Mediterranean Digital Media Observatory

# D5.1

# Report of platform practices and national authorities support

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

| | |
|---|---|
| **Deliverable title** | Report of platform practices and national authorities support |
| **Deliverable number** | D5.1 |
| **Deliverable version** | V1.0 |
| **Contractual Date of delivery** | 30.06.2025 |
| **Actual Date of delivery** | 30.06.2025 |
| **Nature of deliverable** | Report |
| **Dissemination level** | Public |
| **Partner Responsible** | CUT |
| **Author(s)** | Pantelitsa Leonidou, Nikos Salamanos, Theodoros Christophides, Andreas Phylactou, Socratis Ioakim, Paul Christodoulides, Iacovos Mylona, Neophytos Lophitis, Athanasios Sitista, Loukia Taxitari, Kostantinos Papadamou, Panagiotis Ilias, Pashalia Spyridou, Sotirios Chatzis, Michael Sirivianos |
| **Reviewer(s)** | CERTH, UCY |
| **EC Project Officer** | Mark Vella Muscat |


| | |
|---|---|
| **Abstract** | This deliverable outlines the MedDMO's efforts to monitor the platforms' practices and support the national authorities. In the annexes of the deliverable, we included the MedDMO reports produced as the outcome of the two rounds of monitoring the platforms' practices and supporting the national authorities. |


| | |
|---|---|
| **Keywords** | Very Large Online Platforms, Code of Practice on Disinformation, Monitoring platforms practices |


MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s).
For further information, visit www.meddmo.eu

# Copyright

© Copyright 2022 MedDMO Consortium

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the MedDMO Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

## Revision History

| Version | Date | Modified By | Comments |
|---|---|---|---|
| V0.1 | 01/12/2024 | CUT | First Draft Table of Content |
| V0.2 | 15/02/2025 | CUT | Final Table of Contents |
| V0.3 | 30/02/2025 | CUT | First Draft |
| V0.4 | 15/04/2025 | CUT | Second Draft |
| V0.5 | 20/05/2025 | CUT | Final Second Draft |
| V0.6 | 20/06/2025 | CUT | Final Second Draft Review |
| V0.7 | 23/06/2025 | CUT | Review |
| V1.0 | 30/06/2025 | CUT | Final Document |

## Glossary

| Abbreviation | Meaning |
|---|---|
| MedDMO | Mediterranean Digital Media Observatory |
| EDMO | European Digital Media Observatory |
| VLOPs | Very Large Online Platforms |
| CoP | Code of Practice on Disinformation |
| AFP | Agence France Presse |
| EH | Ellinika Hoaxes |
| NCRTV | National Council for Radio and Television |
| CRTA | Cyprus Radio Television Authority |
| MBA | Broadcasting Authority of Malta |
| ERGA | European Regulators Group for Audiovisual Media Services |
| DSC | Digital Service Coordinator |
| DSA | Digital Service Act |

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s).
For further information, visit www.meddmo.eu

## Table of Contents

Revision History ..... 4
Glossary ..... 4
Index of Figures ..... 5
Executive Summary ..... 6
1 Introduction ..... 7
2 Monitoring the Platform's Practices ..... 9
2.1 MedDMO Approach in Monitoring the Platforms' Practices: ..... 9
2.2 Monitoring the Platforms' Practices 2023 ..... 10
2.3 Monitoring the Platforms' practices 2024 ..... 12
3 Supporting the National Authorities ..... 15
4 Conclusion ..... 19
Annex I: 1st Report of platform practices and national authorities support 2023 (for Cyprus, Greece & Malta) 20
Annex II: Mapping Disinformation Circulation in Cyprus, Greece, and Malta and Evaluating Platforms' Moderation Practices: Insights from MedDMO Fact-Checks Archive - Part A of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta) ..... 21
Annex III: MedDMO report on monitoring Meta, TikTok, and Google's practices to combat disinformation and supporting national media authorities 2024 (for Cyprus, Greece & Malta) - Part B of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta) ..... 22

## Index of Figures

Figure 1: Snapshots of the MedDMO co-organized events with the national authorities of Greece, Cyprus, and Malta ..... 18

# Executive Summary

This report presents the results of MedDMO's activities under Task 5.1 (Monitoring Platform Practices) and Task 5.2 (Supporting National Authorities), highlighting its continued efforts to counter disinformation in Cyprus, Greece, and Malta. The monitoring work focused on evaluating the implementation of the EU Code of Practice on Disinformation (CoP) by Meta, Google, and TikTok. Two rounds of monitoring practices were completed, covering January–June 2023 and January–June 2024, resulting in dedicated public reports. These assessments combined structured evaluations of platforms' self-reported commitments with an evidence-based approach using fact-checks from the MedDMO archive.

In parallel, MedDMO established collaborations with the national broadcasting authorities in the three countries: the National Council for Radio and Television (Greece), the Cyprus Radio Television Authority (Cyprus), and the Maltese Broadcasting Authority (Malta). These collaborations focused on sharing findings, supporting media literacy, organizing public events, and offering practical tools for disinformation awareness and regulation readiness. Activities also extended to support emerging regulatory frameworks under the Digital Services Act (DSA) and cooperation with the designated Digital Services Coordinators (DSCs) of the three countries.



MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu

# 1 Introduction

One of the primary roles of the Mediterranean Digital Media Observatory (MedDMO) in combating disinformation across Greece, Cyprus, and Malta is the systematic monitoring of these countries' media landscapes. In particular, MedDMO focuses on two key areas: (i) monitoring online platforms' practices** (Task 5.1) and (ii) supporting national authorities (Task 5.2). This report outlines the work undertaken in both areas, highlighting the outcomes of two rounds of platform monitoring and the corresponding support activities for authorities. Notably, these efforts have produced three public reports that document the findings and progress of the platform monitoring and authority support tasks.

In Task 5.1 (Monitoring Platform Practices), our team focused on evaluating the implementation of the EU Code of Practice on Disinformation by major online platforms widely used in the region, namely Meta, Google, and TikTok. The goal was to assess each platform's compliance with the Code of Practice and determine whether their reported anti-disinformation policies and moderation practices are effectively applied in Greece, Cyprus, and Malta. This evaluation followed an investigative, evidence-based approach. In support of our analysis, we leveraged the MedDMO fact-check archive and engaged with MedDMO fact-checking partners who collaborate directly with these platforms. This enabled us to gather concrete evidence on how disinformation content is handled in practice and to verify whether platform-stated policies translate into tangible actions on the ground.

Additionally, to ensure our methodology aligns with broader European efforts, we have actively participated in several initiatives led by the European Digital Media Observatory (EDMO). This involvement informs us about common frameworks for assessing the Code of Practice signatories' reports and complements our monitoring efforts. These initiatives include:

All outcomes of the platform monitoring tasks have been disseminated to the public in Greece, Cyprus, and Malta. This dissemination raises awareness about how disinformation is being addressed (or overlooked) by social media and online platforms. Importantly, the findings from Task 5.1 directly inform our strategy for Task 5.2 (Supporting National Authorities). In parallel with monitoring platform practices, we have conducted research on content moderation in online social networks, which further enriches our understanding of how disinformation circulates and how it can be countered. The insights gained from these research activities and monitoring reports serve as a knowledge base for advising and assisting national authorities.

Under Task 5.2, we have engaged in collaboration with the broadcasting authorities of all three countries to bolster their efforts against disinformation. Specifically, we established communication and support channels with the National Council for Radio and Television in Greece, the Cyprus Radio Television Authority, and the Maltese Broadcasting Authority. Our collaboration with these bodies involves providing expert support on issues of online disinformation, sharing findings from our platform monitoring (including evidence of how misinformation spreads and is moderated), and jointly exploring strategies to strengthen their responses to disinformation. By working together with the national regulators, we aim to help them develop more effective actions and strategies, ensuring that our research outputs translate into practical measures within each country's media oversight framework.

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). 

The remainder of this report is organized as follows: Section 2 presents the methodology and findings of MedDMO's two rounds of monitoring Very Large Online Platforms (VLOPs) and their implementation of the EU Code of Practice on Disinformation (CoP) across Cyprus, Greece, and Malta. It includes a description of the approach followed, the integration of fact-checking data, and the evaluation of platform-reported commitments, resulting in three detailed reports annexed at the end of the document. Section 3 outlines the actions taken to support national authorities in the three countries. Section 4 concludes the report with key insights and outlines MedDMO's next steps, while the annexes include the full reports produced with the outcomes of the monitoring platforms' practices and supporting the authorities' MedDMO tasks.

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu

# 2 Monitoring the Platform's Practices

In this section, we report MedDMO efforts in Task 5.1. Monitoring the platforms' practices.

## 2.1 MedDMO Approach in Monitoring the Platforms' Practices:

MedDMO's approach to monitoring platform practices follows a structured process consisting of four key steps:

**1. Identifying platforms for monitoring**

Meta, Google, and TikTok were selected as the focus of this monitoring effort due to their significant influence on the digital information environment in Cyprus, Greece, and Malta. Meta, which includes Facebook and Instagram, plays a central role in how users access and engage with news and opinions. Google, through its Search engine and YouTube platform, shapes content visibility and discoverability. TikTok, as a rapidly growing platform, offers a dynamic space for short-form, user-generated content that often drives viral narratives. Given their scale and reach, these platforms are critical to understanding and addressing the spread of disinformation in the region.

**2. Identifying sources of relevant information**

We mapped available sources of information on platform practices, focusing primarily on the Code of Practice (CoP) Signatories' Reports. These were complemented by data from platform transparency centres and public information hubs. Additionally, we integrated internal resources from the MedDMO project, including:

(i) the MedDMO Fact-Checks Archive, and

(ii) contributions from MedDMO's fact-checking partners, who actively collaborate with platforms to address disinformation.

**3. Assessing platform practices and producing reports**

Our monitoring focused on two key CoP Signatories' Reports: Report No.2 (covering 1 January – 30 June 2023) and Report No.4 (covering 1 January – 30 June 2024). This aligns with EDMO-led CoP reports assessment cycles during the same periods. We complemented the CoP reports assessment with evidence gathered through our own fact-checking archive and insights from MedDMO's fact-checking organisations. This dual approach ensured that our assessments were both systematic and grounded in the local disinformation landscape.

