate harm but impacting public safety might be labeled. Even if not violating other rules, Twitter may remove borderline cases involving misleading media (Twitter, 2023, p.26).

Table 6 - *Ratings for Measure 15*

Commitment 15								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 15.1	3	3	2	2	1	3	3	2
Measure 15.2	3	1	3	3	1	1	2	2
Measure 15.3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

Measure 16

Commit to operate channels of exchange between their team in order to proactively share information about cross-platform influence operations, foreign interference in information space and relevant incidents that emerge on their respective services, in full compliance with privacy legislation and with due consideration for security and human rights risk.

Google

Reports on Google's TAG and Trust & Safety Team work to monitor malicious actors around the globe, disable their accounts, and remove the content that they post, including but not limited to coordinated information operations and other operations that may affect EEA member states (Google, 2023, pp. 84-85). TAG also engages with other platform Signatories to receive and, when strictly necessary for security purposes, share information related to threat actor activity – in compliance with applicable laws. **Here more details are needed**. **SLI 16.1.1 not provided in the way requested.** YouTube and Google search say they are committed to providing more granular information regarding 16.1.1 for future reports. QRE 16.2.1 is only applicable for YouTube but describes Internet Research Agency (IRA) linked influence operations (IO) with specific focus on Russian oligarch Yevgeny Prigozhin and the IRA, who is alleged to have peddled influence campaigns around the interests of Russia and Prigozhin's Wagner Group in Africa and disinformation and deception tactics carried out by this group (Google, 2023, pp. 84-86). The disinformation narratives were appealing to things like African pride and empowerment and narratives around the impact of western imperialism on Africa. Some of the authors were likely unwitting accomplices. Google terminated various IRA linked YouTube channels that appear French and supportive of Russian policy objectives in Libya (Google, 2023, pp. 84-86).

Meta

Facebook discusses its work with government authorities, law enforcement, security experts, civil society and other tech companies to stop emerging threats by establishing a direct line of communication, sharing knowledge and identifying opportunities for collaboration (Meta, 2023, p. 68). The information provided also discusses partnerships with tech companies and across civil society, and this effort should continue to expand going forward to study these networks' cross platform behaviours. They also discuss CIB takedowns and IO threats and how these often involve partnerships and peers. The same information is given for Instagram (Meta, 2023, pp. 67-70). **The SLIs are not provided** and it is stated they are seeking to improve SLIs across chapters in their next report. However, **Meta does not mention a dedicated forum for cross-platform information sharing**. As TikTok states in their report "The cross-platform sharing forum referred to in Commitment 16 has not been set up yet." **This information is missing here**.

Microsoft

Report that they look forward to working on this commitment with the other signatories as they develop further cross-platform information sharing (Microsoft, 2023, pp. 70-72).

TikTok

The cross-platform sharing forum referred to in Commitment 16 has not been set up yet. **No time frame predicted** (TikTok, 2023, pp. 51-52).

Twitter

For several years, Twitter has collaborated with various companies and platforms to maintain consistent communication between its Threat Disruption and Site Integrity teams and their counterparts in peer organizations. Twitter mentioned some information operation case studies to show their effort towards continued partnership and cooperation (Twitter, 2023, pp.26-28). This approach and detail does not respond to the specifics of the measure.

Table 7 - *Ratings for Measure 16*

Commitment 16								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 16.1	2	2	2	2	2	2	1	1
Measure 16.2	3	N/A	1	1	1	N/A	1	1

2.2.1. Integrity of Services Conclusion

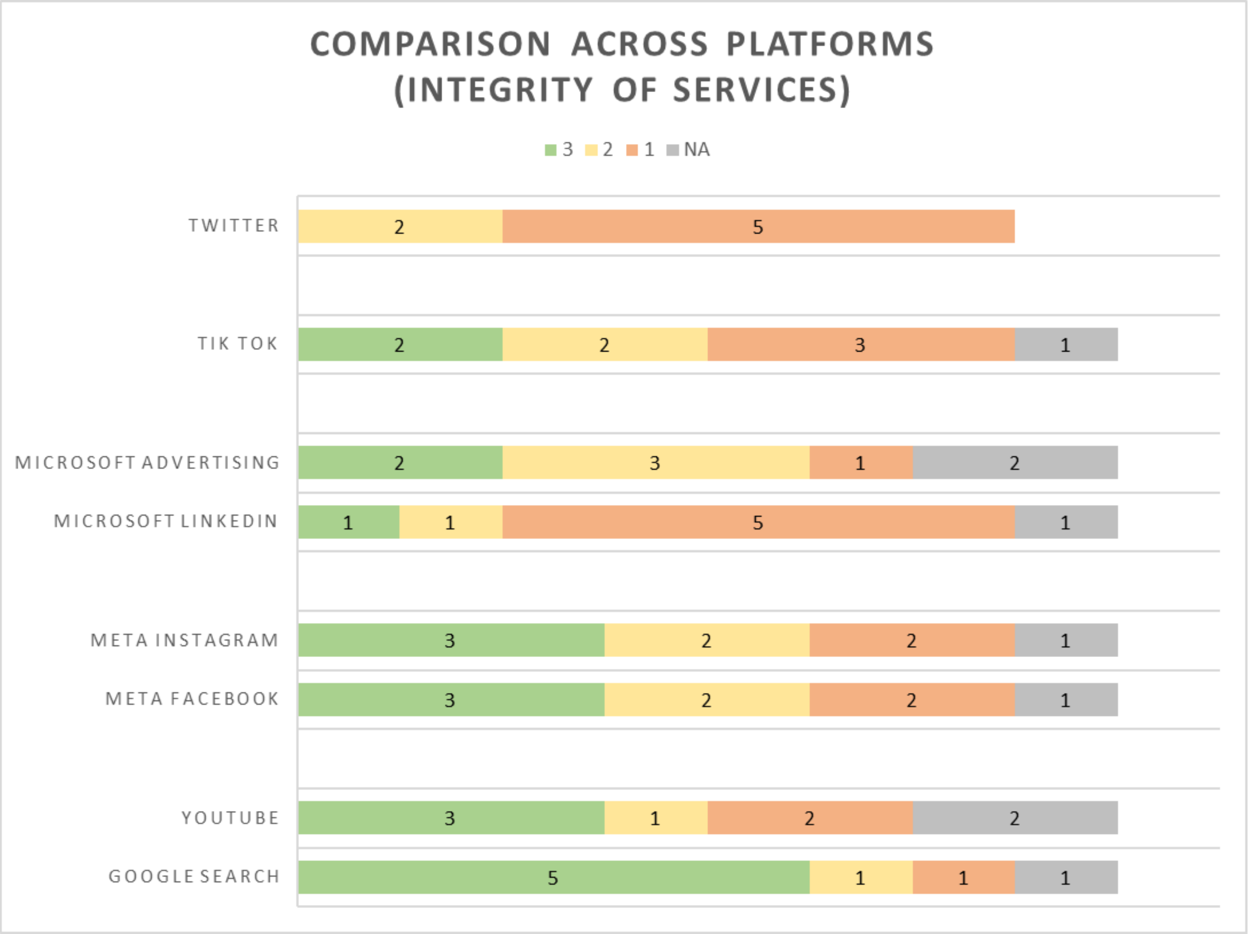
On **Measure 14** there were various low scores of compliance to Measure 14.2, again this was due to missing SLIs. For Google, a lack of information provided regarding "Relevant Signatories will also develop further metrics to estimate the penetration and impact that Fake/Inauthentic accounts have on genuine users and report at the Member State level (including trends on audiences targeted; narratives used etc.)" (CoP, .16). For Microsoft, the response represented limited implementation, in particular regarding the implementation of Bing's anti-abuse policies set out in Measure 14.1. In Terms of Meta, the SLI data was not provided and instead replaced with a Q3 '2022 quarterly report which falls outside the scope of the response. Meta's responses here were also limited in terms of focus of their reporting to specific data around coordinated inauthentic behaviour (CIB), when the TTPs mentioned in the commitment can be applied outside of CIB operations. For TikTok the response was rated low because it contained superficial information regarding implementation of policies provided. There were substantive questions not addressed such as number of people working on CIO or Content Moderation? In terms of SLI 14.2.2 TikTok reports that the number of Fake accounts as a % of MAUs is 0.0067 - which is suspiciously low. In comparison, Meta estimates the number on their platforms to be circa 5 %. **The discrepancy in these claims warrants further**

investigation and precision. Measure 15.1 and 15,2 saw Linked in score low as Microsoft states that QRE 15.1.1 is "not applicable", even though they 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'')." **This creates confusion and would seem to suggest that the QRE does apply to LinkedIn**. On Measure 16, TikTok reports that the cross-platform sharing forum referred to in Commitment 16 has not been set up yet, but neglects to provide details on current information-sharing methods with other stakeholders. Microsoft also has not responded to the SLI.

For Measure 14, **Twitter** responded with information partially taken from the **Twitter** Help Site. The information reported is only related to efforts before Elon Musk took over the platform. Information is missing for all other QRE's and SLI's on this measure. There is no note on whether or not Twitter plans to implement further measures going forward. For Measure 15, **Twitter** provided information on Commitment level only. This answer works for QRE 15.1, however some of it is again taken from the Twitter Help Site. For Measure 16, Twitter responds by references its collaborative work on threat disruption and site integrity. This response, while advocating for continued cooperative efforts did not meet the requirements of the commitments. As with the previous pillar, Twitter's responses are the least complete or detailed with missing data and specifics. There is a **degree of incomplete content on Measure 14.2 in Google’s response**. In QRE 14.2 Google provided the Q3’2022 quarterly report, but did not directly address the measure. So a fair assessment was that the reported actions taken or measures were sometimes relevant and sometimes not in terms of what was outlined in the measures, commitment and pillar overall. As we saw in the Advertising Pillar, there was just **generally a trend of responses which lacked the requested data or which lacked depth in terms of response. State level detail was missing frequently. The types of responses given create questions on the verifiability of some of the information provided.**

In terms of **Bulgaria**, In **SLI 14.2.1-14.2.4** it is noted that the number of fake accounts removed by TikTok in Bulgaria was 7602, these fake accounts had 267,899 followers at the time of removal. Likewise, the number of fake ‘Likes’ removed for Bulgaria was 75,224, with the number of fake ‘Likes’ prevented estimated to be 4.6 million. The number of fake followers removed for Bulgaria is said to be 829,090, while the number of fake followers prevented is said to be 120,064. The number of accounts banned under the impersonation policy in Bulgaria is said to be 118. The number of times the branded content toggle has been used to disclose the existence of a commercial relationship was 7,239 in Bulgaria. For SLI 14.2.1, the number of fake accounts LinkedIn prevented or restricted until December 1-31, 2022 in Bulgaria was: 65,426. The same number is used for the number of actions taken by type and the number of identified TTPs. The number of instances of fake accounts reported in TTP1 that engaged with a feed or post for Bulgaria was 216. The number of instances of fake accounts reported in TTP1 that followed a LinkedIn profile or page was 12,418 for Bulgaria. The number of LinkedIn pages or groups created in December 2022 by the fake accounts reported in TTP 1 SLI 14.1.1 was 1 for Bulgaria. For TikTok, in Bulgaria during this period the number of fake accounts removed was 7,602. The number of followers of fake accounts identified at the time of removal for Bulgaria was 267,899. The number of fake likes removed for Bulgaria was 75,224 and the number of fake likes prevented was 4,600,356. The number of fake followers removed in Bulgaria was 829,090, while the number of fake follows prevented for Bulgaria was 120,064. 118 accounts were banned in Bulgaria under the impersonation policy. 7,239 was the number of times branded content toggle has been used to disclose the existence of a commercial relationship in Bulgaria. Within the area of integrity of services, we can observe comparatively more clarity and again Google offers more detailed explanations. **However there are questions around how this data was compiled and who helped to compile it in Bulgaria.**

Figure 3 - *Integrity of Service Rating Comparison*



As in the previous section, within Figure 3 the green colour stands for responses rated as 'Good', the yellow colour represents a rating of 'Adequate', the red colour covers 'Poor' responses and the grey colour represents N/A. As we can see in Figure 3, Google, through its platforms YouTube and Google Search, has provided the largest number of responses rated as 'Good' under the Integrity of Services pillar with 5. Microsoft's LinkedIn was assessed to contain the largest amount of responses rated as 'Poor' within this pillar, also 5. However, Microsoft Advertising, was only rated as 'Poor' on 1 response. TikTok also received a significant amount of responses rated as 'Poor'. Meta scored the second highest number of 'Good' responses outside of Google, with Instagram and Facebook both scoring 3 'Good' ratings, yet both platforms also scored 2 'Poor' ratings.

2.3. Empowering Users, Empowering researchers & Empowering Fact-Checkers

The CoP highlights the significance of equipping users with the capability to detect and report false and/or misleading content in order to constrain the impact of disinformation. To this end, the Signatories have committed to continuing to improve the discoverability of trustworthy content, and to improving the safety of services and enabling users with specific tools to identify disinformation and to report such content, as specified in the 2021 European Commission's Guidance. This commitment is a vital component of limiting the spread of disinformation and will go a long way in helping to create a safer and more honest online environment. There are significant challenges faced by Bulgaria in terms of political polarisation, concentration of media ownership and editorial independence, ongoing work on media literacy, powerful technological influence on information consumption habits, geopolitical context and its position as a low resource language country. These specific vulnerabilities in terms of disinformation and misinformation make empowering users, researchers and fact-checkers a vital component of strategies to tackle the issue at the local level.

Measure 17

***QRE 17.1.1 -** Relevant Signatories will outline the tools they develop or maintain that are relevant to this commitment and report on their deployment in each Member State.*

***SLI 17.1.1 -** Relevant Signatories will report, at the Member State level, on metrics pertinent to assessing the effects of the tools described in the qualitative reporting element for Measure 17.1, which will include: the total count of impressions of the tool; and information on the interactions/engagement with the tool.*

Google

In terms of **Google Search** the answer here describes relevant tools for reporting on metrics pertinent to assessing the effects of the tools, including total impressions of the tool and information on interactions/engagement with the tool. **Google Search** includes the "About This Result" option next to most results, which includes a menu icon that users can tap to discover more about the result or feature. Within this feature, there is also a "More About This Page" link which provides additional insights about sources and topics users find on **Google Search** (Google, 2023, p. 87). This allows users to see more information about the source, find out what others on the web have said about a site and to learn more about the topic. **Google Search** also contains Content Advisory notices which are helpful for users as they highlight when information is scarce or when information is travelling faster than facts. For **YouTube,** the response is a little vague**.** Outside of describing YouTube's commitment to taking its responsibility seriously through outlining clear policies to moderate content on the platform and provide tools that can leverage or improve their media literacy education and better evaluate content and sources. No specifics are given on this. YouTube also has information panels which may appear alongside search results and videos to provide more context (Google, 2023, pp. 87-90). These seem to be applied only to high-profile social and political events. For Bulgaria, the number of times the "More About This Page" feature was viewed in Q3'2022 was 60,548, the number of times the "About This Result" panel was viewed was 415,804. The number of times Content Advisories for low-relevance results were viewed in Bulgaria was 757,400. Finally, the estimated number of times Content Advisories for low quality and rapidly changing results were viewed was 5,640 (Google, 2023, p. 91).

Meta

Meta underlined that since the invasion of Ukraine the company launched educational media literacy campaigns to raise awareness of how to spot misinformation for users in Poland, Slovakia, Lithuania, Latvia, Estonia, Albania, Bosnia and Herzegovina, Kosovo, Serbia and Bulgaria. All of these campaigns were designed in partnership with our local fact-checking partners as well as expert safety NGOs. In terms of SLIs Meta claims to have reached 16 million users and had 72 million impressions with our campaign in all countries. Meta are looking at launching similar campaigns in 2023. There is no breakdown by country. A Youth campaign in Germany, France, Spain, Italy and the UK is said to have reached 50 million users and had 15 million views. **The same numbers are used for both Facebook and Instagram (Meta, 2023, pp. 70-73), which is again something that stands out as something potentially unlikely. We recommend that the EC sends a written request for clarification from Meta.**

Microsoft

LinkedIn has taken special care to counter low authority information in relation to the COVID-19 crisis and the Russian Invasion of Ukraine, as detailed below and further in the Crisis Reporting appendices. **Bing Search** offers a number of tools to help users understand the context and trustworthiness of search results. Even in circumstances where a user is expressly seeking low authority content (or if there is a data void so little to no high authority content exists for a query), Bing Search provides tools to users that can help improve their digital literacy and avoid harms resulting from engaging with misleading or inaccurate content. The tools and methods listed under SLI 17.1.1 are extensive and include: **NewsGuard Impressions (NGI)**, **Knowledge Cards ("KC'')** – Represents viewership of Knowledge Cards (of all types/topics) during the Reporting Period, **Transparency Hub Viewership ("TH")** – Represents the total views of the Microsoft, Transparency Report Hub during the Reporting Period, **Public Service Announcement ("PSA")** – Represents views of public service, announcement panels (of all types/topics) rendered in Bing to EU users during the Reporting Period (Microsoft, 2023, pp. 74-82). **NewsGuard Extension Downloads ("NGED")** – Reflects total downloads of the NewsGuard extension by Edge users in the EU. In terms of the SLI, Bulgaria is referenced in relation to Bing's COVID-19 Information Hub for the Bulgarian market: Under total impressions for **NGI** were 305 for Bulgaria, under **KC** were 5.22M, under **TH** were 8, under **PSA** were 200 and under **NGED** 19 (Microsoft, 2023, p. 83).

