Mediterranean Digital Media Observatory

D5.1

1st Report of platform practices and national authorities support 2023 (for Cyprus, Greece & Malta)

Project Title	MedDMO
Project Number	101083756
Thematic Priority	DIGITAL-2021-TRUST-01-EDMO
Start of Project	1 December 2022
Duration	30 months

Deliverable title	1st Report of platform practices and national authorities support –2023
Deliverable number	D5.1
Deliverable version	V1.0
Contractual Date of delivery	n/a
Actual Date of delivery	n/a
Nature of deliverable	Report
Dissemination level	Public
Partner Responsible	CUT
Author(s)	Pantelitsa Leonidou, Nikos Salamanos, Michael Sirivianos, Loukia Taxitari, Socrates Ioakeim, Andreas Phylactou
Reviewer(s)	CERTH, UCY
EC Project Officer	Kyriaki Tragouda

Abstract	This is the MedDMO annual report detailing the results of monitoring the online platform practices towards supporting the national authorities of Cyprus, Greece and Malta on combating active disinformation campaigns. In addition, the progress on the implementation of the Code of Practice on Disinformation by online platforms has also been reported.
Keywords	Code of Practice on disinformation, Social media, disinformation campaign, fact checking

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756.
For further information, visit www.meddmo.eu

Copyright

Revision History

VERSION	DATE	MODIFIED BY	COMMENTS
V0.1	01/12/2023	CUT	First Draft Table of Content
V0.2	15/12/2023	CUT	Final Table of Content
V0.3	12/03/2024	CUT	First Draft
V0.4	17/03/2024	CUT	Second Draft
V0.5	27/03/2024	CUT	Final Second Draft
V0.6	07/04/2024	CUT	Final Second Draft Review
V0.7	11/04/2024	AUTH, UCY	Review
V1.0	06/05/2024	CUT	Final Document

Glossary

ABBREVIATION	MEANING
CoP	Code of Practice on Disinformation
VLOPs	Very Large Online Platforms
IFCN	International Fact Checking Network
AFP	Agence France-Presse
EH	Ellinika Hoaxes
QREs	Qualitative Reporting Elements
SLIs	Service Level Indicators (SLIs)

Table of Content

Index of Figures

Index of Tables

Executive Summary

This report presents a qualitative evaluation of the practices of online platforms in Cyprus, Greece, and Malta, with a focus on the implementation of the signed Code of Practice on Disinformation by three very large online platforms (VLOPs): Meta, Google, and TikTok for the period of 1st of January to 30th of June 2023 (based on their CoP reports of July 2023). It also discusses research activities from MedDMO partners and overall activities supporting the national authorities in the three countries.

The global challenge of disinformation has become increasingly pervasive, impacting societies, political landscapes, and public discourse worldwide. Cyprus, Greece, and Malta each face unique challenges within their disinformation landscapes. In Cyprus, misinformation proliferates through social media channels and websites, particularly during election periods and amidst significant news events such as the war in Ukraine. Greece experiences misinformation crises during natural disasters and political decisions, with actors ranging from government mechanisms to far-right movements disseminating false information. In Malta, state-sponsored trolls contribute to a complex disinformation environment, exacerbated by the aftermath of the assassination of investigative journalist Daphne Caruana Galizia.

This study focuses on monitoring the actions of three prominent online platforms: Meta, Google, and TikTok, which are widely used in the three countries. Meta holds a central role in shaping public discourse, Google influences online content visibility, and TikTok provides a unique arena for user-generated content.

The objective is to investigate the policies, practices, and tools implemented by these platforms to combat misinformation within the digital ecosystems of Cyprus, Greece, and Malta. The report assesses the efficiency of these measures and evaluates their accessibility to diverse audiences, aiming to provide a comprehensive understanding of how effectively the strategies address the challenges posed by misinformation.

This analysis marks the initial step in comprehensively examining platform practices in combating misinformation across different digital spaces in Cyprus, Greece, and Malta. Through this exploration, prevalent trends, challenges, and potential strategies for mitigating the impact of misinformation within these online landscapes are identified.

1 Introduction

This report presents a qualitative evaluation of the practices of online platforms in Cyprus, Greece, and Malta. The study primarily focuses on the implementation of the signed Code of Practice on Disinformation by three very large online platforms (VLOPs): Meta, Google, and TikTok. Additionally, the report discusses the research activities from MedDMO partners, and overall activities supporting the national authorities in the three countries.

The global challenge of disinformation has become increasingly pervasive, impacting societies, political landscapes, and public discourse across the world. With the rise of digital platforms and social media, the dissemination of false or misleading information has reached unprecedented levels. This phenomenon not only poses a threat to the integrity of information but also influences public opinion, undermines trust in institutions, and potentially sways political outcomes. The countries of Cyprus, Greece, and Malta are not exempt from this pressing issue, each grappling with unique nuances and challenges within their respective disinformation landscapes. Specifically:

In **Cyprus** the problem of disinformation is notably serious, particularly during election periods and in the midst of hot-button issues dominating the news cycle, such as the war in Ukraine. Misinformation often proliferates through social media channels, websites, and newspapers, with individuals aiming to undermine political figures or parties, instil fear, and mould public opinion through the dissemination of false news. The culprits range from anonymous accounts and pages linked to political interests to foreign websites attempting to tarnish the image of the Republic of Cyprus.

Greece faces a significant challenge with misinformation, often escalating during crises, be they natural disasters or political decisions. Disinformation crises frequently follow real-world events, such as disastrous fires leading to misinformation about causes and actors, or political decisions triggering false narratives about the supposed harmful technology embedded in new digital IDs. Main actors include government mechanisms, political parties, far-right movements, and even celebrities, all contributing to the dissemination of false information, predominantly on social media platforms.

Malta grapples with a major disinformation problem, exacerbated in the years following the assassination of investigative journalist Daphne Caruana Galizia on 16th October 2017. State-sponsored trolls play a substantial role, engaging in coordinated attacks on anti-corruption activists and civil society members who expose wrongdoing. The 2017 assassination triggered an onslaught of disinformation, including widespread claims fuelled by troll armies and false articles and emails supporting conspiracy theories about the circumstances surrounding Daphne’s murder, which continues to this day. The disinformation landscape in Malta involves online troll armies, state-orchestrated propaganda, and targeting of journalists, creating a complex and concerning environment.

The report is organised as follows: **Section 2** presents the approach used for monitoring platform practices based on the CoP signatories report. Following this, **Section 3** gives a summary of the results, covering Meta (Facebook and Instagram), Google (Ads, Search, YouTube), and TikTok, with each platform's practices analysed in detail, including the following pillars: Scrutiny of ad placements, Political advertising and efforts to empower users, research communities, and fact-checking organisations. The report also discusses MedDMO's fact-check activities (see **Section 4**) looking into how platforms elaborate the fact-checking efforts and discussing the MedDMO fact-checking organisations (Agence France-Presse and Ellinika Hoaxes) collaborations with platforms. Additionally, **Section 5** presents two research initiatives aiming at automatic disinformation detection on social

media platforms such as Twitter and YouTube. In **Section 6**, the report examines the established policies and regulations by the three countries Cyprus, Greece, and Malta to combat disinformation. Finally, in **Section 7** we discuss MedDMO steps to support the national authorities in addressing disinformation issues. Each section contributes to a comprehensive understanding of the strategies and actions taken to combat disinformation in the three countries.

2 Methodology for Monitoring the platforms practices

Monitoring the Code of Practice on Disinformation

On June 16, 2022, 34 participants involved in revising the 2018 Code came together to sign and introduce the enhanced Code of Practice on Disinformation[1]. This updated Code builds upon the objectives outlined in the European Commission's May 2021 Guidance, introducing a wider range of commitments and measures to combat online disinformation. Signatories are given the autonomy to select the commitments they support and are responsible for ensuring their effective implementation. Although the Commission does not officially endorse the Code, it aligns with the expectations set forth in the Guidance.

Signatories have committed to taking action in various areas, including demonetizing the spread of disinformation, ensuring transparency in political advertising, empowering users, fostering collaboration with fact-checkers, and facilitating researchers' access to data. To ensure the Code remains adaptable, signatories have established a permanent Task Force for ongoing collaboration. The Code incorporates a robust monitoring framework, incorporating qualitative reporting elements and service-level indicators to assess implementation effectiveness. Signatories will establish a Transparency Centre to provide regular updates to the public on the policies in place to fulfil their commitments.

In our analysis of platform disinformation practices, we will focus on monitoring the actions of three prominent online platforms: **Meta (Facebook and Instagram), Google (YouTube, Search, Advertising), and TikTok,** widely used in **Cyprus, Greece, and Malta**. Meta, encompassing Facebook, Instagram, and WhatsApp, holds a central role in shaping public discourse. Google, as the predominant global search engine, influences the accessibility and visibility of online content, while TikTok, a rapidly expanding short-form video platform, provides a distinct arena for user-generated content.

Our objective is to investigate the policies, practices, and tools implemented by Meta, Google, and TikTok to combat misinformation within the digital ecosystems of Cyprus, Greece, and Malta. We will assess the efficiency of these measures and evaluate their accessibility to the platforms' diverse audiences. This inquiry seeks to provide a comprehensive understanding of the extent to which the implemented strategies effectively address the challenges posed by misinformation, ensuring that users across the three countries can navigate online spaces with greater resilience to false or misleading information.

Furthermore, this analysis marks the initial step in comprehensively examining their practices in combating misinformation, considering the diverse ways users engage with and consume information across different digital spaces. Through this exploration, we aim to identify prevalent trends, challenges, and potential strategies for mitigating the impact of misinformation within the online landscapes of Cyprus, Greece, and Malta.

[1] https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

Our analysis focuses on the CoP Measures outlined in Table 1. We assess and present the **findings based on the July 2023 CoP Signatories reports (covering the period of 1st of January to 30th of June 2023)**, following the **methodology proposed by the EDMO Ireland hub and German-Austrian Digital Media Observatory (GADMO) hub [Park et al., 2023].** To thoroughly assess the performance of each platform, we have utilised an assessment scale (refer to Table 2) to rate the platform's **reported actions and policies for each measure (QREs), as well as the implementation of these actions in Cyprus, Greece, and Malta and the reported quantitative information (SLIs)**.

Pillars	Measures	Issue covered by the measure
II Scrutiny of Ad placements	1.1	Demonetization of disinformation
	2.1	Tackling advertising containing disinformation
III Political Advertising	6.2	Labelling political /issue ads
	10.1	Repositories of political and issue ads
	10.2	
V Empowering Users	17.1	Media Literacy
	17.2	
	18.1	Safe design
	18.2	
	21.1	Better equipping users to identify disinformation
	24.1	Transparent appeal mechanism
VI Empowering the Research Community	26.2	The provided access to platforms' data for researchers
VII Empowering the Fact-Check Community	31.1	Cooperation with fact-checkers
	31.2	

Table 1: CoP measures covered by this analysis

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.

Table 2: CoP measure-level assessment scaling system

In **Section 3**, we present the summary of our evaluation results. The summarised results are followed by a detailed analysis per platform. For each platform and measure, we present the assigned scores (general, Cyprus, Greece, and Malta), followed by summarising the reported actions pertaining to the measure, reported SLIs, and evaluators' comments on the platforms' practices based on their responses and actual implementation.

3 Summary of Results

The final findings from the overall analysis presented in this document are shown in the plots below (Figures 1-6).

The evaluation is referring to the reported materials, specifically:

1. We evaluate the reported platform's practices and policies (QREs - Qualitative Reporting Elements)
2. We evaluate the reported quantitative data together with the related implementation by the platforms, namely, the quality and verifiability of Service Level Indicators (SLIs) together with the overall implementation of the reported policies and practices (QREs) by the platforms for the three countries.

Meta

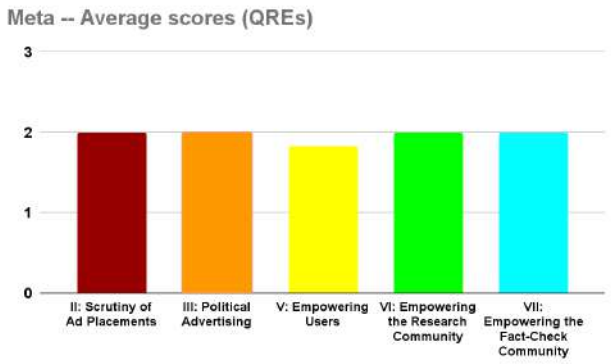


Figure 1: Meta -- Average Scores per pillar

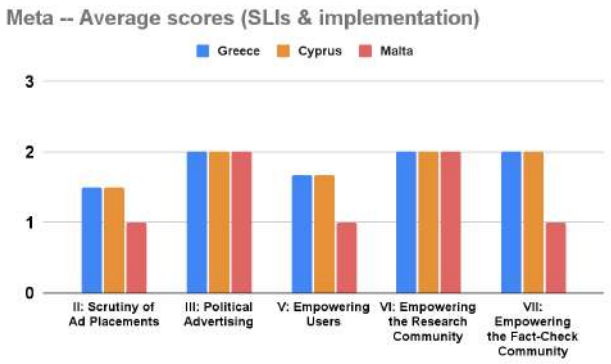


Figure 2: Meta -- Average Scores per pillar and country

Google

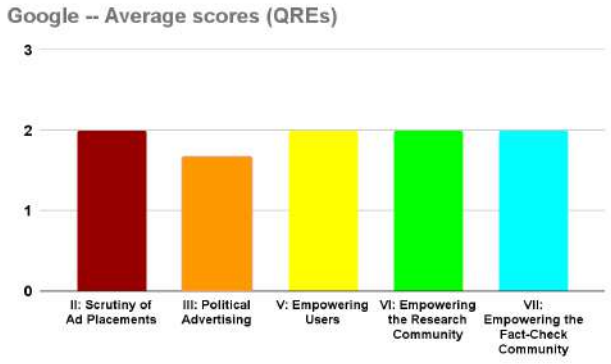


Figure 3: Google -- Average Scores per pillar

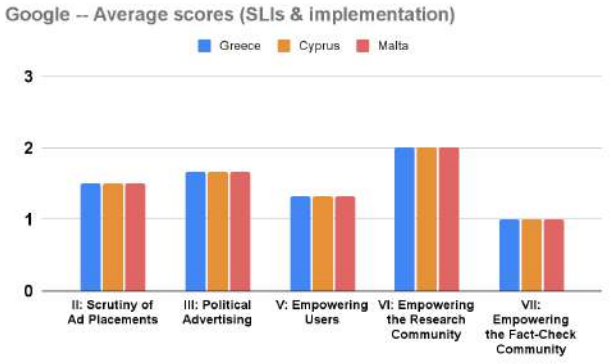


Figure 4: Google -- Average Scores per pillar and per country

TikTok

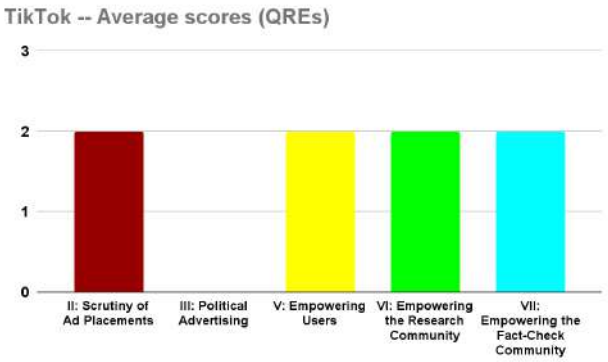


Figure 5: TikTok -- Average Scores per pillar *

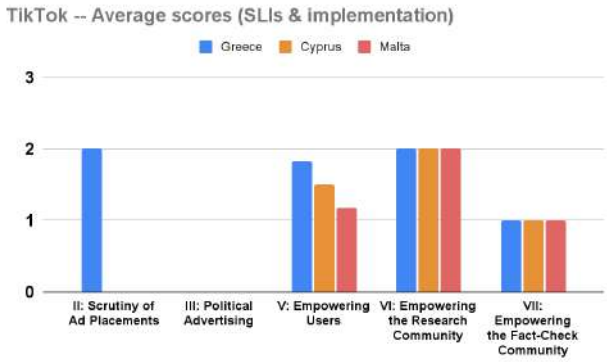


Figure 6: Average Scores per pillar and per country **

* TikTok prohibits political advertising, hence there are no scores – Not Applicable (N/A) is assigned to the relevant CoP measures. In the figures above this is indicated with zero average scores for the ''Political Advertising'' pillar.
** No TikTok ads are available in Cyprus and Malta, so there are no scores – – Not Applicable (N/A) is assigned to the relevant CoP measures. In the figure above this is indicated with a zero average score for the ''Scrutiny of Ad Placements' pillar'' for Cyprus and Malta.

Table 3: Overall evaluation per platform and country. Average scores per CoP pillar.

The plots in Table 3 depict the average score values per CoP pillar[2] per platform which are presented in Tables 4-6.

[2] The average scores per CoP pillar are calculated based on the scores assigned to the measures included in our analysis (refer to Table 2). It's important to note that not all CoP measures are evaluated.

Meta Scores: 1 (“poor”), 2 (“adequate”), 3 (“good”), n/a (“not applicable”)				
	Evaluation of Reported Actions/Policies (QREs)	**Evaluation of Quantitative data (SLIs) & implementation**		
		Greece	**Cyprus**	**Malta**
Pillar-II: Scrutiny of Ad Placements				
Measure 1.1	2	2	2	1
Measure 2.1	2	1	1	1
Average	**2**	**1.5**	**1.5**	**1**
Pillar-III: Political Advertising				
Measure 6.2	2	2	2	2
Measure 10.1	2	2	2	2
Measure 10.2	2	2	2	2
Average	**2**	**2**	**2**	**2**
Pillar-V: Empowering Users				
Measure 17.1	2	2	2	1
Measure 17.2	1	1	1	1
Measure 18.1	2	2	2	1
Measure 18.2	2	2	2	1
Measure 21.1	2	2	2	1
Measure 24.1	2	1	1	1
Average	**1.83**	**1.67**	**1.67**	**1**
Pillar-VI: Empowering the Research Community				
Measure 26.2	**2**	**2**	**2**	**2**
Pillar-VII: Empowering the Fact-Check Community				
Measure 31.1	2	2	2	1
Measure 31.2	2	2	2	1
Average	**2**	**2**	**2**	**1**

Table 4: Meta's CoP Report Summary of Assessment Results

Google Scores: 1 (“poor”), 2 (“adequate”), 3 (“good”), n/a (“not applicable”)				
	Evaluation of Reported Actions/Policies (QREs)	**Evaluation of Quantitative data (SLIs) & implementation**		
		Greece	**Cyprus**	**Malta**
Pillar-II: Scrutiny of Ad Placements				
Measure 1.1	2	2	2	2
Measure 2.1	2	1	1	1
Average	**2**	**1.5**	**1.5**	**1.5**
Pillar-III: Political Advertising				
Measure 6.2	2	2	2	2
Measure 10.1	2	2	2	2
Measure 10.2	1	1	1	1
Average	**1.67**	**1.67**	**1.67**	**1.67**
Pillar-V: Empowering Users				
Measure 17.1	2	2	2	2
Measure 17.2	2	1	1	1
Measure 18.1	2	1	1	1
Measure 18.2	2	1	1	1
Measure 21.1	2	2	2	2
Measure 24.1	2	1	1	1
Average	**2**	**1.33**	**1.33**	**1.33**
Pillar-VI: Empowering the Research Community				
Measure 26.2	**2**	**2**	**2**	**2**
Pillar-VII: Empowering the Fact-Check Community				
Measure 31.1	2	1	1	1
Measure 31.2	2	1	1	1
Average	**2**	**1**	**1**	**1**

Table 5: Google's CoP Report Summary of Assessment Results

TikTok Scores: 1 (“poor”), 2 (“adequate”), 3 (“good”), n/a (“not applicable”)				
	Evaluation of Reported Actions/Policies (QREs)	**Evaluation of Quantitative data (SLIs) & implementation**		
		Greece	**Cyprus**	**Malta**
Pillar-II: Scrutiny of Ad Placements [3]				
Measure 1.1	2	1	N/A	N/A
Measure 2.1	2	1	N/A	N/A
Average	**2**	**2**	**N/A**	**N/A**
Pillar-III: Political Advertising [4]				
Measure 6.2	N/A	N/A	N/A	N/A
Measure 10.1	N/A	N/A	N/A	N/A
Measure 10.2	N/A	N/A	N/A	N/A
Average	**N/A**	**N/A**	**N/A**	**N/A**
Pillar-V: Empowering Users				
Measure 17.1	2	2	1	1
Measure 17.2	2	2	1	1
Measure 18.1	2	2	2	1
Measure 18.2	2	2	2	1
Measure 21.1	2	1	1	1
Measure 24.1	2	2	2	2
Average	**2**	**1.83**	**1.5**	**1.17**
Pillar-VI: Empowering the Research Community				
Measure 26.2	2	2	2	2
Pillar-VII: Empowering the Fact-Check Community				
Measure 31.1	2	1	1	1
Measure 31.2	2	1	1	1
Average	**2**	**1**	**1**	**1**

Table 6: TikTok's CoP Report Summary of Assessment Results

[3] No TikTok ads are available in Cyprus and Malta, so there are no scores assigned to the relevant CoP measures.
[4] TikTok prohibits political advertising, hence there are no scores assigned to the relevant CoP measures.

In the Table 7 below we recorded the reported SLIs per platform/service for the CoP measures included in this analysis.

Service Level Indicators (SLIs) reported per platform/service ✓ - reported SLI x - no SLI reported N/S - platform/service did not subscribe to the relevant Measure N/A - platform/service consider the SLI not applicable						
Measures	**SLIs**	**Meta (Facebook and Instagram)**	**Google Ads**	**Google Search**	**YouTube**	**TikTok**
1.1	1.1.1	x	✓	N/S	N/S	✓
2.1	2.1.1	✓	✓	N/S	N/S	✓
6.2	6.2.1	✓	✓	N/S	N/S	N/S
17.1	17.1.1	x	N/S	✓	✓	✓
17.2	17.2.1	x	N/S	✓	✓	✓
18.1	18.1.1	✓	N/S	N/S	x	✓
18.2	18.2.1	✓	N/S	✓	✓	✓
21.1	21.1.1	✓	N/S	✓	x	✓
	21.1.2	✓	N/S	N/A	N/A	✓
24.1	24.1.1	✓	N/S	N/S	✓	✓
26.2	26.2.1	x	N/S	N/S	✓	x
31.1 and 31.2	31.1.1	✓	N/S	✓	x	✓
	31.1.2	✓	N/S	N/A	N/A	✓
	31.1.3	✓	N/S	x	x	✓
Total Missing SLIs information[5]		4	4			1

Table 7: Missing Service Level Indicators (SLIs) information in Signatories Reports

[5] We report the missing SLIs information only for the CoP measures that are part of this analysis (see Table 7).

3.1 Meta (Facebook and Instagram)

Our analysis on Meta's practices is based on the information provided in **Meta's Code of Practice Report, July 2023, No2** [6].

3.1.1 II. Scrutiny of Ad Placements

Pillar-II: Scrutiny of Ad Placements ***Commitment 1, Measure 1.1*** *QRE 1.1.1., SLI 1.1.1 & 1.1.2, page 11-13*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 1.1	2	2	2	1

Meta users must adhere to ad policies, which include detailed guidelines on monetization, misinformation, branded content, and more. Content Monetization Policies explicitly prohibit the monetization of misinformation, especially content rated false by third-party fact-checkers. The Partner Monetization Policies emphasise the authenticity of shared content and engagement.

Meta **has introduced new Inventory Filters** for Facebook and Instagram Feeds, deploying a multi-stage AI review system to classify content To enhance brand suitability controls[7]. This system complements existing technology that identifies content violating Community Standards, restricting ad placement accordingly. The AI models, aligned with GARM's Suitability Framework, categorise content into high, medium, and low-risk levels, allowing advertisers to choose from three settings for monetizable content. **Currently, the new Inventory Filters are only available to advertisers in English and Spanish-speaking markets.**

Meta has collaborated with Zefr to develop an independent AI-powered solution for third-party verification of content context in Facebook Feed ads. Early testing with Zefr revealed that less than one percent of content falls into the high-risk category, providing advertisers with tools to measure, verify, and understand content suitability near their ads.

Major comments:

1. Meta's policies to defund misinformation dissemination are available in the Greek language, while there is **no option for the Maltese language** (default language: English).
2. Additionally, as mentioned previously the brand suitability controls are not available **in Greek or Maltese speaking markets** – although they are available **for English and Spanish speaking markets**. Those tools do not cover the advertisers. However, Meta expressed its intent to make these tools available for more countries, and languages in the future. Additionally, the aforementioned tools are available only for Facebook and Instagram feeds, not covering content

[6] https://disinfocode.eu/reports-archive/?years=2023

[7] https://www.facebook.com/business/news/brand-safety-suitability-feed-control-verification

such as stories, reels, etc.

3. There is **no reference to any verification mechanism** for ensuring that ads are not placed in apps that disseminate misinformation through Meta Audience Network, where Meta ads are displayed in third party applications.
4. In July 2023, Meta neither **reported quantitative data (SLIs) for the impact of the enforcement** of the above policies, nor brand suitability tools (SLIs 1.1.1 and 1.1.2).

Pillar-II: Scrutiny of Ad Placements ***Commitment 2, Measure 2.1*** *QRE 2.1.1., SLI 2.1.1-page 15-16*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 2.1	2	1	1	1

Disinformation Dissemination through Advertising systems:

Advertising on Meta technologies mandates adherence to the Terms of Service, Community Standards, and Advertising standards. Meta explicitly categorises misinformation as unacceptable content under its Advertising standards, with specific types subject to removal based on Community Standards—these include physical harm or violence, harmful health misinformation, voter or census interference, and manipulated media. Additionally, ads must not feature debunked content verified by third-party fact-checkers. Persistent dissemination of false information by advertisers may result in limitations on advertising privileges across Meta's technologies.