**4. Disseminating the outcomes**

The results of our monitoring activities were shared with the MedDMO community, the Broadcasting Authorities of Cyprus, Greece, and Malta, EDMO, and the broader public via the MedDMO website and social media channels. The findings also informed ongoing internal activities and strategic planning across other areas of the MedDMO project.



MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). 

**More specifically, we published three reports covering the following periods:**

Round 1: Monitoring the platforms' practices 2023 (1st of January - 30th of June 2023):

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

Round 2: Monitoring the platforms' practices 2024 (1st of January - 30th of June 2024):

- "Mapping Disinformation Circulation in Cyprus, Greece, and Malta and Evaluating Platforms' Moderation Practices: Insights from MedDMO Fact-Checks Archive" - Part A of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta)[2]
- "MedDMO report on monitoring Meta, TikTok, and Google's practices to combat disinformation and supporting national authorities 2024 in Cyprus, Greece, and Malta" - Part B of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta)[3]

## 2.2 Monitoring the Platforms' Practices 2023

The MedDMO 1st Report of Platform Practices and National Authorities Support (2023) provides a multi-dimensional overview of how disinformation is addressed in Cyprus, Greece, and Malta, through platform practices, fact-checking collaboration, national policies, and research-driven innovations.

The methodological framework underpinning the report follows the evaluation approach of the Code of Practice on Disinformation (CoP), drawing specifically on the signatories' transparency reports published in July 2023, which cover the period from 1 January to 30 June 2023. The analysis of platform behaviour is based on specific CoP measures and follows the standardised methodology developed by the EDMO Ireland hub and German-Austrian Digital Media Observatory (GADMO). This enables a structured and comparative assessment of Meta, Google, and TikTok across five core areas: scrutiny of ad placements, political advertising transparency, user empowerment, support for researchers, and fact-checker collaboration.

To monitor how platforms respond to verified disinformation, MedDMO conducted an extensive analysis of its Fact-Check Archive, covering fact-checks published between 1 January and 30 June 2023. The archive includes work from all MedDMO fact-checking partners, featuring articles in Greek and English. Greek-language content primarily addresses disinformation circulating in Greece and Cyprus, while English-language fact-checks may pertain to Malta, Greece, and/or Cyprus. Each fact-check entry in the archive includes the article's title, the authoring fact-checking organisation (AFP or EH), its thematic category, date of publication, and a link to the full article. This archive served as a foundation for evaluating how misinformation is propagated across platforms and how platforms respond to verified fact-checks. For each article, MedDMO researchers manually identified

[1] https://meddmo.eu/1st-report-of-platform-practices-and-national-authorities-support-2023-for-cyprus-greece-malta/

[2] https://meddmo.eu/mapping-disinformation-circulation-in-cyprus-greece-and-malta-and-evaluating-platforms-moderation-practices-insights-from-meddmo-fact-checks-archive-january-june-2024/

[3] https://meddmo.eu/monitoring-meta-google-and-tiktoks-practices-to-combat-disinformation-in-greece-cyprus-and-malta-meddmo-2nd-annual-report-part-b-2024/

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu

the country of relevance, topic, and the specific platforms where the disinformation was found. They recorded links to the original or archived content, screenshots when available, and contextual evidence within the fact-check articles to confirm the platforms of origin.

A key part of this analysis determined whether the platforms had applied any moderation actions, such as warning labels, content removals, or redirecting users to fact-checking articles. When platform links were active, MedDMO researchers verified whether the content was labelled and, if so, whether the label credited a specific fact-checking organisation (AFP or EH) and directed users to the full debunk. This manual verification process enabled the report to assess the real-world visibility and impact of fact-checking efforts and whether they were integrated into platform workflows in meaningful ways.

The report also explores the collaborative mechanisms between fact-checking organisations and VLOPs. To this end, AFP and Ellinika Hoaxes were invited to respond to an open-ended questionnaire on their experiences working with platforms. Their responses offered qualitative insights into how third-party fact-checking partnerships function in practice, including the operational tools provided by platforms, responsiveness to fact-check submissions, language coverage, and the transparency of feedback loops. These reflections, presented in Section 4, help contextualise the effectiveness and limitations of current cooperation models between platforms and IFCN-accredited organisations.

The report also refers to two research initiatives developed under MedDMO, which focus on automatically detecting disinformation on platforms like Twitter and YouTube. These projects aim to extend monitoring capacity through scalable data processing tools and early-warning systems for harmful content trends.

Find the full report at the end of this deliverable **(Annex I)**.

**MedDMO Contribution to EDMO-led tasks:**

- MedDMO participated in an exercise led by Verian (formerly Kantar Public) in collaboration with ERGA and EDMO hubs to monitor the CoP reports for the period of January-June 2023. The results of this exercise were integrated into the 2023 MedDMO report.
- MedDMO also contributed to the survey of the EDMO Policy Working Group by responding to the "EDMO disinformation regulation questionnaire" for the three countries, Cyprus, Greece, and Malta. The results of this survey are published in the following EDMO reports: ***i) How is disinformation addressed in the member states of the European Union? – 27 country cases***[4] and ***ii) Literature review on actors of disinformation in the European Union***[5]. Our input to the survey was also integrated into the 2023 MedDMO report.

[4] https://edmo.eu/wp-content/uploads/2025/06/EDMO-Report-How-is-disinformation-addressed-in-the-member-states-of-the-European-Union-—-27-country-cases-.pdf

[5] https://edmo.eu/wp-content/uploads/2025/06/EDMO-Report-Literature-review-on-actors-of-disinformation-in-the-European-Union-.pdf

## 2.3 Monitoring the Platforms' practices 2024

The findings from the second round of platform monitoring in 2024 were presented across two dedicated reports:

1) **Part A of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta): "Mapping Disinformation Circulation in Cyprus, Greece, and Malta and Evaluating Platforms' Moderation Practices: Insights from MedDMO Fact-Checks Archive"**

MedDMO conducted a systematic analysis of fact-checking articles published between January 1 and June 30, 2024, drawn from the MedDMO Fact-Checks Archive. The aim was to investigate how false claims circulated on social media platforms and news outlets, and to assess whether—and how—platforms responded once those claims had been debunked. This analysis served as a foundational step in monitoring platform moderation practices across Greece, Cyprus, and Malta by leveraging the outputs of MedDMO's fact-checking activities.

To enable this evaluation, an extended version of the MedDMO Archive Dataset was created. The baseline archive lacked essential information such as the platforms where false claims were identified, direct links to the misleading content, and data on how the platforms handled the content post-debunking. The enhanced dataset includes a total of 184 fact-checking articles: 150 covering Greece and Cyprus (138 in Greek and 12 in English) and 34 covering Malta (in English). From these, MedDMO researchers manually extracted 1,226 individual links to online content that contained false or misleading information. These links were sourced from platforms such as Facebook, Instagram, YouTube, TikTok, X (formerly Twitter), and a range of local online news websites.

The primary objective of the analysis was twofold: first, to identify which platforms were hosting or spreading the false claims identified by fact-checkers in each country; and second, to evaluate the platforms' moderation responses—specifically whether the false content was labelled, removed, or linked to a relevant fact-checking article, and whether these actions aligned with the platforms' stated practices and commitments under the EU Code of Practice on Disinformation.

For each case, the MedDMO team thoroughly reviewed the fact-check article, accessed the archived or live links to the original false content, and documented any visible moderation actions. These included the presence of warning labels, contextual overlays, removal notices, or direct references to the fact-checking articles produced by MedDMO's partners—Ellinika Hoaxes and Agence France-Presse (AFP), both certified by the International Fact-Checking Network (IFCN) and involved in third-party fact-checking programmes with major platforms.

The findings were analysed and presented separately for Malta and for Greece and Cyprus combined, allowing for clearer insights into the platform dynamics and disinformation landscapes of each context. This extended dataset and accompanying analysis not only provide a clearer picture of how disinformation circulates across platforms in the region but also highlight inconsistencies and gaps in content moderation practices, informing future efforts for more accountable and transparent responses from digital platforms.

Find the full report at the end of this deliverable **(Annex II)**.

2) **Part B of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta) "MedDMO report on monitoring Meta, TikTok, and Google's practices to combat disinformation and supporting national media authorities 2024 (for Cyprus, Greece & Malta)"**

This report builds on the findings and groundwork established in the Part A report, which focused on mapping the circulation of disinformation and evaluating platform responses using fact-checks from the MedDMO archive. While Part A relied on fact-checks as a source of ground-truth evidence to assess how platforms moderated false or misleading content, this second report shifts the focus toward assessing the platforms' own self-reported commitments and practices under the EU Code of Practice on Disinformation (CoP).

The report offers a comprehensive evaluation of how major online platforms—Meta (Facebook and Instagram), Google (Search and YouTube), and TikTok—implemented their disinformation-related commitments across Cyprus, Greece, and Malta during the first half of 2024. The analysis is structured around three key CoP pillars: Empowering Users, Empowering the Research Community, and Empowering the Fact-Checking Community.

Findings are based primarily on the CoP Signatories' Reports submitted in September 2024, covering the reporting period from 1 January to 30 June 2024. These reports were evaluated both qualitatively and quantitatively through analysis of Qualitative Reporting Elements (QREs) and Service Level Indicators (SLIs). The assessment followed the methodology developed by the European Digital Media Observatory (EDMO) CoP Reports Analysis Subgroup, ensuring a consistent and structured approach to evaluating platform transparency and implementation at the national level.

In addition to assessing reported measures, the report examines the tools and resources made available by platforms to users, researchers, and fact-checkers, with particular attention to issues of linguistic accessibility and localisation. It highlights persistent disparities in the availability and visibility of platform features across different languages—particularly Greek and Maltese—and emphasizes how such gaps can undermine effective disinformation response efforts.

As in the first monitoring round in 2023, AFP and Ellinika Hoaxes were invited to complete an open-ended questionnaire. Their responses capture updates in their perspectives on the depth, quality, and operational transparency of their collaboration with VLOPs, including how their fact-checks are used, labelled, and integrated into platforms' moderation systems. These insights are incorporated into the report to reflect the evolving role of fact-checkers in platform accountability.

The report concludes with a series of recommendations for platforms and national stakeholders, aligned with the three CoP pillars. These recommendations aim to strengthen the effectiveness, inclusivity, and transparency of disinformation responses in the Mediterranean region. They are also designed to support national authorities in Cyprus, Greece, and Malta as they work toward enhanced oversight and compliance under the DSA. Finally, the report outlines MedDMO's ongoing cooperation with national media authorities in the three countries.