TikTok

TikTok notes that in conjunction with removing content which violates policies, they have dedicated significant resources to increasing the number of in-app measures which show users additional context on certain content, or direct them to authoritative information. They also claim they have made these tools available in 21 EU official languages (plus, for EEA users, Norwegian and, as the spoken language of Liechtenstein, German). In 2020, TikTok deployed a combination of a number of in-app intervention tools (including video notice tags, search interventions, public service announcement, online and in-app information hubs and safety centre pages) around Covid-19, Covid-19 Vaccine, Holocaust Denial, Monkeypox and the War in Ukraine). Also, TikTok applies labels, irrespective of the topic, to encourage users to consider the reliability of the content or the source, such as: Unverified content label. State-controlled media label – TikTok has taken steps to restrict access to content from Russia Today, Sputnik, Rossiya RTR / RTR Planeta, Rossiya 24 / Russia 24 and TV Centre International. TikTok piloted the state-controlled media label policy by applying it to media entities in RU, UA and BY. The policy was then expanded to media entities across 40 countries in January 2023. Users across all EEA countries can view the label when they come across the content or profile pages of labelled entities. TikTok claims to have continued its extensive in-app interventions (including video tags, search interventions and in-app information hubs) around Covid-19, Covid-19 Vaccine, Holocaust Denial, Monkeypox and the War in Ukraine. TikTok also claims to be expanding the application of state-controlled media labels to additional countries, along with working with fact-checking partners to identify specific disinformation trends in countries and develop tailored, localised media literacy campaigns to tackle those trends. In addition, notifying creators when their content has been made ineligible for recommendation and enabling them to appeal - is being undertaken in the context of TikTok's obligations under the DSA (Articles 17 and 20), with an implementation date during the summer of 2023. TikTok is also rolling out three media literacy campaigns in Europe in partnership with trusted organisations, starting with a campaign to address disinformation related to the war on Ukraine in certain Eastern European countries - ultimately aiming to improve the digital literacy of users. Where possible and where appropriate, they will aim to scale these campaigns across Europe later in the year (TikTok, 2023, pp. 52-55).

In terms of the number of impressions of Video Notice tags, with relation to COVID-19, 10 million impressions are noted for Bulgaria, along with 8,621 for the number of clicks of video tags covered by the COVID-19 intervention the click-through rate was 0.09% for Bulgaria. Likewise, under the COVID-19 intervention, the number of impressions of Video Notice Tag for Bulgaria was 1,384,591, the number of clicks of Video Notice Tag were 94 and the click through rate was 0.01%. The number of impressions of Video Notice Tag coverage by the intervention on Holocaust denial was 625,320 in Bulgaria, the number of clicks of Video Notice Tag on this intervention was 1,498 and the Click Through rate was 0.24%. For Monkeypox, the number of impressions for Bulgaria was 255, the number of clicks of search interventions was 0, and the click-through rate of search interventions was 0%. The number of impressions of Public service announcements on COVID-19 was 5,111,700, and the number of impressions of the same was 599,191, the number of impressions for public service announcements on Holocaust denial was 562, on Monkeypox 54. The number of impressions of the Safety centre page on COVID-19 was 12,581 and on the safety centre page on Election integrity was 999 (TikTok, 2023, pp. 55-70).

Twitter

The response here does not comply with the requested information with respect to member states or with respect to tools Twitter develops or maintains related to media literacy. A global pre-existing partnership with UNESCO is cited and linked to (Twitter, 2023, p.29). There is a flagship piece of work mentioned called 'Teaching & Learning with Twitter'. Response seems to misunderstand media literacy.

***QRE 17.2.1 -** Relevant Signatories will describe the activities they launch or support and the Member States they target and reach. Relevant signatories will further report on actions taken to promote the campaigns to their user base per Member States targeted.*

***SLI 17.2.1 -** Relevant Signatories report on number of media literacy and awareness raising activities organised and or participated in and will share quantitative information pertinent to show the effects of the campaigns they build or support at the Member State level (for instance: list of Member States where those activities took place; reach of campaigns; engagement these activities have generated; number of interactions with online assets; number of participants).*

Google

For **Google Search**, they offer a detailed response which captures efforts to work with information literacy experts to help design tools which allow users to have confidence and control with respect to the information they consume. There is a list of partnerships provided in response to QRE 28.3.1 which includes the European Media & Information Fund (EMIF)(Google, 2023, pp. 93). Additionally, **Google Search** is seeking to build capacity with Librarians to empower their patrons and the general public with information literacy. These measures and partnerships are listed. It is also noted that in 2022, **Google Search** partnered with YouGov and Poynter on a report that summarised findings from a survey of over 8,500 respondents from 7 countries around the world, looking at consumer habits and practices related to misinformation, search literacy and information journeys. There is no response to the SLI for **Google Search** as the Super Searchers program was launched in September 2022. **YouTube** reaffirms its commitment to efforts that deepen users' collective understanding of misinformation. YouTube invests in media literacy campaigns, the most recent as of this report was launched in 2022 and as of December 2022, this campaign was live in more than 50 countries. Including 20 EU member states. In 2023 the campaign, known as 'Hit Pause', is supposed to launch in the remaining EU member States. **To this point there is no data for Bulgaria on this campaign** (Google, 2023, pp. 93-94).

Meta

It is noted that Meta is partnered with Poynter on a Media Literacy program, designed to help seniors detect misinformation, which is currently live in France and Spain. The figures are very general, but it appears the campaign is more active in France, where the response claims has reached 11 million people, while in Spain it has reached 440,000 people (Meta, 2023, pp. 74-75). **It would be very interesting and helpful to know and understand the data and measurements used here.**

Microsoft

This response notes that Microsoft has worked with three organisations to develop and promote media literacy campaigns, once of which was created by Verify, which is a collaboration between Purpose and the United Nations, targeted at the EU market. Microsoft claims it will continue to work with existing and new partners in order to create, disseminate, and report on expanded media literacy campaigns in EU markets and Languages. **Nothing in the SLI for any country but Spain and that data is also not provided. The same response is applied for both LinkedIn and Bing** (Microsoft, 2023, pp. 84-85).

TikTok

On this commitment, several campaigns are described in detail, with examples from elections in Italy, France and Denmark which are relevant. In addition TikTok highlights its elections integrity page, further related to the War in Ukraine, TikTok discusses its rollout of an in-app digital literacy campaign. With respect to COVID-19 the company implemented a comprehensive anti-disinformation campaign in conjunction with its partners that is ongoing. However, for this measure, **no SLIs are provided and they are stated as "N/A" without an explanation as to why it should not apply** (TikTok, 2023, pp. 76-77).

Twitter

No attempted response. Assuming response to overall measure is applicable. In which case, the response here does not comply with the requested information (Twitter, 2023, p. 30).

QRE 17.3.1 - *Relevant Signatories will describe how they involved and partnered with media literacy experts for the purposes of all Measures in this Commitment.*

Google

Refers to the previous QRE for Google Search, i.e. Google Search is seeking to build capacity with Librarians to empower their patrons and the general public with information literacy. These measures and partnerships are listed. It is also noted that in 2022, **Google Search** partnered with YouGov and Poynter on a report that summarised findings from a survey of over 8,500 respondents from 7 countries around the world, looking at consumer habits and practices related to misinformation, search literacy and information journeys (Google, 2023, pp. 93-94). There is a lack of information on partnerships for measure 17.1 related to tools. **YouTube** responds by elaborating on its partnerships with media literacy experts across markets. An example given is the 'Hit Pause' campaign, another is the National Association for Media Literacy Education (NAMLE), a US based organisation. YouTube states it will continue to evolve its media literacy program and add more markets (Google, 2023, p. 96). **Once again there is a lack of information on partnerships in terms of tools here**.

Meta

Here Meta points out that it is working in partnership with experts, educators, civic society, and governments globally. It is noted that Facebook has developed media literacy campaigns in the CEE region. In Romania, they are partnered with one of the members of the Bulgarian-Romanian Observatory on Digital Media (BROD) consortium, Funky Citizens (Meta, 2023, p. 75). **There is no Bulgarian partner in this list, however we are aware that an initiative covering Bulgaria was announced in June 2023**.

Microsoft

Microsoft here mentions it has worked with three organisations to develop and promote media literacy campaigns and reiterate the one created by Verify, a collaboration between Purpose and the United Nations, targeting the EU Market (Microsoft, 2023, p. 85).

TikTok

TikTok here offers a response which sets out the contents of its Safety Partners page. The QRE goes on to discuss the DigitalMente (Italy) campaign carried out in partnership with Unione Nazionale Consumatori (Italian Consumer Association). In terms of promoting election integrity, TikTok states that they partner with various media organisations and fact checkers in the context of election campaigns and they provide details of partnership activities in the Italian and Danish elections, along with the French parliamentary and presidential elections. This included work with the national ministry of education on a digital literacy program. BROD partner Agence France Presse is mentioned as a fact-checking partner in relation to election related content (TikTok, 2023, p. 78). **The response only sets out partnerships established in response to 17.2, no mention of 17.1.**

Twitter

No attempted response. Assuming response to overall measure is applicable. In which case,The response here does not comply with the requested information (Twitter, 2023, p. 30).

Table 8 - *Ratings for Measure 17*

Commitment 17								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 17.1	3	2	2	1	2	3	3	1
Measure 17.2	3	3	2	2	1	1	2	1
Measure 17.3	2	2	2	1	2	2	2	1

Commitment 18

QRE 18.1.1 *– Relevant Signatories will report on the risk mitigation systems, tools, procedures, or features deployed under Measure 18.1 and report on their deployment in each EU Member State.*

SLI 18.1.1 - *Relevant Signatories will provide, through meaningful metrics capable of catering for the performance of their products, policies, processes (including recommender systems), or other systemic approaches as relevant to Measure 18.1 an estimation of the effectiveness of such measures, such as the reduction of the prevalence, views, or impressions of Disinformation and/or the increase in visibility of authoritative information. Insofar as possible, Relevant Signatories will highlight the causal effects of those measures.*

Google

Google Search is not subscribed to this measure. YouTube is, and cites its policies of removing content that violates YouTube policies as quickly as possible, surfacing high-quality information in ranking and recommendations, and rewarding trusted, eligible creators and artists (Google, 2023, pp. 98-100).

Meta

The response provides an overview and links to the Content Distribution Guidelines, Community Guidelines, and Meta Technologies. The answer also explains the process of the algorithm, specifically referring to misinformation. **Clarity is needed here if this still applies to disinformation or if there is another process in place**. QRE 18.1.3 is identical to Instagram and discusses fact-checking labels and their success (Meta, 2023, pp. 76-81). **The answer does not tackle what Meta might be doing at a systemic level**.

Microsoft

In this response, LinkedIn referenced additional QRE responses in order to demonstrate compliance. For AI, LinkedIn refers to QRE 18.2.3, additional tools, procedures and features relevant are also mentioned. The measure is also not relevant for Bing as per Microsoft. **No SLI data and instead Microsoft state they will report on this in the next period** (Microsoft, 2023, pp. 86-90).

TikTok

TikTok's response is around safety design, but there is no explicit focus on targeting disinformation outside of collaboration with fact-checkers and experts, user reporting of content or accounts, a review of trending videos to ensure compliance with TikTok's CGs. TikTok claims that safety is by design built into the TikTok platform and all its features. This is said to be a continuing practice and includes specialised prompts or label share warnings on unverified content (TikTok, 2023, pp. 79-81). While Section 19.1.1 of the document outlines the parameters of the algorithm, **a more in-depth explanation should have been provided here** as Section 19.1 is intended for making the parameters transparent, not to provide details about those parameters themselves. It was also noted that the Share Cancel Rate, which is the percentage of users who do not share a video after seeing the label pop up, was 21.05% in Bulgaria (TikTok, 2023, p. 82).

Twitter

The response here does not comply with the requested information. It only discusses 'Community Notes' as their sole tool. There is no discussion of recommender systems or any processes in place to address this measure (Twitter, 2023, pp.30-33).

***QRE 18.2.1 -** Relevant Signatories will report on the policies or terms of service that are relevant to Measure 18.2 and on their approach towards persistent violations of these policies.*

***SLI 18.2.1** - Relevant Signatories will report on actions taken in response to violations of policies relevant to Measure 18.2, at the Member State level. The metrics shall include - Total number of violations; Meaningful metrics to measure the impact of these actions.*

Google

YouTube's response to this is to refer to QRE 14.1.1. YouTube also offers a methodological approach and the number of videos removed from EU member states. Bulgaria had more than 95 videos removed. **Google Search** lays out its policies which complement this measure, including Media Content Policy, Misleading Content Policy. Google Search also states that it removed content that has been determined to be unlawful under applicable law, in response to a notification from a third party (Google, 2023, pp. 101-102).

Meta

Meta responds here by discussing how their policies and approach to tackling misinformation have been provided previously. It is noted that these include specific actions taken against actors that repeatedly violate these policies. There is a need for clarity on this point as to how we define a repeat violator? Meta states that the company takes action against pages, groups, accounts and domains repeatedly shared or published content that is rated *false* or *altered*, near-identical to what fact-checkers have debunked as *false* or *altered*. They also have policies which enable them to act against accounts that spread misinformation on COVID-19 and vaccine information. The SLI provides a breakdown of Member States. **All scores are generalised. It would be helpful to see this data in more detail.**

Microsoft

LinkedIn again offers a link to its Professional Community policies, offers a definition of misinformation as: "as "specific claims, presented as fact, that are demonstrably false or substantially misleading." This is a definition which applies globally and not just in the EU and is used for the purposes of content moderation and for publicly reporting figures on misinformation. Examples are provided in the Help Centre. The process of content moderation is discussed in detail for LinkedIn. It is also noted that LinkedIn has automated defences to identify and prevent abuse, inauthentic behaviour, which includes misinformation. These measures include regularly rolling our scalable technologies like machine learning models to keep our platform safe. For Bing, the response highlights it's How Bing delivers search results and Microsoft Bing Webmaster Guidelines which provide an overview on how Bing Search's algorithms endeavour to deliver high authority and highly relevant content while minimising negative impacts of spam and sources of low credibility. Bing's AS/RS functionalities work by suggesting queries to users as they type to facilitate a more efficient search experience. In terms of the SLI for LinkedIn, there were 19 pieces of content removed as Misinformation from 1-31 December 2022. None of these removals were appealed (Microsoft, 2023, pp. 91-93).

TikTok

TikTok offers numbers of videos removed because of violation of harmful misinformation policy. In Bulgaria 38 videos were removed for this reason and it is claimed that the number of views of videos removed due to violation of the harmful misinformation policy was 760,281 (TikTok, 2023, p. 85-86).

Twitter

The response here does not comply with the requested information. No responses provided for specific QRE.

***QRE 18.3.1 -** Relevant Signatories will describe research efforts, both in-house and in partnership with third-party organisations, on the spread of harmful Disinformation online and relevant safe design practices, as well as actions or changes as a result of this research. Relevant Signatories will include where possible information on financial investments in said research. Wherever possible, they will make their findings available to the general public.*

Google

The response here described both Google and YouTube as working with industry leaders across various sections to set good policies, remain aware of emerging challenges and to collaboratively learn from best practices and research (Google, 2023, pp. 104-105). Jigsaw, a unit within Google, is leading research exploring threats to open societies and building technology that inspires scalable solutions. Jigsaw has reportedly contributed to research and technology designed to build resilience to disinformation notable efforts are Accuracy Prompts and Prebunking Messages. Additional information on these can be found here.

Meta

Here Meta provides a list of the key initiatives they have supported to empower the independent research community and to gain a better understanding of what our users want, need and expect. There is social science research, Data for Good, Research Platform for CIB Network disruptions, Research and Grant awards, and Research on Misinformation and Polarisation. The answer applies to both Facebook and Instagram (Meta, 2023, pp. 83-84).