The ad review system at Meta utilises automated tools to assess ads and business assets against policies, starting before ads go live and typically concluding within 24 hours. This process evaluates various ad components, such as images, videos, text, targeting, and associated landing pages, while also reviewing business accounts and assets for policy compliance. In case of violations, ads may be rejected, and restrictions can be imposed on business accounts or assets.

Major Comments:

1. Meta provides the links to the specific policies. The information on the reported webpages is available in the Greek language but **there is no option for the Maltese** (by default the information is in English).
2. The reported numbers for SLI 2.1.1 (see Table 8) regarding the removed ads due to Meta misinformation policy are very low for the three countries compared to the overall ads removed.
    - In **Cyprus and Greece**, < 0.005% of overall ads were removed because of misinformation policy violations. In **Malta** <0.02% of overall ads were removed because of violation of misinformation policy.
    - Meta's practice of combining numbers for Instagram and Facebook in their reports may pose limitations for future assessments and diminish the transparency of the reported information.
    - For assessing the numbers provided, there must also be a description of how the numbers were derived and more specifically of how they differentiate the numbers per country (e.g., advertiser's location). Additionally looking at the Meta Ads Library, there is no filter to get the ads removed because of violation for misinformation related policies or for another kind of policy violation — again making it **impossible to assess the reported numbers**.
3. Meta's reported numbers of removed ads do not differentiate the fraction of ads removed due to the misinformation policy or due to ads content similarity with debunked misinformation content from third-party fact-checkers.
4. It would be insightful if Meta shares the numbers of ads that got rejected at the ad review process and ads that got removed at a post-publishing stage.
5. Meta could also share quantitative information on the **number of accounts restricted** from Meta advertising technologies **as a result of repeatedly sharing misinformation**.

SLI 2.1.1	Number of Ads removed on **Facebook and Instagram** combined **for violating the Meta Misinformation policy**	Overall number of **ads removed** on **Facebook and Instagram** combined
Cyprus	**Less than 500**	**Over 110,000**
Greece	**Less than 500**	**Over 110,000**
Malta	**Less than 500**	**Over 33,000**
Total EU	**Over 24,000**	**Over 6,900,000**

Table 8: Meta's reported quantitative information for SLI 2.1.1

3.1.2 III. Political Advertising

Pillar-III: Political Advertising *Commitment 6, **Measure 6.2** QRE 6.2.1-page 27-28*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 6.2	2	2	2	2

Labelling of Political or issue Ads

Meta has implemented a policy for political and issue ads on Facebook and Instagram. Advertisers are required to include a verified "Paid for by" disclaimer on ads related to social issues, elections, or politics, commonly referred to as **"SIEP ads."** This disclaimer serves the purpose of transparently showcasing the entity or person responsible for running the advertisement. Further information on how these disclaimers function in ads about social issues, elections, or politics is available in Meta's help centre, providing users with a comprehensive understanding of the disclosure mechanisms associated with these specific ad categories.

Examples of political or issue ads labelling can be found in Meta's Ad Library (see Figure 7).

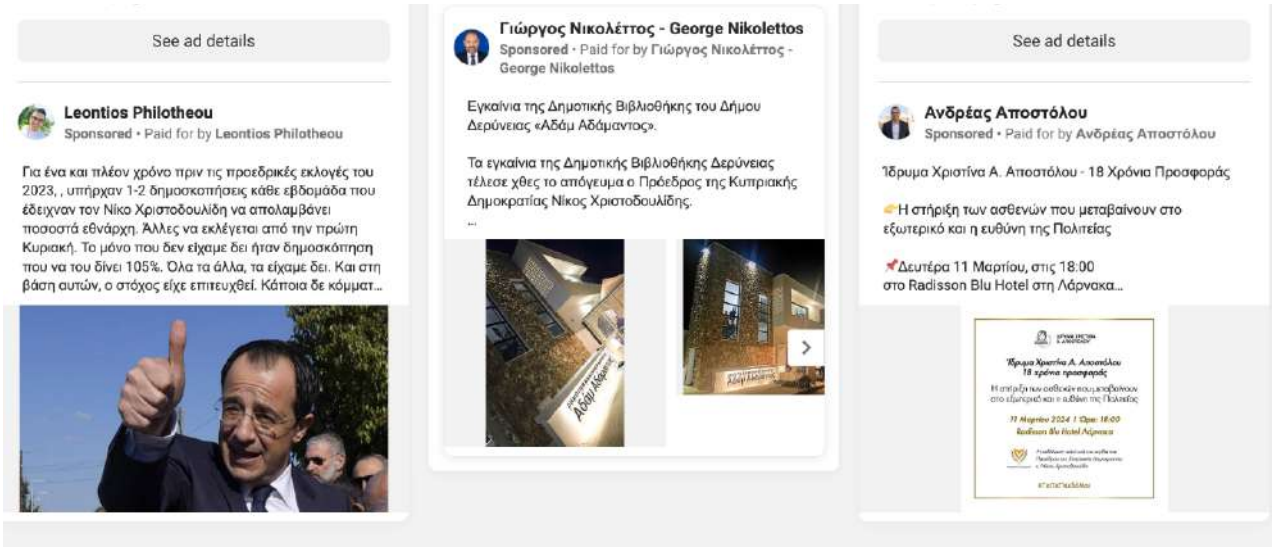


Figure 7: Examples of political ad labelling – Sponsored, ''paid for by'' disclaimers as obtained from Meta Ad Library

Major Comments:

1. Meta requires authorizations and a "Paid for by" disclaimer for political and issue ads. **This is in line with the CoP.** The disclaimer is easily accessible to the user since it is displayed at the beginning of the ad. The labelling text language is also available in **Greek, and Maltese**. **However,** the users may **not be familia**r with what this disclaimer means; additional info must accompany the disclaimer such as label/category note of *"this is a political or issue ad"*.
2. During the reporting period, **Greece** recorded one of the highest numbers of accepted political and issue ads on both Instagram and Facebook among all EU member states. This surge in ad volume aligns with expectations, considering the reported period (January-June 2023) coincided with a pre-election period in Greece[8]. The elevated activity in political and issue-related advertising may be a result of the intensified engagement and communication efforts leading up to the elections in the country.
3. For **Cyprus**, the number of political and issue ads is **low** (compared to other countries) for the specific period, considering the Cyprus presidential election[9] taking place in February 2023 (see Table 9 below).
4. Meta's practice of combining numbers for Instagram and Facebook in their reports may pose limitations for future assessments and diminish the transparency of the reported information.
5. Meta did not report any numbers for ***"amounts spent by labelled advertisers"*** as referred to in the SLI 6.2.1 description.

SLI 6.2.1	**Number of unique SIEP ads** on **Facebook and Instagram** combined **displaying "paid for by"** disclaimers from 01/01/2023 to 30/06/2023
Cyprus	**Over 3,600**
Greece	**Over 44,000**
Malta	**Over 1,400**
Total EU	**Over 680,000**

Table 9: Meta's reported quantitative information for SLI 6.2.1

[8] https://en.wikipedia.org/wiki/June_2023_Greek_legislative_election

[9] https://en.wikipedia.org/wiki/2023_Cypriot_presidential_election

Pillar-III: Political Advertising ***Commitment 10, Measure 10.1 &2*** QRE *10.2.1 page 37-38*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 10.1	2	2	2	2
Measure 10.2	2	2	2	2

Advertising transparency:

Meta's **Ad Library**[10] functions as a tool for advertising transparency, housing a searchable repository of all current ads across Meta technologies for a duration of 7 years. Users can leverage the Ad Library to search for ads (see Figure 8-10) based on various criteria, including categories such as social issues, elections, and politics (SIEP ads), specific countries, keywords, or advertiser names. Additionally, the library offers diverse filters, allowing users to narrow down searches based on factors like time periods, active or inactive status of ads, and media types featured in the ads. Notably, the library is equipped with sorting and storage functionalities for efficient management of search results. Over the last six months from January to June 2023, Meta has introduced features allowing users to save and name frequent queries for more efficient access to filtered results when logged in.

The **Ad Library API** provides programmatic access to data about ads related to social issues, elections, or politics in countries where the Ad Library is active, such as European Union nations. This API enables users to perform customised keyword searches for both active and inactive ads, offering a solution for those familiar with programmatic tools. For users less versed in API usage, a simpler research solution is provided through the Ad Library report[11], facilitating easier exploration of advertising data. Individuals within the EU with a Facebook account have access to the Ad Library API.

[10] https://www.facebook.com/ads/library/

[11] https://www.facebook.com/ads/library/report/

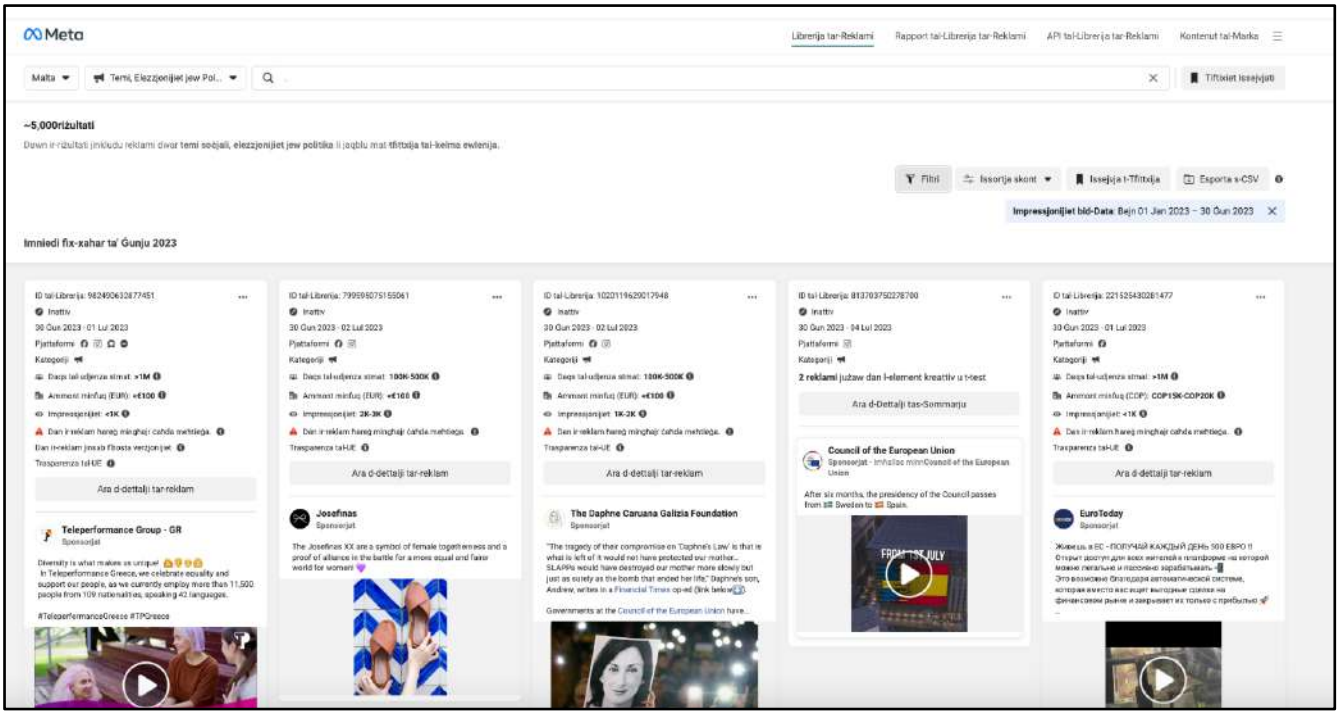


Figure 8: Meta's Ad Library snapshot when searching for ads in Malta for the period of January-June 2023

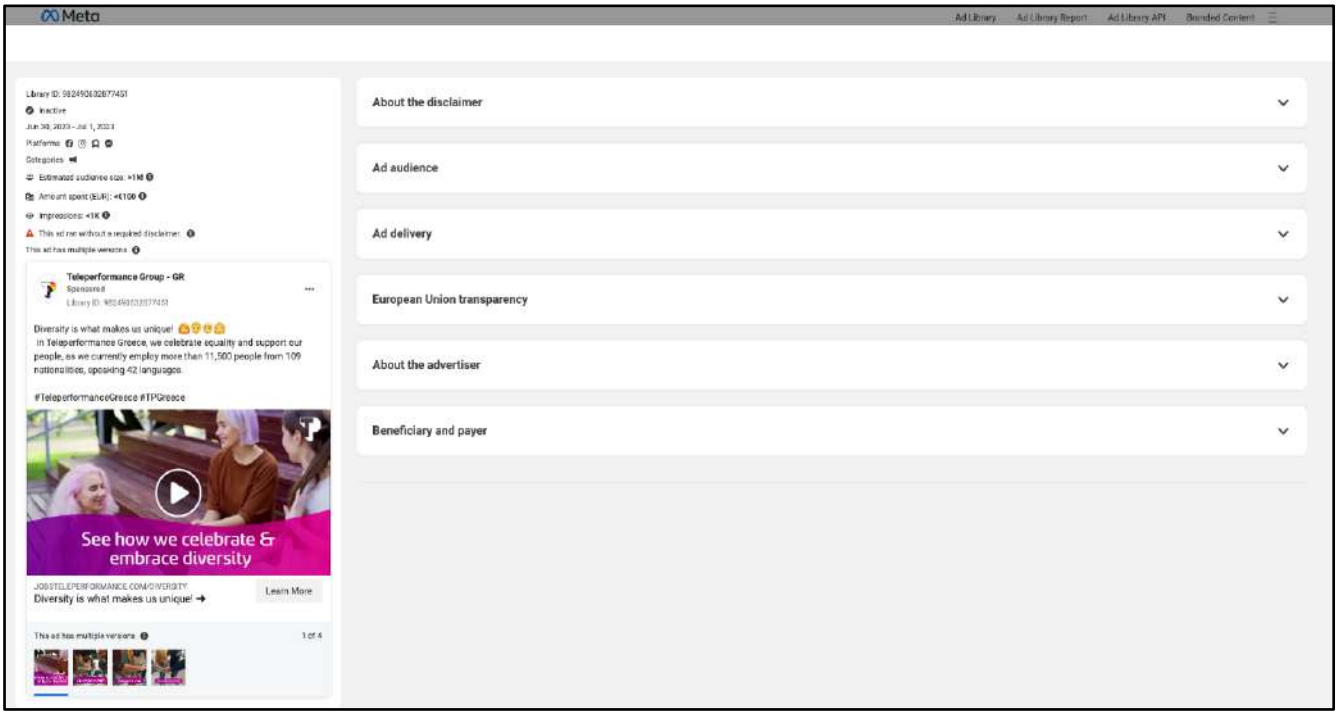


Figure 9: Meta's Ad Library example of the information given per ad

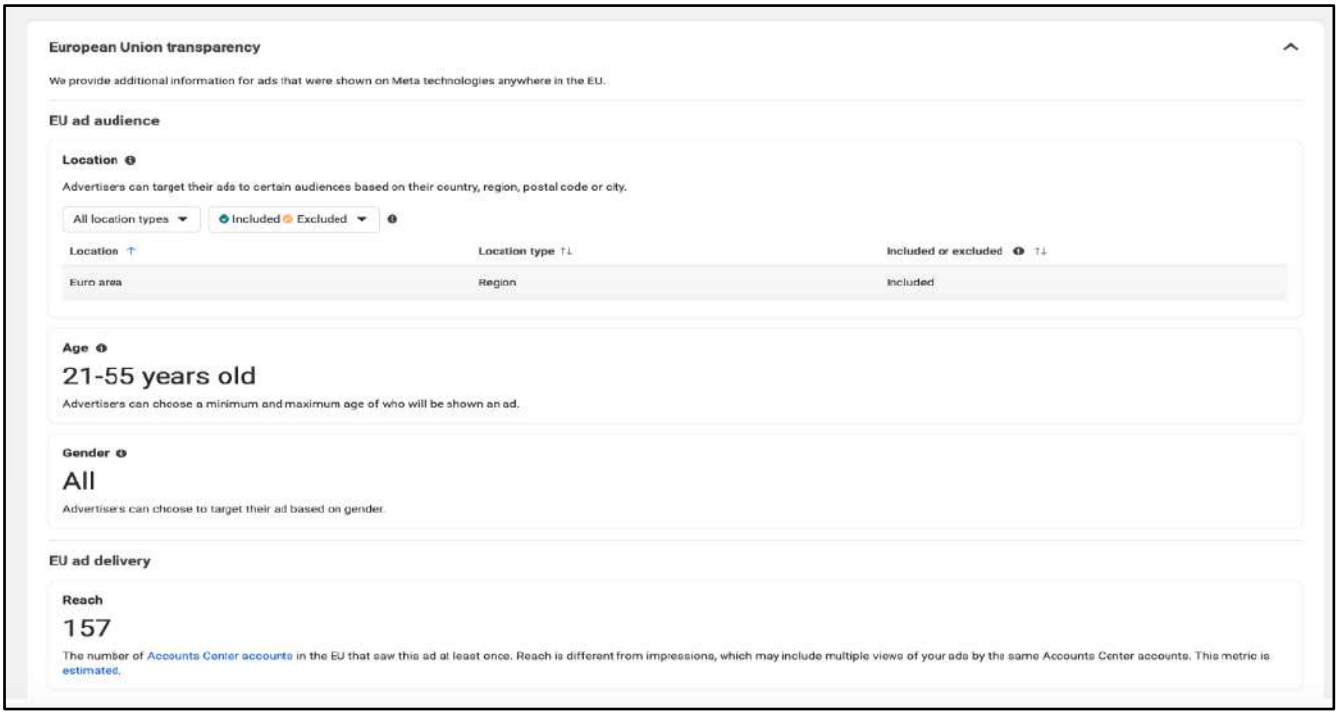


Figure 10: Meta's Ad Library example of the European Union Transparency related information per ad

Major Comments:

1. The Meta Ads Library effectively fulfils Measure 10.1&2 requirements, offering comprehensive transparency for advertising across Meta technologies. Regular updates within 24 hours ensure current information availability.
2. **Quantitative information on library users (or usage)** is not provided either at Member State level or EU level. This information is required by the Code of Practice on Disinformation in QRE 10.2.1 (for Measures 10.1 and 10.2).
3. Suggestions for improvement include implementing a filter for searching for removed ads due to policy violations. This could help researchers assess the efficiency of ad related policies and ad reviewing tools. Additionally, the Ad library should extend the results sorting options beyond impression numbers.
4. The Ad library site is accessible in **Greek and Maltese** languages as well.

3.1.3 V. Empowering Users

Pillar-V: Empowering Users ***Commitment 17, Measure 17.1*** *QRE 17.1.1, SLI 17.1.1. &* ***Measure 17.2*** *QRE 17.2.1, SLI 17.2.1* *page 59-61*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 17.1	2	2	2	1
Measure 17.2	1	1	1	1

Meta reported that over time it has created tools and resources, such as online tutorials, lesson plans for educators, tips for detecting misinformation, and awareness-raising ad campaigns. Although no further information is available such as the audience, the language, or the impact of those tools.

A fundamental element of Meta's strategy involves providing users with specific and relevant context when they encounter flagged posts. This approach aims to enhance user awareness and enable more informed decision-making regarding the content they come across, especially in situations involving sensitivity or potential misinformation. More specifically they reported the following established mechanisms:

Warning Screens (*Services: Facebook and Instagram)*

Warning screens are implemented for debunked misinformation content, potentially sensitive content, including violent imagery, posts describing bullying (shared for awareness), certain forms of nudity, and content related to suicide or suicide attempts.

Verified Badges for Authenticity (*Services: Facebook and Instagram)*

To counter impersonation and scammers posing as high-profile individuals, Meta introduces verified badges on Pages and profiles. These badges serve as confirmation of the authentic presence of public figures, celebrities, or global brands, ensuring users can trust the legitimacy of the accounts they engage with.

Notification Screens for Ageing Articles (*Services: Facebook only*)

To provide additional context before sharing news articles, a notification screen is implemented for articles exceeding 90 days. Users can still share older articles if they wish, emphasising transparency about the article's age and source. Notably, content from government health authorities and recognized global health organisations is exempt, ensuring the unimpeded dissemination of credible health information.

Meta's Media literacy campaigns reported for the period January-June 2023:

1.Misinformation Awareness Campaigns (Lithuania and Bulgaria):

Instagram campaigns in Lithuania and Bulgaria during summer 2023, such as "Facts in Focus," enlisted creators to share tips on identifying misinformation through creative storytelling. The initiative continues into the second half of 2023, and a Youth Summit in November further educates on well-being tools.

2.Slovakia pre-election media literacy campaigns:

In Slovakia, Meta's pre-election media literacy campaigns on Facebook and Instagram, including "Facts in Focus," involved local creators showing how to critically assess information. Collaborating with DigQ, a digital literacy NGO, a short video campaign runs until September 30 on 'spotting and reacting to fake news'. Ongoing efforts aim to dispel misconceptions, and a workshop with the Slovak Media Council shares media literacy best practices with local organisations.

Meta partnership with media literacy experts:

- Meta collaborates with global media literacy experts, educators, civic society, and governments for its digital citizenship initiatives.
- Partnerships include various government bodies (ministries of education, media regulators), third-party fact-checkers, parent-teacher associations, the European Association for Viewers Interests (EAVI), the UNESCO Institute for Information Technologies in Education (UNESCO IITE), Yale University, Harvard University, Micro:bit Educational Foundation, and more.
- Meta participates in the Steering Committee of the EU Digital Citizenship Working Group, contributing multidisciplinary expertise to the ongoing EU debate on digital citizenship since its launch in December 2020.

Major Comments:

1. Meta's established mechanisms (warning screens, verified badges, and ageing articles notifications) are available for users in Cyprus, Greece, and Malta. However, warning screen text is not available in the Maltese language.
2. **No media literacy campaigns** specially designed for **Cyprus, Greece, and Malta.** The criteria for determining which Member States receive media literacy campaigns on the platform are unclear. Notably, during election periods in Cyprus and Greece in 2023, where disinformation was a significant concern, campaigns akin to those conducted in Slovakia would prove beneficial, particularly during significant events. The involvement of the Slovak Media Council in the pre-election campaign **underscores the importance of national authorities actively participating and engaging in such initiatives**. Campaigns can be organised for specific Member States, or languages when the numbers show that the established mechanisms do not stop users from interacting with questionable content.
3. There is no information if Meta collaborates with media literacy experts in Cyprus, Greece and Malta other than third-party fact-checking collaborations.
4. Meta did not include any quantitative information for the effectiveness of the established mechanisms and media literacy campaigns (as required in SLIs 17.1.1 and 17.2.1).

Pillar-V: Empowering Users *Commitment 18, **Measure 18.1** QRE 18.1.1, SLI 18.1.1. & **Measure 18.2** QRE 18.2.1, SLI 18.2.1* *page 62-69*				
	Evaluation of Reported Actions	**Evaluation of Implementation**		
		Greece	**Cyprus**	**Malta**
Measure 18.1	2	2	2	1
Measure 18.2	2	2	2	1

Risk mitigation systems, tools, procedures, or features:

- **Harmful Content Prevention:** Meta employs both advanced technologies and human review teams to prevent the spread of harmful content, including misinformation.
- **Content Distribution Guidelines[12]: (Facebook)**
  - The Content Distribution Guidelines are pivotal in shaping content visibility in the Feed.
  - In March 2023, Meta summarised changes to these guidelines, specifying adjustments made to types of content that receive reduced distribution.
- **Content Distribution Guidelines: (Instagram - published in May 2023)**
  Content Removal and Lowering Guidelines:
  - Removal of posts violating Community Guidelines, with predicted violations also shown lower in feed and stories.
  - Addressing content that may contravene guidelines on Hate Speech, Bullying, Adult Nudity, Violence, and Trading of regulated products.

  Fact-Checked Misinformation:
  - Commitment to reducing the spread of misinformation, lowering posts rated false by fact-checking partners to limit their visibility.
  - Consistent posting of false information may also result in lower visibility for accounts.

  Imminent Violence or Physical Harm:
  - Removal of misinformation and unverifiable rumours under Community Guidelines if likely to contribute to an imminent risk of violence or physical harm.
  - User-Reported Content: Consideration of content based on user reports, influencing its visibility on the platform.

Meta has released system cards for both Facebook[13] and Instagram[14] [15], to provide users with accessible insights into how AI shapes their product experiences. These cards detail how AI systems rank content, make predictions, and allow users to customise their experience on platforms such as Feed, Stories, and Reels. The cards cover connected content from followed accounts and recommendations for unconnected content. In addition to the

[12] https://transparency.fb.com/policies/improving/prioritizing-content-review/
[13] https://transparency.fb.com/en-gb/features/ranking-and-content/
[14] https://ai.meta.com/tools/system-cards/instagram-feed-ranking/
[15] https://ai.meta.com/blog/ai-unconnected-content-recommendations-facebook-instagram/

system cards, Meta has shared information on the types of inputs (signals) and predictive models that inform content ranking. These signals, found in the Transparency Centre, represent the majority of those used in the overall ranking process. Meta also uses signals to identify and address harmful or low-quality content in alignment with their Content Distribution Guidelines.

Meta's General Approach to Misinformation:

- Removal of content likely to contribute to imminent physical harm, interference with political processes, and highly deceptive manipulated media.
- Focus on reducing prevalence and fostering productive dialogue for other forms of misinformation.
    - misinformation labels, fact-checking warnings, giving context to flagged content (fact-checking articles). See examples of the misinformation warning labels and giving context to the user in Figures 11-15.
- Collaboration with third-party fact-checking organisations.
- Promotion of media and digital literacy resources for users.