Find the full report at the end of this deliverable **(Annex III)**.



MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu

**MedDMO Contribution to EDMO-led tasks:**

- As part of its contribution to the broader European effort to monitor the implementation of the EU Code of Practice on Disinformation (CoP), MedDMO actively participated in the methodology developed by the EDMO CoP Reports Analysis Subgroup for assessing platform transparency and accountability. The evaluation framework was structured around a detailed survey, divided into three sections for each CoP commitment: Section A focused on the quality and completeness of reported information; Section B examined the supporting evidence provided by platforms; and Section C assessed the actual impact of the reported actions. The assessment covered major Very Large Online Platforms and Search Engines (VLOPs & VLOSEs), including Meta, Google, TikTok, and Microsoft. MedDMO voluntarily responded to Sections A and B for Meta (specifically Facebook and Instagram), evaluating the clarity, completeness, and substantiation of Meta's self-reported commitments. Additionally, MedDMO supported Section C by disseminating the expert survey to media literacy professionals and researchers, as well as fact-checking organisations across Cyprus, Greece, and Malta. MedDMO then analysed the expert responses and developed the Country Briefs for each of the three countries, providing insights into the availability and effectiveness of platform tools and resources, the nature of local collaborations, and the perspectives of national experts. The results of this analysis will be published by EDMO in the near future, contributing to a broader EU-wide understanding of how platforms are addressing disinformation at the national level. The suggested methodology was used by MedDMO in the assessment of the CoP Signatories' reports in the second round of MedDMO platforms' monitoring.
- As part of the EDMO EU Election 2024 Political Ads Exercise led by EDMO Ads Transparency working group, MedDMO contributed to a coordinated EU-wide effort to evaluate the transparency and use of political advertising on Meta and Google during the run-up to the European Parliament elections. Focusing on Cyprus, Greece, and Malta, MedDMO followed a three-step process: first, identifying relevant political parties and their official Facebook and Google advertising accounts; second, collecting data on political ads from 1 April to 5 June 2024, including total ads, spending, and keyword-based distribution from both the Meta Ad Library and Google's Ad Transparency Reports; and third, performing a qualitative content analysis of EU-related ads, selecting one ad per political party per month per platform (three per platform, six total per party) and assessing them using a 19-point transparency evaluation framework. While the results of this exercise have yet to be published, MedDMO awaits further guidance from EDMO regarding the next steps. In parallel, MedDMO is also exploring how this methodology could be integrated into MedDMO II to assess platforms' commitments related to political advertising under the Code of Practice on Disinformation.

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu

# 3 Supporting the National Authorities

MedDMO Task 5.2. aims to support the national authorities of the three countries in combating disinformation. In particular, MedDMO established collaborations with the Greek National Council for Radio and Television, the Cyprus Radio Television Authority, and the Broadcasting Authority of Malta.

MedDMO initiated the collaboration by introducing the MedDMO project and its objectives within the broader context of EDMO and offering support to the three authorities in matters related to disinformation and media literacy. More specifically, MedDMO offered to support the authorities in sharing information and in the context of the Code of Practice on Disinformation, the Digital Service Act, the AI-Act, and readiness to support the authorities in their reporting to ERGA. Additionally, the authorities were encouraged to request on-demand surveys on emerging online disinformation narratives. MedDMO is committed to sharing the outcomes of the project with the authorities and has invited them to organize campaigns and events to spread disinformation awareness and promote media literacy in the three countries.

The authorities of the three countries engaged with the MedDMO objectives and were keen to work together with the MedDMO partners to get support for their own agendas and tasks relevant to combating disinformation, as well as their role in promoting media literacy. Within the context of the MedDMO project, the following supporting and awareness activities were conducted in collaboration with the authorities:

**National Council of Radio and Television (Greece):**

- MedDMO supported the NCRTV representatives in the exercise relevant to monitoring the Code of Practice at the Member State Level, organised by Kantar Public, and informed the authority of the relevant tasks. NCRTV participated in the exercise as an ERGA member group.
- NCRTV requested support from MedDMO in contributing fact-checks of claims circulated in Cyprus relevant to the Middle East Crisis in response to the request of ERGA.
- MedDMO invited NCRTV to attend the Press Conference event co-organized by MedDMO and CRTA on the **14th of March 2024**.
- MedDMO, National Council for Radio and Television (NCRTV), and the Journalists' Union of Athens Daily Newspapers (ESIEA) hosted a seminar on disinformation and modern propaganda at ESIEA premises in Athens on the **18th of October 2024**. Experts discussed the challenges of combating fake news, verification tools, and the role of journalism in promoting media literacy. The event attracted journalists, media professionals, and journalism students, concluding with an open discussion on strengthening fact-checking and transparency in the digital era.

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). 

**Cyprus Radio Television Authority (CRTA):**

- MedDMO, represented by Dr. Michael Sirivianos, participated in the discussion panel of the Media Literacy Day[6] on the **3rd of November 2023**, which was moderated by a representative from the CRTA.
- MedDMO co-organized a Press Conference event[7] with the Cyprus Radio Television Authority in the context of the MedDMO plenary meeting in Cyprus on the **14th of March 2024.** The Greek and Maltese Broadcasting authorities were invited to (virtually) attend the event.
- MedDMO worked together with CRTA to respond to the request for assistance with the Election-Watch.EU Election Assessment Mission 2024: Cyprus.
- MedDMO informed the CRTA for the Verian exercise to assess the CoP Signatories' reports No.2, and offered support for their potential participation as ERGA members.
- MedDMO organised a workshop on the fact-checking process for representatives of the CRTA staff at the Cyprus University of Technology premises on the **5th of July 2024**.
- CRTA suggested the production of a disinformation awareness video or a series of videos. CRTA offered to use its TV and radio channels to disseminate videos. The production of the videos is a work in progress.
- MedDMO offered a fact-checking process workshop for all the CRTA staff on their premises on the **29th of November 2024**. The workshop ended with a Q&A session, in which the CRTA staff had the opportunity to learn more about the MedDMO work, and further activities of collaboration were discussed.
- MedDMO and CRTA co-organized a series of workshops on the topic of fact-checking and disinformation awareness for journalists and media professionals in Cyprus:
    - On the **27th of March 2025**, the workshop for journalists and media professionals took place at the Antenna media organisation's premises.
    - On the **3rd of April 2024**, the workshop took place at the CRTA premises, inviting journalists and media professionals from all the local media organisations.
    - On the **22nd of May 2025**, the workshop took place at the RIK media organisation's premises for the organisation’s journalists and media professionals.

**Broadcasting Authority of Malta (MBA):**

- MedDMO informed the MBA of the Verian exercise to assess the CoP Signatories' reports No.2, and offered support for their potential participation as ERGA members.
- MedDMO invited the MBA to attend the Press Conference event co-organized by MedDMO and CRTA on the **14th of March 2024**.
- MedDMO and MBA co-organised a webinar for journalists, journalism students, and media professionals in Malta on the **19th of June 2024**. During the webinar, there was an introduction of the MedDMO

6 https://medialiteracy.pi.ac.cy/events/medialiteracy2023/?fbclid=IwAR0d7FBZ5M2uvqWyjLsSJzgWuTa8wtu6At_8OUsGbGVIERFF8Y2jv1FJZE0

7 https://www.socialcomputing.eu/2024/03/19/the-meddmo-press-conference-concluded-successfully/

project and then a series of case study presentations from MedDMO fact-checking partners: Times of Malta, Fact-Check-Cyprus, Ellinika Hoaxes, and Agence France Presse. The webinar concluded with a Q&A session, fostering discussion among journalists and fact-checkers.

- On the **16th of January 2025, MedDMO and MBA participated in the** Digital Services Forum, which was held online. At this event, MedDMO and EDMO initiatives were introduced.
- MedDMO and MBA co-organised a webinar[8] for journalists, media professionals, and the public audience in Malta on the **17th of June 2025**. The webinar began with an overview of the MedDMO project and its regional objectives and outcomes. This was followed by a presentation on the role and functions of the Media Literacy Development Board within Malta's national media literacy strategy. The webinar continued with an overview of fact-checking practices in Malta's media landscape and concluded with a presentation on how fake copyright complaints are being used to remove journalistic content from Google's search results. The event closed with an open discussion and Q&A session. The webinar recording is available online[9].

[8] https://meddmo.eu/webinar-disinformation-in-malta/

[9] https://www.youtube.com/watch?v=Qbw5uBzO3E8

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). 




*Figure 1: Snapshots of the MedDMO co-organized events with the national authorities of Greece, Cyprus, and Malta*

**MedDMO Collaboration with the Digital Service Coordinators of the three countries:**

MedDMO aims to establish collaboration with the Digital Service Coordinators (DSCs) in the three countries under the Digital Service Act. MedDMO is already collaborating with CRTA, Cyprus's DSC. MedDMO organised a meeting with the DSCs of Greece and Malta, the Hellenic Telecommunications and Post Commission (EETT) on the **25th of November 2024**, and the Maltese Communication Authority (MCA) on the **3rd of December 2024** accordingly. MedDMO presented its efforts and objectives within the broader context of EDMO to the DSCs, and a discussion followed for further collaboration. MCA invited the MedDMO coordinator for a 15-minute intervention at the Digital Services Forum, which was held online on the **16th of January 2025**. The Broadcasting Authority of Malta also participated in the event. In the second round of MedDMO, we expect to further engage with the DSCs of the three countries.

# 4 Conclusion

MedDMO's platforms' monitoring and supporting the authorities' work over the reporting period has reinforced its role as a key actor in the fight against disinformation in the Mediterranean region. Through systematic platform practices assessments and data-driven insights from fact-checking archives, the project has provided a clearer view of how disinformation circulates and how online platforms respond—or fail to respond—to it in practice. MedDMO's involvement in joint exercises for CoP monitoring has aligned its work with broader European standards.

MedDMO's collaborations with national authorities and Digital Services Regulators have laid the groundwork for more coordinated and informed responses to disinformation challenges. Awareness campaigns, joint workshops, and strategic support activities have deepened ties with regulators and highlighted the importance of localised expertise in policy and enforcement.

Looking ahead, MedDMO aims to build on these foundations by strengthening its cooperation with the Digital Services Coordinators, expanding the scope of its monitoring, and enhancing its methodologies, particularly in political advertising and automated detection. The knowledge and tools developed during this phase will directly inform the continuation of the project in MedDMO II, supporting an even more robust ecosystem for disinformation resilience in Cyprus, Greece, and Malta.