Microsoft

Reports that LinkedIn has acquired Miburo in July 2022, which is the internal research team, referred to the Digital Threat Analysis Centre (DTAC). This team conducts research on information influence operations and publishes both internal and public reports on its findings. It is additionally noted that Microsoft funds and works with external organisations, the Global Disinformation Index, and Newsguard and the alliance for securing Democracy to ingest data and research they conduct into Microsoft products, including Bing and LinkedIn. Bing is said to regularly review and consider safe design practices and research and conduct user studies as part of its product and new feature development processes. Bing is also partnered with Microsoft research and third-party research organisations to contribute to novel research concerning safe design practices and disinformation. They give examples of pioneering research into 'data voids' (Microsoft, 2023, pp. 94-95). **For both responses no outcomes or actions mentioned outside these.**

TikTok

TikTok responds by highlighting partnerships with third party experts and researchers in relation to the creation of warning and labelling systems designed to reduce the spread of disinformation. Research is described in brief and the response also refers to other QREs (17.1.1 & 21.3.1) (TikTok, 2023, pp. 87-88). Actions described for implementation. 17.1.1 describes actions, but not research. 21.3.1 describes one of the research partnerships and findings.

Twitter

The response here does not comply with the requested information. No responses provided for specific QRE.

Table 9 - *Ratings for Measure 18*

Commitment 18								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 18.1	N/A	2	1	1	2	N/A	2	1
Measure 18.2	2	2	2	2	3	1	2	1
Measure 18.3	2	2	2	2	2	1	3	1

Measure 19

QRE 19.1.1 - *Relevant Signatories will provide details of the policies and measures put in place to implement the above-mentioned measures accessible to EU users, especially by publishing information outlining the main parameters their recommender systems employ in this regard. This information should also be included in the Transparency Centre.*

Google

Refers to Google Search's ranking system, how it sorts through hundreds of billions of web pages and other content in the Search index to present the most relevant and useful results in a fraction of a second (Google, 2023, p. 106). The main parameters that help determine which results are returned for a user's query are included. These are the Meaning of your query, Relevance of content, Quality of content, Usability and context and settings. The Search settings, SafeSearch, how results are automatically generated and How Search Works webpages are cited and referenced here (Google, 2023, pp. 106-108). **This response does not directly answer the question**. It is worth noting that only one link goes to a usable site for users to understand the recommender - this content should have been specified in the answer. The response to YouTube refers to QRE 18.1.2 on YouTube's recommendation systems. However **this response is incomplete and says nothing, for example, on transparency.**

Meta

Refers to 18.1.3 about the previously discussed parameters of the recommender system (Meta, 2023, p. 85), but this Measure is asking how they make this transparent (so 18.1.3 is not relevant). The relevant information is included in the Transparency Centre link but **this should have been included in the text.**

Microsoft

Notes that LinkedIn has published a variety of articles to explain to users how its recommender systems work. During the reporting period LinkedIn collated and expanded on existing resources to further explain the main parameters of LinkedIn recommender systems and options for users to influence and control these. The main parameters of Bing's search ranking algorithms are published in the How Bing Ranks your Content section of the Bing Webmaster Guidelines. The answer (Microsoft, 2023, p. 96) refers to QREs 18.3.1, 14.1.1 and 22.2.1. **The comprehensiveness of the response is somewhat questionable as there are just links provided**. **In addition, Blog Posts as a suitable medium for this important information might be of questionable effectiveness.**

TikTok

Information about the recommender system given in response here does not relevant, says "make clear to users in ToS and CGs" (TikTok, 2023, pp. 89-90) but when you click on links you don't get a straight answer - it is not making transparent to users the same answers given by the company in their data sheet so there is a discrepancy. **Strictly speaking, TikTok is not meeting the first part of the commitment**; it does meet the second part about making available information about the options which users are provided about recommender systems.

Twitter

The response here does not comply with the requested information. The response does not give details on implementation. There is a promise of open-sourcing the recommendation algorithm over coming months. No responses provided for specific QRE's or SLI's (Twitter, 2023, pp. 34-35).

SLI 19.2.1 *- Relevant Signatories will provide aggregated information on effective user settings, such as the number of times users have actively engaged with these settings within the reporting period or over a sample representative timeframe, and clearly denote shifts in configuration patterns.*

Google

Here Google Search provides the number of impressions on the personal results control of logged in users in Q3'2022 broken down by EEA Member State Bulgaria scored 6.875 under number of impressions. YouTube provide the percentage of daily active users that are signed in to the platform. That is 70% for Bulgaria (Google, 2023, p. 109).

Meta

Meta is unable to deliver this SLI in the time provided for the baseline reported. They say they are working to improve their SLIs across chapters in the January-June 2023 report (Meta, 2023, p. 88). The response offers data on the number of content removed for violating Meta's harmful health misinformation or voter, or census or interference policies. Bulgaria scored "less than 500" for both Facebook and Instagram (Meta, 2023, pp. 81-82).

Microsoft

The report states the number of LinkedIn users from EU member states who used the feed "sort by" functionality within 1-31 December 2022. For Bulgaria the number was 1,037 and the number of times that members used the functionality in Bulgaria was 5,354. **For Bing, the response does not have specific data for Bulgaria in terms of the AS/RS Settings** (Microsoft, 2023, p. 98).

TikTok

Here TikTok have provided data stating that 8,902 users have filtered hashtags and engaged with the settings laid out in SLI 19.1.1 in Bulgaria. Additionally the number of users in Bulgaria who clicked on the "Not Interested" message was 591,195 (TikTok, 2023, p. 91).

Twitter

The response here does not comply with the requested information. No response provided (Twitter, 2023, p. 35).

Table 10 - *Ratings for Measure 19*

Commitment 19								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 19.1	1	1	1	1	2	2	2	1
Measure 19.2	3	2	1	1	3	3	3	1

Commitment 20*: Signatories commit to empower users with tools to assess the provenance and edit history, authenticity, or accuracy of digital content.*

Google

Not subscribed to this commitment (Google, 2023, pp. 111-112).

TikTok

Not subscribed to this commitment as claims it would be imprudent to commit to this measure at a time when the underlying technology remains unproven and the standards to be complied with are not yet finalised (TikTok, 2023, p. 93).

Meta

Meta claims it is not subscribed to this measure as assessing provenance and edit history of digital content are one of several ways to make more informed decisions about the content they see online. They also claim other tools to achieve this objective, such as those set out in Measure 21 are what they call relevant and pertinent to their subscribed products at this time (Meta, 2023, p. 88).

Microsoft

No response to this measure, claimed as not applicable (Microsoft, 2023, pp. 100-101).

Twitter

The response to this measure notes that Twitter has recently introduced the capacity to edit tweets for those subscribed to Twitter Blue. It is noted that when a tweet is edited, an annotation appears on the content to show you when it was last edited. It can be found here. Response does not satisfy the measure (Twitter, 2023, pp.35-36).

QRE 20.2.1 *- Relevant Signatories will provide details of global initiatives and standards bodies focused on the development of provenance tools (for instance, C2PA) that signatories have joined, or the support given to relevant organisations, providing links to organisation websites where possible.*

Microsoft

Reports on a partnership Microsoft is a founding member of, the Coalition for Content Provenance and Authenticity (C2PA). The coalition aims to address the prevalence of disinformation, misinformation and online content fraud through developing technical standards for certifying the source and history of provenance of media content. Standards are still in development (Microsoft, 2023, pp. 102-103).

Measure 21

QRE 21.1.1 - *Relevant Signatories will report on the policies, features, or programs they deploy to meet this Measure and on their availability across Member States. When cooperating with independent fact-checkers to label content on their services, Relevant Signatories will report on: - Independent fact-checkers they work with to label content on their services (unless a fact-checking organisation opposes such disclosure on the basis of a reasonable fear of retribution or violence), the languages they operate in, the policies they work under, and any labelling applied - any tools or features available to inform users that content they interact with has been rated by an independent fact-checker.*

SLI 21.1.1 - *Relevant Signatories will report through meaningful metrics on actions taken under Measure 21.1, at the Member State level. Depending on the policies, features or programs in question, this could include reporting on actions taken under relevant policies; on reach of labels or fact-checks and other authoritative sources; or other similarly relevant metrics. At the minimum, the metrics will include: total impressions of fact-checks; ratio of impressions of fact checks to original impressions of the fact-checked content–or if these are not pertinent to the implementation of fact-checking on their services, other equally pertinent metrics and an explanation of why those are more adequate*.

Google

Fact checks on **Google Search** discussed in a very general way before discussing the labelling of such items. In addition Google also provides tools like Fact Check Explorer and the Google FactCheck Claim Search API. Google search also enables any fact-checker to signal their fact-checks so that they can be indexed for free and provides training to fact check organisations on how to use the ClaimReview mark-up. The ClaimReview mark-up is not restricted to any set of organisations that partner with Google search, so the remaining elements of QRE 21.1.1 are said not to apply (Google, 2023, pp. 114-116). **YouTube'**s fact-checking information panels are said to provide context by highlighting relevant, third-party fact-checked articles above search results for relevant queries. The response also outlines what factors for YouTube will determine whether or not a fact-check information panel will appear (intent to seek accuracy of claim). It is noted that these panels rely on a network of third-party publishers and leverage the ClaimReview tagging system. At the beginning of Q3'2022 there were 135 articles available related to Bulgaria in the Google Search Fact Check Explorer. By the end of Q3'2022 that number had grown to 173 (Google, 2023, pp. 114-116). **This small increase is indicative of the very limited fact-checking capacity which Bulgaria has - an issue that needs to be addressed urgently**.

Meta

Meta claims that it partners with over 26 independent third-party fact-checkers certified through the IFCN, covering 22 languages, in the EU. The response goes on to describe fact-checking partners, and their global impact. **Bulgaria has had two IFCN-certified fact checker but we cannot see any evidence that this fact checker was involved**. Additionally the response refers to 30.1.2 and what they do, which fulfil the QRE requirement. **SLI only reports a number of labels and only ranges of them - this answers SLI 21.1.2 but not SLE 21.1.1.** It is noted that the number of labels applied to content in Bulgaria was "Over 510,000" (Meta, 2023, pp. 90-92). **In order to verify this, researchers in Bulgaria need access to this key data**.

Microsoft

LinkedIn reiterates its prohibition on misinformation and disinformation on its platform as outlined in QRE 1.1.1 and QRE 17.1.1 and again points to its Professional Community Policies. They note that they do not label misinformation once identified but it is removed and this includes situations where LinkedIn personnel leverage the conclusions of fact checkers to determine whether the content at issue violates these policies. Additionally it is noted that Microsoft has partnered with NewsGuard and provides a free plug-in for the Microsoft Edge web browsers, as well as an opt-in news rating feature for the Edge mobile app. Bing Search provides a fact check feature which offers credible ways to assess the reliability of content displayed in its search results by providing fact-check flags and warnings on certain results and by directing users to fact check articles. Additionally ClaimReview is mentioned as a tagging system. The SLI states that there were no impressions of fact checks for Bulgaria and no reach of labels/fact-checkers and other authoritative sources (Microsoft, 2023, pp. 105-108).

TikTok

TikTok claims to have 8 IFCN accredited fact checking organisations providing coverage in Europe, in 10 official European Languages. It is not stated if Bulgaria's recognised IFCN fact checkers are among these. QRE 30.1.2 sets out the specific organisations. In terms of unverified content labelling. These fact-checkers support certain of the in-app tools they have designed in order to bring users additional context on certain content or provide access to authoritative information. On the COVID-19 pandemic, TikTok partnered with a number of EU based fact-checkers to prevent the spread of harmful misinformation related to the pandemic including AFP, Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit and DPA. With respect to election integrity, TikTok have launched campaigns in advance of several major elections as we discussed above. Some of these included the set-up of an in-app Elections Centre, the campaign was launched with a blog post which explained the content labelling and fact-checking process. TikTok has also published blog posts in over 25 languages and created a hub on their Safety Centre to raise user awareness of the program, labels and work of fact-checking partners (TikTok, 2023, pp. 94-95). The document also provides figures on the share cancel rate after the unverified content label share warning pop-up, for Bulgaria this figure stood at 21.05%. The share of removals under the harmful misinformation policy was 0.03% for Bulgaria. The Share of proactive removals under misinformation policy was 0.01%, the share of removals before any views under misinformation policy was 0.00% and the share of removals within 24 hours by misinformation policy was 0.01% (TikTok, 2023, p. 96).

Twitter

The response here is a direct copy of the response in QRE 18.2, which is focused on Community Notes, that are unavailable to EU member states with the exception of Ireland. There is no information provided on how these community notes distinguish if a tweet is misleading or how people are rating them. The response is more appropriate here than when previously provided, but is a poor response overall (Twitter, 2023, p.38-39).

QRE 21.2.1 *- Relevant Signatories will report on the research or testing efforts that they supported and undertook as part of this commitment and on the findings of research or testing undertaken as 23 part of this commitment. Wherever possible, they will make their findings available to the general public.*

QRE 21.3.1 *- Relevant Signatories will report on their procedures for developing and deploying labelling or warning systems and how they take scientific evidence and their users' needs into account to maximise usefulness.*

Google

Google Search discusses the content advisory notices that help to alert users when they have encountered a query and results set that might not yet include high quality information from reliable sources (Google, 2023, p. 118). Google search also highlights two content advisories and its consultations with independent experts on the effectiveness and possible risks of the content advisory feature ahead of launch. Final paragraph needs more detail about what scientific experts are being utilised. YouTube works with authoritative information providers around the world to create information panels that provide additional context about the content they are searching for and watching on the platform (Google, 2023, pp. 118-119). **Some good examples here are used but could be more specific**.

Microsoft

Responds that LinkedIn has not undertaken and/or supported separate research and testing on the potential efficacy of warnings or updates targeted to users that have interacted with content that was later actioned upon violation of Professional Community Policies. LinkedIn also reiterates that it removes content that violates their policies. QRE 21.2.1 is not applicable to Bing search. QRE 21.3.1 sees Bing note that it regularly consults research and evidence including from internal Microsoft research and data science teams related to safe design practices, labelling, and user experience. Such research is considered part of product design and testing. Bing's participation in the W3C organisation that helped to design and promote schema.org and ClaimReview, and regularly meets with stakeholders to discuss common issues, including necessary updates (Microsoft, 2023, p. 109).

Meta

For Facebook it is noting that the fact-checking program's rating as well as its labels were developed in close consultation with fact-checkers and misinformation experts. As Meta comments, Facebook continues to engage with fact-checkers and content moderation experts, including for instance by consulting the Oversight Board on approach to COVID-19 misinformation, and Instagram's fact-checking program is developed in close consultation with fact-checkers and misinformation experts. Meta insists that also works with independent experts who possess knowledge and expertise to determine what constitutes misinformation (Meta, 2023, p. 94). **For both products the answer seems rather generic, it would be helpful to have more detail about what scientific evidence / scientific experts they are consulting with.**

TikTok

Refers to response QRE 17.1.1 around unverified content and state-controlled media labels. The labels are developed and deployed in accordance with scientific evidence by partnering with fact-checkers and working with external experts. This work included unverified content label, a partnership with behavioural scientists on the roll-out of specialised prompts, the state-controlled media labels. TikTok also claim they are taking user feedback into consideration in order to identify new topics and consider which tools might be suited to raising awareness of that topic and combating misinformation (TikTok, 2023, pp. 100-101). The other examples outside of Irrational Labs are not described as being scientific evidence based. So **this aspect is unclear.**

Twitter

The response here is rated based on an inference from the above answer. It appears that research and testing has been then, which is enough to satisfy the QRE in principle (Twitter, 2023, pp. 38-39).

Table 11 - *Ratings for Measure 21*

Commitment 21								
	Google		Meta		Microsoft		TikTok	Twitter
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 21.1	2	1	2	2	2	3	3	1
Measure 21.2	2	N/A	1	1	1	N/A	1	2
Measure 21.3	1	2	1	1	1	2	2	2

Measure 22

QRE 22.1.1 *- Relevant Signatories will report on how they enable users of their services to benefit from such indicators or trust marks.*

***SLI 22.1.1** - Relevant Signatories will report on Member State level percentage of users that have enabled the trustworthiness indicator.*

Google

Not subscribed to this commitment (Google, 2023, p. 121).

Meta

Meta claims that they use several of the products and features listed under Measure 22.7 (in particular information panels, banners, pop-ups, and prompts) as already outlined under Commitment 21, as well as in their crisis monitoring reports on both Covid-19 and Ukraine (Meta, 2023, pp. 94-95).

Microsoft

Offers a description of its partnership with NewsGuard to provide a free plug-in for Microsoft Edge web browser for providing reliability. The measure is said to be not applicable to Bing Search. The SLI here states that the number of users in Bulgaria who used the "About this profile" feature were 2,235. The number of times those members used the feature in the period of the report was 2,905 (Microsoft, 2023, pp. 111-113).