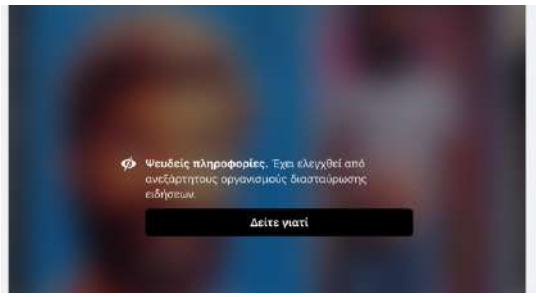


Figure 11: Example of Fact-checked post that contains misinformation warning (on Facebook in Greek).

Figure 12: More details on the misinformation warning, with option to view the fact-check article and to anyway view the post (on Facebook in Greek.).

Figure 13: View fact-check option – list the fact-checking articles related to the post with links to them and fact-checking organisation name (on Facebook in Greek).

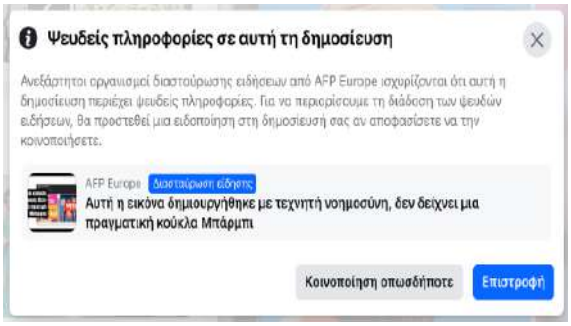


Figure 14: When user clicks to “Share” a post that contains misinformation (on Facebook in Greek).

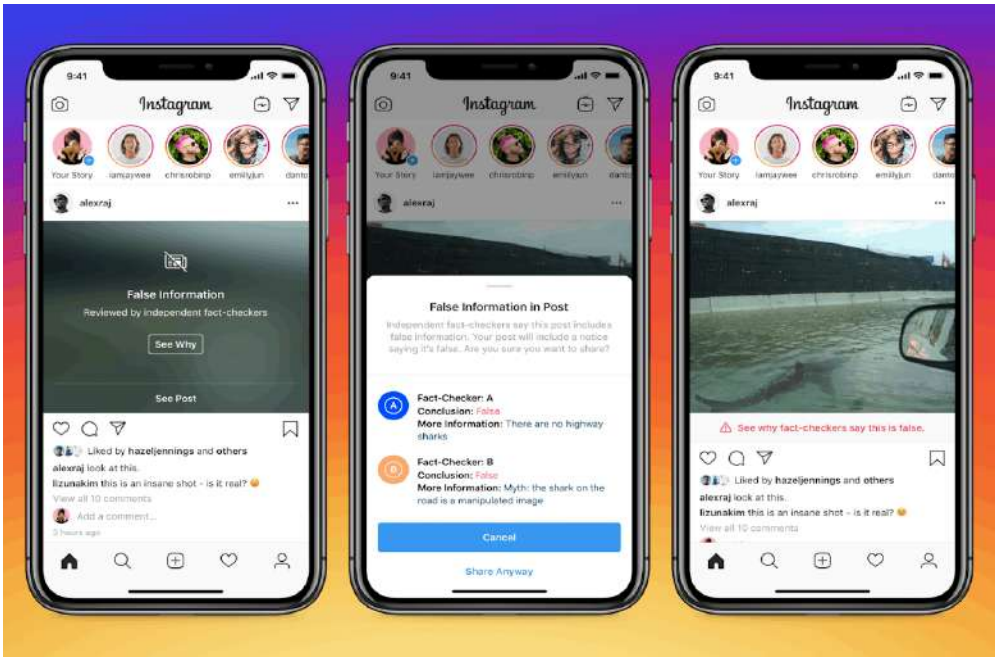


Figure 15: Misinformation warning screens on Instagram in English. Source: https://petapixel.com/2019/12/17/instagrams-new-false-information-warning-will-tell-you-if-a-photo-is-fake/

Meta reported user behaviour data on Fact-Checking Warning Screens: 95% of users scrolling through Feeds don't click to view content with fact-checking warnings on Facebook and Instagram. On average, **37% and 38% of users intending to share fact-checked** content **do not proceed** after receiving a warning **from Facebook and Instagram** accordingly.

Tackling Misinformation:

In response to Measure 18.2, **Meta** outlines its strategy, including transparent publication of policies and Community Standards and Content Distribution Guidelines, specific actions against repeat offenders, and a tiered approach to account restrictions for persistent violations. These policies extend uniformly across all EU Member States.

Actions Against Repeat Offenders:

- Pages, groups, accounts, and domains repeatedly sharing or publishing False or Altered content face reduced distribution.

Account Restriction Policy (as of February 2023):

- First strike: Warning with no further restrictions.
- Subsequent strikes lead to escalating restrictions, ranging from limited feature access to extended content creation bans.
- Severity-based additional restrictions for severe policy violations.
- Persistent violations may result in account disabling.

**** Note that while Meta counts strikes on both Facebook and Instagram, these restrictions only apply to Facebook accounts.***

Applicability:

- Restrictions primarily apply to Facebook accounts.
- May extend to Pages representing individuals (e.g., celebrities, political figures).

Major Comments:

1. In the case of fact-checked content, the user will also get access to the fact-checked article related to the content they consume and at the same time if the user tries to share this content, they receive a warning as a reminder that this information is false. The warnings and context information are available **in Greek and English, but not in Maltese.**
2. The warnings on fact-checked content on Meta appears to have a good impact on users not sharing misinformation content in the three countries (see Table 10: Meta's reported quantitative information for SLI 18.1.1 below - higher than 38% of shares is not completed by users when there is a fact-checking warning i**n Malta, Cyprus, and Greece,** while in average for the EU, 37% of shares are not completed.)
3. Regarding the **number of removed content as a result of violating our harmful health misinformation or voter or census interference policies in the EU**, currently there is no way to assess the reported numbers. The number of removed content (see Table 11: Meta's reported quantitative information for SLI 18.2.1 below) **for Cyprus, Greece, and Malta (less than 500)** is the same with multiple other EU member states. Someone would expect that countries with comparably different population size – active Meta accounts have also higher or lower volume of removed content based on the policy.
4. Meta reports **no information on the number of users/accounts** that were restricted because of repetitively posting misinformative content.

	Facebook	Instagram
SLI 18.1.1	**Rate of reshare non-completion among the unique attempts by users to reshare a content on Facebook to feed/groups that is treated with a fact-checking label in EU member state countries from 05/01/2023 to 30/06/2023.**	**Rate of reshare non-completion among the unique attempts by users to reshare a content on Instagram to feed/groups that is treated with a fact-checking label in EU member state countries from 05/01/2023 to 30/06/2023.**
	% of reshares attempted that were not completed on treated content on Facebook between 05/01/2023 to 30/06/2023.	**% of reshares attempted that were not completed on treated content on Instagram between 05/01/2023 to 30/06/2023.**
Cyprus	**43%**	**43%**
Greece	**46%**	**44%**
Malta	**50%**	**38%**
Total EU	**37%**	**38%**

Table 10: Meta's reported quantitative information for SLI 18.1.1

	Facebook	Instagram
SLI 18.2.1	**Number of unique contents that were removed from Facebook for violating our harmful health misinformation or voter or census interference policies in EU member state countries from 01/01/2023 to 30/06/2023.**	**Number of unique contents that were removed from Instagram for violating our harmful health misinformation or voter or census interference policies in EU member state countries from 01/01/2023 to 30/06/2023.**
Cyprus	**Less than 500**	**Less than 500**
Greece	**Less than 500**	**Less than 500**
Malta	**Less than 500**	**Less than 500**
Total EU	**Over 140,000**	**Over 6,900**

Table 11: Meta's reported quantitative information for SLI 18.2.1

Pillar-V: Empowering Users ***Commitment 21, Measure 21.1.*** *QRE 21.1.1, SLI 21.1.1. page 73-77*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 21.1	2	2	2	1

User benefit from independent fact-checkers:

Meta collaborates with more than **26 certified independent third-party fact-checkers**, accredited by the non-partisan International Fact-Checking Network (IFCN), covering **22 languages within the EU.**

This partnership has a global impact, with the treatment of false-rated posts, including demotion, notification, and warning, being applied globally. The Third-party Fact-checking (3PFC) program encompasses over 90 organisations across more than 60 languages globally.

Fact-checkers independently review content, assessing its accuracy using ratings such as ''***False'', ''Altered'', ''Partly false'', ''Missing context'', ''Satire'', and ''True''.***

Additional information on these ratings is available on Meta Transparency Centre[16].

Upon **receiving ratings from fact-checkers,** Meta takes action by:

(1) labelling the content, (2) reducing its visibility, and (3) implementing sanctions for repeat offenders.

[16] https://transparency.fb.com/en-gb/features/content-ratings-fact-checkers-use/

Major Comments:

1. Meta's 3PFC programme operates to label content on Instagram and Facebook. For **Malta and the Maltese language no fact-checking organisation is assigned, while for Cyprus and Greece there is.**
2. Meta's report provides the **global count of fact-checking articles (see** Table 12: Meta's reported quantitative information for SLI 21.1.2 **below)**, lacking specific data at the Member State or language level. While it's challenging to assess these numbers, it's noteworthy that the quantity of articles used to rate content on **Instagram** is **3.7 times less** than that on **Facebook**. The lower number of fact-checking articles used for rating content on Instagram compared to Facebook may be attributed to differences in user engagement, content sharing patterns, or the nature of the platforms. Instagram primarily focuses on visual content, such as photos and short videos, which might require different fact-checking approaches than the predominantly text-based content on Facebook. Additionally, user behaviour and the prevalence of misinformation may vary between the two platforms, influencing the need for fact-checking resources. The specific reasons for the observed difference would likely require a more detailed analysis of platform dynamics and content characteristics.
3. Likewise, the **quantity of labelled content following fact-checker ratings**, as reported at the EU level, is **36.4 times lower on Instagram compared to Facebook**.
4. Meta also discloses the **percentage of attempted reshares that were not completed on treated content for both Facebook and Instagram**, illustrating the effectiveness of their labelling. In **Malta, Cyprus, and Greece,** users r**efrained from completing reshares on treated content at a rate exceeding 38%**. This suggests that treated content plays a substantial role in curbing the dissemination of misinformation on these platforms.
5. While Meta's report highlights the efficacy of warning labels in deterring resharing of content after fact-checking, there is an opportunity for further investigation. The inclusion of additional metrics (such as number of impressions of the labels, number of clicks on fact-checking articles, etc.) and exploration of different presentation approaches could offer a more comprehensive evaluation of the effectiveness of these measures on Facebook and Instagram.
6. Meta did not report all the required numbers in SLI 21.1.1. (*total impressions of fact-checks; ratio of impressions of fact-checks to original impressions of the fact-checked content*).
7. Meta's report, in addressing QRE 21.3.1, highlights their collaboration with fact-checkers but falls short in elucidating the incorporation of user needs and current scientific evidence into the development and deployment of labelling or warning systems. A more comprehensive understanding of these aspects is crucial for evaluating the overall robustness and impact of Meta's approach.

SLI 21.1.2	**Number of Articles written by third party fact checkers to justify rating**		**Content treated with fact checks on Instagram due to violating assessment by third party fact checkers**		**% of reshares attempted that were not completed on treated content - Instagram**	
	Instagram	**Facebook**	**Instagram**	**Facebook**	**Instagram**	**Facebook**
Cyprus	**-**	**-**	**Over 37,000**	**Over 380,000**	**43%**	**43%**

Greece	-	-	**Over 72,000**	**Over 1,500,000**	**44%**	**46%**
Malta	-	-	**Over 15,000**	**Over 160,000**	**38%**	**50%**
Global	**52,000**	**190,000**				
Total EU			**Over 1,100,00**	**Over 40,000,000**	**38%**	**37%**

Table 12: Meta's reported quantitative information for SLI 21.1.2

Pillar-V: Empowering Users ***Commitment 24, Measure 24.1.1*** *QRE 24.1.1, SLI 24.1.1.* *page 80-83*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 24.1	2	1	1	1

Appeals systems:

For violations of Community Standards: If a post is removed on Facebook/Instagram due to policy violations, the poster is notified. They can either accept the decision or disagree, opting for a further review. **In cases of fact-checked content**, an appealing process is available for a user to ask for a correction or dispute a fact-checking rating. Users have the option to appeal a fact-check rating provided by a third-party fact-checker or identified by Meta's technology. This can be done within the product or by contacting the third-party fact-checking organisation directly via email. The fact-checkers assess the accuracy of each correction.

Service: Facebook

Transparency: Facebook's Account Status feature empowers users with restricted accounts by providing detailed insights into the nature, timing, and duration of imposed restrictions. This encompasses Profile, Page, Group, and Recommendation interfaces. Users gain a centralised resource for a comprehensive overview of their violation history and their account's compliance with policies. This includes potential restrictions on personal profiles or managed Pages, along with guidance on the appeal process.

Notification System:

- **Proactive Notification**: A novel pop-up notification feature has been introduced on Facebook to alert users if the content they are about to post might infringe upon our Community Standards. This enables users to opt for post deletion before publication.
- **Post-violation** pop-ups have been implemented to notify users about the removal of their content, ensuring clarity on the reasons behind the action.

- Over the years, Facebook introduced friction in its products to provide users with additional context for making informed decisions about what to read, trust, and share. For instance, pop-up notifications are now in place to alert users attempting to follow or share content from Pages, groups, or accounts known for disseminating misinformation.

Service: Instagram

Transparency: In June 2023, Instagram enhanced Account Status, providing users with more policy-related details and potential account impacts. Beyond assessing content eligibility for recommendations in Explore, Reels, and Feed Recommendations, users can now also check if their account is recommended in Search or as a suggested account under "Accounts You May Follow."

Major Comments:

1. Meta offers users an appeals and notifications procedure for instances where their content is removed or treated due to misinformation.
2. Related information in the Transparency centre is **available in Greek but not Maltese.**
3. Meta did not report the number of appeals, successful appeals and other metrics (described in SLI 24.1.1) for the swiftness of the appeals reviewing procedure neither for Facebook nor Instagram. Instead, Meta reported the same numbers *(Number **of unique contents that were removed from Facebook for violating our harmful health misinformation or voter or census interference policies in EU member state countries***) as in SLI 18.2.1 (see Table 11) which is not relevant to the specific SLI.

3.1.4 VI. Empowering the Research Community

Pillar-VI: Empowering the Research Community ***Commitment 26, Measure 26.2.1*** *QRE 26.2.1, SLI 26.2.1.* *page 92-94*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 26.2	2	2	2	2

Meta's tools and processes to provide access to data for research purposes:

Meta Content Library and API[17]:

- Rolled out in June 2023.

[17] https://developers.facebook.com/docs/content-library-and-api

- Includes data from public posts, pages, groups, and events on Facebook.
- Enables searching, exploration, and filtering through a graphical user interface or a programmatic API.
- **Content Library API features:**
    - Searching and filtering with sorting options.
    - Multimedia exploration for photos, videos, and reels.
    - Customizable producer lists for refining search results.
    - API code generation in Python or R.
    - Designed for computational researchers familiar with R or Python.
    - Developer Documentation and technical guides are available, and a specific help centre[18] to support the API users
    - **Eligibility and Application:**
        - Open to researchers from qualified academic and research institutions.
        - Applicants focused on scientific or public interest research topics.
        - Apply for access [19]through partners with expertise in secure data sharing, such as the University of Michigan's Inter-university Consortium for Political and Social Research.

Ad Library Tools:

- Dedicated Ad Library website and API.
- Enables searching through all currently active ads across Meta technologies.
- Provides comprehensive information on ad content, start date, advertiser details.
- Additional transparency for EU ads active within the past year.
- Displays spend, reach, and funding entity information for social issues, elections, or political ads in the last seven years.
- Ad library is open to public
- All Facebook users can access the Ad Library API

Meta Research Support[20] for Academics and Independent Researchers:

Meta has a dedicated team focused on providing academics and independent researchers with the necessary tools and data to analyse Meta's impact globally.

Meta Datasets Available for Independent Researchers:

Meta offers various data sets for independent researchers, and access opportunities are centralized and logged for easy reference. **Key datasets include:**

- **Ad Targeting Data Set:** Provides detailed targeting information for social issue, electoral, and political ads globally since August 2020. Over 70 researchers globally have access to the Ads Targeting API since its public launch in September 2022.
- **URL Shares Data Set:** Offers differentially private individual-level counts of interactions with URLs on Facebook from January 2017 to September 2022. Access is granted by Social Science One, and more than 250 researchers globally have access since its release in February 2020.
- **Research Platform for CIB Network Disruptions**

[18] https://developers.facebook.com/docs/content-library-api/get-help

[19] https://developers.facebook.com/docs/content-library-api/get-access

[20] https://research.facebook.com/

- **CrowdTangle[21]:**
  - Content discovery and social monitoring platform.
  - Provides access to a subset of public data on Facebook and Instagram.
  - Offers engagement metrics and analytics for public pages, groups, and verified profiles.
- **Data for Good:** Offers dashboards for easier understanding of Meta's data, enhancing accessibility for researchers.

Major Comments:

1. Meta tools to provide access to researchers to its data are presented in their report with reference links to the relevant websites to get more details which appear to be of good quality.
2. Meta's provided information does not include specific metrics for the uptake, swiftness, or acceptance level of the tools and processes outlined in Measure 26.2 (i.e., SLI 26.2.1). Metrics related to the number of monthly users, application statistics, and average response time are not provided for the reporting period.

3.1.5 VII. Empowering the fact-checking community

Pillar-VII: Empowering the fact-checking community ***Commitment 31, Measure 31.1*** *QRE 31.1.1, SLI 31.1.1. & **Measure 31.2.** QRE 31.2.1, SLI 31.2.1* *page 105-109*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 31.1	2	2	2	1
Measure 31.2	2	2	2	1

Meta's Third-Party Fact-Checking Program

Meta's fact-checking program[22] is a crucial initiative designed to combat misinformation across Facebook, Instagram, and WhatsApp. In collaboration with independent third-party fact-checkers certified by the International Fact-Checking Network (IFCN), Meta has built a global network of over 90 organisations in 60 languages. This network focuses on debunking viral misinformation, particularly hoaxes lacking factual basis.

[21] CrowdTangle will no longer be available after August 2024, as announced at the following link: https://help.crowdtangle.com/en/articles/9014544-important-update-to-crowdtangle-march-2024.

[22] https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking

The fact-checking process involves three main stages. First, fact-checkers identify potential hoaxes through their reporting and Meta's technology, which flags posts likely to contain misinformation. Second, fact-checkers review and rate the accuracy of stories through original reporting, consulting primary sources, public data, and media analyses. Importantly, fact-checkers do not remove content; removal only occurs when content violates Community Standards.

Upon rating content as false, Meta significantly reduces its distribution, notifies previous sharers about the misinformation, and applies a warning label linking to the fact-checker's article. The platform takes action against accounts that consistently share misinformation, treating them as repeat offenders. This approach is triggered by content flagged by third-party fact-checkers (3PFC), indicating that repeated sharing of misinformation can lead to consequences for the account involved.

In the context of rating AI-generated content, Meta's fact-checking program policies extend to cover such content. Fact-checkers tasked with assessing AI-generated media leverage expertise in artificial intelligence, visual analysis techniques, and metadata analysis to effectively identify and evaluate the accuracy of this type of content. AI is utilised to scale the fact-checking process by applying warning labels to duplicate false claims reducing their distribution.

This fact-checking program is part of Meta's broader three-part approach to problematic content. Content violating Community Standards and Ads policies, such as hate speech, fake accounts, and terrorist content, is promptly removed for safety, authenticity, privacy, and dignity. When identified by fact-checkers, the distribution of misinformation is reduced within Feed and other surfaces, striking a balance between enabling user expression and promoting authenticity. Strong warning labels and notifications are applied to fact-checked content, allowing users to see conclusions from fact-checkers and make informed decisions about what to read, trust, and share.

Major Comments:

1. **Meta has agreements with fact-checking organisations in Cyprus and Greece. Agence France-Presse (AFP) fact-checking organisation** covers **Greece** and **Cyprus,** while **Ellinika Hoaxes** has an agreement with Meta to cover **Greece** (as reported by Meta in QRE 30.1.2.).
2. **There is no agreement with a fact-checking organisation specifically assigned for Malta or the Maltese language.**
3. As it, the numbers reported could not be assessed, however the numbers reported (**SLI 31.1.2 same with numbers reported in SLI 21.1.2 - see Table 12)** show that fact-checkers work has an impact since significant percentages of users' reshare-attempts were not completed when content is treated with a fact-checking label.
4. There is **no information** for the fact-checking articles published at a member-state or language level. Further information on the number of clicks on the fact-checking articles could help in assessing the impact of the fact-checkers work. An estimation of how many users clicked/read the fact-checking articles could give an indication of the user access to the fact-checking content and impact.
5. Further details on the overall volume of content flagged by Meta's automated mechanism as potentially containing misinformation, coupled with information on the proportion of these flagged instances that are confirmed to be disinformation, would provide valuable insights into the efficacy of the technologies employed by the platform.

6. Regarding the contextual info given in SLI 31.1.3 (see Table 13 below), having the average of monthly active users in the EU on its own is not useful for the assessment of the numbers.

SLI 31.1.3	Facebook	Instagram
	Average of monthly active users on Facebook in the European Union between 1/01/2023 and 30/06/2023	**Average of monthly active users on Instagram in the European Union between 1/01/2023 and 30/06/2023**
Total Global	258 million average monthly active users on Facebook in the European Union	257 million average monthly active users on Instagram in the European Union

Table 13: Meta's reported quantitative information for SLI 31.1.3

3.2 Google (Ads, Search, YouTube)

Our analysis on Google's practices is based on the information provided in **Google's Code of Practice Report, July 2023, No2**[23].

3.2.1 II. Scrutiny of Ad Placements

Pillar-II: Scrutiny of Ad Placements ***Commitment 1, Measure 1.1*** *QRE 1.1.1., SLI 1.1.1 & 1.1.2, page 2-8*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 1.1	2	2	2	2

Google AdSense enables online content creators to earn revenue through targeted ads, matching ads with publisher sites based on content and visitor demographics. Google has implemented the following policies to deter malicious actors from exploiting the platform's monetization features, aiming to disrupt incentives for deceptive practices among publishers and ensure a safer advertising ecosystem.

- **Unreliable and Harmful Claims:** AdSense prohibits content containing demonstrably false claims that could undermine trust in democratic processes, including health misinformation and climate change denial.

[23] https://disinfocode.eu/reports-archive/?years=2023

- **Replicated Content:** AdSense doesn't allow ads on screens with copied content unless there's additional value or commentary provided.
- **Manipulated Media:** AdSense bans content that deceives users through manipulated media concerning politics, social issues, or public concerns.
- **Dangerous or Derogatory Content:** AdSense prohibits content promoting hatred, discrimination, or harm based on various characteristics, including harassment and threats.
- **Deceptive Practices:** AdSense bars deceptive practices such as false pretences, stealing personal information, or misrepresenting identity, especially in content related to politics or social issues.
- **Shocking Content:** AdSense restricts monetization on content containing graphic violence, obscenity, or profanity.

Major Comments:

1. The policies mentioned above are available online on Google's support page. The information is also translated in Greek, but not Maltese.
2. Google reports the *Number of Actioned AdSense Pages and Domains* in SLI1.1.1 and the *Estimated Cost of Blocked Requests on Pages and Domains in SLI1.1.2* (see Table 14 below) for all the aforementioned policies combined. It would be **more insightful if Google shares the numbers per policy for better understanding and transparency.** Additionally, the total number of AdSense pages and domains per member state will give content to the numbers reported.
3. The numbers for **Cyprus** are very high, **479,632 actioned pages and 340 actioned domains.** Someone would expect that based on the population of the country the numbers for bigger countries i.e., Greece will be higher. However, that requires further investigation of these numbers, which at the moment is not possible. The numbers may indicate that the ads traffic in Cyprus is higher or that ads are more prone to violate those policies in cases where the payment country is Cyprus.
4. For **Greece, ~90,5K pages were actioned**: the numbers here are in line with other EU member states such as Lithuania, Denmark, Belgium, Portugal, Slovakia.
5. The number of **actioned pages for Malta is the second lowest with 689** – after Liechtenstein with 4.
6. Google in SLI 1.1.2, reported the estimated cost of blocking ads requests on pages, and domains. The numbers looked to be in line with the number of action pages, and domains in each country, meaning that higher numbers in SLI1.1.1. are expected to reflect higher cost in SLI 1.1.2. However, when comparing **Greece (€137,394.51)** to **Cyprus (€137,394.51)**, a higher number would be expected for the estimated cost in Cyprus, since the number of actioned pages is 5 times higher while the estimated cost is only 2.6 times higher. This could indicate differences of ads pricing in the specific countries. It is hard to assess these numbers. For **Malta,** the estimated costs are very low (**€2,701.51**) as expected from the numbers reported in SLI 1.1.1.