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu

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



MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu

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. The content of this document is © the author(s).
For further information, visit www.meddmo.eu

# Copyright

© Copyright 2022 MedDMO Consortium

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the MedDMO Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

## 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

Executive Summary....8
1 Introduction....9
2 Methodology for Monitoring the platforms practices....10
3 Summary of Results....13
3.1 Meta (Facebook and Instagram)....19
3.1.1 II. Scrutiny of Ad Placements....19
3.1.2 III. Political Advertising....22
3.1.3 V. Empowering Users....27
3.1.4 VI. Empowering the Research Community....36
3.1.5 VII. Empowering the fact-checking community....38
3.2 Google (Ads, Search, YouTube)....40
3.2.1 II. Scrutiny of Ad Placements....40
3.2.2 III. Political Advertising....44
3.2.3 V. Empowering Users....46
3.2.4 VI. Empowering the Research Community....56
3.2.5 VII. Empowering the fact-checking community....57
3.3 TikTok....61
3.3.1 II. Scrutiny of Ad Placements....61
3.3.2 III. Political Advertising....63
3.3.3 V. Empowering the users....66
3.3.4 VI. Empowering the Research Community....79
3.3.5 VII. Empowering the fact-checking community....80
4 MedDMO fact-check activities....82
4.1 Analysis of MedDMO fact-checks....82
4.2 MedDMO Fact-Checking Partners Collaboration with VLOPs....85
5 Research Activities: Towards Automatic Disinformation Detection in online social platforms....89
5.1 HyperGraphDis - Disinformation Detection on Twitter with Graph Neural Networks....89
5.2 Disinformation Detection on YouTube: with Large Language Models (LLMs)....90
6 Policies to Regulate Disinformation in Cyprus, Greece, and Malta....90
7 Supporting the national authorities....93
8 Conclusions....95
9 References....96

Annex I: Questionnaire for MedDMO fact-checking organisations ..... 97

Annex II: Platforms' Data Repositories ..... 98

## Index of Figures

Figure 1: Meta -- Average Scores per pillar ..... 13
Figure 2: Meta -- Average Scores per pillar and country ..... 13
Figure 3: Google -- Average Scores per pillar ..... 13
Figure 4: Google -- Average Scores per pillar and per country ..... 13
Figure 5: TikTok -- Average Scores per pillar ..... 14
Figure 6: Average Scores per pillar and per country ..... 14
Figure 7: Examples of political ad labelling from Meta Ad Library ..... 22
Figure 8: Meta's Ad Library snapshot when searching for ads in Malta for the period of January-June 2023 ..... 25
Figure 9: Meta's Ad Library example of the information given per ad ..... 25
Figure 10: Meta's Ad Library example of the European Union Transparency related information per ad ..... 26
Figure 11: Example of Fact-checked post that contains misinformation warning (on Facebook in Greek). ..... 30
Figure 12: Option to view the fact-check article and to anyway view the post (on Facebook in Greek.). ..... 30
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). ..... 30
Figure 14: When user clicks to "Share" a post that contains misinformation (on Facebook in Greek). ..... 30
Figure 15: Misinformation warning screens on Instagram in English. ..... 31
Figure 16: Examples of political or issue ads labelling can be found in Google's Ads Transparency Centre[24] ..... 45
Figure 17: Google's "About this result" examples ..... 47
Figure 18: ClaimReview schema Googles Example from page ..... 53
Figure 19: Examples of Ellinika Hoaxes and AFP using ClaimReview schema ..... 53
Figure 20: YouTube Information Panels description (screenshot) ..... 54
Figure 21: TikTok's reported quantitative information for SLI 1.1.1 ..... 63
Figure 22: Screenshots of TikTok Commercial Content library when searching for ads displayed in Greece for the period of January to June 2023 ..... 65
Figure 23: Screenshot of TikTok Commercial Content Library when searching for ads targeting Cyprus and Malta ..... 65
Figure 24: Example of TikTok's Unverified Content Label taken from TikTok CoP report, No2 ..... 67
Figure 25: Screenshot of TikTok's media literacy campaign for Greek elections 2023 from TikTok CoP report No2 ..... 71
Figure 26: MedDMO collaboration with the national media authorities in Cyprus, Malta, and Greece ..... 93

# Index of Tables

Table 1: CoP measures covered by this analysis ..... 11
Table 2: CoP measure-level assessment scaling system ..... 12
Table 3: Overall evaluation per platform and country. Average scores per CoP pillar. ..... 14
Table 4: Meta's CoP Report Summary of Assessment Results ..... 15
Table 5:Google's CoP Report Summary of Assessment Results ..... 16
Table 6: TikTok's CoP Report Summary of Assessment Results ..... 17
Table 7: Missing Service Level Indicators (SLIs) information in Signatories Reports ..... 18
Table 8: Meta's reported quantitative information for SLI 2.1.1 ..... 21
Table 9: Meta's reported quantitative information for SLI 6.2.1 ..... 23
Table 10: Meta's reported quantitative information for SLI 18.1.1 ..... 32
Table 11: Meta's reported quantitative information for SLI 18.2.1 ..... 33
Table 12: Meta's reported quantitative information for SLI 21.1.2 ..... 35
Table 13: Meta's reported quantitative information for SLI 31.1.3 ..... 40
Table 14: Google's reported quantitative information for SLIs 1.1.1 and 1.1.2 ..... 42
Table 15: Google's reported quantitative information for SLI 2.1.1 ..... 44
Table 16: Google's reported quantitative information for SLI 6.2.1 ..... 46
Table 17: Google's reported quantitative information for SLI 17.1.1 ..... 49
Table 18: Google's reported quantitative information for SLI 17.1.1 and 17.2.1 ..... 50
Table 19: Google's reported quantitative information for SLI 18.2.1 ..... 52
Table 20: Google's reported quantitative information for 21.1.1 ..... 54
Table 21: Google's reported quantitative information for SLI 24.1.1 ..... 56
Table 22: Google's reported quantitative information for SLI 26.2.1 ..... 57
Table 23: Google's reported quantitative information for SLI 31.1.1 ..... 60
Table 24: TikTok's reported quantitative information for SLI 17.1.1 ..... 69
Table 25: TikTok's reported quantitative information for SLIs 18.1.1 and 18.2.1 ..... 75
Table 26: TikTok's reported quantitative information for SLI 21.1.1 ..... 77
Table 27: TikTok's reported quantitative information for SLI 21.1.2 ..... 77
Table 28: TikTok's reported quantitative information for SLI 24.1.1 ..... 78
Table 29: TikTok's reported quantitative information for SLIs 31.1.1 – 3 ..... 82
Table 30: Main findings from the MedDMO fact-checks analysis ..... 85
Table 31: MedDMO Fact-Checking Partners Collaboration with Meta ..... 87
Table 32: MedDMO Fact-Checking Partners Collaboration with TikTok ..... 88
Table 33: MedDMO Fact-Checking Partners Collaboration with Google and other platforms ..... 89
Table 34: Policies to Regulate Disinformation in Cyprus, Greece, and Malta ..... 93