TikTok

Not subscribed to this commitment (TikTok, 2023, p. 102).

Twitter

The response here describes the use of Community Notes as the trust/quality indicator. There is no other response, data or information provided under the measure (Twitter, 2023, p.44).

***QRE 22.2.1** - Relevant Signatories will report on whether and, if relevant, how they feed signals related to the trustworthiness of media sources into their recommender systems, and outline the rationale for their approach.*

Microsoft

LinkedIn does not prioritise any new sources in our feed, but in crisis situations, (e.g., Covid-19 or Ukraine), they would use search banners to point members to reputable sources of information (e.g., when members searched for COVID, they pointed members to "trusted storylines'' where they provided trustworthy information about those topics, including links to global health organisations) (please also see QRE 17.1.1). While **Bing Search** is not a recommender system, the section "How Bing Ranks Your Content" in the Microsoft Bing Webmaster Guidelines details the parameters Bing uses in its ranking algorithms and provides an overview of how Bing works to ensure that its ranking algorithms can determine the trustworthiness of a given website and rank that the website accordingly (Microsoft, 2023, p. 114). Answer to the QRE for LinkedIn is basically no, it doesn't feed it into the recommender systems. The response around Bing is complete.

***QRE 22.3.1** - Relevant Signatories will provide details of the policies and measures put in place to implement the above-mentioned measures accessible to EU users, especially by publishing information outlining the main parameters their recommender systems employ in this regard. This information should also be included in the Transparency Centre.*

Microsoft

In addition to the LinkedIn User Agreement, LinkedIn has established and published (a) the LinkedIn Professional Community Policies to set out and elaborate on LinkedIn's requirements and expectations for its member base; and (b) help center content that collates and expands upon existing resources to further explain the main parameters of LinkedIn recommender systems and options provided to users to influence and control these recommender systems (Microsoft, 2023, pp. 115-116). A full response which provides locations of policies and measures. For Bing the QRE is not responsive to the specific question and only references Bing Webmaster Guidelines. Measures 22.4-22.6 are said not to be relevant to Microsoft.

QRE 22.7.1 *- Relevant Signatories will outline the products and features they deploy across their services and will specify whether those are available across Member States.*

SLI 22.7.1 *- Relevant Signatories will report on the reach and/or user interactions with the products or features, at the Member State level, via the metrics of impressions and interactions (clicks, click-through rates (as relevant to the tools and services in question) and shares (as relevant to the tools and services in question).*

Google

In QRE 22.7 SOS alerts are mentioned as what Google Search deploys in the case of a crisis. They respond that there have been special features created to provide information about COVID-19. YouTube highlights information from authoritative third-party sources using information panels. These panels include COVID-19 Information panels and Crisis resource panels (Google, 2023, pp. 122-123). **Two tools are described but more information is needed** about "special features providing information" - it doesn't say how these appear to people to lead them to good information. **SLIs provided as an overview, not by member state.**

Microsoft

LinkedIn has taken special care to counter low authority information in relation to the COVID-19 crisis and the Russian Invasion of Ukraine, as detailed below and further in the Crisis Reporting appendices. In addition to broader measures, Bing Search has taken special care to counter low authority information and misinformation in relation to the COVID-19 crisis and the Russian Invasion of Ukraine, as detailed below and further in the Crisis Reporting appendices (Microsoft, 2023, pp. 116-118). **In terms of the SLI there is no data provided for LinkedIn** and is promised in the next reporting period. For Bing, the number of visits to the COVID-19 Hub in Bulgaria was 2,328, the number of users of this hub in Bulgaria is listed as 1,934 (Microsoft, 2023, pp. 118-119).

Twitter

The response here refers back to the previous response, the response seems reasonable as an answer for Twitter's approach to leading people to information. However, there are no SLIs or additional information provided (Twitter, 2023, p.45).

Table 12 - *Ratings for Measure 22*

Commitment 22								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 22.1	N/A	N/A	N/A	N/A	3	3	N/A	

Measure 22.2	N/A	N/A	N/A	N/A	1	1	N/A	
Measure 22.3			N/A	N/A	3	1		
Measure 22.4								
Measure 22.5								
Measure 22.6								
Measure 22.7	2	2			1	2	2	2

Measure 23

***QRE 23.1.1** - Relevant Signatories will report on the availability of flagging systems for their policies related to harmful false and/or misleading information across EU Member States and specify the different steps that are required to trigger the systems.*

***QRE 23.2.1** - Relevant Signatories will report on the general measures they take to ensure the integrity of their reporting and appeals systems, while steering clear of disclosing information that would help would-be abusers find and exploit vulnerabilities in their defences.*

Google

Google Search aims to make the process of submitting removal requests as easy as possible, and has built reporting tools, which allow users in all EU Member States to report potentially violative content for review under search Content Policies. The Report Content on Google tool offer assistance for users to the right reporting form (Google, 2023, pp. 125-129). Google Search has reporting tools for search features, such as knowledge panels and featured snippets. The response provides links to tool, the process for reporting search results is more complete though it generally doesn't seem as user friendly. YouTube discusses how community members have the opportunity to report or flag content they believe violate YouTube's Community Guidelines. YouTube's flagging feature is also discussed, along with a Trusted Flagger program (Google, 2023, pp. 125-129). **Not all content is relevant in this response**.

Meta

Meta claims that they try to provide the full content reporting process, where the community standards are referenced. Meta defines three pillars of enforcement practices: artificial intelligence, human review, and user reports, where users are also able to report content that they specifically identified as false information through a process outlined on their website. **Very similar answer has been given for both Facebook and Instagram, which merits an EC-led request for details and clarification** (Meta, 2023, pp. 96-98).

Microsoft

LinkedIn provides a complete answer discussing Professional Community Policies which encourages users to flag and report content they believe violates these. Bing does not have a reporting function for user generated content as it is a search engine. However it does have a Report a Concern Form which permits users to report third-party websites for a variety of reasons including disclosure of private information, spam and malicious pages and illegal materials. Bing's feedback tool is also mentioned. For QRE 23.2.1, the response references the previous QRE and reaffirms LinkedIn's processes for flagging and removing disinformation (Microsoft, 2023, pp. 122-124). Also **more detail needs to be provided about how the QA team makes decisions to safeguard the moderation process**. Bing claims they do not experience issues with mass flagging of content or abuse of reporting features.

TikTok

Respond to say they provide users with simple intuitive ways to report/flag content in app for any breach of terms of service or CGs including for harmful misinformation in each EU member state and in an official language of the

EU. The answer is comprehensive and they discuss which violations get automated responses. The Appeals systems and the Reporting Systems are also discussed in depth (TikTok, 2023, pp. 105-107).

Twitter

Once again the response from Twitter copies the Community Notes content from earlier and does not address the specifics of the measure. These responses create the impression that no steps are taken with respect to appropriate, proportionate follow-up actions (Twitter, 2023, pp. 46-48).

Table 13 - *Ratings for Measure 23*

Commitment 23								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*Linked In*	*Bing Search*		
Measure 23.1	2	2	3	3	3	2	3	1
Measure 23.2	1	1	2	2	2	1	3	
Measure 23.3								1

Measure 24

QRE 24.1.1 - *Relevant Signatories will report on the availability of their notification and appeals systems across Member States and languages and provide details on the steps of the appeals procedure.*

SLI 24.1.1 - *Relevant Signatories provide information on the number and nature of enforcement actions for policies described in response to Measure 18.2, the numbers of such actions that were subsequently appealed, the results of these appeals, information, and to the extent possible metrics, providing insight into the duration or effectiveness of processing of appeals process], and publish this information on the Transparency Centre.*

Google

YouTube discusses the process of when content is removed for violating Community Standards. There are other reasons outside Community Guidelines where content may be removed, for example, a first-party privacy complaint or a court order. There are also explanations of the appeals process for a strike, for a video removal and to appeal the age restriction of a video. The response gives details about various forms of appeals that can be made by users (Google, 2023, pp. 130-133). **Overview of SLIs only, no member-level data, and no specific numbers, only a range.** Average score for poor SLI and good QRE.

Meta

Meta reports that in line with its policies on Community Standard violations, the company lets the user know when a piece of content is removed. It can be found in both the user's feed and the support inbox. The notification refers to which part of the Community Standards the user did not follow. Meta shows to publishers when their content is fact-checked, and has an appeals process in place for publishers who wish to issue a correction or dispute a rating with the fact checker. **The SLI response is very broad and does not have details about the number of appeals**. (Category numbers only) In Bulgaria "less than 500" contents were removed from Facebook for violating the harmful health misinformation or voter or census interference policies. For Instagram, the number was “less than 100” (Meta, 2023, pp. 99-101).

Microsoft

When a post, comment, reply, or article, is reported and found to go against LinkedIn’s Professional Community Policies, they reaffirm that they take appropriate actions to remove it and/or restrict accounts depending on the severity of violation (Microsoft, 2023, pp. 126-127). **More detailed information is needed on how users access these reporting flows and appeals systems**. The SLI notes that the number of pieces of content removed as misinformation in Bulgaria during this period was 19, there were no appeals and no appeals granted (Microsoft, 2023, p. 128).

TikTok

There were no new implementation measures here. TikTok utilise their CGs and users are notified by an in-app notification in all EU member states and in an official language of the European Union when an account has been banned or their content has been removed for violating said CGs, as well as where the unverified label has been applied to their video. Appeals are raised, queued, reviewed manually by human moderators. Users can also share feedback when they do not agree with the finding or the result of the appeal. In Bulgaria, the number of accounts removed under TikTok’s I&A policies was 8. The number of appeals of videos removed for violation of harmful misinformation policy was 11. The number of successful appeals of this was 8 and the appeal success rate of videos removed for violation of the harmful misinformation policy was 72.73% (TikTok, 2023, pp. 107-109). **Given the low numbers reported here, further details need to be requested from TikTok, as these beg the question of how comprehensive are TikTok’s actions and their verifiability.**

Twitter

The response reads as if the commitment has not been met as it discusses future actions and enforcement. The link provided requires a twitter log in and there is also a link which outlines how Twitter approaches enforcement actions (Twitter, 2023, p.49).

Table 14 - *Ratings for Measure 24*

Commitment 24								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 24.1	N/A	2	2	2	3	N/A	3	1
Measure 24.2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

Measure 25

Google

Not subscribed (Google, 2023, pp. 124-125).

Microsoft

Claim this commitment is not relevant or applicable to LinkedIn or Bing (Microsoft, 2023, p. 129).

TikTok

Claim this commitment is not relevant or applicable as TikTok is not a messaging app (TikTok, 2023, pp. 110-111).

Twitter

Claim that this commitment is not relevant.

***QRE 25.1.1 -** Relevant Signatories will report on the tools, policies, partnerships, programs, and campaigns in place to meet this Measure and on their availability across Member States, including, where possible, relevant details on the civil entity and their results.*

***SLI 25.1.1** - When in compliance with local law, and subject to any necessary information being made available by third-parties, Relevant Signatories will to the extent possible report on use of select tools (e.g. number of claims submitted by users to fact-checkers or reach of fact checks produced from claims submitted on the platform)*

Meta

Meta refers here to the measures it has in place for content which has been identified as misinformation on Facebook and shared directly in Messenger. Meta reports the use of two tools: misinformation labels and sharing warnings, which are fact-checking labels (Meta, 2023, pp. 103-104). The response appears to demonstrate limited efforts to meet this Commitment. This refers to section 17 for SLI but does not provide the information needed, which is the number of claims submitted by users to fact-checkers on the platform - this particular data is missing, and the response covers both Messenger and WhatsApp. For WhatsApp it is noted that the app partners with organisations certified by the IFCN around the world (Meta, 2023, pp. 104-105). Thus, there are **missing SLI data, plus barely adequate QRE, and so the overall rating of this response is poor**. WhatsApp only lists Croatia, France, Germany, and Greece. Ireland, Italy, Portugal and Spain as fact checking organisations using WhatsApp products between June 2022 and January 2023. **Implementation in Bulgaria is urgently needed.**

***QRE 25.2.1** - Relevant Signatories will report on the tools and features available to limit the propagation of viral Disinformation on their services, and to empower users to think about the messages they receive.*

Meta

The response notes Meta's steps to limit the number of chats that a message can be forwarded to at one time in Messenger. It is noted they also have additional protections in place for content that has been identified as misinformation on Facebook and shared on Messenger. These protections are not elaborated. Likewise, for WhatsApp it is claimed the platform provides end-to-end encryption by default. They note that they include forwarding labels, Limits to messaging forwarding and search the web tool (Meta, 2023, pp. 106-107). Examples for WhatsApp are sufficiently relevant. **The SLI and QRE are inadequate responses to this request.**

Table 15 - *Ratings for Measure 25*

Commitment 25			
	Meta		Twitter
	Messenger	*Whatsapp*	
Measure 25.1	1	2	Not applicable
Measure 25.2	1	2	Not applicable

Measure 26

QRE 26.1.1 *- Relevant Signatories will describe the tools and processes in place to provide public access to non-personal data and anonymised, aggregated and manifestly-made public data pertinent to undertaking research on Disinformation, as well as the safeguards in place to address risks of abuse.*

QRE 26.1.2 *- Relevant Signatories will publish information related to data points available via Measure 25.1, as well as details regarding the technical protocols to be used to access these data points, in the relevant help centre. This information should also be reachable from the Transparency Centre. At minimum, this information will include definitions of the data points available, technical and methodological information about how they were created, and information about the representativeness of the data.*

SLI 26.1.1 *- Relevant Signatories will provide quantitative information on the uptake of the tools and processes described in Measure 26.1, such as number of users.*

Google

This response covers both Google Search and YouTube and notes that both provide publically available data via Google Trends, providing a largely unfiltered (what kind of filtering is done beside anonymization and categorisation?) sample of actual search requests made on both YouTube and Google Search. The Fact Check Explorer and Google FactCheck Claim Search API are also listed in response (Google, 2023, pp. 137-138). For QRE 26.1.2 Google Fact Check Explorer using the Claim Review mark-up is referenced and includes the following information: Claim Made By, Rating Text, Fact Check article, Claim reviewed and Tags. The number of Fact Check explorer tool users in Q3'2022 in Bulgaria was 96. Keeping in mind that this tool is free to use, this number is quite low. The number of Google Trends users from Google Search was more than 16,000. The number of Google Trends users from YouTube was more than 800 in Bulgaria (Google, 2023, pp. 137-139).

Meta

The QRE responds by discussing the publishing of integrity reports from the transparency centre on a quarterly basis. It is also noted that for both Facebook and Instagram there are extensive public reports about coordinated behaviour in the Quarterly Adversarial Threat Report. **The SLIs are not delivered** and it is noted that they will be improved across chapters in the next report from January-June 2023 (Meta, 2023, pp. 111-112).

Microsoft

LinkedIn reports that it is dedicated to supporting research and regularly provides information and data to the research community in a variety of ways. Of those named, LinkedIn highlights API Access for data related to various issues. To date they say they have not received requests for access to such cases. They promise to publish information as it continues to build further data research infrastructure pertinent to these commitments (Microsoft, 2023, pp. 133-135). **No data on Help Centre and SLI 26.1.1**.

TikTok

TikTok is in the process of developing an API designed to provide researchers with access to relevant data on harmful misinformation. There is no data in this Commitment. TikTok add that they are engaging with EDMO on this priority. In parallel, over the past months, they have been working on developing a global and separate transparency API that will provide selected researchers with access to various public and anonymized data from its platform (TikTok, 2023, p. 112).

Twitter

First, among VLOP and VLOSE platforms, Twitter developed a data access API-program in 2006 for the needs of the academic community. Twitter explains here that researchers can apply for various levels of API access. In addition, Twitter has made several disclosures regarding government-backed information operations. Researchers evaluate and analyze this data to determine the strategies and tactics of state actors on the platform (Twitter, 2023, p. 72). Against the background of this information, it is striking that no measures relating to this obligation were named or filled out. Information about access to real-time data, tools, number of data sets, processing of reporting are missing. No data provided for the SLI.

QRE 26.2.1 *- Relevant Signatories will describe the tools and processes in place to provide real time or near real-time access to non-personal data and anonymised, aggregated and manifestly made public data for research purposes as described in Measure 26.2.*

QRE 26.2.2 *- Relevant Signatories will describe the scope of manifestly-made public data as applicable to their services.*

SLI 26.2.1 - *Relevant Signatories will provide meaningful metrics on the uptake, swiftness, and acceptance level of the tools and processes in Measure 26.2, such as: - Number of monthly users (or users over a sample representative timeframe) - Number of applications received, rejected, and accepted (over a reporting period or a sample representative timeframe) - Average response time (over a reporting period or a sample representative timeframe)*

Google

Google Search is not subscribed to this measure, while YouTube provides information on the YouTube Researcher program (Google, 2023, pp. 140-142). **There seems to be very little use of this program from academics and researchers**.