Google - Advertising		
SLI 1.1.1-2	**SLI 1.1.1 (page 4-5)**	**SLI 1.1.2 (page 6-8)**

	Google reported the AdSense Pages and Domains that were actioned for any of the policy topics (**QRE 1.1.1**) in scope for reporting by EEA Member State **payment countries** in the first half of 2023 (1 January 2023 to 30 June 2023).		**Google determined the financial value per EU Member State by combining internal data on blocked AdSense bids with an estimate of Cost Per Thousand Impressions (CPM) for Display Ads provided by a third party, Ebiquity. This value represents an unrealized monetary value for the first half of 2023 (from January 1st, 2023, to June 30th, 2023).**	
	Number of Actioned AdSense Pages	**Number of Actioned AdSense Domains**	**Estimated Cost of Blocked Requests on Pages**	**Estimated Cost of Blocked Requests on Domains**
Cyprus	**479,632**	**340**	**€361,624.43**	**€447,377.11**
Greece	**90,542**	**17**	**€137,394.51**	**€3,310.73**
Malta	**689**	**4**	**€2,701.51**	**€68.58**
Total EU	**20,129,069**	**390**	**€29,431,265.24**	**€1,919,407.95**

Table 14: Google's reported quantitative information for SLIs 1.1.1 and 1.1.2

Pillar-II: Scrutiny of Ad Placements ***Commitment 2, Measure 2.1*** *QRE 2.1.1., SLI 2.1.1, page 12-17*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 2.1	2	1	1	1

Google reported the following policies:
Destination Requirements Policies: (Insufficient Original Content)
Google Ads enforces strict destination requirements to ensure a positive user experience and combat spam tactics used by malicious actors. These policies mandate that ad destinations must offer unique value, be functional, easy to navigate, and adhere to the Better Ads Standards. Additionally, Google Ads prohibits destinations with insufficient original content, inaccurate redirects, non-functional or inaccessible destinations, unacceptable URLs, and unverified phone numbers.

Inappropriate Content Policies: (Dangerous or Derogatory Content, Shocking Content, Sensitive Events)
To maintain diversity and respect, Google Ads prohibits ads containing inappropriate content such as dangerous or derogatory content, shocking content, and content potentially profiting from sensitive events. This includes ads featuring hacked political materials, ensuring that the platform is not used to disseminate unauthorised or misleading content.

Misrepresentation Policies: (Unacceptable Business Practices, Coordinated Deceptive Practices, Misleading Representation, Manipulated Media, Unreliable Claims, Misleading Ad Design, Clickbait Ads, Unclear Relevance, Unavailable Offers, Dishonest Pricing Practices)

Through its Misrepresentation Policy, Google Ads prohibits ads or destinations that deceive users by excluding relevant product information or providing misleading information about products, services, or businesses. This includes ads with unacceptable business practices, coordinated deceptive practices, misleading representation, manipulated media, unreliable claims, misleading ad design, clickbait tactics, unclear relevance, unavailable offers, and dishonest pricing practices. These policies aim to maintain transparency and integrity in advertising, benefiting both users and advertisers.

Major Comments:

1. Google reports the implemented policies together with some qualitative data. Google indeed applies efforts in line with the CoP and the implemented policies are sufficient and in the right direction.
2. That said, the reported SLIs (see Table 15) are not adequate for a quantitative verification of the reported policies and actions. The reported SLIs are not rich enough for a comparative analysis between countries, especially in the presence of the 2023 national elections in Greece and Cyprus. The major lack of information is that Google does not report the overall advertising traffic i.e. the total number of AdSense pages and domains that have been advertised per country. This lack of information does not allow a conclusive remark on the overall effectiveness of the implemented policies.

Google Advertising			
SLI 2.1.1 page 15-17	**Creatives that were actioned for any of the policy topics in scope for reporting, by EEA Member State billing country and policy in H1 2023 (1 January 2023 to 30 June 2023).** **To ensure a safe and positive experience for users, Google requires that advertisers comply with all applicable laws and regulations in addition to the Google Ads policies. Ads, assets, destinations, and other content that violate Google Ads policies can be blocked on the Google Ads platform and associated networks.** **Ad or asset disapproval - Ads and assets that do not follow Google Ads policies will be disapproved. A disapproved ad will not be able to run until the policy violation is fixed and the ad is reviewed.** **Account suspension - Google Ads Accounts may be suspended if Google detects violations of its policies or the Terms and Conditions.** **Policies in scope: Destination Requirements (Insufficient Original Content); Inappropriate Content (Dangerous or Derogatory Content, Shocking Content, Sensitive Events); Misrepresentation (Unacceptable Business Practices, Coordinated Deceptive Practices, Misleading Representation, Manipulated Media, Unreliable Claims, Misleading Ad Design, Clickbait Ads, Unclear Relevance, Unavailable Offers, Dishonest Pricing Practices).**		
	Number of Creatives actioned for:		
	Destination Requirements Creative	**Inappropriate Content Creative**	**Misrepresentation Creative**
Cyprus	**11,774,216**	**2,725,767**	**360,511**
Greece	**1,365,842**	**1,176**	**129,651**
Malta	**1,753,214**	**23,366**	**223,834**

Total EU	**589,714,971**	**4,200,862**	**9,986,202**

Table 15: Google's reported quantitative information for SLI 2.1.1

3.2.2 III. Political Advertising

Pillar-III: Political Advertising *Commitment 6, **Measure 6.2** QRE 6.2.1-page 29-31*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 6.2	2	2	2	2

Pillar-III: Political Advertising *Commitment 10, **Measure 10.1 &2** QRE 10.2.1 page 43-45*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 10.1	2	2	2	2
Measure 10.2	1	1	1	1

Labelling of Political or issue Ads

In summary, verified election ads in regions with verification requirements must include a disclosure of the sponsor's identity. While Google Ads/DV360 automatically generate disclosures (*"'Paid for by' disclosure"*) for most formats, some formats require advertisers to include their own disclosure. Verified ads also feature 'About This Ad' and 'Why this Ad' options for user transparency. Google enhances transparency by providing additional information about advertisers and their ads, including recent ads. This includes updates to the 'About This Ad' feature, which now includes verified advertiser name, location information, and a link to other recent ads. Additionally, the majority of EU impressions now include a '*See more ads by this advertiser*' link.

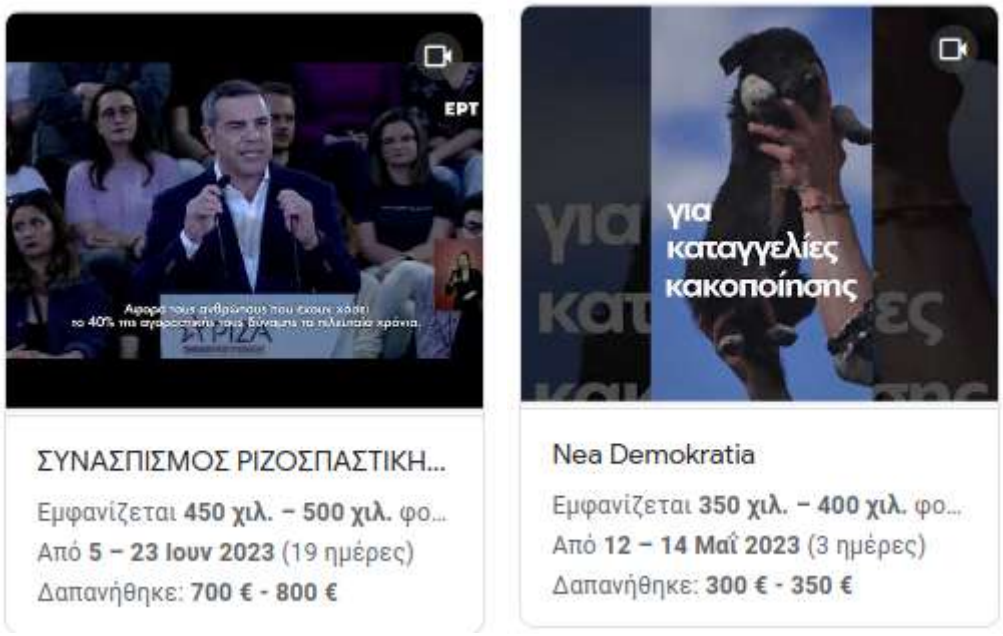


Figure 16: Examples of political or issue ads labelling can be found in Google's Ads Transparency Centre[24].

Major Comments:

1. In this pillar, Google has reported quite extensively the implemented policies. Political ads related information is publicly available [24] **with the option to translate in Greek**. There is however, **no such option** for the **Maltese language.**
2. Political ads include a 'Paid By For' disclosure. Indeed, the labelling design of political ads is clear, Google reports that the disclosures sometimes are visible after clicking the 'About this ad' button. In that case, for better accessibility, the disclosure should be visible directly in the ad.
3. For **Cyprus, Greece and Malta advertisers** are required to have verified Google accounts to run political ads, and those ads are subject to disclosure and targeting restrictions. However, there are more restrictions set for political advertisements in specific countries such as Canada, Taiwan, etc.
4. In SLI 6.2.1, Google reports the *Number of creatives from verified advertisers labelled for EU election Ads and the Amount spent by verified advertisers on Creatives labelled for EU elections ads* (see Table below). **Greece** is the country with the **third highest number of creatives for EU election ads** (after Spain and Netherlands) with **2,390 creatives**. **Malta** has **zero** and **Cyprus** has **641** creatives for EU election ads. Google did not report any numbers on political ads per member state in this SLI in the general term of political ads based on the disclaimer "paid for by".

[24] https://support.google.com/adspolicy/

5. Measure 10.2 requires the political ads to be publicly available for 5 years. Google did not report relevant information. The ads available in the repository are starting from 21/03/2019.
6. Google reported that in the period of 01/01/2023 to 30/06/2023 the Political Advertising Transparency Report had ~44,000 pageviews globally.

Google Advertising		
SLI 6.2.1 page 30-31	**(1) Creatives belonging to Google Ads/DV360 accounts that have completed the verification process for EU Election Ads and that were labelled as EU Election Ads, by EU Member State billing country in H1 2023 (1 January 2023 to 30 June 2023);** **(2) Amounts spent related to those ads in EUR, by EU Member State serving country in H1 2023.**	
	Number of Creatives from verified advertisers labelled for EU Election Ads	**Amount spent by verified advertisers on Creatives labelled for EU Election Ads**
Cyprus	**641**	**€45,077.36**
Greece	**2,390**	**€924,993.06**
Malta	**0**	**€329.72**
Total EU	**20,441**	**€4,411,563.81**

Table 16: Google's reported quantitative information for SLI 6.2.1

3.2.3 V. Empowering Users

Pillar-V: Empowering Users *Commitment 17, **Measure 17.1** QRE 17.1.1, SLI 17.1.1. & **Measure 17.2** QRE 17.2.1, SLI 17.2.1 page 103-113*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 17.1	2	2	2	2
Measure 17.2	2	1	1	1

Media Literacy Tools:

Google Search:

Google Search provides users the following tools to support their ability to evaluate the search results:

(1) **'About This Result:'**[25] which offers additional context about search results, including Wikipedia descriptions, HTTPS security status, and indexing dates.

(2) **'More About This Page:'**[26] link further allows users to access information about the source and topic, including self-descriptions, external opinions, and related sources.

(3) **Content Advisory Notices:** These are used to inform users when information is scarce or evolving rapidly, helping them navigate data voids and unreliable information.

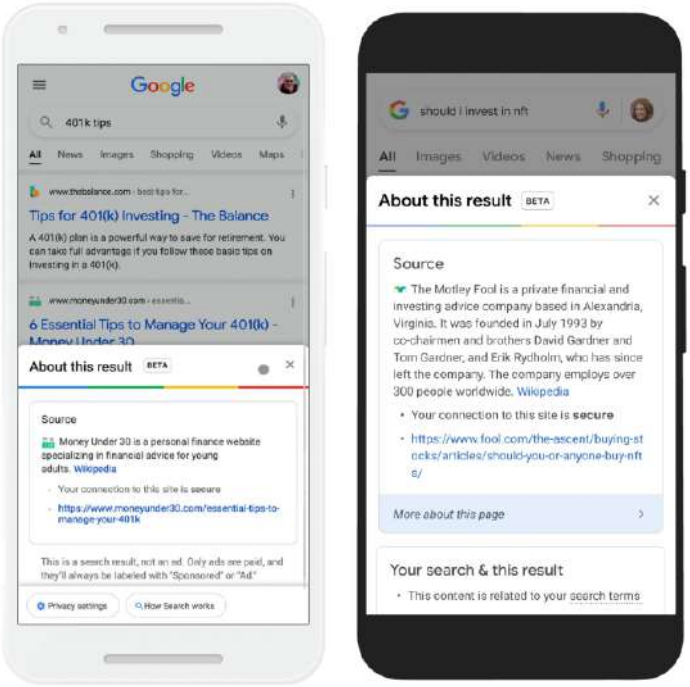


Figure 17: Google's "About this result" examples

YouTube:

YouTube offers policies and tools for users towards responsible content moderation and media literacy. Authoritative sources are prominently featured on the **platform's homepage and in search results,** with **information panels** providing additional context to help users evaluate content. For instance, during developing

[25] https://blog.google/products/search/about-search-results/

[26] https://blog.google/products/search/evaluating-information-online-tools/

news events or election periods, YouTube displays information panels linking to authoritative sources for users to access reliable information.

Media Literacy Campaigns:

In February 2023, Jigsaw revealed the results[27] of a significant **prebunking experiment on social media**, which ran from September 2022 to January 2023 and reached nearly a third of the populations in **Poland, Czech Republic, and Slovakia**, garnering over 38 million views. The experiment aimed to help individuals identify common strategies used to spread false claims online and in this way be less prone to manipulation. The success led Google to expand the initiative to **Germany** with Moonshot and local NGOs. Additionally, Google funded the **'BuloBús' project**[28] in Spain, aiming to improve media literacy by travelling to 20 towns to provide citizens with tools to combat misinformation. Moreover, grants were allocated to Facts Matter in 2023 for a study on a 'harms-framework' around misinformation, involving Google and YouTube.

Google Search:
Google Search collaborates with information literacy experts to design tools that empower users to feel confident and in control of their information consumption. Partnerships with organisations like the European Media & Information Fund contribute to this effort. Additionally, Google Search invests in building the capacity of librarians through programs like 'Super Searchers', aiming to enhance information literacy among the general public. Training sessions have been conducted in several countries, including Ireland, Italy, Portugal, and the UK, in partnership with 'Public Libraries 2030'. Since the launch of the Super Searchers Program, there have been training sessions in Portugal (12 library staff trained), Italy (30 library staff trained), and three sessions in Ireland (totalling 150 library staff trained).

YouTube:
The **'Hit Pause' campaign,** launched in 2022 and now **live in all EEA Member States**, delivers engaging public service announcements and educational content via YouTube channels and ads. This initiative, led by the YouTube Trust & Safety team, educates users on identifying manipulation tactics and safeguards against misinformation.

Major Comments:

1. **Google Search** tools to provide further information on the search results are available in the **three countries and in the Greek and partly in the Maltese (partly) languages.**
2. **YouTube's information panels are available in Cyprus, Greece and Malta in Greek, no Maltese option.** However, it is not clear if the **fact-check information panels** are available in the three countries.
3. For reported numbers for **Google Search tools in SLI 17.1.1 (see table below)**, it is not possible to assess these numbers. There is need for further information such as the total number of searches per Member State or the number of distinct users/device impressions of the specific features. Similarly, the reported numbers for YouTube's information panels require more context information to be assessed. Information regarding the fact-check information panels is

[27] https://medium.com/jigsaw/defanging-disinformations-threat-to-ukrainian-refugees-b164dbbc1c60
[28] https://bulobus.com/

missing in this report which is very important to have.

4. Google Search media literacy campaigns are of quality, but there is no effort for establishing any relevant campaign in the three countries.
5. **YouTube's Hit Pause Campaign:** videos are available with Greek subtitles but no Maltese. Some videos are translated also in other languages i.e. Spanish. The videos are of quality, with interesting content and short to attract the user's attention. The numbers reported in SLI 17.2.1 (see table below) regarding the impression number of the campaign show that there was reach of the campaign in the three countries.

Google Advertising							
SLI 17.1.1 page 107-110	**Impression proportion estimate of content advisories for:** (1) **low relevance results i**n H1 2023 (1 January 2023 to 30 June 2023), broken down by EEA Member State; (2) **rapidly changing results** in H1 2023, broken down by EEA Member State; (3) **potentially unreliable** sets of results in H1 2023, broken down by EEA Member State; *Note metrics 1-3 are estimated proportions; metric 1 represents the number of content advisories for low relevance results out of all queries over the reporting period; metric 2 and 3 follow the same logic but are for content advisories for rapidly changing results and content advisories for potentially unreliable sets of results, respectively. **Number of times the:** (4) **'More About This Page'** feature was viewed in H1 2023, broken down by EEA Member State; (5) **'Source' section of the 'About This Result'** panel was viewed in H1 2023, broken down by EEA Member State; (6) **'Your Search and this result' section of the 'About This Result'** panel was viewed in H1 2023, broken down by EEA Member State; (7) **'Personalization' section of the 'About This Result'** panel was viewed in H1 2023, broken down by EEA Member State.						
	Impression proportion estimate of content advisories for….			**Number of times the …**			
	low relevance results	**rapidly changing results**	**potentially unreliable set of results**	**'More About This Page' feature was viewed**	**'Source' section of the 'About This Result' panel was viewed**	**'Your Search and this result' section of the 'About This Result' panel was viewed**	**'Personalization' section of the 'About This Result' panel was viewed**
Cyprus	**0.169%**	**0.00079%**	**0.0000934%**	**67,234**	**718,672**	**615,818**	**421,530**
Greece	**0.128%**	**0.00062%**	**0.0000163%**	**166,446**	**5,206,194**	**4,595,204**	**4,029,026**
Malta	**0.198%**	**0.00083%**	**0.0001513%**	**46,400**	**421,500**	**364,800**	**216,300**
Total EU	**0.099%**	**0.00072%**	**0.00072%**	**28,820,492**	**365,503,914**	**313,707,910**	**185,256,434**

Table 17: Google's reported quantitative information for SLI 17.1.1

YouTube		
	SLI 17.1.1, page 107-110 Impressions of information panels (excluding fact-check panels, crisis resource panel, health information panels) in H1 2023 (1 January 2023 to 30 June 2023), broken down by EEA Member State.	**SLI 17.2.1, page 111-113 Media Literacy campaign impressions in H1 2023 (1 January 2023 to 30 June 2023), broken down by EEA Member State.**
	Impressions of information panels	**Number of impressions from YouTube's European media literacy campaign, 'Hit Pause'**
Cyprus	**1,242,349**	**476,270**
Greece	**12,079,641**	**4,922,903**
Malta	**1,020,762**	**361,286**
Total EU	**4,018,088,701**	**404,875,148**

Table 18: Google's reported quantitative information for SLI 17.1.1 and 17.2.1

Pillar-V: Empowering Users ***Commitment 18, Measure 18.1*** *QRE 18.1.1, SLI 18.1.1. &* ***Measure 18.2*** *QRE 18.2.1, SLI 18.2.1 page 114-123* **Google Search is not subscribed to Measure 18.1.*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 18.1	2	1	1	1
Measure 18.2	2	1	1	1

Risk mitigation systems, tools, procedures, or features:

YouTube's Recommendation System:

YouTube claims that it has followed the following strategies to **mitigate misinformation in recommendations** (these efforts are applied globally, including across the EU):

1. Removal of **policy-violating content,**
2. **Promotion of high-quality information in rankings and recommendations**,
3. **Recognition of trusted creators and artists**.
4. Responsible recommendations, i.e. recommendations to connect users with reliable information while reducing exposure to problematic content.

5. **Recommendations** are guided by **classifiers** that **assess the authority of videos, with input from human evaluators and certified experts such as medical professionals.** These evaluators consider various factors, including the expertise of content creators, the reputation of channels, and the main topic of videos, to determine their authority.
6. The recommendation system relies on signals like watch history, search history, and user feedback to personalise suggestions. Users have control over their recommendations and can manage their watch and search history. YouTube also limits low-quality and borderline content to maintain a responsible platform.
7. Human Evaluators assess **borderline content** (that nearly violates community guidelines) for **inaccuracies, misleading information, insensitivity, or potential harm**, which helps train YouTube systems to identify and address such content automatically.
8. YouTube does not recommend low-quality or borderline content.

YouTube Community Guidelines Enforcement:

1. YouTube issues strikes to creators whose content violates its policies. If a creator receives three strikes within 90 days, their channel may be permanently removed. Severe abuse may result in immediate termination. Additionally, YouTube may remove content for other reasons, such as privacy complaints or court orders, without issuing strikes.
2. Creators are notified via email, mobile, and desktop notifications, as well as alerts in their channel settings, if their channel receives a strike. These notifications detail the action taken and the policy violation.
3. YouTube reserves the right to restrict a creator's content creation privileges, which may include turning off certain features or prohibiting the creation of new channels to bypass restrictions. Violating this restriction constitutes circumvention under YouTube's Terms of Service, leading to the termination of existing and new channels associated with the user.

Policies:

Google Search Content policies:[29]

1. **The policies include prohibiting deceptive practices, manipulated media, and promoting transparency in news sources**.
2. These measures are enforced through spam protection tools and guidelines for search features to reduce the spread of low-quality content and protect users from deceptive practices.
3. These include the **Medical Content Policy**, which prohibits content contradicting scientific or medical consensus, and the **Misleading Content Policy**, which prevents the display of preview content that misleads users. Users are provided guidance on **reporting policy-violating content**, which Google Search removes based on user reports and internal processes.

[29] https://support.google.com/websearch/answer/10622781#zippy=%2Cspam

Major Comments:

1. YouTube recommendation and Google search policies are in the right direction on mitigating disinformation by the removal of harmful disinformation content and identifying borderline content and low-quality content to not be recommended.
2. On the other hand, no quantitative data are reported in SLI 18.1.1.
3. YouTube **number of videos removed for misinformation policy violation** (reported in SLI 18.1.2, see table below) is **very low for the three countries**. The numbers cannot be assessed.

SLI 18.2.1	YouTube	
	Videos with 101–1,000 views removed for violation of misinformation policies	Videos with >1,001 views removed for violation of misinformation policies
Cyprus	9	5
Greece	74	60
Malta	5	5
EU total	2,920	1852

Table 19: Google's reported quantitative information for SLI 18.2.1

Pillar-V: Empowering Users *Commitment 21, Measure 21.1. QRE 21.1.1, SLI 21.1.1. page 73-77*				
	Evaluation of Reported Actions (QREs)	Evaluation of Implementation (SLIs)		
		Greece	Cyprus	Malta
Measure 21.1	2	2	2	2

Google Search:

Google claims that **fact-check articles in Google Search results** play a crucial role in **providing context and information to users.** Google relies on **machine-readable ClaimReview markup** on websites to enhance these search results, making it easier for users to understand **through ''rich snippets'' (see photo * below) what is being fact-checked and the assessment by fact-checkers. The fact-checking organisation in order to use the ClaimReview markup is required to meet Google Search's eligibility and technical criteria.**

The **'Fact Check' label in Google Search** applies to **published stories with fact-checked content indicated by the schema.org ClaimReview markup.** Google Search enables any fact-checker to signal their fact-checks for

indexing by implementing this markup on their content. The use of ClaimReview markup is open to all organisations, and **specific partnerships do not apply to Google Search**.

Google offers tools such as **Fact Check Explorer**[30] and the **Google FactCheck Claim Search API**[31] to support fact-checking efforts. All the fact-check articles following the ClaimReview markup can be found in Fact Check Explorer that allows the user to search for fact-checks with a language filter available.

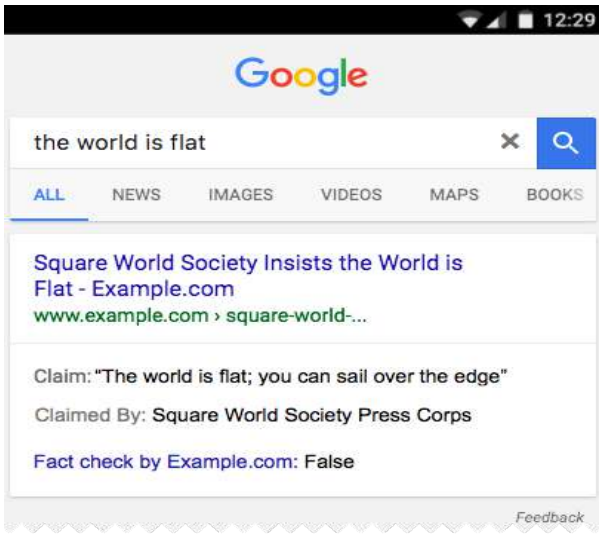


Figure 18: ClaimReview schema Googles Example from page

Figure 19: Examples of Ellinika Hoaxes and AFP using ClaimReview schema

YouTube:

Fact-checkers are able to post and share both short- and long-form video content on YouTube. Fact-check content made available on YouTube can be surfaced through relevant search results, via recommendations, or linked directly from other websites and online platforms.

Fact-check information panels are displayed above search results for relevant queries, providing context with links to third-party fact-checked articles. See **Measure 26.1 below**, for more information for the fact-check information panels.