# 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**

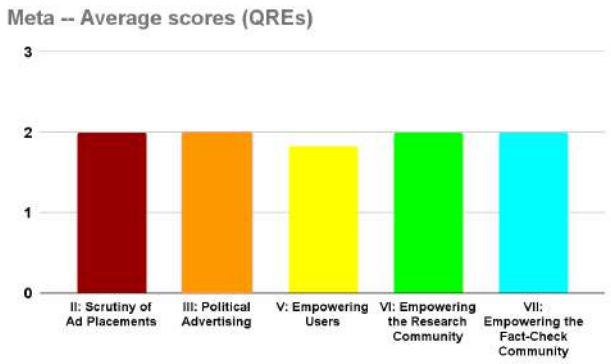



Figure 1: Meta -- Average Scores per pillar

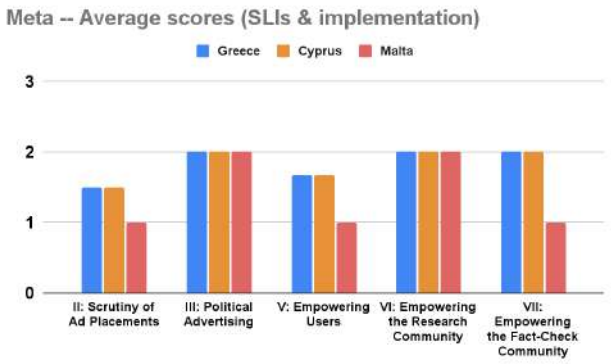



Figure 2: Meta -- Average Scores per pillar and country

**Google**

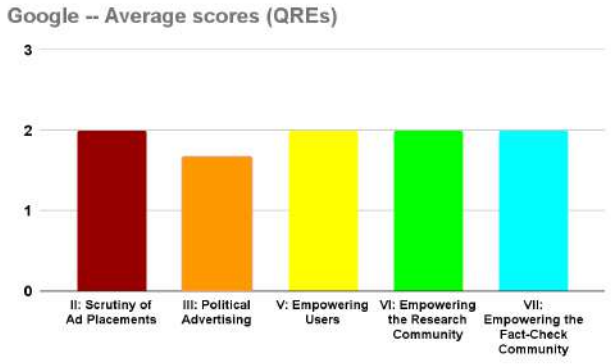



Figure 3: Google -- Average Scores per pillar

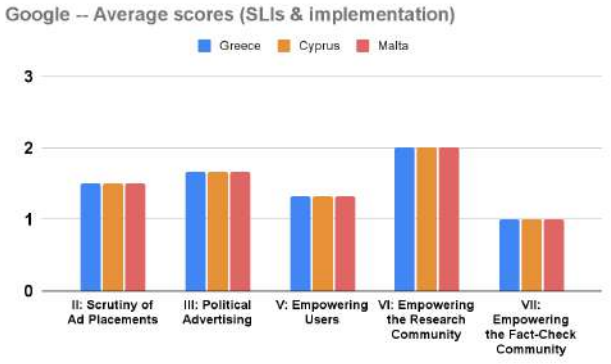



Figure 4: Google -- Average Scores per pillar and per country

**TikTok**

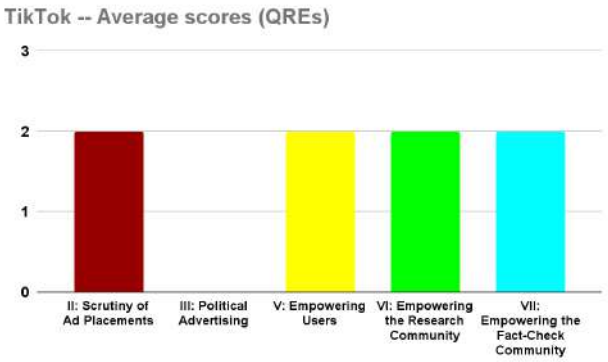



Figure 5: TikTok -- Average Scores per pillar *

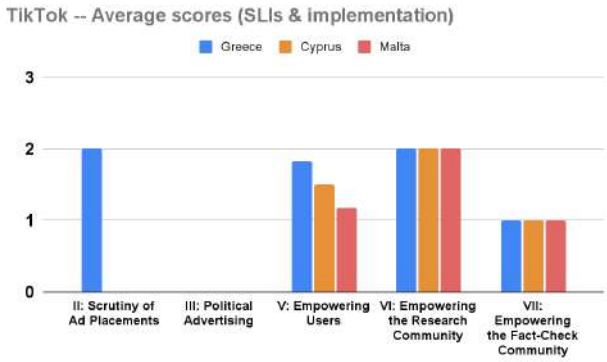



Figure 6: Average Scores per pillar and per country **

* 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<br>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<br>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<br>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<br>✓ - reported SLI<br>x - no SLI reported<br>N/S - platform/service did not subscribe to the relevant Measure<br>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*** <br> ***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**<br>***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<br>*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).

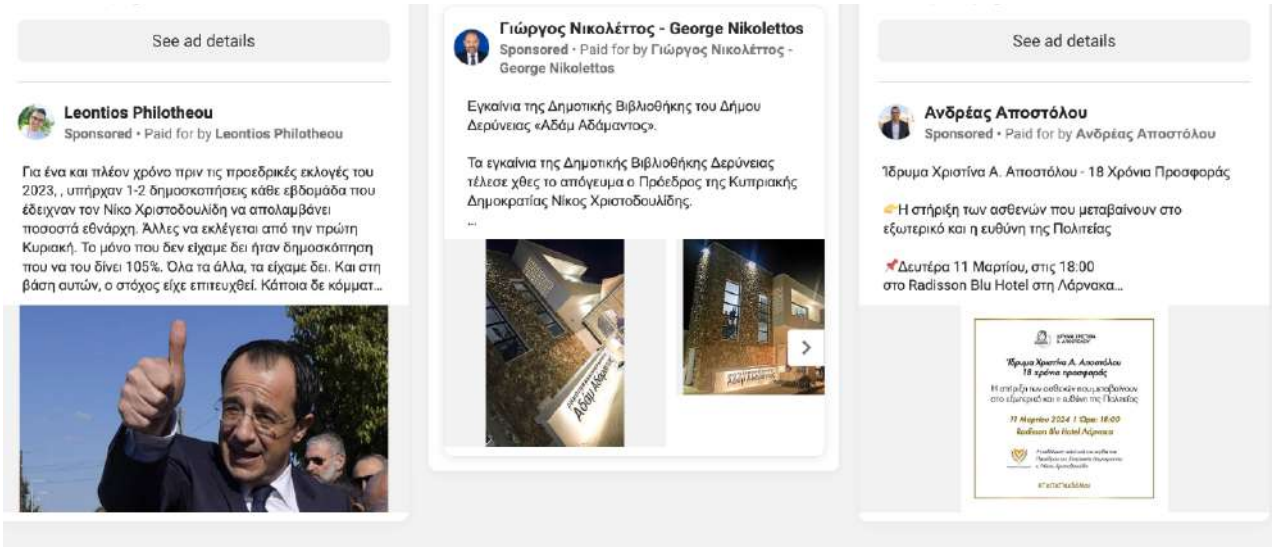



Figure 7: Examples of political ad labelling – Sponsored, ''paid for by'' disclaimers as obtained from Meta Ad Library

**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<br>***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/

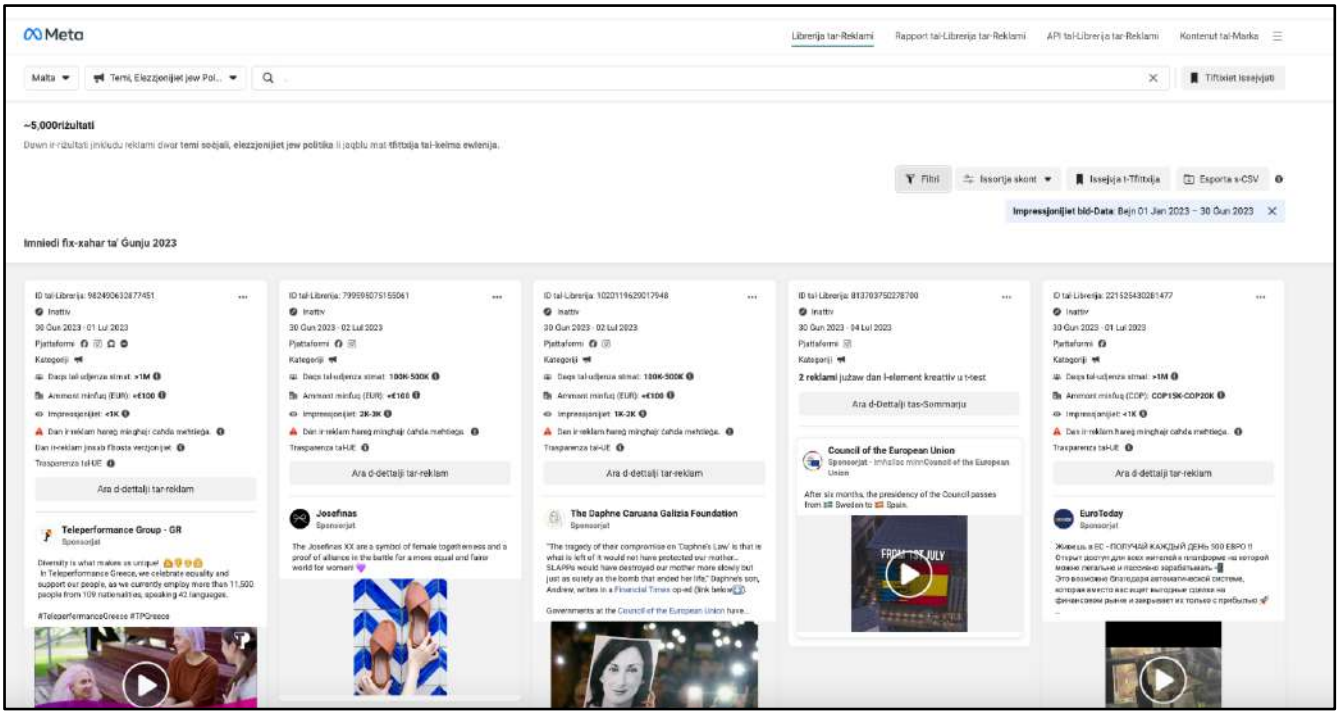



Figure 8: Meta's Ad Library snapshot when searching for ads in Malta for the period of January-June 2023