Meta

In its response Meta highlights the CrowdTangle platform which provides access to a small subset of public data on Facebook. There are over 1000 academic accounts which have access as of January 2023. The rest of the response covers the core products of CrowdTangle, including the search function, Live Displays, Intelligence and Notifications (Meta, 2023, pp. 112-114). **The SLI is only partially responsive.**

Microsoft

LinkedIn reaffirms its support for the research community and states that it regularly provides information and data in a variety of ways, including non-personal, aggregated data and API access. The SLI response notes that LinkedIn will publish information as it continues to build further data research infrastructure pertinent to these commitments. Bing also cites its various means of supporting the research community by highlighting MS MARCO datasets, the Bing Search related ORCAS: Open Resource for Click Analysis in Search. Additionally, Bing notes that in 2020 it shared a search dataset for Coronavirus intent, its Keyword Tools and Backlinks and the use of BING APIs. Finally, it is noted that Microsoft Research maintains a public portal of codes, APIs, software development kits and data sets, along with Microsoft Research Open Data (Microsoft, 2023, pp. 135-136).

TikTok

TikTok is in the process of developing an API designed to provide researchers with access to relevant data on harmful misinformation. There is no data in this Commitment. TikTok add that they are engaging with EDMO on this priority. In parallel, over the past months, they have been working on developing a global and separate **transparency API that will provide selected researchers** with access to various public and anonymized data from its platform (TikTok, 2023, pp. 112-113). **However, EDMO and the EC need to work closely with TikTok to ensure a transparent and equitable process for vetting researchers and their provision of access to data.**

Twitter

Twitter highlights **t**he following links for the API program (Twitter, 2023, p. 51): **API program, Information Operations.** Twitter has also conducted its own research into issues such as political bias in algorithmic content recommendations. See **here.** Against the background of this information, it is striking that no measures relating to this obligation were named or filled out. Information about access to real-time data, tools, number of data sets, processing of reporting are missing. No data provided for the **SLI**.

QRE 26.2.3 - *Relevant Signatories will describe the application process in order to gain the access to non-personal data and anonymised, aggregated and manifestly-made public data described in Measure 26.2.*

Google

The response here concerns only YouTube and notes that the YouTube Researcher Program has a 3-step application process. 1. YouTube verifies the applicant is an academic researcher affiliated with an accredited, higher-learning institution. 2. The researcher creates an API project in the Google Cloud Console and enables the relevant YouTube APIs. See more here. 3. The researcher applies with their institutional email, includes lots of detail and confirms their information is accurate (Google 2023, p.141). **A major issue here is that over 80% of applications are rejected. Count of unique researchers that access the data API during the timeframe: under 15** (Google 2023, pp. 141-142). **This needs to urgently be investigated by the European Commission.**

Meta

Despite the focus of the request, Meta does not provide a description of the application process nor any qualifier on the current freeze on registering for CrowdTangle (Meta, 2023, p. 113). **Thus current and planned provisions for data access by Meta are highly insufficient and need to be improved significantly in a very short time, due to the forthcoming EU elections.**

Microsoft

Accessing non-public data through LinkedIn APIs requires the development of a developer application and complete additional requirements specific to particular APIs that are being sought for access. They would then agree with the terms and conditions of use of these APIs. An email is offered to contact as well as a more general message stating that LinkedIn is open to contact from researchers. No specific information on reporting procedures. For Bing the response states there is not an application process in place in order to access the MS MARCO, ORCAS, or Bing coronavirus query datasets. These are freely accessible. The same is said for Bing's Keyword Research Tool. Bing's APIs may be accessed by signing up for an account (Microsoft, 2023, pp. 135-136).

TikTok

TikTok claim they are in the process of building a dedicated API in order to provide researchers with access to relevant data on disinformation. They state they are engaging with EDMO on this. In parallel they are working on developing a global and separate transparency API for select researchers and anonymise data from their platform (TikTok, 2023, p. 113). **No data provided.**

Twitter

Twitter provide researchers' access to its data through the **API-programme** (Twitter, 2023, Executive Summary).

QRE 26.3.1 *- Relevant Signatories will describe the reporting procedures in place to comply with Measure 26.3 and provide information about their malfunction response procedure, as well as about malfunctions that would have prevented the use of the systems described above during the reporting period and how long it took to remediate them.*

Google

Highlights the options to report issues for Google Trends and the Trends Help Centre, Google Fact Check Explorer and YouTube Researcher program. No Malfunctions response system in place as Google is unaware of any pertinent malfunctions that could prevent access to these reporting systems (Google, 2023, p. 142).

Meta

It is noted that CrowdTangle users can receive direct support through submitting requests via help.crowdtangle.com and/or accessing their library of available resources. The answers are the same for both Facebook and Instagram (Meta, 2023, p. 114). **Limited answer with no data provided.**

Microsoft

Discusses LinkedIn's API support via StackOverflow and LinkedIn's Help Centre. Bing offers an email address to deal with issues related to MS MARCO and ORCAS dataset. There is a reporting issues request ticket system available for Bing's Keyword Research Tool. There is also another email for reporting issues related to Microsoft Open Data and a support ticket request process for Bing APIs (Microsoft, 2023, p. 138).

TikTok

TikTok claim they are in the process of building a dedicated API in order to provide researchers with access to relevant data on disinformation. They state they are engaging with EDMO on this. In parallel they are working on developing a global and separate transparency API for select researchers and anonymise data from their platform (TikTok, 2023, p. 113). **No Bulgaria-specific data provided**.

Twitter

Twitter responds by highlighted that it published its first transparency report in 2012. Since then, the Twitter Transparency Centre is said to have become more detailed with almost every subsequent publication, now offering country-level data on both legal requests and Terms of Service violations. (Twitter, 2023, Executive Summary)

Table 16 - *Ratings for Measure 26*

Commitment 26								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 26.1	2	2	1	1	1	1	1	N/A
Measure 26.2	N/A	1	1	1	1	1	1	
Measure 26.3	2	2	1	1	1	2	1	

Measure 27

QRE 27.1.1 - *Relevant Signatories will describe their engagement with the process outlined in Measure 27.1 with a detailed timeline of the process, the practical outcome and any impacts of this process when it comes to their partnerships, programs, or other forms of engagement with researchers.*

Google

Google Search and YouTube continue to engage constructively with other Signatories, the European Commission, EDMO, and civil society, as part of the Code of Practice's Permanent Task-force, in order to satisfy Commitment 27. As of the filing of this report, there is no agreed-upon timeline to report on (Google, 2023, p. 144).

Meta

Meta discusses its ongoing engagement with EDMO working group on platform to researcher data sharing to develop standardised processes for sharing data with researchers. The same answer applies for both Facebook and Instagram (Meta, 2023, pp. 115-116). No third party body has been established.

Microsoft

Not Applicable. Work has not yet started on the permanent task force (Microsoft, 2023, pp. 139-140).

TikTok

TikTok has been engaging with EDMO as part of this process and is committed to participating in the working group that is being set up in order to put in place the independent third-party body that is referred to above, including by nominating a TikTok representative. Same responses for QRE 27.2.1 and QRE 27.3.1 (TikTok, 2023, p. 114).

Twitter

Twitter claims its API is used widely among academic researchers. To date, academic researchers are one of the largest groups of people using the Twitter API (Twitter, 2023, p. 72).

QRE 27.4.1 *- Relevant Signatories will describe the pilot programs they are engaged in to share data with vetted researchers for the purpose of investigating Disinformation. This will include information about the nature of the programs, number of research teams engaged, and where possible, about research topics or findings.*

Google

Google Search is reportedly exploring options to engage in pilot programs towards sharing data with vetted researchers for the purpose of investigating mis-/disinformation. YouTube offers details on its program for academic researchers interested in using YouTube's global Data API (Google, 2023, p. 145).

Meta

Meta reports that the company has a Research Platform for CIB Network Disruptions and since 2018, that they have been sharing information with independent researchers about our network disruptions relating to coordinated inauthentic behaviour (CIB). The response is almost the same for both Facebook and Instagram. However, the Facebook response mentions the launch of an early access version of the research API (Meta, 2023, p. 117).

Microsoft

Mentions Microsoft's partnership with Princeton University and the Carnegie Endowment for International Peace to fund and provide data to the Institute for Research on the Information Environment (IRIE). The same response is given for both LinkedIn and Bing Search. However Bing also references responses in QRE's 26.1-2 and QRE 28.1.1 (Microsoft, 2023, pp. 140-141). **The answer does not appear to satisfy the question**. Limited information provided.

TikTok

TikTok claim they will be launching a pilot phase as part of their efforts to develop an API for researchers to engage with the research community and make sure that their tools, and the data provided through the API, suit researcher needs. In the meantime, TikTok have asked the members of their Content and Safety Advisory Councils with expertise in various subject matters, including misinformation, to test an early version of the global API for researchers that they are developing in parallel (See QRE 26.1.1). The aim is to gather their feedback on usability and the overall experience of accessing public data through this API (TikTok, 2023, p. 115).

Twitter

Twitter notes that they have a long tradition of cooperation with academic society because its API started in 2006. Academic researchers have used data from the public conversation to study topics as diverse as the conversation on Twitter itself - from state-backed efforts to disrupt the public conversation to floods and climate change, from attitudes and perceptions about COVID-19 to efforts to promote healthy conversation online. (Twitter, 2023, p. 72).

Table 17 - *Ratings for Measure 27*

Commitment 27								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 27.1	2	2	1	1	1	1	1	N/A
Measure 27.2	N/A	1	1	1	1	1	1	N/A
Measure 27.3	2	2	1	1	1	2	1	N/A

Measure 28

QRE 28.1.1*: Relevant Signatories will describe the resources and processes they deploy to facilitate research and engage with the research community, including e.g. dedicated teams, tools, help centres, programs, or events.*

Google

Reiterates Google's long standing commitment to transparency. States that Google's products, processes, and practices via the Lumen Database, Google Trends, and Fact Check Explorer demonstrate some of the ways Google provides tools to support not only researchers but journalists and others (Google, 2023, pp. 147-148). Google Search refers to QRE 26.1.1 and QRE 26.1.2 and links to Google Fact Check Tool APIs and Google Trends. They also note that Google's partnership with Lumen is an independent research project managed by the Berkman Klein Centre for Internet & Society at Harvard Law School. YouTube mentions the YouTube Researcher Program and provides links to support options such as Issue Tracker and YouTube API Code Samples at GitHub (Google, 2023, pp. 147-148).

Meta

Meta's response covers both Facebook and Instagram and notes that it has a team dedicated to providing academics and independent researchers with the tools and data they need to study Meta's impact on the world (Meta, 2023, p. 119). It also states that there are current models in existence to support independent external research which include onboarding support, training and education for researcher products and datasets, community meets-ups and office hours, and the promotion of research opportunities through newsletters and educational materials.

Microsoft

Responses to this QRE and SLI are missing quantitative data. For LinkedIn, the answer refers to QRE 26.1-2 above and notes that LinkedIn is regularly exploring potential partnerships and that they are working to explain access to data for research purposes consistent with the goals of the CoP and the applicable requirements of the Digital Services Act (DSA). A similar answer is provided for Bing. LinkedIn responds by saying it facilitates research, engages with the research community and provides data to the research community in a variety of ways laid out in QRE 26.1-2. which covers CrowdTangle (Microsoft, 2023, pp. 142-143).

TikTok

TikTok's Outreach & Partnerships Management Team engages regularly with, and sets up partnerships with, the academic and research community. This team, together with subject matter experts within TikTok's product team, are central to their fact-checking programme from identifying new partners and on boarding them to regularly meeting with them. Trust and Safety teams are said to regularly consult and engage with the research community, including on harmful misinformation and deceptive behaviours, when updating or launching new policies or features on the platform. TikTok also has many teams committing time to facilitating research. Individuals with backgrounds in product, data science, outreach and legal are working together to build an API to share information on harmful misinformation as well as a global transparency API (see QRE 26.1.1) (TikTok, 2023, pp. 115-116).

Twitter

Twitter highlights that the main resource for the research is its API. They have Information Operations, too. Twitter has also conducted its own research into issues such as political bias in algorithmic content recommendations. (Twitter, 2023, p. 54).

QRE 28.2.1 *- Relevant Signatories will describe what data types European researchers can currently access via their APIs or via dedicated teams, tools, help centres, programs, or events.*

Google

Refers to QRE 28.1.1 (Google, 2023, p. 148).

Meta

Meta outlines a variety of data sets for researchers and offers the opportunity to consult a chart to verify if the data would be available for request. The cited main data available only to researchers are: Ad Targeting Data Set, URL Shares Data Set, Researcher API. Meta underlines that 30+ researchers in Europe have access to the Researcher API Beta. 70+ researchers globally have access to Ads Targeting API (Meta, 2023, pp. 119-120).

Microsoft

Refers to QRE 26.1.1 and QRE 26.2.3 (Microsoft, 2023, p. 143).

TikTok

TikTok is currently working on developing APIs to allow researchers to access transparent public and anonymized data about content and activity on the platform. Same answer for QRE 28.3.1 and 28.4.1 (TikTok, 2023, p. 116).

Twitter

No response.

Table 18 - *Ratings for Measure 28*

Commitment 28								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 28.1	2	1	2	2	1	1	2	
Measure 28.2	2	2	2	2	N/A	N/A	1	
Measure 28.3	2	2	1	1	1	1	2	
Measure 28.4	2	2	1	1	1	1	1	

Measure 29

Google

Not Subscribed (Google, 2023, pp. 149-151).

Meta

Meta says not applicable, only applicable to research organisations (Meta, 2023, p. 121).

Microsoft

Microsoft states not applicable to them (Microsoft, 2023, p. 144).

TikTok

TikTok are not committed to this measure (TikTok, 2023, pp. 117-118).

Twitter

Twitter refers to response to measure 26 and also points to the transparency offered in the datasets and research related to Community Notes under commitment 18 (Twitter, 2023, p. 55).

Commitment 30

QRE 30.1.1 - *Relevant Signatories will report on and explain the nature of their agreements with fact-checking organisations; their expected results; relevant quantitative information (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); and such as relevant common standards and conditions for these agreements*.

SLI 30.1.1 *- Relevant Signatories will report on Member States and languages covered by agreements with the fact-checking organisations, including the total number of agreements with fact-checking organisations, per language and, where relevant, per service.*

Google

Globally, Google and YouTube work with publishers and journalists to support quality journalism and global media literacy. Google's digital tools, training and resources are helping newsrooms to find, verify and tell stories. Google News Initiative has provided training, including digital verification techniques, to over 100,000 European journalists since 2015, and Google's free online curriculum has been visited over 400,000 times. As mentioned in response to QRE 21.1.1, Google Search and YouTube enable any fact-checkers to mark up their content for the purpose of indexation in Google's and others' services for free using the publicly available schema.org ClaimReview mark-up (Google, 2023, p. 153).

Fact-checkers must also be either a verified signatory of the International Fact-Checking Network's Code of Principles or an authoritative publisher to be eligible on YouTube. Accordingly, Google and YouTube agreements and partnerships with fact-checking organisations differ from those of services that would rely upon proprietary tools or closed partnerships. In 2021, Google contributed €25M EUR to help launch the **European Media and Information Fund** (EMIF) 'to strengthen media literacy skills, fight misinformation and support fact checking' over 5 years (2021-26). The EMIF was established by the European University Institute and the Calouste Gulbenkian Foundation(Google, 2023, p. 153).

The **European Digital Media Observatory** (EDMO) agreed to play a scientific advisory role in the evaluation and selection of projects that will receive the fund's support, but does not receive Google funding. Google has no role in the assessment of applications. To date, at least 33 projects have been granted €5.6M EUR with the list of selected grantees from this fund available here (Google, 2023, p. 153).

Additionally, on 29 November 2022, Google and YouTube announced they will work with the International Fact-Checking Network (IFCN), to provide $13.2M USD over 2.5 years to 135+ organisations via in-direct payments (Google, 2023, p. 152). Within the funding provided, $1.2M USD will be used by IFCN to operate the fund, manage the application process and outreach. The goal is to reach fact-checking organisations of differing maturity: Build: fact-checkers with little or no online presence; Grow: fact-checkers with a basic digital presence looking to expand reach; Engage: digitally mature fact-checkers, looking to invest in new technologies. The International Fact-Checking Network already includes Signatory organisations present in the following EEA Member States: Austria, Belgium, Bulgaria, Croatia, Estonia, France, Germany, Greece, Italy, Latvia, Lithuania, Norway, Poland, Portugal, Slovenia, Spain, and Sweden (Google, 2023, p. 154).