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

[31] https://toolbox.google.com/factcheck/apis

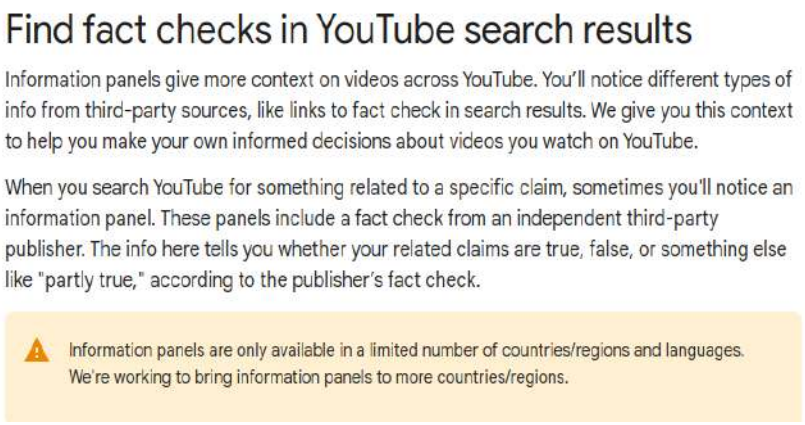


Figure 20: YouTube Information Panels description (screenshot)

Major Comments:

1. Google Search does not label results as misinformation/misleading/etc. but employs ClaimReview markup schema to enhance the information on fact-check articles in search query results. This initiative by Google is in the right direction to fulfil Measure 26.1 and is available in the three countries. Ellinika Hoaxes and AFP already follow the ClaimReview schema for their fact-checking articles.
2. In the SLI21.1.1, Google reported the number of fact-check articles available in the Fact Check Explorer at language level. For the **Maltese** language there are **zero fact-check articles**. There are nearly **2K fact-check articles in Greek. There is no information if the fact-check articles in Greek refer to events in Cyprus or Greece.**
3. YouTube **fact-check information panels** aim to provide context to users based on their search query; However, it is not clear how it is used and in which countries are available. Additionally, Google does not report any quantitative information to assess the effectiveness of this feature.

SLI 21.1.1	Google Search	
	Number of articles available in Google Search Fact Check Explorer	
	At the beginning of H1 2023	At the End of H12023
English	73,093	71,891
Greek	2,018	2,014
Maltese	0	0

Table 20: Google's reported quantitative information for 21.1.1

Pillar-V: Empowering Users ***Commitment 24, Measure 24.1.1*** *QRE 24.1.1, SLI 24.1.1. page 151-155* **Google Search has not subscribed to this Measure*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 24.1.1	2	2	2	2

YouTube Notification and Appeal Processes:

Notification:

When content on a creator's channel violates YouTube's Community Guidelines, they may receive a strike. However, content may also be removed for reasons other than guideline violations, such as privacy complaints or court orders, without resulting in a strike. If a strike is issued, the creator is notified via email, mobile and desktop notifications, and an alert in their channel settings upon signing into YouTube. The notification explains the action taken and which policy was violated.

Appeal Process:

Creators have the option to appeal strikes, video removals, age restrictions on videos, playlist or thumbnail removals, and channel terminations. The appeals process can be initiated through YouTube Studio or via email notification. After an appeal is submitted, YouTube reviews it and notifies the creator of the outcome. If content is found to comply with Community Guidelines, it may be reinstated, and the strike removed. If content is deemed inappropriate for all audiences, an age-restriction may be applied. Content found to violate guidelines will remain removed, with no additional penalty for rejected appeals.

Major Comments:

1. YouTube notifies users in case of a strike, a video removal, etc. The users can appeal in these cases by email or through YouTube specified pages. The appeal process is clearly reported. However, it is not clear how YouTube reviews the appeals for misinformation policies violation, e.g., if fact-checking organisations are involved in the process or not.
2. The number of appeals for video removal due to misinformation policy violation in **Cyprus – 3 (1 successful)** and **Malta – 6 (0 successful)** is **very low (see Table below)**. In **Greece**, there were **83 appeals for videos with 7 of those being successful.** For the numbers reported in SLI 24.1.1, it is not clear how YouTube derived the location, e.g., based on the account location.

YouTube		
SLI 24.1.1 page 154-155	**(1) Appeals following video removal for violations of YouTube's misinformation policies in H1 2023 (1 January 2023 to 30 June 2023), broken down by EEA Member State;** **(2) Video reinstatements following a successful appeal against content removals for violations of YouTube's misinformation policies in H1 2023, broken down by EEA Member State.**	
	Number of videos removed that were subsequently appealed	**Number of videos removed that were then reinstated following a creator's appeal**
Cyprus	**3**	**1**
Greece	**83**	**7**
Malta	**6**	**0**
Total EU	**3,059**	**356**

Table 21: Google's reported quantitative information for SLI 24.1.1

3.2.4 VI. Empowering the Research Community

Pillar-VI: Empowering the Research Community ***Commitment 26, Measure 26.2.1*** *QRE 26.2.1, SLI 26.2.1.* *page 163-167*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 26.2	2	2	2	2

Google provides the following tools to empower the researchers:

1) **YouTube Researcher Program:** The YouTube Researcher Program[32] provides scaled, expanded access to global video metadata across the entire public YouTube corpus via a Data API[33] for academic researchers affiliated with an accredited, higher-learning institution.
2) **Google Trends:** Google Search and YouTube provide publicly available data via Google Trends[34], which provides access to a largely unfiltered sample of actual search requests made to Google Search and YouTube's search function (anonymised data).
3) **Google Fact Check Explorer:** Fact Check Explorer and the Google FactCheck Claim Search API[35] allow

[32] https://research.youtube/

[33] https://developers.google.com/youtube/v3/getting-started/?target=_blank

[34] https://trends.google.com/trends/

[35] https://toolbox.google.com/factcheck/apis

anyone to explore the Fact Check articles that are using the ClaimReview markup.

Major Comments:

1. The report from Google presents tools aimed at granting researchers access to its data, with accompanying reference links to pertinent websites for further information, suggesting a high standard of quality.
2. Google reported the quantitative information in SLI 26.2.1 (see table below), relevant to the number of applications received for the YouTube Researcher program. Regarding the three countries, there was **only one application from Cyprus received and approved** for the reporting period. Looking into the numbers reported at EU level, YouTube received 40 applications with 25 of those being approved. 35% of the applications got rejected, which is considered to be high. The median application resolution time in the EU (10 days) is reasonable.

YouTube						
SLI 26.2.1 page 107-167	**Applications**				**Number of unique researchers accessing the API**	**Median application resolution time**
	Received	**Approved**	**Rejected**	**Under Review**		
Cyprus	**1**	**1**	**0**	**0**	**1**	**-**
Greece	**0**	**0**	**0**	**0**	**0**	**-**
Malta	**0**	**0**	**0**	**0**	**0**	**-**
Total EU	**40**	**25**	**14**	**1**	**33**	**10 days**

Table 22: Google's reported quantitative information for SLI 26.2.1

3.2.5 VII. Empowering the fact-checking community

Pillar-VII: Empowering the fact-checking community ***Commitment 31, Measure 31.1*** *QRE 31.1.1, SLI 31.1.1. &* ***Measure 31.2.*** *QRE 31.2.1, SLI 31.2.1 page 189-191*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 31.1	2	1	1	1
Measure 31.2	2	1	1	1

Google: claims that support fact-checking organisations globally and also within the EU though direct and in-direct funding and grants. Key support actions:

1. On 14 April 2023, the International Fact-Checking Network's (IFCN) Global Fact Check Fund opened its Phase 1 (Build) application process for grants, through its $13.2M partnership with Google, including YouTube.
2. 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).
3. Additionally, on 29 November 2022, Google, including 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 Search & YouTube:

1. Google's main partnerships are with the **European Media and Information Fund** and the **International Fact-Checking Network**. Both organisations provide indirect payments to fact-checking members.
2. Additional partnerships include **deCheckers, German Press Agency DPA, CTK (Czech Press Agency), and Demagog Poland**. These organisations were either provided direct grants or will provide indirect payments to fact-checking members.

3. Additional collaborations – **YouTube:**

    - The following EU based fact-checking organisations participate in the YouTube Partner Program (YPP) YouTube's monetisation program: **Observador, AFP Sprawdzam, Perikasa Fakta, Fact Check Myanmar, Faktantarkistus, AFP Checamos, Bayerischer Rundfunk, France Info, EFE Verica, The France 24 Observers**.

Policies:

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**. 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.
2. Google Search and YouTube's use of fact-checks **does not include specific actions regarding content that has been fact-checked (such as labelling it as false or removing it).** For YouTube it is clearly mentioned that harmful misinformation identification and removal is prioritised.

<u>YouTube collaboration with fact-checkers involves:</u>

1) YouTube as a platform for fact-checking organisations to integrate fact-checking content.

Fact-checkers can upload short- and long-form videos, which are surfaced through search results, recommendations, and external links. Users can subscribe to fact-checking channels to receive notifications of new content. The YouTube Studio provides tools for creators to manage their presence, interact with audiences, and access analytics. Fact-checking organisations can view data about their video performance through the Channel Analytics Dashboard. YouTube supports fact-checkers through regular meetings with EU-based organisations and provides guidance through the Creator Support teams.

2) Fact-check Information Panel:

Fact-check information panels are displayed above search results for relevant queries, providing context with links to third-party fact-checked articles. If a user's search indicates a need for accuracy information, relevant and recent fact-checks are displayed from eligible publishers. Information panels may include the **publisher's name, a link to the article, and the publication date**. YouTube aims to continue supporting and integrating fact-checker content to provide users with accurate information.

These panels rely on a network of third-party publishers adhering to **ClaimReview tagging guidelines.** YouTube works with fact-checking organisations that meet **specific eligibility criteria** to ensure the quality and reliability of fact-check content on its platform. These eligibility criteria include:

- **ClaimReview Tagging System**: Fact-checking organisations must adhere to the publicly available ClaimReview structured data guidelines. This tagging system helps to identify and categorise fact-check content on the web.
- **Membership in International Fact-Checking Network (IFCN)**: Fact-checking organisations must either be part of the International Fact-Checking Network (IFCN) or **be an authoritative publisher** recognized for their fact-checking efforts.
- Adherence to Guidelines: Publishers must follow the guidelines set by the IFCN or equivalent authoritative bodies regarding fact-checking methodologies, transparency, and accountability.

Major Comments:

1. Google Search and YouTube ***do not label or remove*** fact-checked content. They report that they remove harmful misinformation. Additionally, **a fact-check information panel** may be displayed in search results.
2. YouTube is actively encouraging fact-checking organisations to utilise its platform for disseminating their fact-checks, and it provides support by offering guidance on producing content that can enhance their reach.
3. Google **does not collaborate directly with fact-checking organisations in Cyprus, Greece, and Malta**. However, **Ellinika Hoaxes** in their response (see MedDMO Fact-checking Partners Collaboration with VLOPs section) **mentioned that their content is featured on Google Search results through ClaimReview and Fact Check Explorer. AFP** collaborates with Google for **developing media literacy training material.**
4. **YouTube** has also **established fact-check information panels** to provide users context for the videos in their search results. However, the information on their page[36] explains that the **information panels are available only in specific regions/countries and languages** without giving the list of the countries or languages covered. There is **no proof that the fact-check information panels** are available in **Cyprus, Greece, and Malta or in Greek and Maltese.** Google has not shared any numbers in SLI 31.1.1. regarding the number of distinct fact-check information panels or the impression number of the fact-check information panels.

[36] https://support.google.com/youtube/answer/9229632?hl=en&ref_topic=9257092

5. No numbers reported in SLI 31.1.1 and 2, Google's report refers to SLI 21.1.1 as a response to SLI 31.1.1.

YouTube			
SLI 21.1.1 page 154-155 (referenced as a report for SLI 31.1.1)	**Number of impressions** on Fact Check Rich Snippets, by EEA Member State	**Number of articles** available in Google Search Fact Check Explorer at the beginning of H1 2023, broken down by EEA language	**Number of articles** available in Google Search Fact Check Explorer at the end of H1 2023, broken down by EEA language
Cyprus	**287,855**	**Language: Greek: 2,018**	**Language: Greek: 2,014**
Greece	**2,399,480**		
Malta	**140,633**	**Language: Maltese: 0**	**Language: Maltese: 0**
Total EU	**104,670,544**		

Table 23: Google's reported quantitative information for SLI 31.1.1

3.3 TikTok

Our analysis on TikTok's practices is based on the information provided in **TikTok's Code of Practice Report, July 2023, No2**[37].

3.3.1 II. Scrutiny of Ad Placements

Pillar-II: Scrutiny of Ad Placements ***Commitment 1, Measure 1.1*** *QRE 1.1.1., SLI 1.1.1 & 1.1.2, page 2-7*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 1.1	2	1	N/A	N/A

Pillar-II: Scrutiny of Ad Placements ***Commitment 2, Measure 2.1*** *QRE 2.1.1., SLI 2.1.1, page 13-15*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 2.1	2	1	N/A	N/A

TikTok approach to defund disinformation dissemination:

Policies and Approach:

TikTok follows a set of policies within its **Community Guidelines (CGs)** to address harmful misinformation and deceptive behaviours on its platform. All users are required to comply with these guidelines, and **content violating them is removed**. Paid ads are also subject to strict ad policies, prohibiting misleading, inauthentic, and deceptive behaviours. Violative ads are not permitted, and repeated violations may result in **account suspension or banning.**

TikTok has expanded its ad policies[38] with the introduction of four granular policies covering **Medical Misinformation, Dangerous Misinformation, Manipulated Media, and Dangerous Conspiracy Theories.** The platform is continuously assessing and considering additional policy areas. Specific ad policies target topics with a higher risk of disinformation, such as the Covid-19 ad policy, which prohibits distasteful presentations of Covid-19. To enforce the Covid-19 policy, TikTok also promoted authoritative sources of information and also provided free ad credits to health authorities, governments, etc. to promote true information for Covid-19 related issues.

[37] https://disinfocode.eu/reports-archive/?years=2023

[38] https://ads.tiktok.com/help/article/tiktok-advertising-policies-ad-creatives-landing-page?redirected=1

TikTok is a member of the GARM Framework and offers brand safety tools to advertisers so they can choose to place their ads near content that fit their brand image.

Verification in the Context of Ads:

TikTok's ad policies require **advertisers** to meet **landing page requirements,** ensuring accuracy and completeness of information.

Verified badges are granted to certain accounts, including advertisers, to aid users in making informed choices and establishing trust. Factors considered for verification include authenticity, uniqueness, and activity of the account.

TikTok is conducting trials for mandatory **verification for government, politician, or political party accounts** in the US, with verification **already available (but not mandatory) for these accounts in the EU**. Various policies are in place to prevent misuse of features, including restrictions on access to advertising features and solicitation for campaign fundraising.

Major Comments:

1. TikTok reported specific ad policies related to disinformation and that all ads are reviewed for compliance with those policies before being published.
2. The platform provides data on ads removed due to the **Covid-19 and political ads policies** at Member State level, but for **Greece, Cyprus, and Malta** no such ads have been removed. Looking at the **available Placements and Locations**[39] based on the country/region where the ad account was registered/created there is no information about **Cyprus and Malta**. Also, there are **no ads in the TikTok ads library**[40] **when selecting Cyprus or Malta as location**. This implies that the **TikTok ads feature is not enabled for accounts registered in Cyprus and Malta**, and no ads are displayed for accounts located in the two countries.
3. The TikTok ad policies[41] are not available in the **Greek and Maltese language**.
4. TikTo**k did not report numbers** regarding **ads removed due** to violation of the policies covering **Medical Misinformation, Dangerous Misinformation, Manipulated Media, and Dangerous Conspiracy Theories.**
5. TikTok **did not report any numbers for the estimated financial value of ads removed** due to violations of ads policies.

TikTok	
SLI 1.1.1	**Methodology of data measurement:** We have set out the number of ads that have been removed from our platform for violation of our Covid-19 misinformation and political

[39] https://ads.tiktok.com/help/article/placements-available-locations?lang=en#anchor-15

[40] https://library.tiktok.com/ads

[41] https://ads.tiktok.com/help/article/tiktok-advertising-policies-ad-creatives-landing-page?lang=en

	content policies respectively. Note that numbers have only been provided for monetised markets and are based on where the ads were displayed. As mentioned above, in order to improve our existing ad policies, we have recently developed four more granular policies and as a result also expanded our existing policy coverage. As these policies were launched towards the end of the reporting period, we do not have meaningful data to share for this report, but we expect to be able to provide this data in the next report. We are pleased to be able to include the impressions data for ads removed for the below policies in this report.	
	Number of **ad removals** under the **Covid-19 misinformation ad policy**	Number of **ad removals** under the **political content ad policy**
Cyprus	**0**	**0**
Greece	**0**	**0**
Malta	**0**	**0**
Total EU	**20**	**390**

Figure 21: TikTok's reported quantitative information for SLI 1.1.1

3.3.2 III. Political Advertising

Pillar-III: Political Advertising ***Commitment 6, Measure 6.2*** *QRE 6.2.1, page 27-28*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 6.2	N/A	N/A	N/A	N/A

Labelling of Political or issue Ads:

TikTok explicitly prohibits political advertising, addressing the recognized potential for disinformation in this Code with a dedicated chapter on Political Advertising. The platform's ads policies restrict political actors from placing ads, encompassing content that references, promotes, or opposes candidates, political parties, government officials, elections, referenda, and related merchandise featuring prohibited elements. Cause-based and public service advertising is permitted if devoid of partisan political motives.

While political ads are prohibited, TikTok provides users with enhanced information through the "About this Ad" functionality for permitted ads. This feature has been refined and improved to comply with DSA Article 26(1) transparency obligations, offering details on the ad's presenter, payer, parameters used for user targeting, and guidance on adjusting those parameters.

Major Comments:

1. **TikTok prohibits political ads on its platform.**
2. However, it is important to note that videos with political content are available on TikTok. For example, in Greek elections 2023[42], many politicians used TikTok to reach younger voters. Similarly, we observe the same in Cypriot elections in 2023[43].

Pillar-III: Political Advertising ***Commitment 10, Measure 10.1 & 2,*** *QRE 10.2.1, page 33-34*				
	Evaluation of Reported Actions	**Evaluation of Implementation**		
		Greece	**Cyprus**	**Malta**
Measure 10.1	N/A	N/A	N/A	N/A
Measure 10.2	N/A	N/A	N/A	N/A

As mentioned above TikTok prohibits political advertisements, however they launched the **Commercial Content Library**[44], to comply with other DSA regulations, which is publicly available (no need for a TikTok account or other authorization). The user can search for advertisements for a selected country, and a specific time frame (TikTok ads are available from October 2022). The search can also include keywords or specific advertisers' names. TikTok library allows users to sort the search results by published date, last show date, and number of unique users seen the ads (impressions). There is no functionality to store the search results so far. For each ad TikTok provides further information such as the actual ad content, advertiser information, the publication data, last shown date, number of unique users seen, and other ad target audience information set by the advertiser such as gender, age, user's interests, video interactions, creator interactions, etc. The ads also provide information on the number of user impressions per country. The library is updated on a 24h schedule.

Figure 22 shows on the left an example of search results for ads that targeted Greece for the period of 01/01/2023-30/06/2023 and on the right the advertisement's information and metadata that is available through TikTok ads library. Figure 23 show that there are no advertisements when searching for ads targeting Cyprus and Malta.

[42] https://www.kathimerini.gr/opinion/562487494/oi-politikoi-archigoi-sto-tiktok/

[43] https://limassoltoday.com.cy/stiles/filoksenoumena/proedrikes-2023-nea-ergaleia-kai-praktik/

[44] https://library.tiktok.com/ads

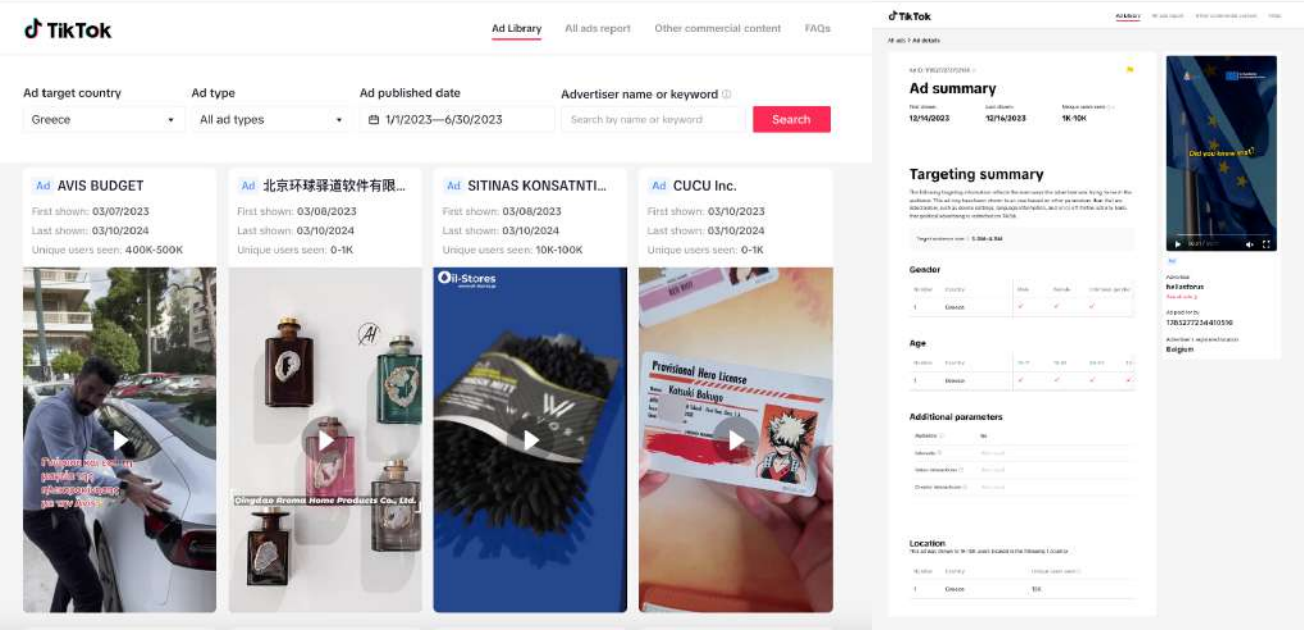


Figure 22: Screenshots of TikTok Commercial Content library when searching for ads displayed in Greece for the period of January to June 2023

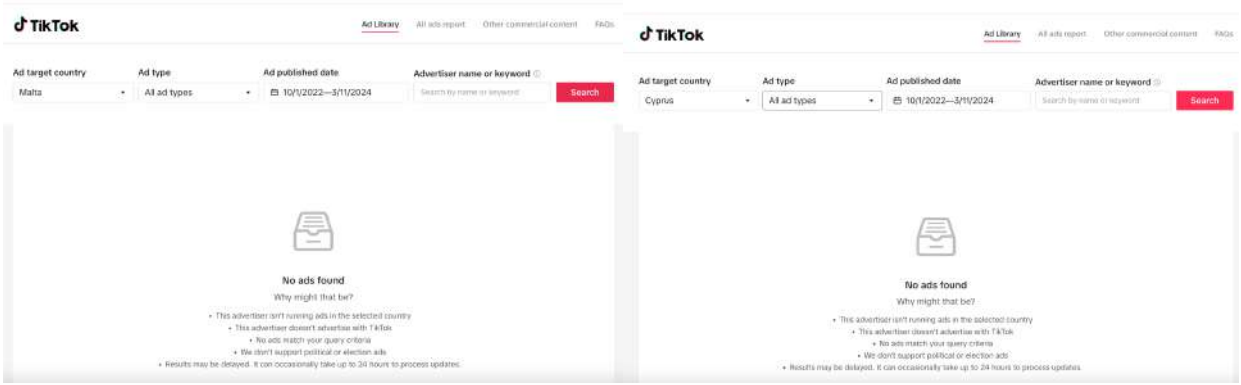


Figure 23: Screenshot of TikTok Commercial Content Library when searching for ads targeting Cyprus and Malta

Major Comments:

1. TikTok Commercial Content Library is of good quality, providing ads transparency, however since TikTok prohibits political ads there is no specific ads category for political ads.
2. **No ads are available for Cyprus, and Malta.** Searching for **Greece** as an ad target country, for the period of 01/01/2023-30/06/2023 resulted in **781,196 ads.**
3. **TikTok did not provide any numbers regarding its usage at Member State or EU level.**

3.3.3 V. Empowering the users

Pillar-V: Empowering Users ***Commitment 17, Measure 17.1*** *QRE 17.1.1, SLI 17.1.1. &* ***Measure 17.2*** *QRE 17.2.1, SLI 17.2.1* *page 77-111*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 17.1	2	2	1	1
Measure 17.2	2	2	1	1

TikTok removes content violating its policies and implements **in-app measures.** These measures aim to offer additional context or guide users to authoritative information, **available in 23 EU official languages and, for EEA users, Norwegian and Icelandic.** Collaborating with external experts and fact-checker partners, TikTok considers their feedback, along with user input, to identify new topics and deploy tools for awareness and misinformation prevention.

TikTok deployed the **following in-app intervention tools** on the topics of ***Covid-19, Covid-19 Vaccine, Holocaust Denial, Monkeypox, War in Ukraine, Climate change*** (only search intervention tool):

- **Video Notice Tags**: Applied to videos with relevant words or hashtags, these clickable tags invite users to "Learn more about [the topic]." Clicking the tag redirects the user to a trusted resource page.
- **Search Intervention**: When a user searches for keywords related to the topic, a banner may appear, encouraging them to verify facts and providing a link to a resource page. If the search term is violative, the user won't see results and will be redirected to a trusted resource page.
- **Public Service Announcement:** Searching for a hashtag on the topic displays a public service announcement reminding users of TikTok's Community Guidelines (CGs) and offering links to trusted resource pages.
- **Online and In-App Information Hubs and Safety Centre Pages:** These tools often link to resource pages, guiding users to accurate information from trusted sources. Depending on the topic or EU country, users may be directed to an external authoritative source (e.g., a national government website), an in-app information hub (e.g., War on Ukraine), or a dedicated page on the safety centre website (e.g., Covid-19 and Elections Integrity).

Additionally, TikTok **applies warning labels to content associated with unfolding or emergency events**, irrespective of the topic. These labels, accessible in 23 EU official languages (and for EEA users, Norwegian and Icelandic), are part of TikTok's efforts to prompt users to assess the reliability of content and its sources.