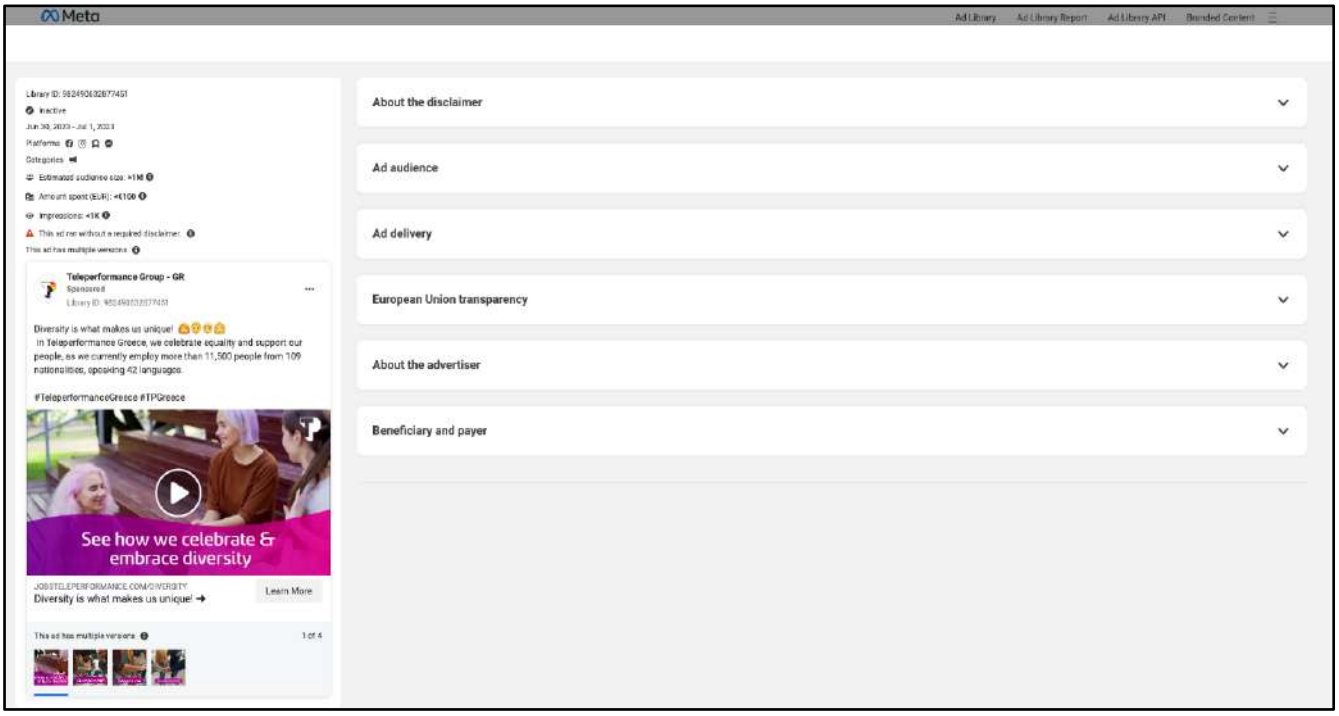



Figure 9: Meta's Ad Library example of the information given per ad

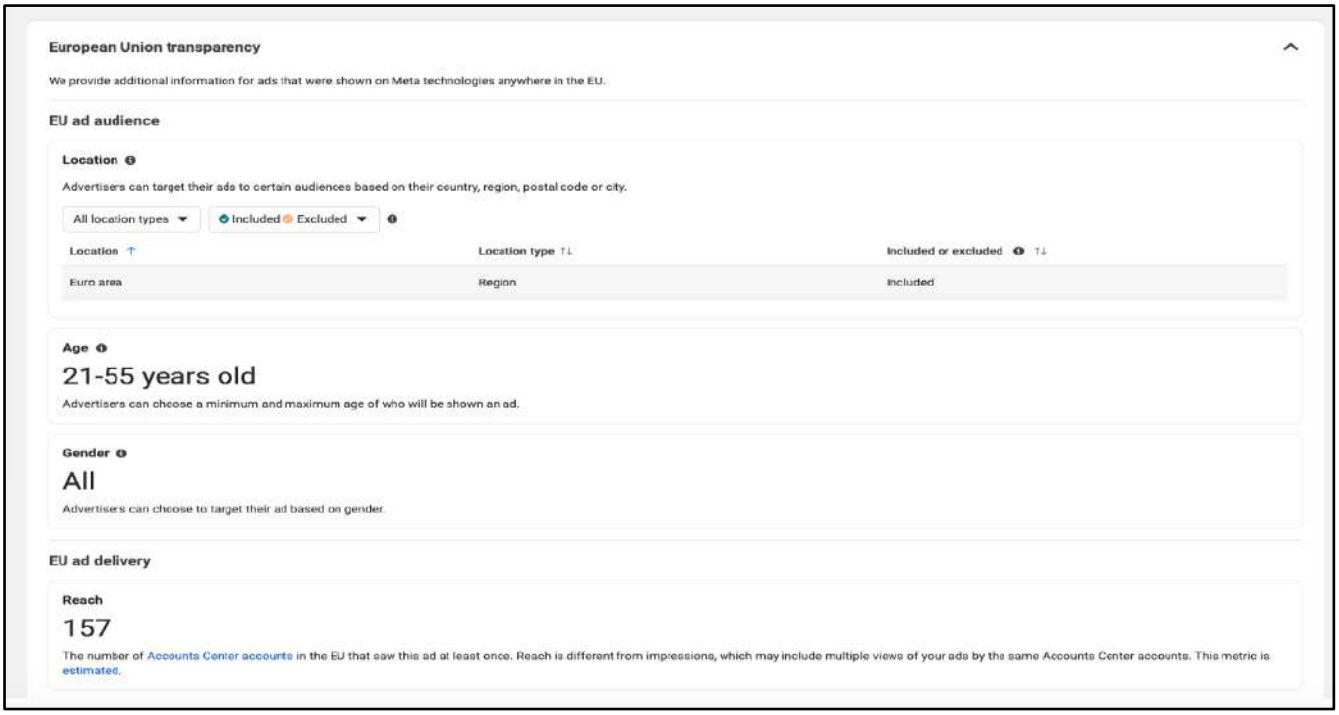



Figure 10: Meta's Ad Library example of the European Union Transparency related information per ad

**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**<br>***Commitment 17, Measure 17.1*** *QRE 17.1.1, SLI 17.1.1. &* ***Measure 17.2*** *QRE 17.2.1, SLI 17.2.1*<br>*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<br>*Commitment 18, **Measure 18.1** QRE 18.1.1, SLI 18.1.1. & **Measure 18.2** QRE 18.2.1, SLI 18.2.1*<br>*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.

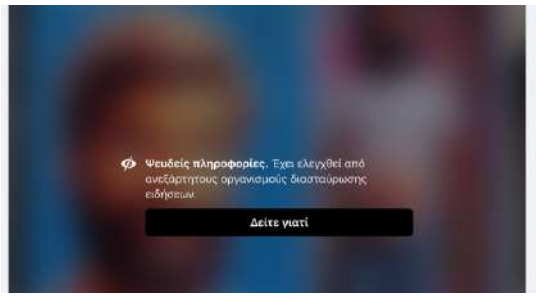



Figure 11: Example of Fact-checked post that contains misinformation warning (on Facebook in Greek).




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).

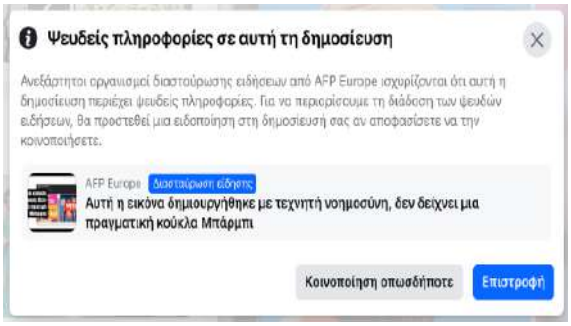



Figure 14: When user clicks to “Share” a post that contains misinformation (on Facebook in Greek).

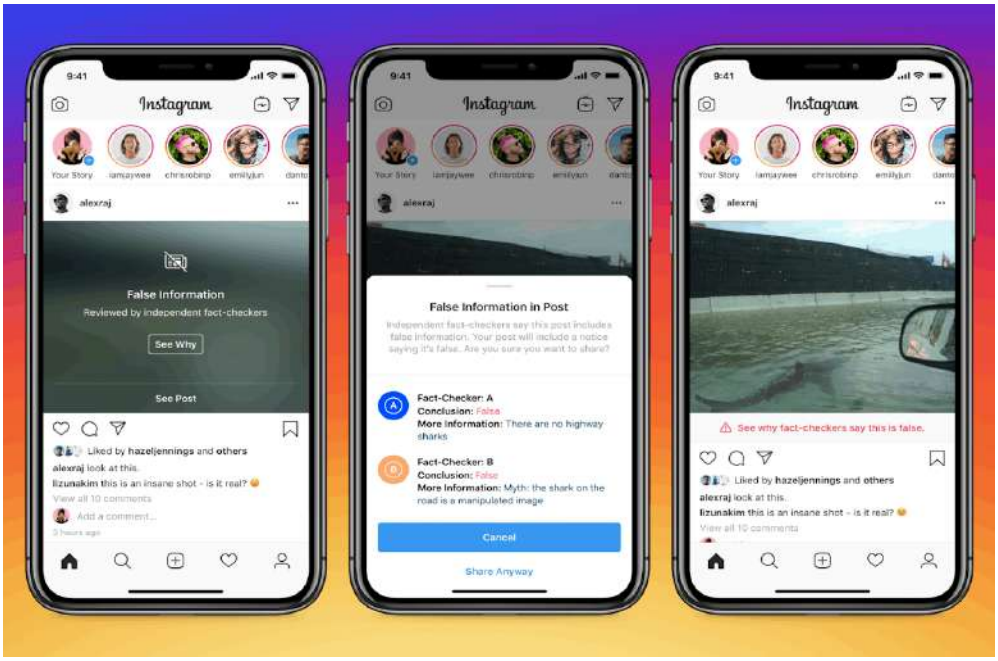



Figure 15: Misinformation warning screens on Instagram in English. Source: https://petapixel.com/2019/12/17/instagrams-new-false-information-warning-will-tell-you-if-a-photo-is-fake/

**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**<br>***Commitment 26, Measure 26.2.1*** *QRE 26.2.1, SLI 26.2.1.*<br>*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*** <br> ***Commitment 31, Measure 31.1*** *QRE 31.1.1, SLI 31.1.1. & **Measure 31.2.** QRE 31.2.1, SLI 31.2.1* <br> *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*** <br> ***Commitment 1, Measure 1.1*** <br> *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**<br>***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).**<br><br>**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.**<br><br>**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.**<br><br>**Account suspension - Google Ads Accounts may be suspended if Google detects violations of its policies or the Terms and Conditions.**<br><br>**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.

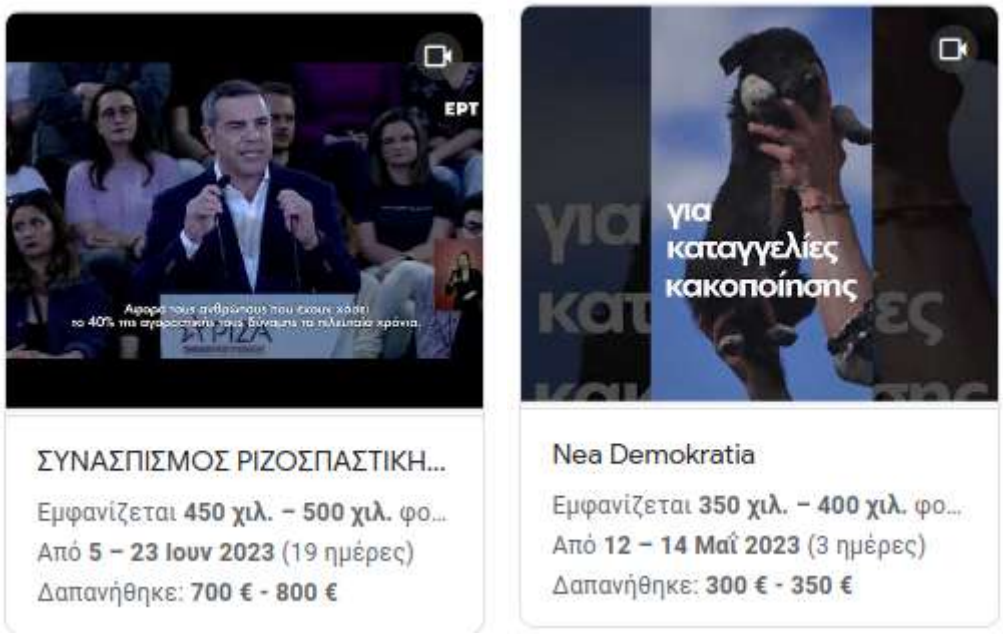



Figure 16: Examples of political or issue ads labelling can be found in Google's Ads Transparency Centre[24].

**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);**<br>**(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**<br>*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.