Meta

It is noted that Meta's fact-checking partners all go through a rigorous certification process with the IFCN. It is noted that the IFCN is dedicated to bringing fact checkers together worldwide. All of the fact-checkers associated with Meta must follow the IFCN's Code of Principles in order to promote excellence in fact checking (Meta, 2023, p. 123).

Microsoft

Notes that LinkedIn has entered into long term and pilot fact checking arrangements with external, independent global news agencies. Bing notes its support of the schema.org ClaimReview fact check protocol. Bing does not maintain formal agreements with any individual fact-checking organisation but continues to evaluate additional fact checking organisations and tools for use by search engines (Microsoft, 2023, p. 147).

TikTok

In terms of cooperation with the EU fact-checking community, TikTok note that they have secured new fact-checking partnerships with Sweden, Hungary, Poland and Romania; plans for Portugal, Denmark, Greece and Belgium (TikTok, 2023, pp. 119-120). No Eastern Europe. **No mention of Bulgaria.**

Twitter

Twitter claims that this measure is not applicable to its platform.

QRE 30.1.2 - *Relevant Signatories will list the fact-checking organisations they have agreements with (unless a fact-checking organisation opposes such disclosure on the basis of a reasonable fear of retribution or violence).*

Google

Response covers both Google Search and YouTube and notes Google's work with publishers and journalists to support quality journalism and global media literacy (Google, 2023, p. 154). Links are offered to resources for story verification. Fact-checking and indexation via the free schema.org ClaimReview mark-up. Fact-Checkers must be a verified signatory of the International Fact-Checking Network's Code of Principles or an authoritative publisher to be eligible on YouTube. Contributions to the European Media and Information Fund (EMIF) to strengthen media literacy skills are mentioned and the announcement that Google and YouTube will work with the IFCN to fund various organisations to reach fact-checking organisations of differing maturity. There are signatory organisations in Bulgaria, the Association of European Journalists-Bulgaria and its website factcheck.bg, and the Bulgarian section of AFP *proveri*.

Meta

It is noted in QRE 30.1.2 that AFP are the fact-checkers for Bulgarian under both Facebook and Instagram platforms (Meta, 2023, p. 124).

Microsoft

Reuters is mentioned along with an unnamed pilot arrangement. Bing notes its support of the schema.org ClaimReview fact check protocol. Bing does not maintain formal agreements with any individual fact-checking organisation (Microsoft, 2023, p. 147).

TikTok

Within Europe, these are IFCN-accredited fact-checking partners of TikTok: 1. Agence France Press; 2. Facta.news; 3. Lead Stories; 4. Logically; 5. Newtral; 6. Science Feedback; 7. dpa Deutsche Presse-Agentur; and 8. Teyit. TikTok have put in place temporary agreements with fact-checking partners to provide additional European language coverage for a period in an unfolding crisis. Hungary is given as an example (TikTok, 2023, p. 120). **No data about evaluation and criteria for distribution of funding.**

TikTok declares they have a standardised service agreement with fact-checkers. The contracts include internal articles for anti-bribery and corruption provisions. According to TikTok, fact-checkers are compensated in a fair, transparent way based on the work done by them. Fact-checkers are independent organisations certified by IFCN and have editorial independence in the fact-checking process (TikTok, 2023, p. 122).

Twitter

Twitter claims that this measure is not applicable to its platform.

***QRE 30.1.3 -** Relevant Signatories will report on resources allocated where relevant in each of their services to achieve fact-checking coverage in each Member State and to support fact-checking organisations' work to combat Disinformation online at the Member State level.*

Google

Reiterates that Google's main partnerships are with the EMIF and the IFCN. Both organisations provide in-direct payments to fact-checking members (Google, 2023, p. 154).

Meta

No data has been provided on financial contributions. It is noted that along with the remuneration of fact-checking partners, Meta also underlines the contributions to programs such as industry initiatives, sponsorships, fellowships, and grant programs. Meta highlights examples such as providing their Ukrainian fact-checking partners emergency funding to help protect their team's safety. The Climate Misinformation Grant program and partnership with France

24 and AFP to share media literacy resources to help identify reliable information are also pointed to (Meta, 2023, pp. 124-125).

Microsoft

LinkedIn has also implemented internal processes empowering hundreds of global internal content reviewers to be able to obtain a fact check from external fact-checker partnerships. Conclusions of fact checkers are reviewed by internal content reviews to determine whether or not the content violates policies. Bing again cites the ClaimReview fact check protocol (Microsoft, 2023, pp. 147-148).

TikTok

TikTok have fact-checking coverage in 10 official European languages (Dutch, English, French, German, Hungarian, Italian, Polish, Romanian, Spanish and Swedish), and, therefore, the spoken language of 15 EEA markets. TikTok can request (and have previously requested) temporary coverage in relation to a number of European languages with a current partner e.g. Hungarian or languages which affect European users, including Azeri, Armenian, Turkish, Russian, Ukrainian and Belarusian (TikTok, 2023, pp. 121-122). They can also put a temporary arrangement in place with a new partner. **Again no mention of Bulgaria or Bulgarian.**

Twitter

Twitter claims that this measure is not applicable to its platform.

Table 19 - *Ratings for Measure 30*

Commitment 30								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 30.1	2	2	2	2	1	1	1	N/A
Measure 30.2	2	2	1	1	1	1	1	
Measure 30.3	1	1	2	2	1	1	1	
Measure 30.4	1	1	1	1	1	1	1	

Measure 31

QRE 31.1.1 - *Relevant Signatories will report on their specific activities and initiatives related to Measures 31.1 and 31.2, including the full results and methodology applied in testing solutions to that end.*

SLI 31.1.1 *- Member State level reporting on use of fact-checks by service and the swift and efficient mechanisms in place to increase their impact, which may include (as depends on the service): number of fact-check articles published; reach of fact-check articles; number of content pieces reviewed by fact-checkers.*

Google

Redirects to response QRE 21.1.1. It is noted that YouTube will explore opportunities to provide more granular information for future reports for the SLI. Google Search's SLI response is to see SLI 21.1.1 (Google, 2023, p. 159).

Meta

The response discusses the process for actioning content rated by fact-checkers as outlined in QRE 21.1.1**.** These actions are to label it, to ensure less people see it, and to sanction repeat offenders. The SLI notes that the number of fact-checked labelled content for Bulgaria was “over 510,000” and for Instagram it the number was “over 18,000” in Bulgaria (Meta, 2023, pp. 129-130). There is no data by market for the % of reshares attempted that were not completed on treated content on Facebook or Instagram. As the work in the Permanent Taskforce on the development of the repository of fact-checking content has not yet started at the time of submission of this report, **no technical solutions as referred to under Measure 31.4 can currently be reported on**.

Microsoft

Response notes that LinkedIn leverages its fact checkers to review user generated content that may violate the Professional Community Policies, which prohibit disinformation. Such content is removed. For Bing Search the response mentions the ClaimReview tags embedded in websites with fact-checked content to help inform its algorithms and to provide useful context and indications of trustworthiness to its users. **The SLI reports that 0 Fact Check Impressions (FCI) are and 0 for Fact Check URLs** (FC URL) (Microsoft, 2023, pp. 151-152).

TikTok

The data provides the **number of fact checked videos per Member State.** TikTok works with 8 certified by IFCN in 10 languages in Europe. **Bulgaria has none among them (TikTok, 2023, p. 125).**

However, response *SLI 31.1.2* - notes that **38 videos were removed in Bulgaria** because of policy guidelines, known misinformation trends and a knowledge-based repository (TikTok, 2023, p. 127). Compared to Austria and Ireland, countries with approximately a similar population, where the number of videos removed is respectively 17 and 28. Romania is three times more in population than Bulgaria, and TikTok has an agreement with a certified fact-checker; the removed videos were 478. (TikTok, 2023, pp. 125-126) QRE 31.3.1 – notes that TikTok are regularly engaged with EDMO on this priority and are committed to participate in the taskforce (TikTok, 2023, p. 130).

Twitter

Twitter claims that this measure is not applicable to its platform.

Table 20 - *Ratings for Measure 31*

Commitment 31								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 31.1	N/A	N/A	2	2	1	1	1	
Measure 31.2	N/A	N/A	2	2	1	1	1	
Measure 31.3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	
Measure 31.4	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

Measure 32

QRE 32.1.1 (for Measures 31.1 and 31.2*) - Relevant Signatories will provide details on the interfaces and other tools put in place to provide fact-checkers with the information referred to in Measure 31.1 and 31.2.*

***SLI 32.1.1 (for Measures 31.1 and 31.2)** - Relevant Signatories will provide quantitative information on the use of the interfaces and other tools put in place to provide fact-checkers with the information referred to in Measures 32.1 and 32.2 (such as monthly users for instance).*

Google

Reports on Search Console, which is a free service offered by Google that includes various tools and reports to help webmasters, including fact-checking organisations, to monitor, maintain, and troubleshoot their sites presence in Google search. The Search Performance report is also linked which shows metrics about how a site performs in Google Search results. Information can be found here. YouTube highlights the YouTube Help Centre which provides details on fact checks on YouTube. YouTube also states it will provide more granular information in future reports (Google, 2023, pp. 162-163).

Meta

Reports on the access which all Meta's fact-checking partners have access to a dashboard that they built in 2016. The dashboard is said to include a variety of content from across Facebook, including links, videos, images and text-only posts. The platform also provides data points to help fact-checkers prioritise what content to review (Meta, 2023, pp. 135-136). **No data for SLI provided.**

Microsoft

Refers to the Professional Community Policies and the prohibition on misinformation, the removal process and consultation with the fact-checker process. This seems at least like a partial adherence. It doesn't really address 31.2. Bing's response focused on the ClaimReview tags embedded in fact-check content posted on websites that are indexed in the Bing Search Index. The Bing Webmaster Tools is a dashboard which is said to provide website operators with a range of data and analytics which can be used by fact checking organisations (Microsoft, 2023, p. 155).

TikTok

Reports that their fact-checking partner's access content which has been flagged for review through an exclusively available dashboard. The dashboard shows the fact-checkers certain quantitative information about the services they provide, including the number of videos queued for assessment at any one time, as well as the time the review has taken. Data is also shared at regular meetings to help them quantify the impact of the fact-checked content over time (TikTok, 2023, p. 131). **No data provided for the SLI.**

Twitter

Twitter responded by referencing previous answers on access to API-programme.

***QRE 32.3.1** - Relevant Signatories will report on the channels of communications and the exchanges conducted to strengthen their cooperation - including success of and satisfaction with the information, interface, and other tools referred to in Measures 32.1 and 32.2 - and any conclusions drawn from such exchanges.*

Google

The answer for both YouTube and Google Search notes that Google is in regular discussions with those such as the International Fact Checking Network (IFCN) to discuss collaborations and efforts to build and support the work of fact-checkers. Additionally, Google and YouTube have worked with the IFCN to provide $13.2M USD over 2.5 years to 135+ organisations via in-direct payments (Google, 2023, p. 164). Going forward Google and YouTube plan to engage in regular discussion on similar and other topics with the European Fact Checking Standards Network (EFCSN) (Google, 2023, p. 164).

Meta

Both Facebook and Instagram provide the same response and note that Meta has a team in charge of their relationships with fact-checking partners, working to understand their feedback and improve fact-checking. Examples given include the new labels “missing context” and “altered” (Meta, 2023, p. 136).

Microsoft

Linked in states it is exploring ways in which it can further support information exchange with its fact-checking partners. Bing Search states that it welcomes continued cooperation with signatories and fact-checking organisations (Microsoft, 2023, p. 156).

TikTok

Reports that TikTok is committed to participate in the taskforce made up of the relevant signatories’ representatives that is being set up for this purpose. They are engaging with EDMO pro-actively on this commitment (TikTok, 2023, p. 131).

Twitter

Twitter responded by referencing previous answers on access to API-programme.

Table 21 - *Ratings for Measure 32*

Commitment 32								
	Google		**Meta**		**Microsoft**		**TikTok**	**Twitter**
	Google Search	*YouTube*	*Facebook*	*Instagram*	*LinkedIn*	*Bing Search*		
Measure 32.1	2	1	1	1	1	1	1	N/A
Measure 32.2	1	N/A	1	1	1	N/A	1	
Measure 32.3	2	2	2	2	1	1	1	

2.3.1. Empowering Users/Researchers/Fact-Checkers Conclusion

On **Measure 17** Microsoft registered low scores of compliance to Measure 17.2, again this was due to missing SLIs along with the campaigns being vague and not fully described. Wording is identical to LinkedIn, so not platform specific. Instagram had a low rating in their response to QRE 17.1 failing to provide the required metric on the tool's impressions and instead gave other types of counts. Combined Facebook and Instagram data which does not seem appropriate if each platform is separately signed up to this Commitment. For QRE 17.3, Instagram was rated low due an identical response from both Facebook and Instagram, with no Instagram specific work. No details provided on how the partnerships address QRE 17.1 and QRE 17.2. There is no response to the SLI for **Google Search** as the Super Searchers program was launched in September 2022. Meta was rated low for both Facebook and Instagram responses to measure 18.1 by offering Links to guidelines and technologies and explaining the process of the algorithm, **the answer seems to be targeted at misinformation and does not answer the question about disinformation**.

QRE 18.1.3 answer identical for Instagram. Talks about fact-checking labels and their success but the Commitment is more focused on systemic level actions. For AI, LinkedIn refers to QRE 18.2.3, additional tools, procedures and features relevant are also mentioned. The measure is also not relevant for Bing as per Microsoft. No SLI data and instead Microsoft states they will report on this in the next period. For 18.3, Microsoft's response was also rated low, LinkedIn referred to their acquisition of Miburo in July 2022, which is the internal research team, referred to the Digital Threat Analysis Centre (DTAC). This team conducts research on information influence operations and publishes both internal and public reports on its findings. Bing is said to regularly review and consider safe design practices and research and conduct user studies as part of its product and new feature development processes.

Google and Meta scored low on QRE 19.1 both because they did not adequately respond to the measure. For Google various resources and webpages were cited and referenced yet these responses do not directly respond to the measure on the issue of recommender systems, only one link goes to a usable site for users to understand the recommender, this content should have been the focus of the answer. YouTube's response was similarly incomplete, not making any mention of transparency. For Meta their response to QRE 19.1 refers to 18.1.3 in reference to the previously discussed parameters of the recommender system, however this measure is specifically asking for both the details and policies in place, and transparency. There is a link to the relevant information included in the transparency centre but this should have been included in the text. Both responses for Microsoft related to 18.3 do not describe adequate information, actions or outcomes to meet the criteria of the measure. For measure 19.2 Meta have not provided any data for the SLI and promise to provide the data in the January-June 2023.

Google's YouTube rated low on Claims that 21.1.2 does not apply and has provided no information for 21.1.1. This is confusing and leads one to ask why not as actions taken should be measurable if they have implemented those actions so there is a need for further elaboration here. For measure 21.3 Meta's response again had a low rating as for both Facebook and Instagram the answer seems rather generic, it would be helpful to have more detail about what scientific evidence / scientific experts they are consulting with. For Measure 21, Meta responded that the measure is not applicable, Google only responded to 22.7 and TikTok only responded to that same measure. Microsoft's response to QRE 22.2.1 related to LinkedIn was a bit too elaborate and could have been clearer. In reference to SLI 22.7.1 there is no data provided for LinkedIn and is promised in the next reporting period. For measure 23, the responses of Google to QRE 23.2 were rated low as they reported content is addressed within legal compliance; not about prevention of abuse of systems and this is the focus of 23.2.1. The descriptions provided are quite general, they could be more specific about what limitations they employ, how they are "handling" affected webmasters and "adjusting automation". For Microsoft, Bing scored a low result for 23.3 again due to a lack of detail around how the QA team makes decisions to safeguard the moderation process. For measure 24, there are several aspects of the commitments which are stated to be non-applicable.