Unverified Content Label[45]:

- Applied to content related to unfolding or emergency events that, despite being assessed by fact-checkers, cannot be verified as accurate.
- Videos with this label become ineligible for recommendation in anyone's For You feed to curb the spread of potentially misleading information.
- Creators receive notifications, informing them that their video has been flagged as unsubstantiated content. Additional information is provided to raise awareness about the content's credibility.

2. **State Affiliated Media Label:**
   - In the EU, Iceland, and Liechtenstein, access to content from specific sources like Russia Today, Sputnik, Rossiya RTR / RTR Planeta, Rossiya 24 / Russia 24, and TV Centre International is restricted.
   - A prominent label is applied to other content or accounts from state-affiliated media.
   - Users are presented with a full-screen pop-up explaining the label's significance and inviting them to click "learn more" for redirection to an in-app page[46]. This measure aims to bring transparency to the community and raise awareness among users about the reliability of the source.

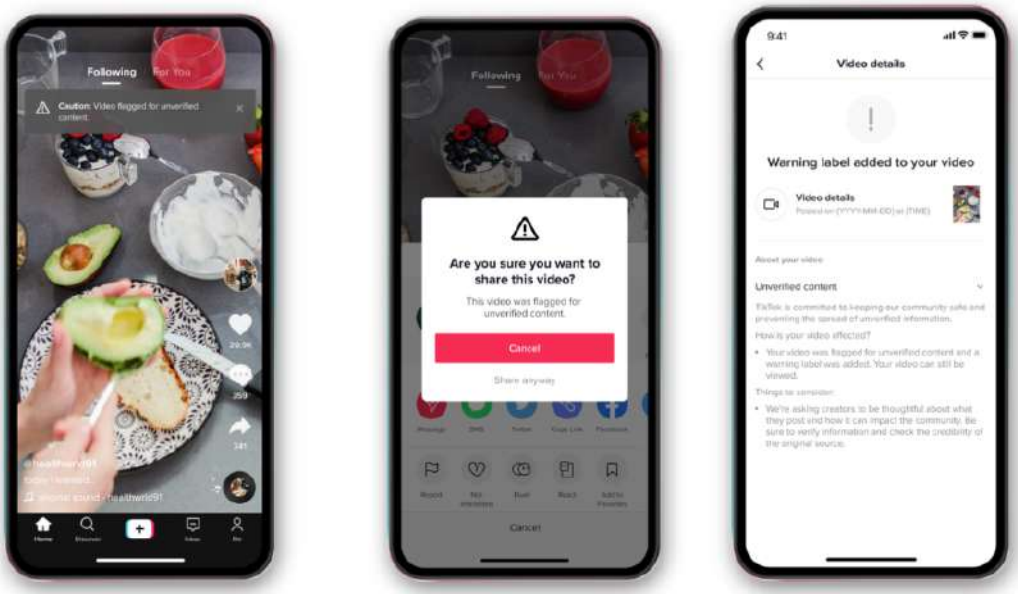


Figure 24: Example of TikTok's Unverified Content Label taken from TikTok CoP report, No2

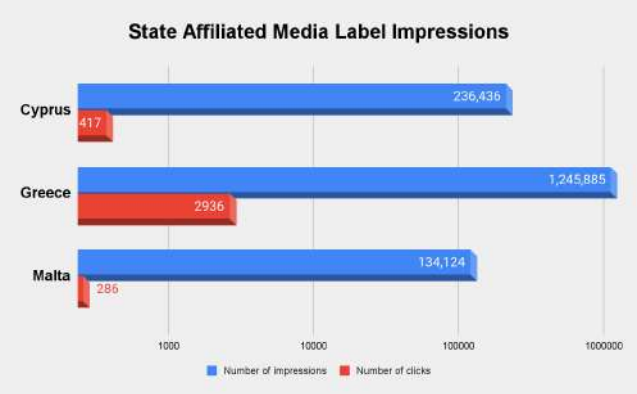
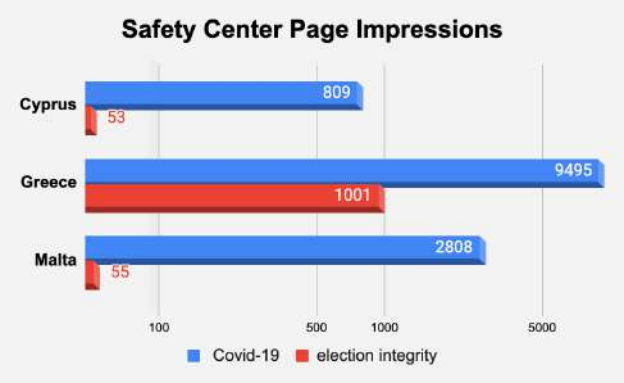
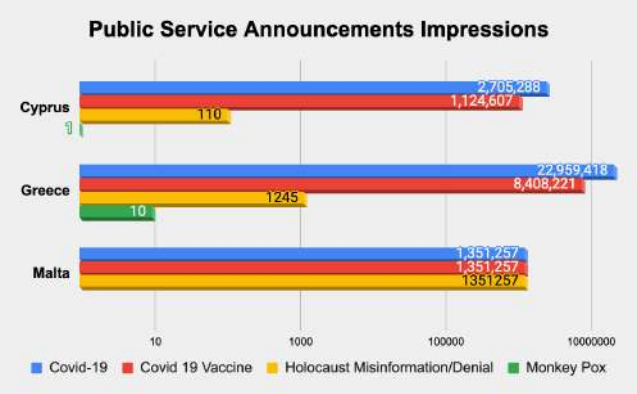
[45] https://newsroom.tiktok.com/fil-ph/know-the-facts

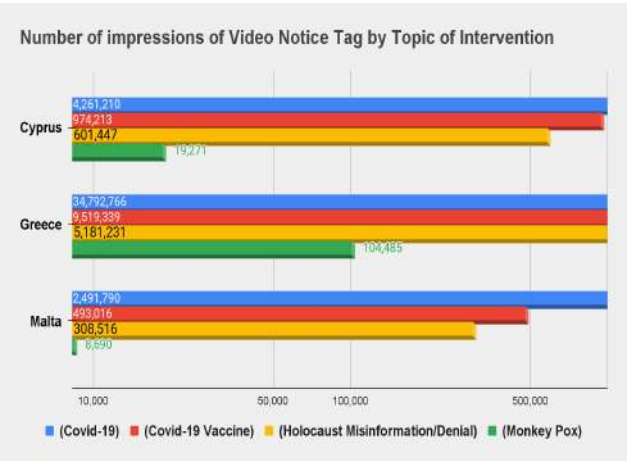
[46] https://www.tiktok.com/tns-inapp/pages/state-affiliated-media

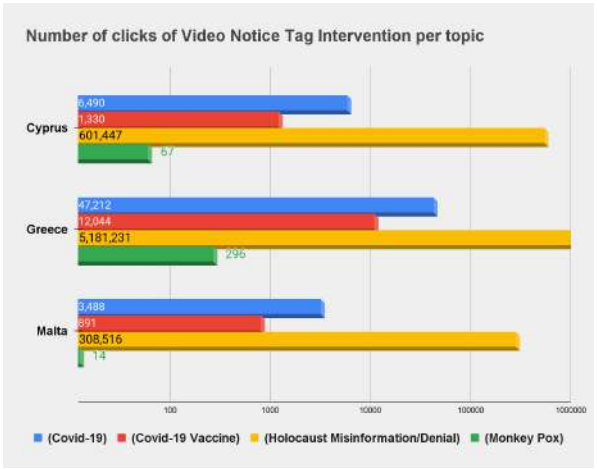
TikTok reported (in SLI 17.1.1.) the following metrics – *number of impressions, clicks, clicks through rate* **– for assessing the impact of the following in-app tools to empower the users:**

- State Affiliated Media label (SAM)
- Safety centre page on the topics of Covid-19 and election integrity *(only number of impressions)*
- Public service announcements (*only number of impressions)*
- Video Notice Tag covered by Interventions
- Search interventions

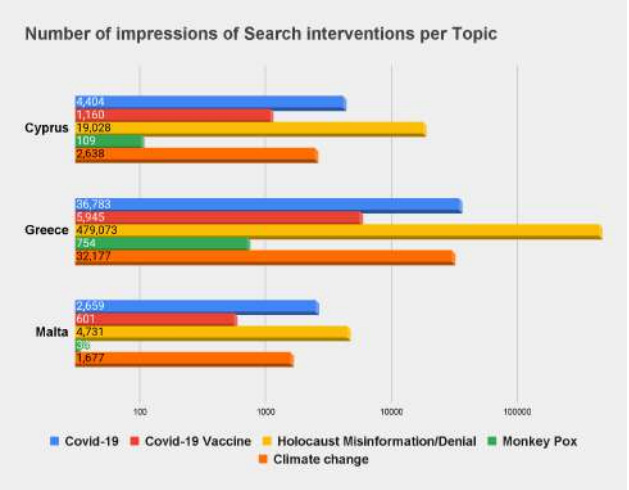
The TikTok reported information for SLI 17.1.1 can be found in the TikTok's July 2023 report pages 84-104, next we present the reported numbers in figures:

State Affiliated Media Label	Safety Centre Page Impressions
	
Public Service Announcements	
	
Video Notice tags intervention	





Search Intervention



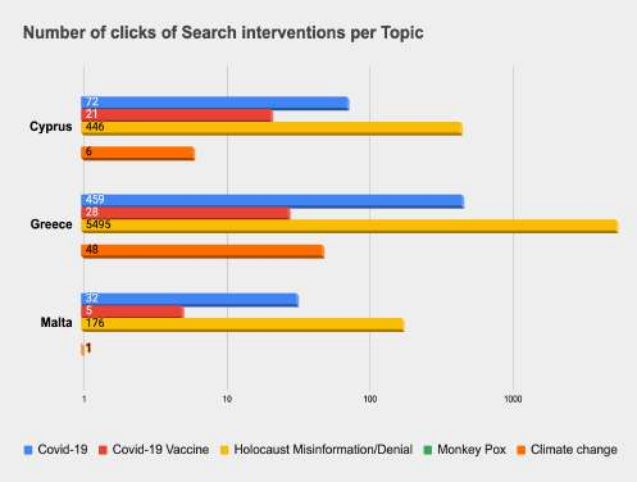


Table 24: TikTok's reported quantitative information for SLI 17.1.1

TikTok Media Literacy Campaigns:

TikTok reported that they conduct various media literacy campaigns both on and off the platform to enhance user awareness on specific topics and provide empowerment. The approach may vary depending on the subject matter. For instance, in campaigns related to elections, TikTok collaborates with national partners and tailors content to resonate with the local audience. In contrast, campaigns like the War on Ukraine focus on connecting users with scalability, safety, and valuable resources. The platform has recently implemented in-app campaigns utilising several intervention tools, including search interventions and video notice tags, as detailed in its response to QRE 17.1.1.

Campaigns for Promoting election integrity:

1. For the **2023 Finnish election**, TikTok launched a search guide on March 6, 2023, providing users with timely information about the Finnish election. The platform collaborated with The National Audiovisual Institute and directed users to their Media Literacy website. For this campaign TikTok collaborated with the National Audiovisual Institute.
2. For the **2023 Greek election,** TikTok rolled out a campaign starting from May 3, 2023. This campaign included a s**earch intervention and an in-app Election Hub** introduced ahead of the May Greek election, as indicated in the provided screenshots. The Election Hub served as a platform connecting users to authoritative information sources and encouraged them to educate themselves on misinformation through AFP's Greek Fact Check page[47]. The campaign persisted leading up to the second Greek election held on June 25. An example of TikTok intervention for the Greek election is in Figure 25.
3. For the **2023 Spanish election**, TikTok initiated a search intervention and an in-app election hub to offer users current information leading up to the Spanish general election on July 23, 2023. Collaborating with Newtral, TikTok's fact-checking partner, and Maldita, a local media literacy organisation, the platform produced educational videos focusing on the electoral process and combating election misinformation.

[47] https://factcheckgreek.afp.com/list

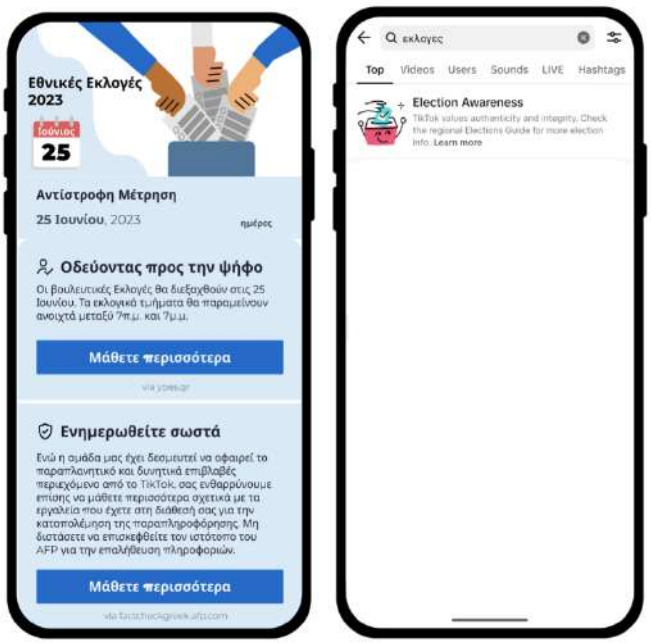


Figure 25: Screenshot of TikTok's media literacy campaign for Greek elections 2023 from TikTok CoP report No2

Furthermore, TikTok conducts **Election Speaker Series** in anticipation of specific elections. TikTok extends invitations to qualified external local/regional experts who share their insights and market expertise with our internal teams.

Campaigns for War in Ukraine:
In collaboration with its fact-checking partners, TikTok has developed and launched eight localised media literacy campaigns this year, addressing the war in **Poland, Slovakia, Romania, Ukraine, Hungary, Estonia, Latvia, and Lithuania.** Users who search for keywords related to the war are directed to informative tips, created in partnership with the platform's fact-checking collaborators, aiming to help users identify and prevent the spread of misinformation on the platform. For the campaign in Poland, TikTok worked with *Fake.pl* while for the other countries the campaign ran, TikTok worked with *Lead Stories*.

Campaigns for Covid-19:
A campaign was enacted across multiple jurisdictions to combat Covid-19-related disinformation. Through the use of dedicated notice tags and search intervention tools, users are now directed to authoritative and localised information from expert organisations. These sources include local public health sites or, in cases where local health sites are unavailable, the World Health Organization (WHO). TikTok actively collaborates with the following entities for the covid-19 campaign: 1) the WHO Tech Taskforce; and 2) European fact-checkers including AFP, Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit, DPA and Reuters.

Major Comments:

1. TikTok did not implement targeted **media literacy campaigns** for Cyprus and Malta. However, in the case of **Greece**, a media literacy campaign was conducted specifically for the 2023 elections (2023 Greek election). Unfortunately, there are no available metrics to gauge the effectiveness and impact of this campaign. Discrepancies in the implementation of media literacy tools across countries are evident; for example, despite the 2023 presidential elections in Cyprus, there was no dedicated campaign for this event. This variance may be attributed to differing levels of emphasis placed by individual countries on such initiatives or TikTok allocating resources unequally among Member States.
2. Political accounts with **verified badges** as a media literacy effort do not have an effect in Cyprus and Malta since several politicians' accounts do not carry the badges. In Greece, several politicians' accounts have the verified badge.
3. **Unverified content labels and redirection to authoritative sources** are available in the Greek and English languages, there is no evidence that is available in Maltese. Users in Cyprus are sometimes redirected to Greek authoritative sources (i.e., Greek governmental pages instead of Cypriot).
4. The click through rate for the aforementioned in-app tools is consistently low for all countries and topics, which can be an indication of low effectiveness of the tools. Though, there is no way to assess the veracity of these numbers.
5. While the existing tools appear to be of high quality, there is currently no mechanism in place to verify that all content or searches receive appropriate labels or context.

Pillar-V: Empowering Users ***Commitment 18, Measure 18.1*** *QRE 18.1.1, SLI 18.1.1. &* ***Measure 18.2*** *QRE 18.2.1, SLI 18.2.1* *page 113-129*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 18.1	2	2	2	1
Measure 18.2	2	2	2	1

Risk mitigation systems, tools, procedures, and features

TikTok is committed to minimising the spread of harmful misinformation across civic processes, public health, and safety, employing a comprehensive approach encompassing policies, products, practices, and partnerships. **Their systematic safety measures include:**

1. **Removal of Violating Content or Accounts:**

- Utilising technology and specialised misinformation moderators to review and assess content against guidelines.
- Proactive moderation through automated technology and targeted sweeps to identify and remove violative content swiftly.
- Community reporting features for users and stakeholders to report potential violations.
- **Collaboration with fact-checkers**. TikTok collaborates with a network of independent fact-checking organisations to identify potential misinformation, taking informed actions based on their assessments. Fact-checking partners do not moderate content but provide input to uphold TikTok's Community Guidelines. This collaboration includes a proactive detection program and a database of previously fact-checked claims to enhance detection and enforcement against misinformation.

2. Safety in For You Feed[48]:

- Reducing prominence or labelling content that negatively impacts authenticity on the *For You feed*[49].
- Making For You feed ineligible for content related to general conspiracy theories or unverified information about emergencies and unfolding events.
- Labelling state-affiliated media accounts to empower users to consider information sources.
- Reviewing viral videos to prevent inappropriate content from entering the recommended system.
- Providing access to authoritative information through information centres, public service announcements, and labelled content to prompt users to seek authoritative information.

3. Design by Safety:

- Collaboration with external partners and local/regional election experts to incorporate expertise into feature and policy development. Examples of collaborations:
    - the implementation of specialised prompts for users to consider before sharing unverified content was developed in collaboration with Irrational Labs.
    - Enrichment program for TikTok Trust and Safety team on the Holocaust in collaboration with Yad Vashem for deeper understanding and combating misinformation related to antisemitism and hatred.
    - Collaboration with local/regional election experts to enhance insights and expertise in preparation for EU elections.
    - As a launch partner of the **Partnership on AI's Responsible Practices for Synthetic Media[50]**, TikTok actively contributed to developing a framework that guides the responsible development, creation, and sharing of synthetic media.

TikTok Information & Authenticity (I&A) policies:

TikTok employs a multi-layered defence against harmful misinformation, anchored in its Information & Authenticity (I&A) policies outlined in the Community Guidelines and Terms of Service. These guidelines,

[48] *The For You feed is the interface users first see when they open TikTok.*

[49] https://www.tiktok.com/community-guidelines/en/fyf-standards/

[50] https://syntheticmedia.partnershiponai.org/

transparently presented to users through a Safety Centre[51], define prohibited content, including misinformation that may cause significant harm to individuals or society. The platform recently updated its CGs, providing detailed examples and enhancing user understanding through user-friendly videos.

The I&A policies prohibit various forms of misinformation, from content posing risks to public safety, medical misinformation, climate change denial, dangerous conspiracy theories, to edited materials that mislead users. TikTok ensures clarity on content ineligibility for the For You feed[52], particularly addressing general conspiracy theories, unverified information during emergencies, and potential high-harm misinformation under fact-checking review.

The enforcement mechanism involves automated technology, human moderation, and specialised misinformation moderators collaborating with external fact-checking partners. TikTok emphasises proactive content moderation, aiming to detect and remove harmful material before user reports. The platform maintains a balance between freedom of expression and user protection by removing false content that causes harm while not penalising simply inaccurate information. Additionally, videos with inconclusive fact checks during unfolding events may become ineligible for the For You feed, limiting the spread of potentially misleading information and labelled with the "unverified content" label.

Major Comments:

1. TikTok outlines its policies and enforcement methods for combating misinformation on the platform. The information is available in **English** and **Greek, but not in Maltese**.
2. TikTok employs both human moderators and technological tools for content moderation. In the reported period, there were **96 moderators for the Greek language**. There are **no moderators for the Maltese language.**
3. TikTok reports the number of share cancel rates after receiving the unverified information warning label (see table below - SLI 18.1.1). For the three countries **Cyprus (44.44%), Greece (39.86%), and Malta (33.33%) the percentage is higher than the EU average (29.93%).** This indicates the effectiveness of the warning label in reducing the misinformation distribution on the platform.
4. The reported figures for removed videos and videos made ineligible for the For You feed due to misinformation policy violations (see table below - SLI 18.2.1) in **Cyprus** and **Malta** are notably low. Assessing these numbers proves challenging, and it appears that they may not accurately represent the prevalence of disinformation in these two countries where disinformation is known to be widespread. The data for **Greece** is comparable with those of other Member States (in terms of population).

[51] https://www.tiktok.com/safety/en/our-approach-to-safety/

[52] https://www.tiktok.com/community-guidelines/en/fyf-standards/

TikTok				
SLI 18.1.1 & 18.2.1	**SLI 18.1.1** **Methodology of data measurement:** **The share cancel rate (%) following the unverified content label share warning pop-up indicates the percentage of users who do not share a video after seeing the label pop up. This metric is based on the approximate location of the users that engaged with these tools.**	**SLI 18.2.1** **Methodology of data measurement:** **We have based the following numbers on the country in which the video was posted: videos removed because of violations of our harmful misinformation policies.** **The number of views of videos removed because of violation of each of the harmful misinformation policies is based on the approximate location of the user.**		
	Share cancel rate (%) following the unverified content label share warning pop-up (users who do not share the video after seeing the pop up)	**Number of videos removed because of violation of harmful misinformation policy**	**Number of views of videos removed because of violation of harmful misinformation policy**	**Number of videos made ineligible for the For You feed under the relevant I&A policies (general conspiracy theories and unverified information related to an emergency or unfolding event)**
Cyprus	**44.44%**	**204**	**1,682,932**	**0**
Greece	**39.86%**	**1,217**	**18,105,562**	**427**
Malta	**33.33%**	**3**	**0**	**7**
Total EU	**29.93%**	**140,635**	**1,012,020,899**	**74,315**

Table 25: TikTok's reported quantitative information for SLIs 18.1.1 and 18.2.1

Pillar-V: Empowering Users ***Commitment 21, Measure 21.1*** *QRE 21.1.1, SLI 21.1.1. page 137-143*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 21.1	2	1	1	1

User benefit from independent fact-checkers

TikTok has expanded its fact-checking program, covering nine additional EEA countries (Denmark, Finland, Norway, **Greece, Cyprus,** Belgium, Czech Republic, Slovakia, and Bulgaria) and reinforcing moderation teams in corresponding 17 languages. These partnerships play a crucial role in enforcing misinformation policies, contributing contextual insights to TikTok specialised misinformation moderators.

Fact-checkers assist in unverified content labelling, addressing inconclusive cases during unfolding events. TikTok engages fact-checkers in specific topic-related tools for Covid-19, election integrity, and climate change, actively involving them in campaigns and in-app interventions (i.e., search intervention tools and redirecting users to

trusted information sources). Following the unverified content label in a video, TikTok established a warning pop-up when a user tries to share the specific video in order to help the users make informed decisions before spreading unverified information.

TikTok publishes blog posts in over 25 languages and maintains a Safety Centre hub[53], informing users about fact-checking programs and labels.

Major Comments:

1. **TikTok fact-checking partnerships do not cover Malta, while it covers Greece and Cyprus, probably through its collaboration with Agence France-Presse (AFP).**
2. TikTok **mainly removes content found to be misinformative/false** and **in cases of inconclusive cases TikTok applies the ''Unverified Content Label''**. TikTok moderators are in charge to assess, remove or label content that its veracity is questionable. Moderators can ask TikTok fact-checking partners to evaluate content and give their analysis result to the moderators. As part of empowering users, **TikTok fails to provide context for labelling content** as unverified, or in cases that content is fully removed. There is no link directing the user to the fact-checking articles provided by the fact-checking organisations.
3. **TikTok informs the creator of content that it is removed due to misinformation policy or if their content receives an unverified content label but again without providing context to the users. There is a lack of information on how the fact-checking organisations' contribution is used in TikTok's mechanisms and tools to combat disinformation.**
4. **TikTok involves some of the fact-checking organisations that they partner with for some of the in-app interventions.**

TikTok					
SLI 21.1.1	**<u>Methodology of data measurement:</u> The share of removals under our harmful misinformation policy, share of proactive removals, share of removals before any views and share of the removals within 24h are relative to total removals under our CGs. The share cancel rate (%) following the unverified content label share warning pop-up indicates the percentage of users who do not share a video after seeing the label pop up. This metric is based on the approximate location of the users that engaged with these tools.**				
	Share cancel rate (%) following the unverified content label share warning pop-up (users who do not share the video after seeing the pop up)	**Share of removals under harmful misinformation policy**	**Share of proactive removals under misinformation policy**	**Share of video removals before any views under misinformation policy**	**Share of video removals within 24h by misinformation policy**

[53] https://www.tiktok.com/safety/en/safety-partners/

Cyprus	**44.44%**	**0.70%**	**0.71%**	**0.76%**	**0.48%**
Greece	**39.86%**	**0.39%**	**0.33%**	**0.29%**	**0.27%**
Malta	**33.33%**	**0.02%**	**0.02%**	**0.03%**	**0.02%**
Total EU	**29.93%**	**0.93%**	**0.93%**	**1.01%**	**0.78%**

Table 26: TikTok's reported quantitative information for SLI 21.1.1

TikTok		
SLI 21.1.2	**Methodology of data measurement:** **The number of videos tagged with the unverified content label is based on the country in which the video was posted.** **The share cancel rate (%) following the unverified content label share warning pop-up indicates the percentage of users who do not share a video after seeing the label pop up. This metric is based on the approximate location of the users that engaged with these tools.**	
	Number of videos tagged with the unverified content label	**Share cancel rate (%) following the unverified content label share warning pop-up (users who do not share the video after seeing the pop up)**
Cyprus	**587**	**44.44%**
Greece	**1,443**	**39.86%**
Malta	**355**	**33.33%**
Total EU	**76,094**	**29.93%**

Table 27: TikTok's reported quantitative information for SLI 21.1.2

Pillar-V: Empowering Users ***Commitment 24, Measure 24.1.1*** *QRE 24.1.1, SLI 24.1.1. page 159-163*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 24.1.1	2	2	2	1

Notification and Appeal systems

In all EU member states, TikTok users receive in-app notifications in their relevant local language when specific actions are taken, such as content removal or access restriction, account bans, feature access limitations (e.g., LIVE), or restrictions on monetization. These notifications are provided in near real-time, typically within seconds or up to a few minutes after the action is taken. In the event of such decisions, users receive an in-app inbox notification detailing the identified violation, accompanied by an option to "disagree" and submit an appeal.