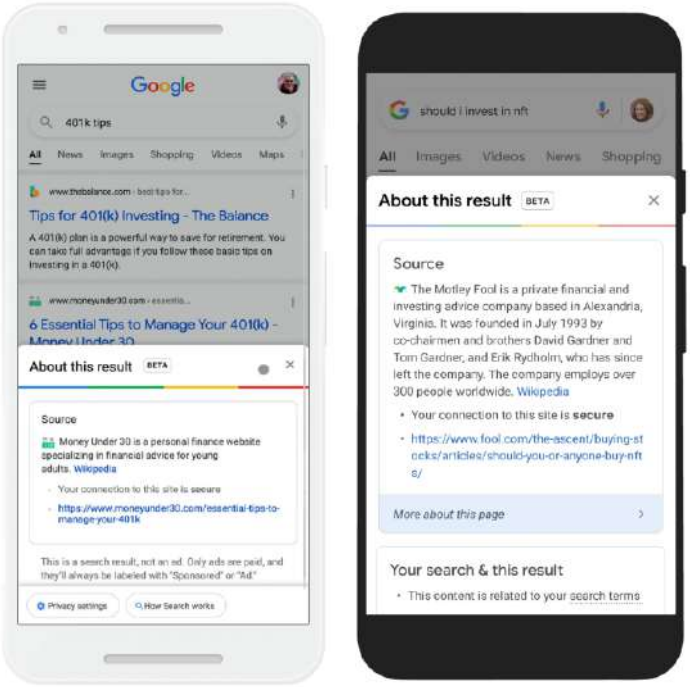



Figure 17: Google's "About this result" examples

***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:**<br>(1) **low relevance results i**n H1 2023 (1 January 2023 to 30 June 2023), broken down by EEA Member State;<br>(2) **rapidly changing results** in H1 2023, broken down by EEA Member State;<br>(3) **potentially unreliable** sets of results in H1 2023, broken down by EEA Member State;<br>*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.<br>**Number of times the:**<br>(4) **'More About This Page'** feature was viewed in H1 2023, broken down by EEA Member State;<br>(5) **'Source' section of the 'About This Result'** panel was viewed in H1 2023, broken down by EEA Member State;<br>(6) **'Your Search and this result' section of the 'About This Result'** panel was viewed in H1 2023, broken down by EEA Member State;<br>(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**<br>***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*<br>**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*<br>*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.

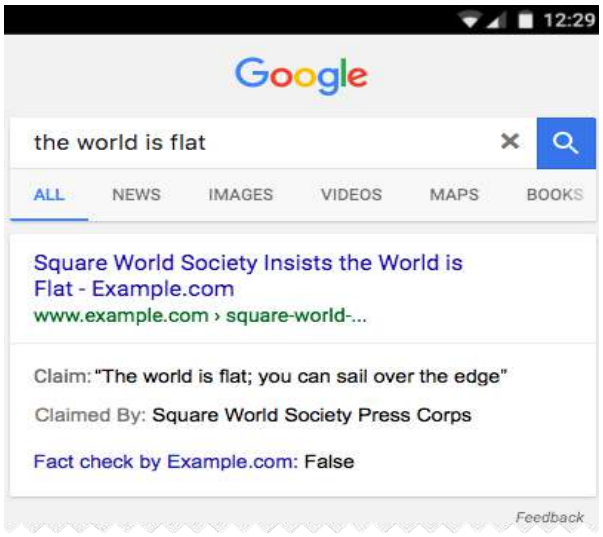



Figure 18: ClaimReview schema Googles Example from page




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

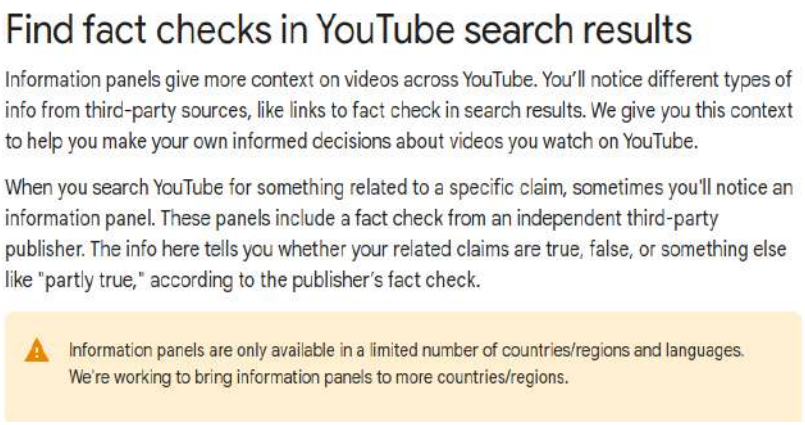



Figure 20: YouTube Information Panels description (screenshot)

**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***<br>***Commitment 24, Measure 24.1.1*** *QRE 24.1.1, SLI 24.1.1. page 151-155*<br>**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;**<br>**(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**<br>***Commitment 26, Measure 26.2.1*** *QRE 26.2.1, SLI 26.2.1.*<br>*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*** <br> ***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*** <br> ***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.<br>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.<br>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**<br>***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**<br>***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

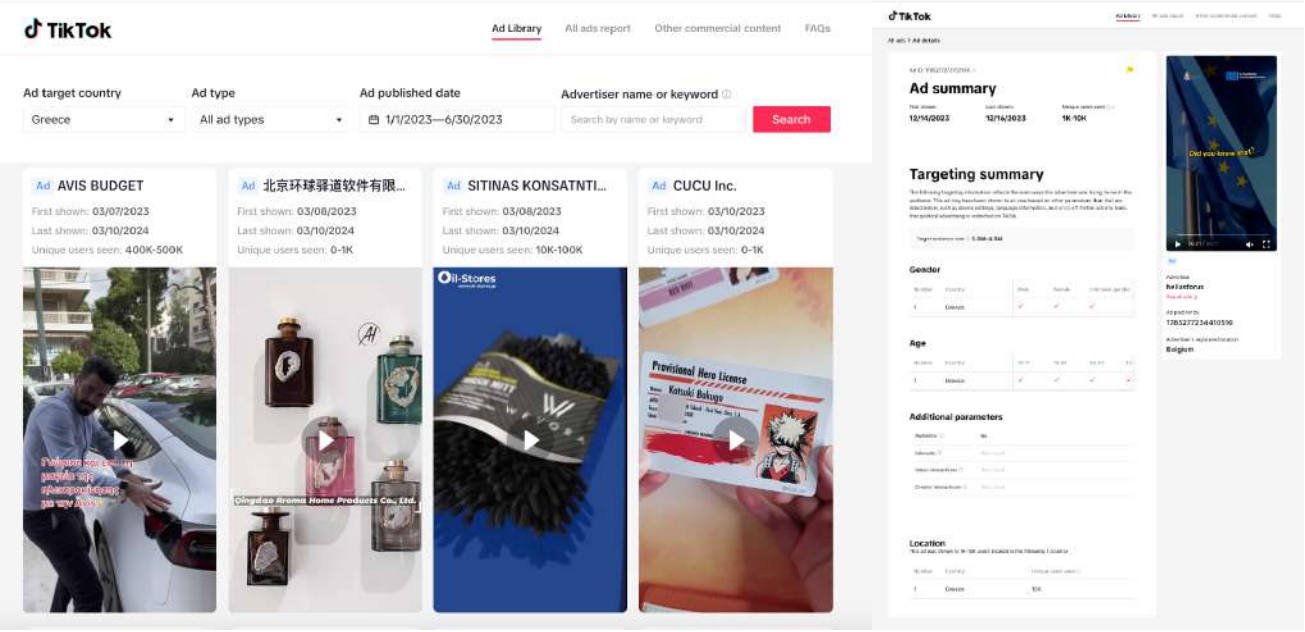



Figure 22: Screenshots of TikTok Commercial Content library when searching for ads displayed in Greece for the period of January to June 2023

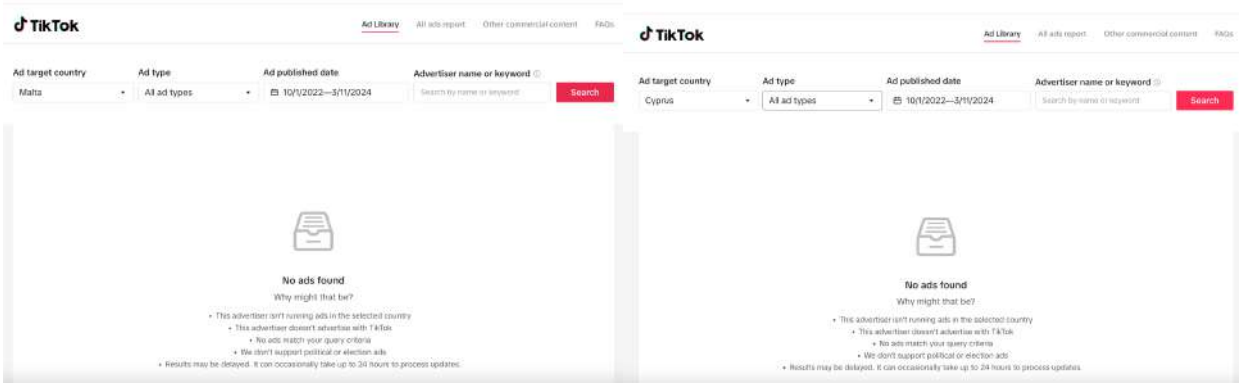



Figure 23: Screenshot of TikTok Commercial Content Library when searching for ads targeting Cyprus and Malta

**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**<br>***Commitment 17, Measure 17.1*** *QRE 17.1.1, SLI 17.1.1. &* ***Measure 17.2*** *QRE 17.2.1, SLI 17.2.1*<br>*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.