For Measure 25 only Meta provided QRE responses to 25.1 and 25.2, messenger scored low ratings here. For Measures 26 and 27, Meta, Microsoft and TikTok gave responses which were rated low due to lack of details referring to the specifics of the measure. For responses to QRE 28.3 and 28.4 both Meta and Microsoft were rated with low scores. For Measure 30 every platform score low on QRE 30.4, For Google this was because they did not outline or describe cooperation with EDMO & European Fact-Checking Standards Network (EFCSN)'s governance body and adhesion procedure launched in November 2022 and December 2022. Facebook and Instagram gave identical answers which note that they are working with regional hubs and look forward to collaborating with EDMO on fact checks. Microsoft responded for both Bing and LinkedIn that they are ready to cooperate on this QRE at the appropriate time. While TikTok provided no data, just noting that they are in continuous dialogue with EDMO and EFCSN. For Commitment 31, Google considers the measure to not be applicable to them, Meta, Microsoft and TikTok responded to both QRE 31.1 and QRE 31.2. Both Microsoft and TikTok were rated low in terms of their response as TikTok provided data but not a report on the mechanisms in place to increase their impact. For Microsoft, while some detail is provided, the answer only refers to claimReview tags and prohibited disinformation removal. For Commitment 32, once again Microsoft and TikTok showed the lowest ratings with 'poor' for each response. In line with this rating, these responses were once again lacking major details such as methodology or are incomplete, irrelevant or fail to address the specific information requests outlined in the measure.

On Measure 17.2 Google's responses could be viewed as vague on the issue of making authoritative sources readily available without specifying how. Some information given which is not appropriate or complete, a lot of non-responses and partial responses. This is the reason why this pillar demonstrates the lowest rated per response. In Measure 17, Missing SLIs and vague or unclear working were an issue. Throughout this pillar the member state level details were missing frequently, often with the promise that this would be included and covered in the next reporting period. Generally, there was confusion around Measure 19 and the issue of recommender systems, again with issues of incomplete information provided. Google's YouTube Claims that 21.1.2 does not apply and that they have provided no information for 21.1.1 is confusing and leads one to ask why not, as actions taken should be measurable if they have implemented those actions. **There is a need for further clear and unambiguous elaboration here and concerns around the verifiability of some of the information provided.**

For 17.1, Google provided data for member states and Bulgaria, the number of times the "More About This Page" feature was viewed in Q3'2022 was 60,548, the number of times the "About This Result" panel was viewed was 415,804. The number of times Content Advisories for low-relevance results were viewed in Bulgaria was 757,400. Finally, the estimated number of times Content Advisories for low quality and rapidly changing results were viewed was 5,640 (Google, 2023, p. 91). For the same measure 17.1, TikTok reported there were 10 million impressions of Video Notice tags in Bulgaria, with relation to COVID-19, along with 8,621 for the number of clicks of video tags covered by the COVID-19 intervention. The click-through rate was 0.09% for Bulgaria. Likewise, under the COVID-19 intervention, the number of impressions of Video Notice Tag for Bulgaria was 1,384,591, the number of clicks of Video Notice Tag were 94 and the click through rate was 0.01%. The number of impressions of Video Notice Tag coverage by the intervention on Holocaust denial was 625,320 in Bulgaria, the number of clicks of Video Notice Tag on this intervention was 1,498 and the Click Through rate was 0.24%. For Monkeypox, the number of impressions for Bulgaria was 255, the number of clicks of search interventions was 0, and the click-through rate of search interventions was 0%. The number of impressions of Public service announcements on COVID-19 was 5,111,700, and the number of impressions of the same was 599,191, the number of impressions for public service announcements on Holocaust denial was 562, on Monkeypox 54. The number of impressions of the Safety centre page on COVID-19 was 12,581 and on the safety centre page on Election integrity was 999 (TikTok, 2023, pp. 55-70).

In the same measure, since the invasion of Ukraine, Meta responded that it has launched its educational media literacy campaigns to raise awareness of how to spot misinformation for users in Poland, Slovakia, Lithuania, Latvia, Estonia, Albania, Bosnia and Herzegovina, Kosovo, Serbia and Bulgaria. All of these campaigns were designed in partnership with our local fact-checking partners as well as expert safety NGOs. For Microsoft, Bulgaria is referenced in relation to Bing's COVID-19 Information Hub for the Bulgarian market: Under total impressions for NGI were 305 for Bulgaria, under KC were 5.22M, under TH were 8, under PSA were 200 and under NGED 19 (Microsoft, 2023, p. 83). On 17.2 for Google, in 2023 the 'Hit Pause' campaign is due to launch in the remaining EU member States. To this point there is no data for Bulgaria on this campaign (Google, 2023, pp. 93-94). For 17.3.1, Meta mentioned that in Romania, they are partnered with one of the members of the Bulgarian-Romanian Observatory on Digital Media (BROD) consortium, Funky Citizens (Meta, 2023, p. 75). There is no Bulgarian partner in this list, however we are aware that an initiative covering Bulgaria was announced in June 2023. For the same measure, TikTok mentions that BROD partner Agence France Presse is a fact-checking partner in relation to election related content (TikTok, 2023, p. 78).

For Measure 17, Twitter's response seems to misunderstand media literacy and what counts as media literacy. There were also some issues with incorrect url's provided. There are again no other details provided for the other QRE's related to measure 17. The response here does not comply with the requested information with respect to member states or with respect to tools Twitter develops or maintains related to media literacy. A global pre-existing partnership with UNESCO is cited and linked to (Twitter, 2023, p.29). There is a flagship piece of work mentioned called 'Teaching & Learning with Twitter'. This is a trend that has continued throughout Twitter's report across pillars with Measure 18, discussing Community Notes as if it was their sole tool. There is no discussion of recommender systems or processes in place which would address the measure directly. The responses of Twitter within the pillar can be summarised as generally non-responsive. No country specific data was provided and a lot of QRE's were left without response.

For Measure 18.1 TikTok noted that the Share Cancel Rate, which is the percentage of users who do not share a video after seeing the label warnings for unverified content pop up, was 21.05% in Bulgaria (TikTok, 2023, p. 82). For 18.2 google notes that Bulgaria had more than 95 videos removed for violations of policies and terms of services, while Meta, offering generalised numbers on the number of content removed for violating Meta's harmful health misinformation or voter, or census or interference policies. Bulgaria scored "less than 500" for both Facebook and Instagram (Meta, 2023, pp. 81-82). Also in response to 18.2 TikTok claims it removed 38 videos in Bulgaria due to violation of their harmful misinformation policy and that the number of views of videos removed due to violation of the harmful misinformation policy was 760,281. In response to SLI 19.2 Google, Meta, Microsoft and TikTok all provide data. Meta offers data on the number of content removed for violating Meta's harmful health misinformation or voter, or census or interference policies. Bulgaria scored “Less than 500” for both Facebook and Instagram (Meta, 2023, pp. 81-82). For Google, Bulgaria scored 6.875 under the number of impressions. YouTube provides the percentage of daily active users that are signed in to the platform. That is 70% for Bulgaria (Google, 2023, p. 109). For Microsoft, specifically LinkedIn, the number of those in Bulgaria using the 'sort by' functionality between 1-31 December was 1,037 and the number of times that members used the functionality in Bulgaria was 5,354. For TikTok, they respond that 8,902 users have filtered hashtags and engaged with the settings laid out in SLI 19.1.1 in Bulgaria. Additionally, the number of users in Bulgaria who clicked on the “Not Interested” message was 591,195 (TikTok, 2023, p. 91).

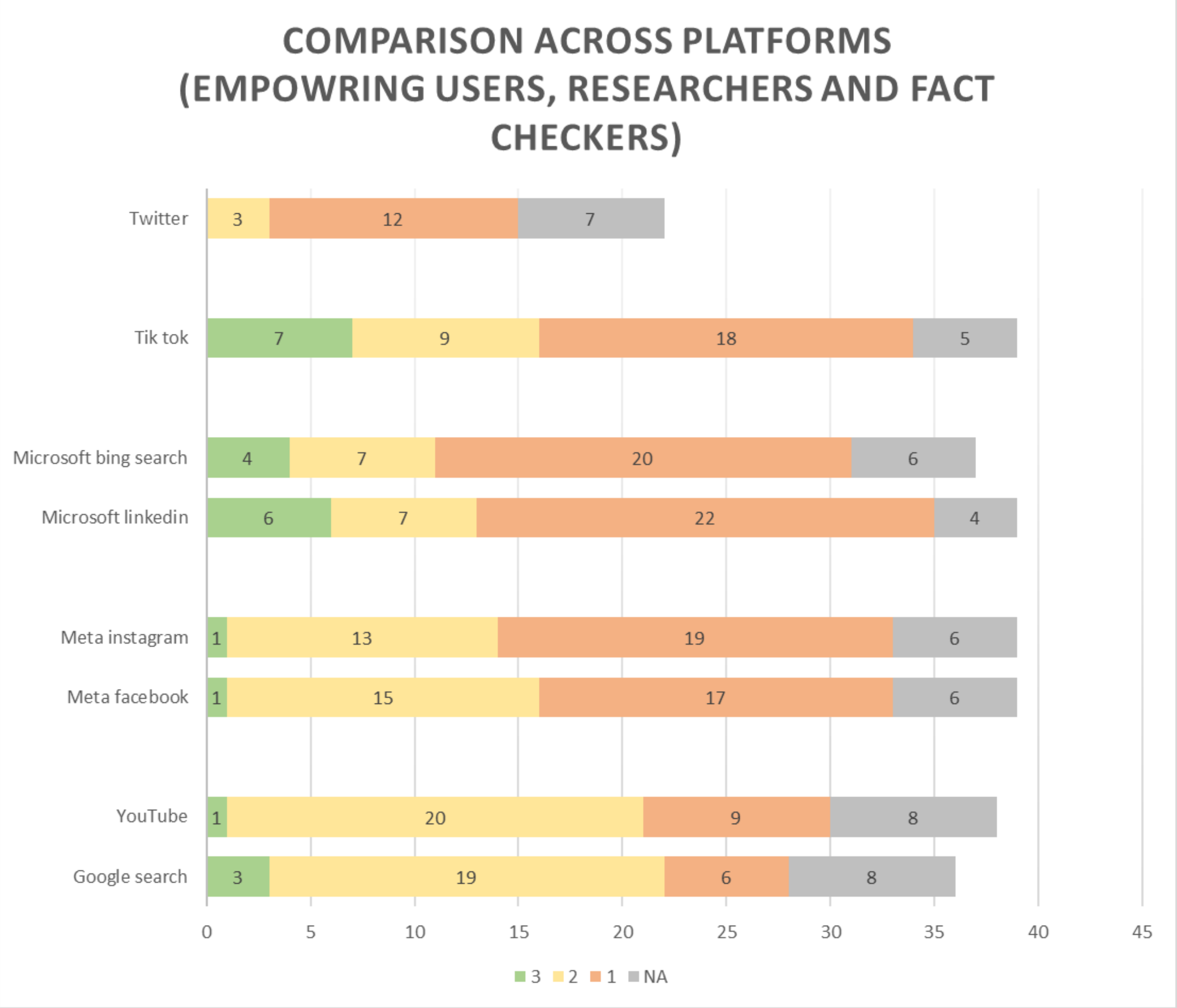
For Measure 21.1.1 Google notes that at the beginning of Q3'2022, there were 135 fact-checking articles related to Bulgaria in the Google Search Fact Check Explorer. For Meta, SLI only reports a number of labels and only ranges of them - this answers SLI 21.1.2 but not SLE 21.1.1. It is noted that the number of labels applied to content in Bulgaria was “Over 510,000” (Meta, 2023, pp. 90-92). For Microsoft, they claim in their SLI that there were no impressions of fact checks for Bulgaria and no reach of labels/fact-checkers and other authoritative sources (Microsoft, 2023, pp. 105-108). TikTok provides figures on the share cancel rate after the unverified content label share warning pop-up, for Bulgaria this figure stood at 21.05%. The share of removals under the harmful misinformation policy was 0.03% for Bulgaria. The Share of proactive removals under misinformation policy was 0.01%, the share of removals before any views under misinformation policy was 0.00% and the share of removals within 24 hours by misinformation policy was 0.01% (TikTok, 2023, p. 96). For Measure 22 was another where responses were limited, Meta claimed the measure was not applicable to them. Google and TikTok only responded to QRE 22.7. Microsoft responded comprehensively to 22.1, and for LinkedIn on 22.3 and adequately for 22.7. In terms of the SLI there is no data provided for LinkedIn and is promised in the next reporting period. For Bing, the number of visits to the COVID-19 Hub in Bulgaria was 2,328, the number of users of this hub in Bulgaria is listed as 1,934 (Microsoft, 2023, pp. 118-119).

For Measure 24.1, Meta reported that it does not have details about the number of appeals. (Category numbers only) In Bulgaria “less than 500” contents were removed from Facebook for violating the harmful health misinformation or voter or census interference policies. For Instagram, the number was “less than 100” (Meta, 2023, pp. 99-101). For the same measure Microsoft noted that the number of pieces of content removed as misinformation in Bulgaria during this period was 19, there were no appeals and no appeals granted (Microsoft, 2023, p. 128) and TikTok noted that in Bulgaria, the number of accounts removed under TikTok's I&A policies was 8. The number of appeals of videos removed for violation of harmful misinformation policy was 11. The number of successful appeals of this was 8 and the appeal success rate of videos removed for violation of the harmful misinformation policy was 72.73% (TikTok, 2023, pp. 107-109). For Measure 26.1.1, Google reported that the number of Fact Check explorer tool users in Q3'2022 in Bulgaria was 96. The number of Google Trends users from Google Search was more than 16,000. The number of Google Trends users from YouTube was more than 800 in Bulgaria (Google, 2023, pp. 137-139).Having in mind this number is free, this is quite a low number. In SLI 30.1.1 Google note that they work since November 2022 with the International Fact-Checking Network (IFCN), they highlight this network has signatory organisations present in various EEA Member States, including Bulgaria. TikTok do not mention any plans for Bulgaria, but they have a new fact-checking partnership with Romania. In QRE 30.1.2, Google also mentions the IFCN signatory in Bulgaria. Meta point out that AFP are the fact-checkers for Bulgarian under both Facebook and Instagram. In Measure 31.1.1, Meta reports that the number of fact-checked labelled content for Bulgaria was “Over 510,000” and for Instagram it the number was “over 18,000” in Bulgaria. There is no data by market for the % of reshares attempted that were not completed on treated content on Facebook or Instagram. TikTok noted, in SLI 31.1.2, that 38 videos were removed in Bulgaria because of policy guidelines, known misinformation trends and knowledge-based repositories. QRE 31.3.1 – notes that TikTok are regularly engaged with EDMO on this priority and are committed to participate in the taskforce.

As in our previous sections, in Figure 4, the green colour stands for responses rated as ‘Good’, the yellow colour represents a rating of ‘Adequate’, the red colour covers ‘Poor’ responses and the grey colour represents N/A.Figure 4 illustrates the distribution of responses across platforms (the stacks have different lengths due to the fact that in several instances the platforms do not confirm their commitment to a specific measure, or we are lacking the rating). What is interesting in the Empowering domain is the lowest amount of detailed responses (smallest number of ratings of the value of 3). In fact the combination of the responses 1 and NA for this group would include the overwhelming number of metrics for all platforms but Google.

Having in mind the importance of properly supporting researchers, fact checkers and users, this illustrates the need to pay more attention to the evidence and further work on tools which support the users, researchers and fact checkers in their different user journeys related to disinformation.

Figure 4 - *Empowering Users Rating Comparison*



3. Overall Findings, Observations and Recommendations

3.1. Summary of Findings

The Bulgarian-Romanian Observatory of Digital Media (BROD) is following the developments around the Code of Practice since the kick off meeting of the project which was accompanied with a round table exploring the regional dimensions relevant to the adoption of the Code. This event took place on 26 January 2023 and showed that there is a long way to go in getting local engagement with the Code; only a few months later we are in a completely different situation where we can explore the first set of reports by signatories of the Code.

This report is only a first and partial exploration of the picture of how VLOP and VLOSE are implementing the strengthened CoP in Bulgaria, based on a manual analysis of the reports of all key VLOP and VLOSE signatories. Thus, this white paper provides a summary of details for anyone who is seeking to complete the picture of the disinformation landscape in Bulgaria adding further evidence from the local practices.