Users have a 180-day window from the notification to submit appeals, with detailed information on how to appeal being available. Appeals are queued for review by specialised human moderators, ensuring thorough

consideration of the context in the decision-making process. Users can monitor appeal status and view results within their in-app inbox. Additionally, users can share feedback through the in-app "report a problem" function, contributing to ongoing efforts to enhance the appeals process on TikTok. All the related information of the notification and appeal systems of TikTok is available in 25 languages[54].

Major Comments:

1. The **information for the TikTok appeal process** is **available in Greek and English language, but not Maltese.**
2. The **number of appeals in Cyprus (71)** appears **relatively high** when compared to the corresponding **number of removed videos due to misinformation policy violations (204)**. However, a **20% success rate in appeals** could suggest potential inaccuracies in TikTok's mechanism for detecting videos that violate misinformation-related policies, both in Greek and more broadly. Assessing the reported numbers is challenging.
3. Similarly, in **Greece**, the reported figures (260 appeals and 1,217 removed videos) raise questions, with a 28% success rate in appeals potentially indicating similar concerns about the accuracy of TikTok's policies and detection mechanisms.
4. For Malta, there **was only one appeal** for the reported period, and it was **successful**. Having **only three videos removed** due to the harmful misinformation policy and a **100% success rate in appeals** (the only appeal being successful) **raises concerns**, especially considering the **absence of moderators for the Maltese language**. Given these circumstances, the reported numbers remain challenging to comprehensively assess.

TikTok				
SLI 24.1.1	**Methodology of data measurement: The number of appeals/overturns is based on the country in which the video being appealed/overturned was posted. These numbers are only related to our harmful misinformation policies. The appeal success rate of videos removed by our harmful misinformation policies is based on the ratio between the number of appeals raised and the number of successful appeals (i.e. overturns).**			
	Number of accounts removed banned under our I&A policies	**Number of appeals of videos removed for violation of harmful misinformation policy**	**Number of successful appeals for violation of harmful misinformation policy (i.e. overturns)**	**Appeal success rate of videos removed for violation of harmful misinformation policy**
Cyprus	**6**	**71**	**15**	**21.13%**
Greece	**73**	**260**	**72**	**27.69%**
Malta	**4**	**1**	**1**	**100%**
Total EU	**6,847**	**37,811**	**14,059**	**37.18%**

Table 28: TikTok's reported quantitative information for SLI 24.1.1

[54] https://www.tiktok.com/legal/page/global/compliant-handling-eea/en

3.3.4 VI. Empowering the Research Community

Pillar-VI: Empowering the Research Community *Commitment 26, Measure 26.2.1 QRE 26.2.1, SLI 26.2.1.* *page 167-170*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 26.2	2	2	2	2

Towards empowering the research community TikTok have implemented the following:

1. TikTok Research API[55]:

TikTok has developed a Research API, providing researchers with access to public data encompassing content, accounts, and comments on the platform. Initially launched in the United States and accessible to US-based academic researchers, the Research API has expanded its reach to include the European Economic Area (EEA), the United Kingdom, and Switzerland. Responding to valuable feedback from researchers, TikTok has already implemented improvements, including a streamlined application process[56] and the introduction of Lab Access. This feature allows research project teams within the same university to provide access for up to 10 researchers, enhancing flexibility and collaboration. Detailed information on the Research API, including data availability and application procedures, can be found on TikTok's dedicated website for developers. A step-by-step guide is also available to researchers[57].

Accessible Data:

- Public account data (user profiles, comments, performance metrics).

- Public content data (comments, captions, performance metrics for videos).

- Public data for keyword search results.

2. TikTok Commercial Content API[58]:

In compliance with the Digital Services Act (DSA), TikTok has developed commercial content-related APIs, focusing on ads, ad metadata, and targeting information. These APIs empower the public and researchers to conduct customised searches based on advertiser names or keywords within the Commercial Content Library repository[59]. The library serves as a searchable database, offering information about paid ads and their metadata, including advertising creative, ad run dates, primary targeting parameters (e.g., age, gender), and audience reach. Like the Research API, TikTok provides detailed information and step-by-step instructions on how

[55] https://developers.tiktok.com/products/research-api/
[56] https://developers.tiktok.com/application/research-api
[57] https://developers.tiktok.com/doc/research-api-get-started/
[58] https://developers.tiktok.com/products/commercial-content-api
[59] https://library.tiktok.com/ads

researchers can access data through the Commercial Content API on the TikTok for Developers website. This ensures transparency and facilitates the application process, guiding researchers on compliance with security measures and data querying procedures.

3. **TikTok Transparency Centre:**

TikTok Transparency Centre hosts the six-monthly Code of Practice reports[60] with 2,500 metrics related to disinformation and the Community Guideline Enforcement Reports[61] for comprehensive insights, providing detailed metrics, actions against violative content, and additional transparency measures in multiple EU languages.

TikTok shared information about receiving over 60 applications for the TikTok Research API from non-profit academic researchers in the US since its launch.

Major Comments:

1. TikTok tools to provide access to researchers to its data seem to be of acceptable quality with all the relevant information available on TikTok website.
2. TikTok did not report any metrics for the uptake, swiftness, or acceptance level of the tools, the number of monthly users, application statistics, and average response time. There is no information for applications coming from researchers from the three countries.

3.3.5 VII. Empowering the fact-checking community

Pillar-VII: Empowering the fact-checking community ***Commitment 31, Measure 31.1.1*** *QRE 31.1.1, SLI 31.1.1. &* ***Measure 31.2.1*** *QRE 31.2.1, SLI 31.2.1* *page 187-195*				
	Evaluation of Reported Actions (QREs)	**Evaluation of Implementation (SLIs)**		
		Greece	**Cyprus**	**Malta**
Measure 31.1.1	2	1	1	1
Measure 31.2.1	2	1	1	1

TikTok Fact-checking partnership in EU:

TikTok has fact-checking coverage in **17 official European languages** (Dutch, **English**, French, German, Hungarian, Italian, Polish, Romanian, Spanish, Swedish, Danish, Finnish, Norwegian, **Greek**, Czech, Slovakian,

[60] https://www.tiktok.com/transparency/en/copd-eu/

[61] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2021-2/

and Bulgarian), covering **22 EEA countries.** TikTok established partnership with the following **nine** fact-checking organisations in Europe:

Agence France-Presse (AFP), *Facta.news, Lead Stories, Logically, Newtral, Science Feedback, dpa Deutsche Presse-Agentur, Teyit, Reuters*

TikTok collaboration with fact-checkers: How does it work?

TikTok uses machine learning to identify potential misinformation but at the same time they have misinformation moderators who assess, confirm, and address harmful misinformation. Fact-checking partners actively contribute to the moderation process in three ways. Firstly, moderators send videos for independent review by fact-checkers, involving assessments of content accuracy through various methods. Secondly, fact-checking partners contribute to a global database of previously fact-checked claims, assisting TikTok's misinformation moderators in making prompt and accurate decisions. Thirdly, a **proactive detection program** involves fact-checkers flagging new claims, enabling moderators to swiftly assess and remove violations.

When content is being fact-checked or cannot be verified, **TikTok may reduce its distribution to limit visibility.** Fact-checkers don't directly act on content; rather, moderators consider their feedback when determining policy violations and appropriate actions.

Additionally, **TikTok uses fact-checking feedback to provide users with context** about specific content. In cases of inconclusive fact-checks or unverifiable content, especially during crises, TikTok **informs viewers through banners** to raise awareness about content credibility and discourage sharing. In such instances, the video may become ineligible for recommendation in anyone's For You feed to curb the spread of potentially misleading information.

Major Comments:

1. TikTok reported that there is fact-checking coverage for Cyprus and Greece. In the report there is no specific information for the fact-checking organisation that covers Greek but AFP in their response to our questionnaire (see Section 4.2: MedDMO Fact-Checking Partners Collaboration with VLOPs - Table 32) mentions the collaboration with TikTok for fact-checking purposes.
2. Malta is not covered under the TikTok fact-checking agreements so far. TikTok reported that efforts are ongoing to expand the fact-checking program.
3. **Cyprus** The discrepancy between the number of videos removed after fact-checking assessments (4) and those removed due to policy violations (204) is notable, with the former being considerably low. The reported numbers raise questions, but without additional means, a thorough assessment is challenging.
4. **Malta** There is no coverage from fact-checking organisations, so the number of removed content after fact-checking assessment is zero. Additionally, the number of videos removed due to policy violations (4) is considered very low. Such assessments appear questionable, yet a conclusive evaluation is hindered by the current lack of means to verify the reported numbers.
5. **Greece** While the platform reports on the number of videos removed due to fact-checking assessments and policy guidelines, the data is deemed low in relation to the platform's penetration. To comprehensively assess the impact of fact-checkers' work, additional metrics such as the impressions of videos before removal are deemed necessary.

TikTok				
SLIs 31.1.1-3	**(SLI 31.1.1)** Methodology of data measurement: The number of fact checked videos is based on the number of videos that have been sent for review to one of our fact-checking partners in the relevant territory.	**(SLI 31.1.2)** Methodology of data measurement: The number of videos removed as a result of a fact checking assessment and the number of videos removed because of policy guidelines, known misinformation trends and our knowledge-based repository is based on the country in which the video was posted. These metrics correspond to the numbers of removals under the harmful misinformation policy since all of its enforcement are based on the policy guidelines, known misinformation trends and knowledge-based repository.		**(SLI 31.1.3)** Methodology of data measurement: The metric we have provided demonstrates the % of videos which have been removed as a result of the fact checking assessment, in comparison to the total number of videos removed because of violation of our harmful misinformation policy.
	Number of fact checked videos (tasks)	**Number of videos removed as a result of a fact checking assessment**	**Number of videos removed because of policy guidelines, known misinformation trends and knowledge based repository**	**Number of videos removed as a result of a fact checking assessment / number of removals under harmful misinformation policy**
Cyprus	**39**	**4**	**204**	**1.96%**
Greece	**89**	**14**	**1,217**	**1.15%**
Malta	**0**	**0**	**3**	**0.00%**
Total EU	**10,181**	**2,848**	**140,635**	**2.03%**

Table 29: TikTok's reported quantitative information for SLIs 31.1.1 – 3

4 MedDMO fact-check activities

4.1 Analysis of MedDMO fact-checks

We collected the fact-checks by MedDMO from the MedDMO fact-checking archive for the period 01-01-2023 to 30-06-2023. There are fact-checks in Greek and English languages. The fact-checks in Greek refer to misinformation that spreads in Greece and/or Cyprus. While fact-checks in English may refer to Malta and/or Greece and/or Cyprus.

From the **MedDMO archive**[62] we have the following information:

1. Fact-check Title
2. Author (and corresponding MedDMO fact-checking organisation that conducted the fact-check)
3. Category of fact-checked information
4. Date of fact-checking article publication
5. Link to fact-checking article

[62] https://meddmo.eu/fact-checking/archives/

The analysis of MedDMO fact-checks aims to explore how the misinformation identified in the three countries was spread and how it was treated from the platforms. By visiting each fact-checking article link we manually collected the following information:

1. Country (Greece, Cyprus, Malta)
2. Topic (Israel Hamas war, Russia/Ukraine war, EU, Climate, Health&Covid-19, 5G, Migration, National elections, LGBTQ+, Other)
3. Platform where misinformation was detected by the fact-checkers
4. Links to the content on platforms
5. Labelled/censored by the platform – did the platform treat the content after fact-checking?
6. Fact-Checking organisation that the platform indicates for users to read the full fact-check article

How we derived on the platform the misinformation spread for a specific fact-checking article:

- screenshots of content on platforms
- links to the actual or archived content on platform
- fact-checking text referred to a specific platform

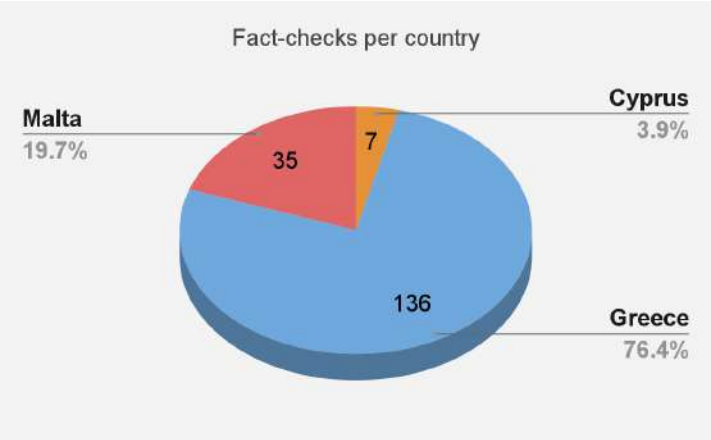
How we derived if the content was labelled by the platform:

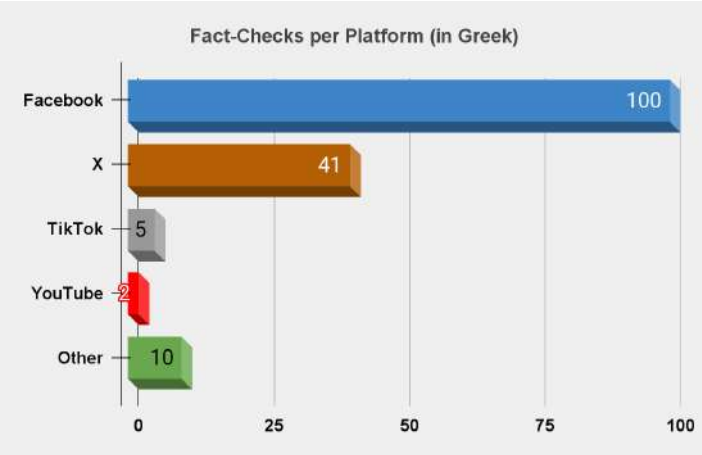
In cases of active content on platform (in cases where the fact-check articles contained links to the posts, etc.) we could verify if the content was treated with a label or not

How we derived the Fact-checking organisation linked to the labelled content on platform:

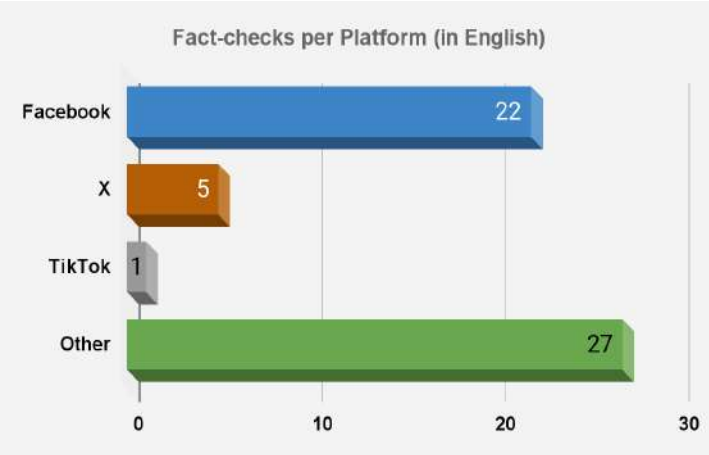
If active content was labelled and there was reference to the fact-checking organisation article for the specific topic.

Main Findings	
123 fact-checks in Greek. The fact-checks refer to misinformation/disinformation spreaded in **Greece and/or Cyprus**	**48 fact-checks in English.** The fact-checks refer to misinformation/disinformation spreaded in **Greece and/or Cyprus or Malta.**

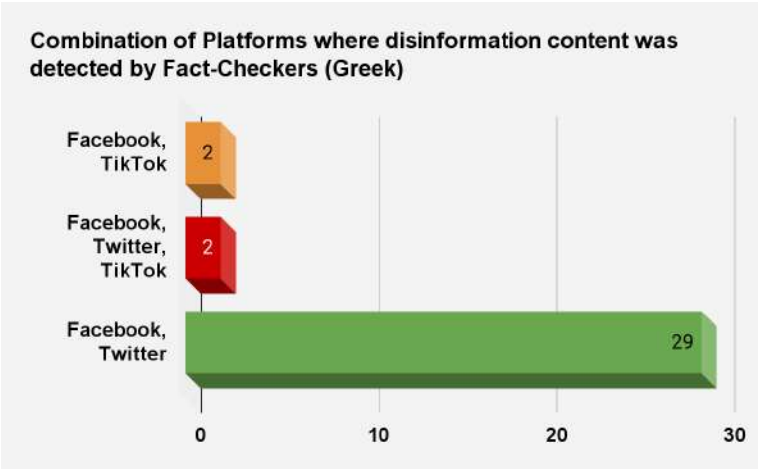


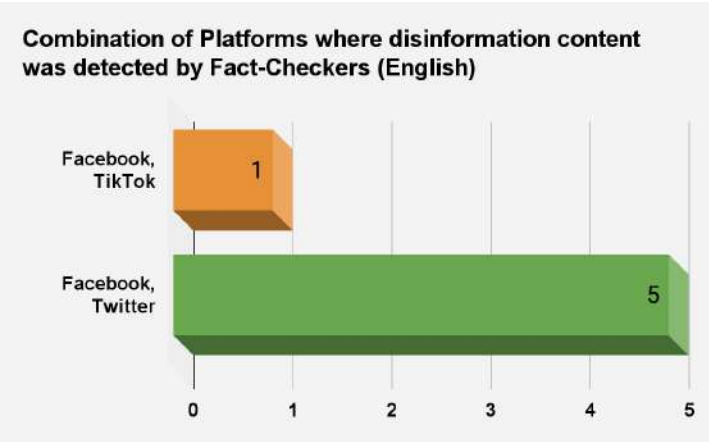


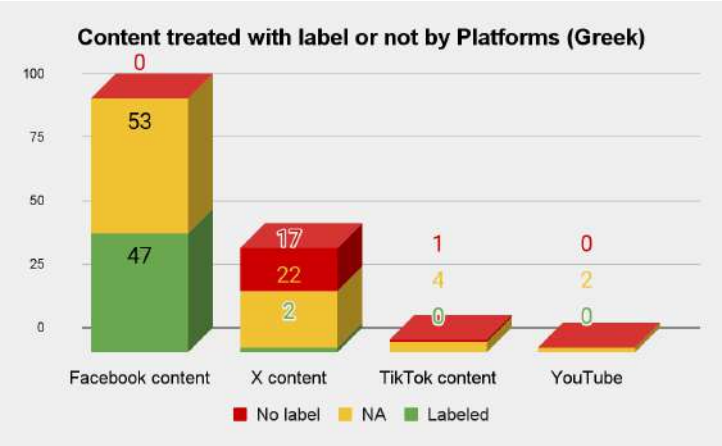
*__Other__ refers to online news outlets



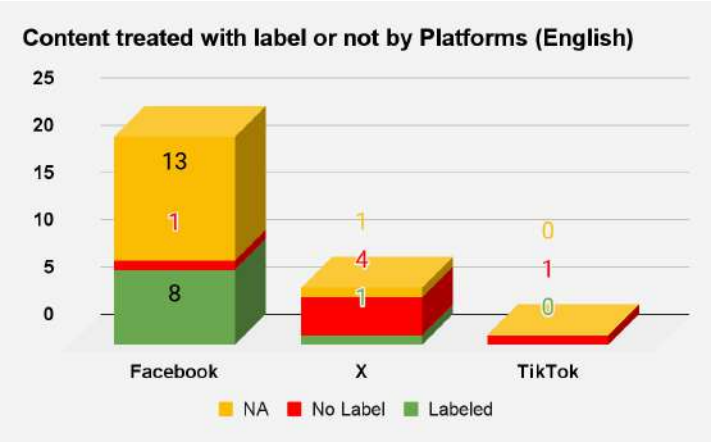
*__Other__ refers to online news outlets



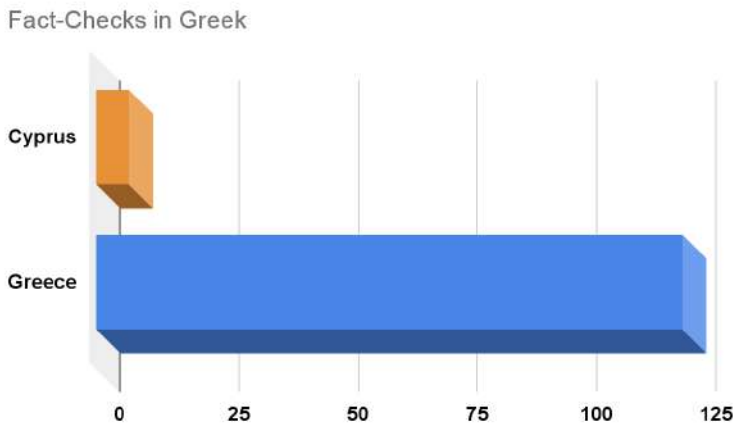




NA:



NA:



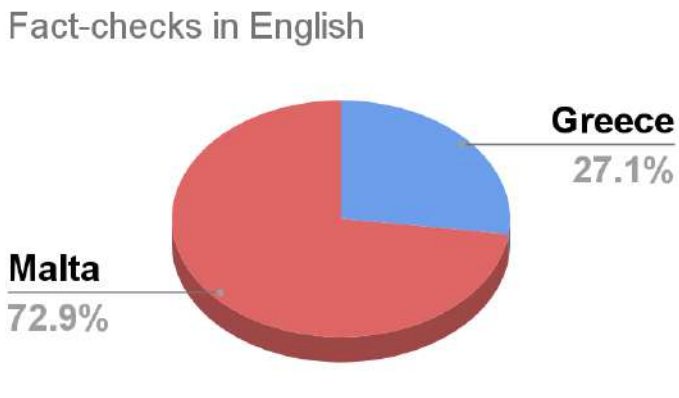


Table 30: Main findings from the MedDMO fact-checks analysis

4.2 MedDMO Fact-Checking Partners Collaboration with VLOPs

MedDMO consortium consists of two International Fact Checking Network (IFCN) certified fact-checking organisations: Agence France-Presse[63] (AFP) and Ellinika Hoaxes (EH)[64]. Our fact-checking partners collaborate with VLOPS to assist them in combating disinformation spread.

We asked the two fact-checking organisations to respond to an open-ended questionnaire regarding their collaboration with the platforms.

The questionnaire is available in **Annex I**. The main findings of the questionnaire follow.

[63] https://www.afp.com/en

[64] https://www.ellinikahoaxes.gr/

Collaboration with Meta: Third Party Fact-Checking Program (3PFC)		
	EH	**AFP**
Collaboration:	Ellinika Hoaxes collaborates with Meta for fact-checking users' content on Instagram and Facebook. The collaboration started in May of 2019.	AFP collaborates with Meta in its Third-Party Fact-Checking (3PFC) program since 2017. The collaboration extends to fact-checking in 26 languages globally, covering various platforms such as Facebook and Instagram.
Fact-Checked Content:	The organisation can review and rate various types of content, including public Facebook and Instagram posts, advertisements, articles, photos, videos, Reels, and text-only posts.	Within the 3PFC program, AFP is authorised to fact-check content on Facebook and Instagram. The organisation fact-checks posts, including images, texts, videos, and sponsored content, monitoring comments to assess potential harm and virality. AFP do not systematically look at online advertisements.
Fact-Checking Process:	The process involves three stages: Detection, Evaluation, and Refutation. They detect misleading information, evaluate submissions, and once identified, publish fact-checks after thorough review.	AFP's fact-checking process involves manual monitoring of social networks, Facebook queues, or WhatsApp queries. The process includes evaluating the fact-checkability and potential harm or virality of content, followed by fact-checking and verification. Fact-checks are written, reviewed, and published, with ratings applied to the original content within the platform's system.
Feedback from Platform:	Ellinika Hoaxes can only see whether content has been rated or not. No specific details on moderation or labelling are provided.	Meta publicly shares information about its collaboration with fact-checking partners, including AFP in their reports to DSA, and the Code of Practice. Metrics about the impact of fact-checking programs are provided, such as the number of non-shares after reading a fact-check. Specific reports by the platform on the impact of AFP's work on the platform are limited.
Number of Fact-Checked Content:	Ellinika Hoaxes produces up to 50 fact-checking articles per month, but the exact number of reports published in Meta services as a result of Meta's third-party fact-checking program is confidential under Meta's partnership NDA.	AFP keeps track of the reports it publishes, though specific figures are considered commercial in confidence and not disclosed. AFP seeks access to the internal archives of the platforms relevant to their fact-checking work, and this is a recurring request.

Meta's Use of Fact-Checking Articles:	Meta utilises warning labels and notifications on content fact-checked by Ellinika Hoaxes. Meta links the fact-checked content on the platform with Ellinika Hoaxes fact-check articles.	Meta extensively uses AFP's fact-checks, connecting them to user posts for moderation actions, including labelling and adding context information. Users are informed about the fact-checks, and AFP's work is integrated into Meta's moderation actions.
Number of Fact-Checkers:	Ellinika Hoaxes has **11 editorial members** participating in Meta's third-party fact-checking program.	AFP has 2 fact-checkers assigned for Meta activities in Greece, contributing to the 3PFC program.
User Requests:	Ellinika Hoaxes receives requests from Meta users for fact-checking content. The exact number is not tracked, but users can seek a review of a fact-check rating or request a review of corrections made to their content.	AFP receives emails and suggestions for fact-checking via its various channels, including WhatsApp tiplines. While direct contact with Meta users is limited, AFP actively engages in reviewing and fact-checking content, contributing to a more informed online environment.

Table 31: MedDMO Fact-Checking Partners Collaboration with Meta

Collaboration with TikTok: Fact-Checking Program	
	AFP
Collaboration:	AFP collaborates with TikTok since 2020 as part of their fact-checking program, embedded in the platform's moderation process. The collaboration extends across several regions globally, including Latin America, Europe, and the Asia Pacific.
Fact-Checked Content:	AFP fact-checks videos on TikTok, often including text within the content. The organisation monitors the platform independently and writes fact-checks based on the content resulting from this monitoring process.
Fact-Checking Process:	The fact-checking process involves manual monitoring of the platform or TikTok's back-office queues or fact-checking queries on WhatsApp, evaluating the fact-checkability, potential harm, and virality of content. AFP independently fact-checks videos, and the resulting fact-checks are published on the platform's back office.
Feedback from Platform:	TikTok publicly explains its collaboration with AFP in its global fact-checking program. The platform shares some metrics in their Code of Practice and DSA reports for the impact of the fact-checking. However, the final moderation decisions remain with TikTok's moderators after the rating of the fact-checkers.
Number of Fact-Checked Content:	AFP keeps track of the reports it publishes but does not disclose specific figures, considering them commercial in confidence. AFP seeks access to the internal archives of the platforms relevant to their fact-checking work, and this is a recurring request.

Use of Fact-Checking Articles:	TikTok shares links to AFP's fact-checks in specific information pages created around events, such as elections[65]. The links are used to provide additional context and information to TikTok users.
Number of Fact-Checkers:	AFP has 2 fact-checkers assigned for TikTok activities, covering the same team members involved in Meta fact-checking for Greece.
User Requests:	While AFP is not in direct contact with TikTok users, the organisation actively monitors and fact-checks content on the platform, contributing to the fight against disinformation.

Table 32: MedDMO Fact-Checking Partners Collaboration with TikTok

Other Collaborations		
	EH	**AFP**
Google	Ellinika Hoaxes collaborates with **Google**, where its content is featured on **Google Search** results **through ClaimReview and Fact Check Explorer.**	AFP does not engage in specific fact-checking of content with **Google**. Instead, the collaboration involves the development of training tools for journalists, journalism students, and wider audiences on investigating disinformation online. This training program operates at a global level and covers multiple languages, including French[66], English[67], Spanish[68], and Portuguese[69]. AFP also create tips and techniques videos in this context (French[70], English[71], Spanish[72]).
Other platforms	While no collaborations beyond Meta and Google currently exist, the organisation is open to discussions with other platforms. The organisation emphasises the importance of various platforms engaging more with fact-	**AFP do not have contracts** with the following platforms, however, **they still fact-check content** on these platforms: Telegram, V-Kontakte, X, LinkedIn, Weibo, Snapchat, YouTube, Naver, Google and Bing Search, etc.

[65] https://activity.tiktok.com/magic/eco/runtime/release/64400a0478c79d0360a77740?appType=tiktok&magic_page_no=1
[66] https://fr.digitalcourses.afp.com/?_gl=1*x...
[67] https://digitalcourses.afp.com/
[68] https://es.digitalcourses.afp.com/?_gl=1*78krj...
[69] https://br.digitalcourses.afp.com/?_gl=1*1lb6m...
[70] https://www.youtube.com/playlist?list=PLo9T0OZu4qjk7MVqK7VxTFoiI5LTM4FYF
[71] https://www.youtube.com/playlist?list=PLo9T0OZu4qjk7MVqK7VxTFoiI5LTM4FYF
[72] https://www.youtube.com/playlist?list=PL3oLC6iSclxCCkNtfzRpxw9Fh0hsKzDe4

	checking initiatives for an enhanced impact.	

Table 33: MedDMO Fact-Checking Partners Collaboration with Google and other platforms

5 Research Activities: Towards Automatic Disinformation Detection in online social platforms

In envisioning the future of automatic fact-checking, the integration of advanced technologies such as Large Language Models (LLMs) and Graph Neural Networks (GNNs) holds promise in creating highly accurate and efficient systems. By continuously refining these models and leveraging vast amounts of data, we can develop automated fact-checking tools capable of swiftly identifying and correcting misinformation, thereby fostering a more informed and discerning society.

Leveraging Language Model (LLMs) and Graph Neural Networks (GNNs) for automatic fact-checking presents a powerful approach to combating misinformation. LLMs excel in understanding and generating human-like text, enabling them to analyse claims and cross-reference them with vast amounts of textual data to detect inconsistencies or falsehoods. By integrating GNNs, which can model complex relationships within data, fact-checking systems can construct knowledge graphs to represent the interconnectedness of information. This allows for the verification of claims against a network of trusted sources, enhancing the accuracy and reliability of automated fact-checking processes. Together, LLMs and GNNs offer a promising solution to the ongoing challenge of combating misinformation in the digital age.

Towards implementing this vision, CUT worked on two research directions. The first, HyperGraphDis, introduces a novel approach for detecting disinformation on Twitter using a hypergraph-based representation to capture social structures, relational features among users, and semantic nuances. HyperGraphDis outperforms existing methods in accuracy and computational efficiency, particularly achieving an impressive F1 score of approximately 89.5% on a COVID-19-related dataset. The second direction focuses on disinformation detection on YouTube, where CUT introduced a methodology leveraging large language models (LLMs) and transfer learning techniques to classify video content based on veracity. The approach yielded promising results, with fine-tuned models achieving high accuracy and F1 scores, while few-shot learning models exhibited even greater potential, particularly in scenarios with limited training data.

5.1 HyperGraphDis - Disinformation Detection on Twitter with Graph Neural Networks

In light of the growing impact of disinformation on social, economic, and political landscapes, accurate and efficient identification methods are increasingly critical. **[Salamanos et al. 2024]** introduces HyperGraphDis, a novel approach for detecting disinformation on Twitter that employs a hypergraph-based representation to capture (i) the intricate social structures arising from retweet cascades, (ii) relational features among users, and (iii) semantic and topical nuances. Evaluated on four Twitter datasets -- focusing on the 2016 U.S. Presidential election and the COVID-19 pandemic -- HyperGraphDis outperforms existing methods in both accuracy and computational efficiency, underscoring its effectiveness and scalability for tackling the challenges posed by disinformation dissemination. HyperGraphDis displays exceptional performance on a COVID-19-related dataset,

achieving an impressive F1 score (weighted) of approximately 89.5%. This result represents a notable improvement of around 4% compared to the other state-of-the-art methods. Additionally, significant enhancements in computation time are observed for both model training and inference. In terms of model training, completion times are accelerated by a factor ranging from 2.3 to 7.6 compared to the second-best method across the four datasets. Similarly, during inference, computation times are 1.3 to 6.8 times faster than the state-of-the-art.

5.2 Disinformation Detection on YouTube: with Large Language Models (LLMs)

Misinformation on YouTube is a significant concern, necessitating robust detection strategies. In **[Christodoulou et al. 2023]**, we introduce a novel methodology for video classification, focusing on the veracity of the content. We convert the conventional video classification task into a text classification task by leveraging the textual content derived from the video transcripts. We employ advanced machine learning techniques like transfer learning to solve the classification challenge.

Our approach incorporates two forms of transfer learning: (a) fine-tuning base transformer models such as BERT, RoBERTa, and ELECTRA, and (b) few-shot learning using sentence-transformers MPNet and RoBERTa-large. We apply the trained models to three datasets: (a) YouTube Vaccine--misinformation related videos, (b) YouTube Pseudoscience videos, and (c) Fake-News dataset (a collection of articles). Including the Fake-News dataset extended the evaluation of our approach beyond YouTube videos.

Using these datasets, we evaluated the models distinguishing valid information from misinformation. The fine-tuned models yielded Matthews Correlation Coefficient > 0.81, Accuracy > 0.90, and F1 score > 0.90 in two of three datasets. Interestingly, the few-shot models outperformed the fine-tuned ones by 20% in both accuracy and F1 score for the YouTube Pseudoscience dataset, highlighting the potential utility of this approach -- especially in the context of limited training data.

6 Policies to Regulate Disinformation in Cyprus, Greece, and Malta

This section examines the legal frameworks implemented by Greece, Cyprus and Malta to address the proliferation of disinformation within their respective jurisdictions. As disinformation continues to pose significant challenges to democratic processes, social cohesion, and public trust in the digital age, governments are increasingly adopting legislative measures to mitigate its impact. This section provides an overview of the laws and regulations enacted by Cyprus, Greece, and Malta aimed at combating disinformation, with a focus on their key provisions and implications for media and online platforms.

Is disinformation tackled with legislative or non-legislative (but state-coordinated) tools in your country? Or neither?
Cyprus: legislative – at a level **Greece:** legislative **Malta:** There is no formal national framework seeking to combat disinformation in Malta. Additionally, Malta does not have a media literacy policy, and media literacy is not a compulsory subject at any level of the educational curriculum. Although Article 82 of the Criminal Code deals with the spread of false news and hate speech, it is another story when it comes to putting it into practice. Some laws – like libel law, for instance – may carry the potential of tackling disinformation indirectly, but such an option is often abused rather than

used for such a purpose. For example, lawsuits are filed as an intimidatory measure to silence critics – in other words, SLAPPs (strategic lawsuits against public participation).
Does the Criminal Code deal with issues related to disinformation/ untrue statements?
Cyprus: Yes, Criminal Code (Cyprus), Art. 50.[73] It deals with the dissemination of false news. The term disinformation is not there. **Greece:** Yes **Malta:** Article 82 of Malta's Criminal Code deals with the spread of false information and hate speech – although it does not directly mention the term 'disinformation'.
In what cases does the Criminal Code provide remedies to disinformation?
Greece: In cases where someone publicly or via the internet disseminates false news with the result of causing fear in an indefinite number of people or in a certain circle or category of persons who are thus forced to carry out unplanned acts or their cancellation, with the risk of causing damage to the economy, the country's defence capability or public health. **Cyprus:** The Criminal Code of Cyprus makes it an offence to disseminate 'false news' or 'news that can potentially harm civil order or the public's trust towards the State or its authorities or cause fear or worry among the public or harm in any way the civil peace and order,' and the offence carries a possible two-year prison sentence. **Malta:** According to Article 82 of the Criminal Code, anyone to 'maliciously spread false news which is likely to alarm public opinion or disturb public good order or the public peace or to create a commotion among the public or among certain classes of the public' shall, upon conviction, be liable to a prison term and a fine.
What are the relevant provisions in the Criminal Code?
Greece: Article 191 of the Greek Criminal Law, as modified by Article 36 of 4855/2021 law. **Cyprus:** Criminal Code (Cyprus), Art. 50[74]. The maximum penalty under the Criminal Code that can be imposed is two years imprisonment or a fine not exceeding €2,500 or both. **Malta:** Article 82 of the Criminal Code[75]
Are there provisions on untrue statements related to public order?
Greece: The provisions refer to damage to economy, the country's defence capability or public health. **Cyprus:** Yes, see previous responses. **Malta:** Yes. As noted above, Article 82 of the Criminal Code has to do with false news that 'is likely to alarm public opinion or disturb public good order or the public peace or to create a commotion among the public or among certain classes of the public'.
Do(es) the media law(s) deal with disinformation or untrue statement? If yes, then what are the relevant provisions?
Greece: No **Cyprus:** Press Law does not deal with disinformation. The term of disinformation is nor there but for inaccurate publication.

[73] http://www.cylaw.org/nomoi/enop/non-ind/0_154/full.html

[74] http://www.cylaw.org/nomoi/enop/non-ind/0_154/full.html

[75] https://legislation.mt/eli/cap/9/eng/pdf

The Press Law of 1989 (145/1989)[76] (38) The law appears to require newspapers to correct inaccurate information related to specific public servants by publishing a free correction, whether it is submitted by the public servant or the relevant authority (39) this provision requires the owner, legally responsible person, or director of a newspaper to register and publish the response of any individual, whether natural or legal, who is named or implied in a publication, regardless of the source of that publication (whether from the newspaper's management or a third party). **Malta:** Yes - false / untrue statement. The Media and Defamation Act in the consolidated laws of Malta[77]
Are there laws specifically aimed at disinformation?
Greece: No **Cyprus:** No **Malta:** No, unless you take into account Article 82 of the Criminal Code, which deals with the spread of false information (kindly refer to responses to previous questions).
Are there laws regulating online platforms that may deal with disinformation? (Which ones?)
Greece: No; **Cyprus:** No; **Malta:** No.
Are there other relevant policy initiatives?
Greece: No **Cyprus:** The Journalistic Ethics Committee does not have the authority to impose penalties or award compensation. However, the committee can investigate claims of violations or threats to freedom of the press[78], either on its own or in response to complaints. Parties involved have the opportunity to present their positions. If the committee finds a violation or threat, it discloses its findings, leading to ethical satisfaction rather than legal conviction for those accused of spreading false news. The decisions and findings of the committee are made public. **Malta:** No, there are not. In early 2021, the Maltese Government appointed a 'Media Literacy Development Board'; however, to date, no official working papers – let alone policy – have been published. It is worth noting that the last known appointed chairperson of this board is the former editor of a pro-Labour Party newspaper.
What definitions are used by policymakers / in policies in your country to define disinformation/misinformation/related concepts?
Greece: There is no specific definition for dis/mis- information. Criminal law refers to dissemination of false news. **Cyprus:** dissemination of fake news, false news, inaccurate publications, no definition is used, only the common terms. **Malta:** As stated in responses to earlier questions, the Criminal Code does not use the term 'disinformation' - it talks about 'false news'.
Do laws differentiate between disinformation and misinformation (intentional vs. unintentional)?
Greece: No **Cyprus:** Not clearly. It can be implied, in Criminal Code (Cyprus), Art. 50., it is mentioned "the accused is granted the right to prove to the Court that the publication was made in good faith and based on facts justifying the publication." However, the provision is not complete, needs to be revised to cover disinformation and misinformation.

[76] http://www.cylaw.org/nomoi/enop/non-ind/1989_1_145/full.html

[77] https://legislation.mt/eli/cap/579/eng/pdf

[78] https://cmec.com.cy/en/rulings/

Malta: The Criminal Code only uses the phrase 'false news'. The law does not formally distinguish between misinformation and disinformation.

Table 34: Policies to Regulate Disinformation in Cyprus, Greece, and Malta

7 Supporting the national authorities

In the context of the MedDMO project, our mission is to support the media authorities in Cyprus, Malta, and Greece in the challenging task of combating disinformation. In pursuit of this objective, we have initiated communication with the respective regulatory bodies, namely the **Cyprus Radio Television Authority (CRTA), the Broadcasting Authority of Malta (BA), and the National Council for Radio and Television (NCRTV)**, with the aim of establishing collaborative partnerships.

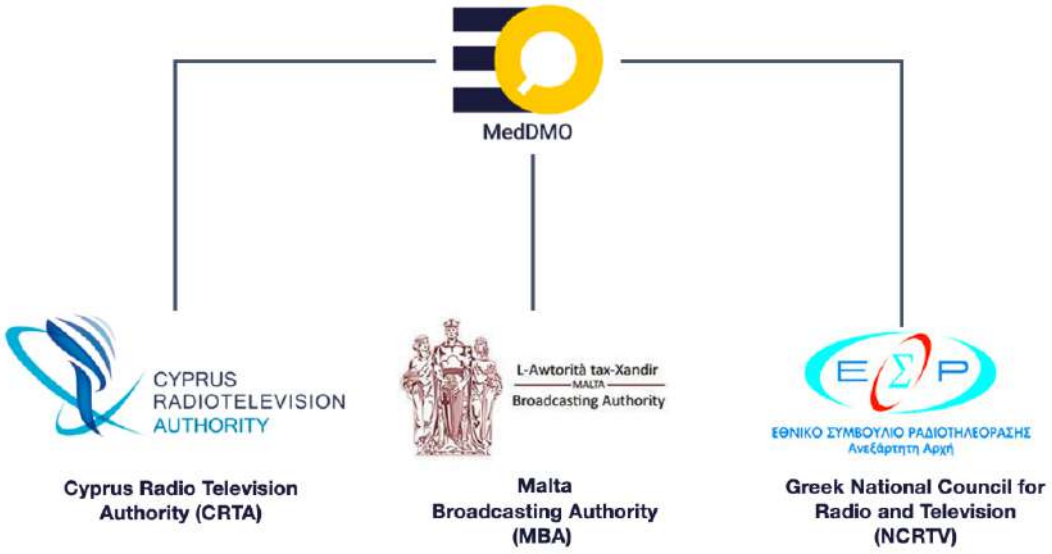


Figure 26: MedDMO collaboration with the national media authorities in Cyprus, Malta, and Greece

Key Areas of Collaboration Between National Media Authorities and MedDMO:

- Collaborative organisation of media literacy campaigns with a focus on combating disinformation.
- Joint facilitation of seminars addressing disinformation and fact-checking specifically tailored for journalists.
- MedDMO's capability to perform on-demand fact-checking upon authorities' requests.
- Dissemination of project outcomes, including research findings, fact-checks, platform monitoring results, educational materials on disinformation, with relevance at national, European, and global levels.
- Assistance provided by MedDMO to authorities in their reporting to the European Regulators Group for Audiovisual Media Services (ERGA) concerning disinformation.
- Promotion of awareness regarding European initiatives addressing disinformation, such as the Code of Practice on Disinformation, AI-ACT, and Digital Services.

- Facilitation of communication among the three authorities, fostering common practices, awareness, and collaboration in addressing disinformation challenges.

The following initiatives illustrate the ongoing support and collaborative efforts with each of the three authorities:

The Case of Cyprus

National Media Authority: Cyprus Radio Television Authority[79]

Relevant actions towards supporting the authority:

- Introducing MedDMO project, its objectives and ways to collaborate with the authority.
- CRTA announced the collaboration with MedDMO through their official channels[80] [81]
- Participation of MedDMO and CRTA representatives in the online event "Advancing Media Literacy – Contemporary approaches in pedagogy"[82] organised by the Cyprus Pedagogical Institute.
- MedDMO participated on behalf of Cyprus in an exercise relevant to Monitoring the Code of Practice at Member State Level organised by Kantar Public and informed the Authority for the tasks.
- MedDMO shared with the authority fact-checks relevant to the crisis in the Middle East - after the request of ERGA SG1 & SG3
- MedDMO shared reports of other EDMO hubs related to monitoring the Code of Practice on Disinformation
- CRTA and MedDMO co-organized a press conference for the MedDMO project in Limassol, Cyprus in 2024 to disseminate the project and announce their collaboration.

The Case of Greece

National Media Authority: National Council for Radio and Television (NCRTV)[83]

Relevant actions towards supporting the authority:

- Introducing MedDMO project, its objectives and ways to collaborate with the authority.
- MedDMO supported the NCRTV representatives in the exercise relevant to Monitoring the Code of Practice at Member State Level organised by Kantar Public and informed the Authority for the tasks.
- MedDMO shared with the authority fact-checks relevant to the crisis in the Middle East - after the request of ERGA SG1 & SG3
- MedDMO shared reports of other EDMO hubs related to monitoring the Code of Practice on Disinformation
- The authority requested the organisation of a seminar on the topic of disinformation and fact-checking for Greek journalists.

The Case of Malta

[79] https://crta.org.cy/en/
[80] https://crta.org.cy/assets/uploads/pdfs/5.2023SinergasiaCRTA_MedDMO.pdf?fbclid=Iw...
[81] https://www.facebook.com/CyRadioTVautority/posts/pfbid0w7LWyrjZxvctQXXfMzzM1fxYPDt1QA1vzs4tW5smtassVg.....
[82] https://medialiteracy.pi.ac.cy/en/events/medialiteracy2023/
[83] https://www.esr.gr/information/

National Media Authority: Broadcasting Authority of Malta (MBA)[84]

Relevant actions towards supporting the authority:

- Introducing MedDMO project, its objectives and ways to collaborate with the authority.
- MedDMO participated on behalf of Malta in an exercise relevant to Monitoring the Code of Practice at Member State Level organised by Kantar Public and informed the Authority for the tasks.
- MedDMO expressed availability to support the authority for sharing fact-checks relevant to the crisis in the Middle East - after the request of ERGA SG1 & SG3
- MedDMO shared reports of other EDMO hubs related to monitoring the Code of Practice on Disinformation

8 Conclusions

In conclusion, this report underscores the pressing need for effective strategies to combat the proliferation of misinformation across online platforms in Cyprus, Greece, and Malta. The analysis has shed light on the diverse challenges faced by these countries within their respective disinformation landscapes, ranging from political manipulation during election periods to state-sponsored disinformation campaigns following tragic events. By examining the practices of Meta, Google and TikTok in implementing the Code of Practice on Disinformation, this report has provided valuable insights into the efforts made by these very large online platforms (VLOPs) to address misinformation within the digital ecosystems of the three nations. However, it is evident that more comprehensive measures are required to counter the pervasive influence of false information, safeguard public discourse, and restore trust in institutions. Moving forward, collaborative efforts between online platforms, national authorities, civil society organisations, and researchers are essential to developing robust strategies that effectively mitigate the impact of misinformation and uphold the integrity of information online.

[84] https://ba.org.mt/

9 References

[Park et al. 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/wpcontent/uploads/2023/09/CoP-Monitor-Report.pdf **(Accessed 2 April 2024).**

[Salamanos et al. 2024] N. Salamanos, P. Leonidou, N. Laoutaris, M. Sirivianos, M. Aspri, M. Paraschiv (2024). ``HyperGraphDis: Leveraging Hypergraphs for Contextual and Social-Based Disinformation Detection''. 18th International AAAI Conference on Web and Social Media (ICWSM'24) **(In press)**
Preprint available at arXiv:2310.01113 http://arxiv.org/abs/2310.01113

[Christodoulou et al. 2023] C. Christodoulou, N. Salamanos, P. Leonidou, M. Papadakis, M. Sirivianos (2023).``Identifying Misinformation on YouTube through Transcript Contextual Analysis with Transformer Models''. Preprint available at arXiv:2307.12155 https://arxiv.org/abs/2307.12155

Annex I: Questionnaire for MedDMO fact-checking organisations

Fact-checking collaboration with platforms

Questions for MedDMO fact-checkers partners

Meta's Third-Party Fact-Checking program:

1. Do you have a collaboration with Meta for fact-checking users' content?

2. When did your collaboration start?

3. Which Meta services' content are you authorised to fact-check within the 3PFC?
   - Facebook,
   - Instagram,
   - Messenger,
   - WhatsAPP

   If there are similar programmes to 3PFC but with other platforms (which offer multiple services) please provide info for which services, you are authorised to fact-check within the specific programmes.

   Please provide further info on which platform services you fact-check generally (not in the context of the fact-checking programmes) or other information you consider to be useful in Question 14.

4. What content are you authorised to fact-check (posts, images, comments, advertisements, ads other:...) within the Meta's 3PFC? If there are similar programmes to 3PFC but with other platforms, please provide info for what content you are authorised to fact-check within the specific programmes. Please provide further info of what content you fact-check generally (not in the context of the fact-checking programmes) or other information you consider to be useful in Question 13.

5. What is the process of reporting disinformation/fact-checking?

6. Did you receive any feedback from the platform related to the flagged content? (if the content you reported is moderated/labelled, how many users see the label, how many shared the content anyway, the time between the reported content and the flagging of the content from Meta, others)

7. What is the amount of fact-checked content (number of reports) by your organisation for each year? Do you keep track of those reports?

8. Did Meta publish or use any fact-checking article from your organisation?

9. How many fact-checkers in your organisation are assigned to participate in Meta third party fact-checking programme for the specific country (if applicable)? Greece: ….. Malta:…….. Cyprus:……

10. Didyou receive any requests from Meta users by email for fact-checking content? How many requests? What is the procedure you follow to reply to these requests?
''From CoP Measure 23.1. - Meta report: *Fact-check: users are also able to request review of a fact-check rating issued by a third-party fact-checker or matched by Meta's technology. They can do this by appealing in-product.* ***In addition, they can reach out directly to the third-party fact-checking organisation via email. Fact-checkers are responsible for evaluating the validity of each correction.''***

11. What are the penalties for accounts/pages/groups that spread disinformation by Meta?

12. What about Meta advertisements fact-checking? Is your organisation report also concerned with disinformation in ads? Please explain.

13. Please add any other information not covered from the previous questions, or comments for your collaboration you consider useful.

Annex II: Platforms' Data Repositories

As reported in CoP VLOPs and VLOSEs signatories reports July 2023, the platforms open access to their public data and ads relevant information through the following libraries:

	Advertisements library	**Content library**
Meta	**Meta Ad Library** https://www.facebook.com/ads/library/	Meta Content library and API https://developers.facebook.com/docs/content-library-api/
TikTok	**TikTok Commercial Content API** https://developers.tiktok.com/products/commercial-content-api	**TikTok Research API** https://developers.tiktok.com/products/research-api/
Google	**Google Ads Transparency Centre** https://adstransparency.google.com/political	**YouTube:** https://research.youtube/ **Google Trends:** https://trends.google.com/trends/ **Fact-check Explorer:** https://toolbox.google.com/factcheck/apis

Mediterranean Digital Media Observatory

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756.
For further information, visit www.meddmo.eu