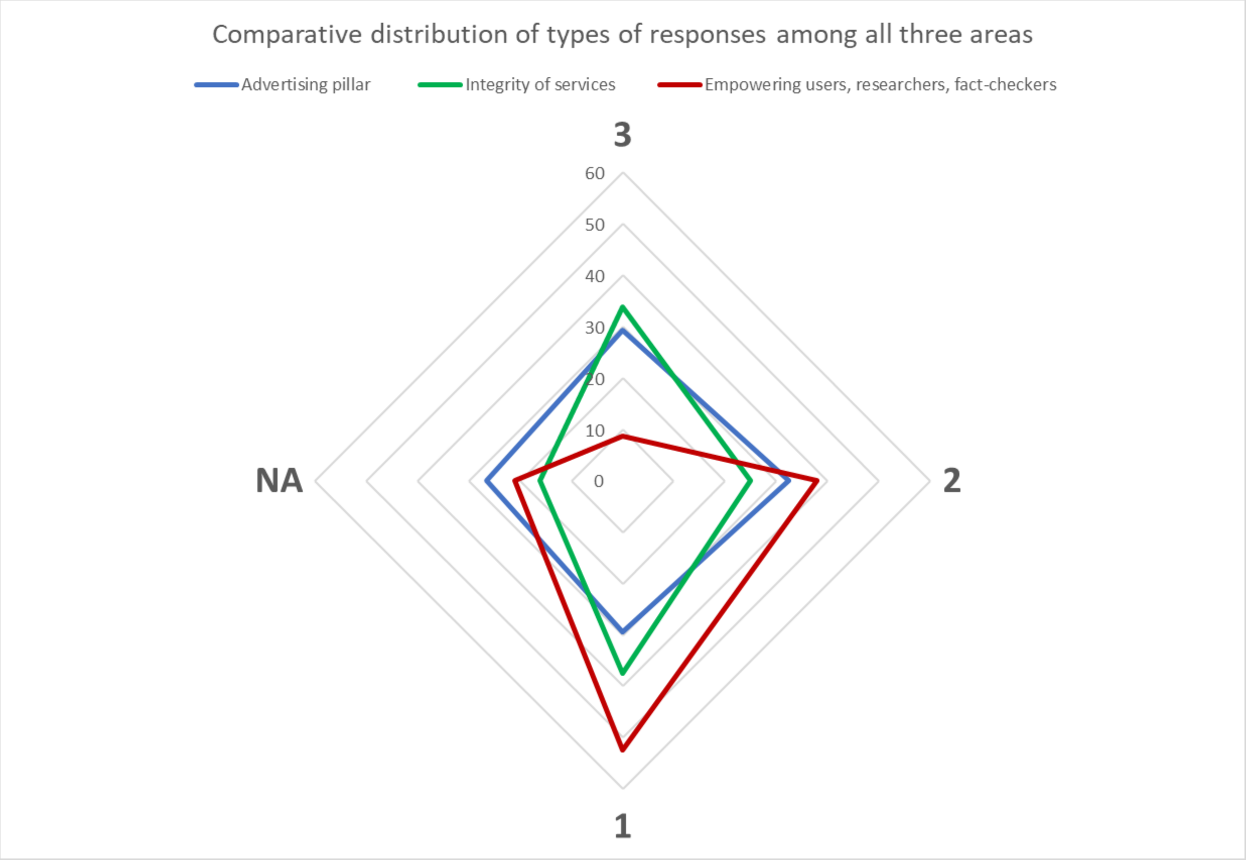
This report is providing an analysis of the data on Bulgaria from five signatories: Google, Meta, Microsoft, TikTok and Twitter. After an initial analysis of the appearance of Bulgaria within the first set of reports, these were determined to be the reports providing most details on Bulgaria.

In particular, the report is structured around compliance with three major pillars: Advertising and Political advertising (areas 2 and 3 of the Strengthened code); Integrity of services (Area 4 of the Strengthened code); and Empowering Users, the research community and Fact-Checkers (areas 5-7 of the Strengthened code).

For the **Advertising and Political Advertising Pillar**. The major recurring theme in terms of compliance was brief responses which lacked detail or information relevant to the requirements set out in the text. Measure 6 was not applicable to Microsoft and TikTok, while Google provided a partial response without describing current operations and procedures. Measure 7 had low compliance from Google and Meta, mainly due to a lack of data and promises to provide more in future reports. TikTok did not respond to Measure 7.4. There is some confusion regarding Measure 7.4, as Meta refers to other measures, indicating it finds Measure 7.4.1 irrelevant. **Overall, many responses lacked requested data, particularly at the state level**. There are also concerns about the **verifiability of the information provided**. **In terms of Bulgaria, Google provided detailed data on various measures, while Microsoft mentioned Bulgaria once, and TikTok, Twitter and Meta did not provide any country-level data for Bulgaria**.

We also observe different patterns in the level of detail in providing explanation to the different areas. The next diagram presents the responses in the three areas as percentages for the respective area. We can observe that the area with the highest number of responses where the platforms declared the metric is not relevant to them is the advertising pillar. Integrity of services shows most detailed responses among the three areas while the combined areas of empowering users, researcher and fact checkers shows the smallest amount of responses with substantial amount of detail. It is not surprising to discover such differences but we can expect that over time the reports will move in the direction of providing more details across all pillars. We can also see that generally Twitter, Microsoft and TikTok see more measures as non-applicable while Meta offers commentary on all measures; Google seems to be handling more measures than others with a sufficient level of detail. However, at times that detail is not appropriate or relevant to the response or measure.

Figure 5 - *Response Rating Distribution Comparison*



In terms of **Integrity of Services,** we can observe more detailed explanations from Google. In general responses here, while detailed, highlight the lack of verifiable information, this is true to some degree for all platforms, but specifically by Microsoft, TikTok and Meta. **Twitter**'s responses continued to be generally superficial or often indirect in terms of relevancy and lacked depth and any additional data or QRE/SLI responses. Responses in this pillar highlight specific data related to fake accounts removed, fake likes, fake followers, and accounts banned in Bulgaria provided by TikTok and LinkedIn. However, **researchers need more granular data, as well as details on how it was collected and analysed with more transparency and detail**. Overall, the responses from the companies **lack the requested clarity and depth, creating doubts about the verifiability of the information**. There are various aspects of the measures which have not been in place or which are said to be in some stage of implementation for which **no data was provided**. It is our hope this data and information will appear in the next reports provided.

Empowering Users, the research community and Fact-Checkers as a pillar is a vital component of limiting the spread of disinformation. The very significant challenges faced by Bulgaria in terms of political polarisation, concentration of media ownership and editorial independence, ongoing work on media literacy, powerful technological influence on information consumption habits, geopolitical context and its position as a low resource language country mean that cooperation between users. As such, platforms, researchers and fact-checkers must develop transparent, cooperative, clear and mutually empowering relationships to help tackle this serious social issue. The specific vulnerabilities in terms of disinformation and misinformation within the discursive landscape across various topics and issues make empowering users, researchers and fact-checkers a vital component of strategies to analyse or tackle the problem at the local level. Under this pillar several measures were addressed which tackled aspects such as Search Interventions and Public Service Announcements, Media Literacy campaigns, Partnerships and Initiatives with fact-checkers and researchers, share cancel rates and content removal, fact-checking and labelling, appeals and content removal and other measures and functions designed to satisfy the requirements of these measurements. This pillar repeats a pattern from the others in that there were various examples of missing SLI's along with a general sense of vagueness or lack of clarity with respect to details on some of the responses. There were often examples of platforms, such as Microsoft in QRE 17.1 failing to provide the metric required by the measure and yet responding with other types of counts. Generally speaking, there are significant gaps in the information provided which should be addressed in future reporting periods. In terms of reporting on specific member state numbers, this pillar provided more data than the others, and **unlike the other pillars data was provided by Google, Meta, Microsoft, and TikTok with respect to Bulgaria**. What is not made entirely clear is the relationship between the data for Bulgaria and the certified IFCN member in Bulgaria – the website factcheck.bg of the Association of European Journalists – Bulgaria and the Bulgarian fact-check section of AFP *proveri*. Moreover, **relatively little detail is provided about the methods used to calculate the figures provided for Bulgaria** and who was involved on the local level in this process. This creates an **alarming knowledge gap that demands greater attention and access to the data** in order to aid the research and fact-checking community. Bulgaria is a vulnerable member state and efforts must be made by all signatories to ensure that it is not overlooked. **Direct access to the data, as well as a fair and proportionate focus on Bulgaria and Bulgarian, are instrumental and badly needed for research-led policy making, as well as to help the Bulgarian research and fact-checking community in their work**. **Twitter** provided no country specific data on any of the measures under any of the pillars.

3.2. Barriers to automating VLOP and VLOSE report analysis

Prior to embarking on our detailed manual analysis effort, we first tried an automatic data analysis approach. The first step was to attempt automatic download of the data from the report website, however this highlighted several issues:

1. Firstly, the website's dynamic nature poses a challenge as any script created for automated download of the CSV and JSON versions of the reports may break in the future, requiring frequent adaptations. A simpler and more basic page format for sharing the reports would be preferable as it will ease access and simplify the automatic data download process.
2. Additionally, the website does not provide information about the file list, such as upload dates for the SCV and JSON files or their last modification dates. This would be extremely helpful to know, to ensure consistency of analysis. In particular, the problem is that if all files are downloaded on a given date and then the database of report files is updated, our data download programme cannot determine reliably which files were modified since our last download. As the database grows over time, manually checking for updates would become an extremely time-consuming task.
3. The process of creating the CSV and JSON files is not well-documented, making it unclear how they were generated. Moreover, one of the CSV files contains a warning and refers to the full PDF file, implying that the uploaded CSV or JSON files cannot be relied upon, and necessitating manual verification each time. It would be helpful to have assurances that the CSV files contain all the relevant information from the full PDF files.
4. There seem to be some discrepancies between the content of the CSV files vs the PDF versions of the reports. We identified this when we counted the occurrences of a keyword in one of the CSV files and then compared the resulting number against the different count obtained through manual searching in the PDF file.
5. Automated extraction of information from the PDF formatted reports is a challenging task, because the documents are formatted as tables. Our team experimented with various Python libraries, but the results were unsatisfactory.

The bottom line is that extracting information automatically from the VLOP and VLOSE reports is very difficult and urgently needs to be improved.

3.3. Recommendations

Based on our analysis, we conclude by making the following set of recommendations for urgent actions on behalf of the European Commission with the purpose of improving VLOP and VLOSE report quality, compliance, verifiability, transparency, and data provision:

1. **The European Commission could provide all VLOP and VLOSE with standardised reporting templates** in CSV and JSON formats, which are designed in consultation with EDMO and Hub researchers. This is needed to facilitate automated processing and cross-platform analysis.
2. Other than formatting issues, we recommend **EC-led standardisation of the reporting periods** (ideally weekly or bi-weekly), **units of reporting** (e.g. ad spend band harmonisation across all platforms), and **required detail of reporting to enable transparency and accountability**.
3. To enable effective Code monitoring and independent research, **the European Commission should investigate funding and establishing a shared, EU-wide large-scale data processing infrastructure**, as well as a mechanism for researchers to share know-how. This is urgently needed, as the current data sharing provisions by VLOP and VLOSE are inadequate, both for the purpose of transparency and independent evaluation by researchers, and for enabling research in the public interest (e.g. during the forthcoming EU elections).
4. We recommend that the **EU invests in the development of shared, free, and comprehensive open-source data cleaning, harmonisation, storage, and analysis tools** for data shared by VLOP and VLOSE. This is badly needed by researchers from less resourced EU countries (such as those from Central and Eastern Europe) to carry out effective monitoring of the Code implementation and VLOP and VLOSE measures against disinformation, as platforms are currently fairing the worst there in terms of effective enforcement of their policies against online abuse and disinformation.
5. We believe the EC should lead in defining **common research data access policies**, rather than leave this to each VLOP and VLOSE independently. These policies need to stipulate that **research data access is freely available to independently vetted researchers across the EU**, and that **this includes researchers not only from academia but also from NGOs, public media, and independent fact-checkers**, all of whom are working on unique and highly valuable disinformation research.
6. **Cross-platform structured data sharing standards and common data access APIs**, need to be created and adopted by all VLOP and VLOSE. Multi-stakeholder action is needed to define and implement these, in order to enable cross-platform, quantitative comparative studies on key compliance issues such as cross-platform spread of disinformation, cross-platform political ad campaigning during elections, etc.
7. **VLOP and VLOSE should face consequences from their withdrawal from the CoP**, in order to discourage other VLOP and VLOSE to follow Twitter's lead in introducing unaffordable data access charges for researchers and refusal of CoP compliance.
8. **The main criteria where VLOP and VLOSE fail is in the verifiability and transparency** of their self-reported numbers under SLIs. We urge **the EC to ensure that data sharing for compliance by VLOP and VLOSE includes large and representative samples of moderated, demonetized and removed content**. This is the only way for researchers to verify (based on these samples) whether platforms are enforcing their policies correctly and that legitimate content and accounts are not being silenced due to biases or errors in algorithmic moderation and/or human moderation error. This is especially critical around global emergencies such as the COVID-19 pandemic and the Ukraine

war, as well as key EU-wide events such as the European elections, as these generate tens of millions posts a day on big platforms such as Facebook, YouTube, Twitter, and TikTok.

9. **Data access for researchers should be free** and the provided data needs to be sufficiently large scale to **enable computational (not just small sample social science) research and monitoring of disinformation spread at scale** and across VLOP and VLOSE. Researchers also need access to sufficiently large volumes of data, to enable longitudinal, large-scale monitoring of VLOP and VLOSE compliance to the CoP.
10. Where VLOP and VLOSE reports have provided very similar answers (especially statistics) for two different platforms (see some of Meta's reports on Facebook and Instagram flagged above), we recommend that the EC submits a formal request for details and clarification.
11. Likewise, the **EC could formally notify VLOP and VLOSE where reports are lacking member-level data and/or specific numbers**. Use of vague ranges and broad, generic responses to SLIs should be considered non-compliant.
12. The **EC could consider VLOP and VLOSE non-compliant also when they only implement measures only in a small number of EU member states** and, in particular, insist on **immediate roll out of measures in Bulgaria** and other similarly vulnerable EU countries where Russian propaganda and political, health, and climate disinformation are not only flourishing but already causing very significant real-world harms. Multiple examples have been flagged in this white paper, e.g. QRE 25.1.1, QRE 30.1.3.
13. The EC should work closely with VLOP and VLOSE to **ensure a transparent and equitable process for vetting researchers and capacity to access data for research purposes.** Currently there is a danger of researchers from Eastern and Central European countries being marginalised, with most effort and funds being focused on research labs in bigger countries and markets. Moreover, **some companies are currently rejecting as many as 80% of requests for access to data by researchers** (see Google under QRE26.2.3), which demonstrates that **researcher vetting and data access policies cannot be left at the discretion of individual VLOP and VLOSE**.

There are also a number of recommendations aimed at actions that VLOP and VLOSE need to undertake in the short to medium term, in order to improve their implementation of and compliance with the CoP:

1. All VLOP and VLOSE should start **working with information Integrity experts and researchers**, including organisations such as NewsGuard and GDI, and use those as source and references of disinformation domains.
2. VLOP and VLOSE should **improve their use of Bulgarian fact-checks** (e.g. see Microsoft and TikTok under SLI 31.1.1, also implementation of QRE 32.1.1) **and work with independent third-party fact-checkers in Bulgaria, as well as provide funding, training, data and tool access, and knowledge sharing** aimed at improving the very limited fact-checking capacity which Bulgaria has - an issue that needs to be addressed **urgently**.
3. Companies that own more than one online social platform should **ensure sufficiently detailed and verifiable answers are provided under each platform**.
4. **VLOP and VLOSE should cooperate closely with each other** to ensure effective cross-platform measures to combat disinformation are put in place.
5. VLOP and VLOSE should **significantly improve the quantitative reporting they provide under SLIs**, broken down on a monthly basis and on a per EU-member state basis. In addition to comprehensiveness and rigour, **special attention needs to be paid to verifiability and transparency of reporting**.
6. VLOP and VLOSE should work with the EC and all other relevant stakeholders, in order to **ensure fit-for-purpose provision of access to data for independent research and monitoring of disinformation**.

References

EC. *The Strengthened Code of Practice on Disinformation 2022*. Available at: https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation (Accessed: 21 July 2023).

Google (2023) *Code of Practice on Disinformation – Report of Google for the period 1 July 2022 - 30 September 2022.* Available at: https://disinfocode.eu/reports-archive/?years=2023 (Accessed: 20 July 2023).

Meta (2023) *Code of Practice on Disinformation – Meta Baseline Report.* Available at: https://disinfocode.eu/reports-archive/?years=2023 (Accessed: 20 July 2023).

Microsoft (2023) *Code of Practice on Disinformation Baseline Report – January 2023 Microsoft.* Available at: https://disinfocode.eu/reports-archive/?years=2023 (Accessed: 20 July 2023).

Park, K. & Mündges, S.(2023). *CoP Monitor. Baseline Reports: Assessment of VLOP and VLOSE Signatory reports for the Strengthened Code of Practice on Disinformation*. Available at: https://fujomedia.eu/wp-content/uploads/2023/09/CoP-Monitor-Report.pdf (Accessed 7 September 2023).

TikTok (2023) *Code of Practice on Disinformation – Report of TikTok for the period 16 June - 16 December 2022.* Available at: https://disinfocode.eu/reports-archive/?years=2023 (Accessed: 20 July 2023).

Kantar Public and Visionary Analytics Methodology (2023) *A Monitoring Framework for the Code of Practice on Disinformation*.