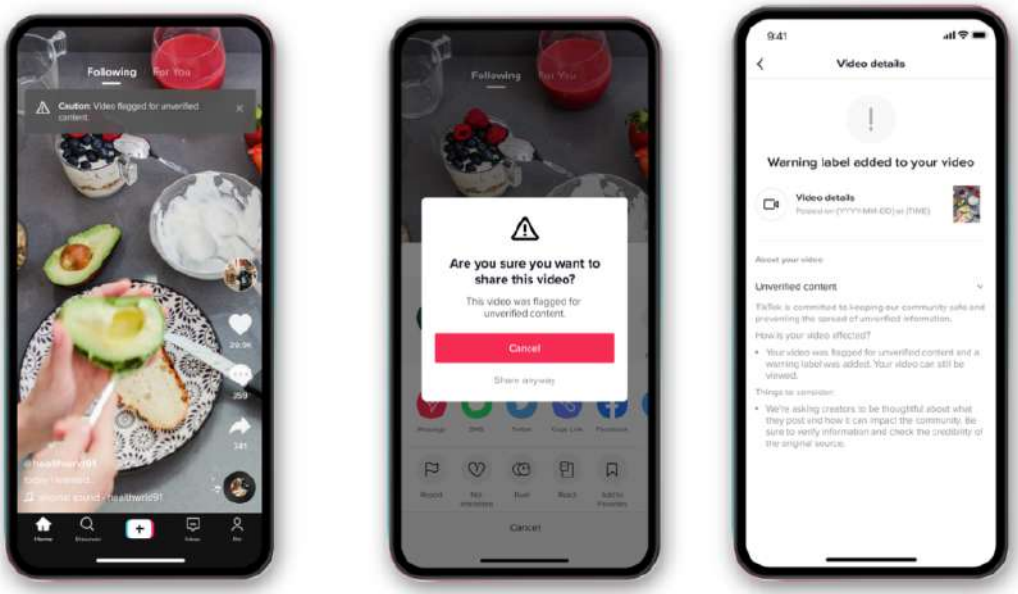



Figure 24: Example of TikTok's Unverified Content Label taken from TikTok CoP report, No2

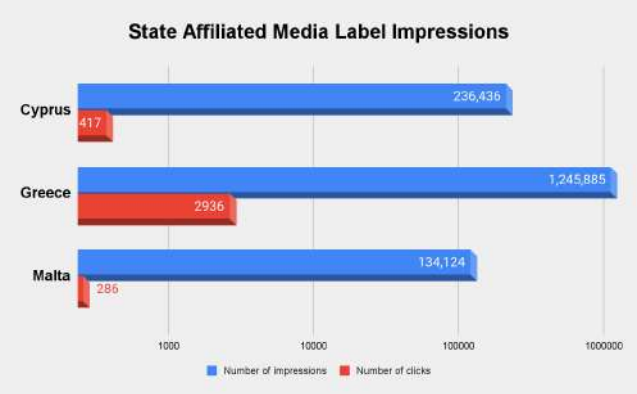
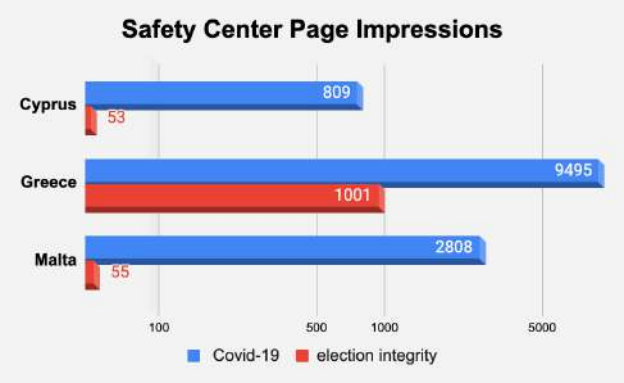
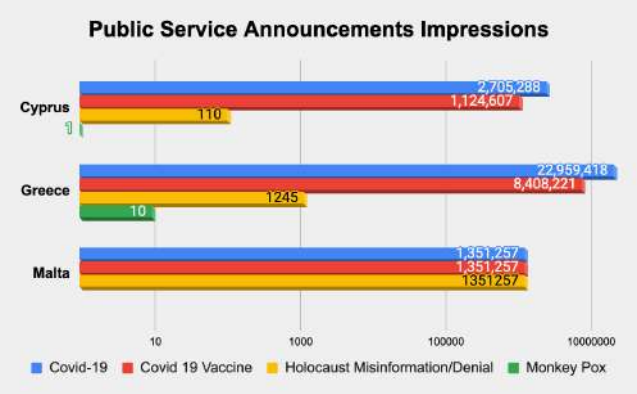
[45] https://newsroom.tiktok.com/fil-ph/know-the-facts

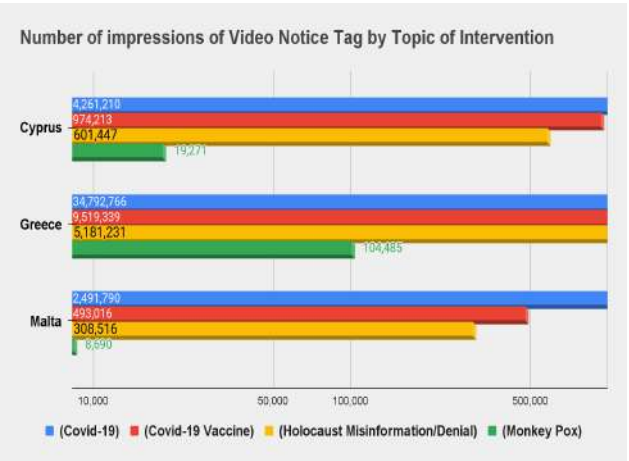
[46] https://www.tiktok.com/tns-inapp/pages/state-affiliated-media

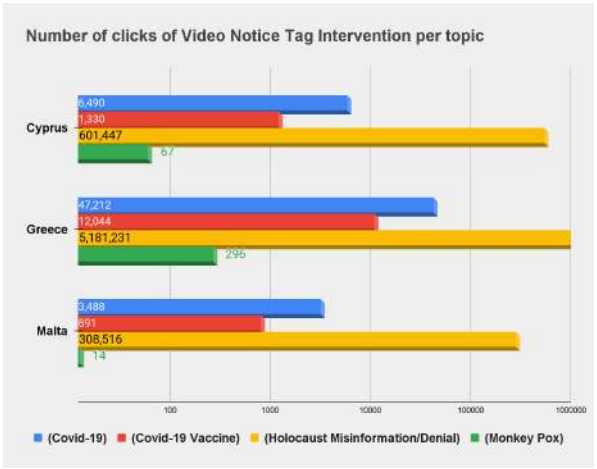
**TikTok reported (in SLI 17.1.1.) the following metrics –** *number of impressions, clicks, clicks through rate* **– for assessing the impact of the following in-app tools to empower the users:**

- 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

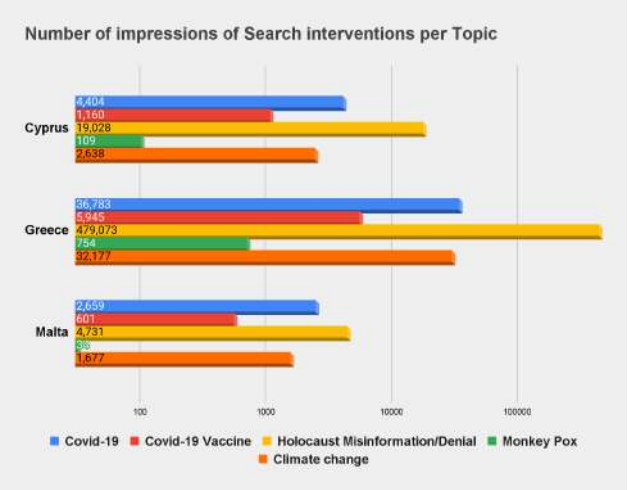
**The TikTok reported information for SLI 17.1.1 can be found in the TikTok's July 2023 report pages 84-104, next we present the reported numbers in figures:**

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




**Search Intervention**



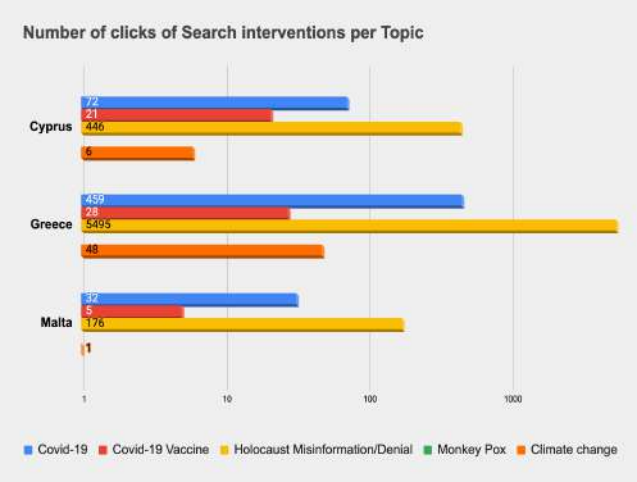




Table 24: TikTok's reported quantitative information for SLI 17.1.1

**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

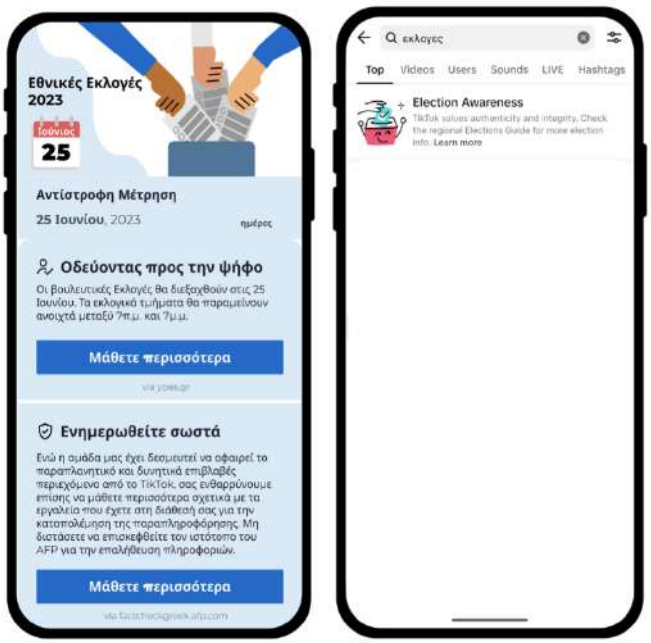



Figure 25: Screenshot of TikTok's media literacy campaign for Greek elections 2023 from TikTok CoP report No2

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**<br>***Commitment 18, Measure 18.1*** *QRE 18.1.1, SLI 18.1.1. &* ***Measure 18.2*** *QRE 18.2.1, SLI 18.2.1*<br>*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**<br>**Methodology of data measurement:**<br>**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**<br>**Methodology of data measurement:**<br>**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.**<br>**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***<br>***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><br>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.<br>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:**<br>**The number of videos tagged with the unverified content label is based on the country in which the video was posted.**<br>**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***<br>***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*<br>*Commitment 26, Measure 26.2.1 QRE 26.2.1, SLI 26.2.1.*<br>*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***<br>***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*<br>*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

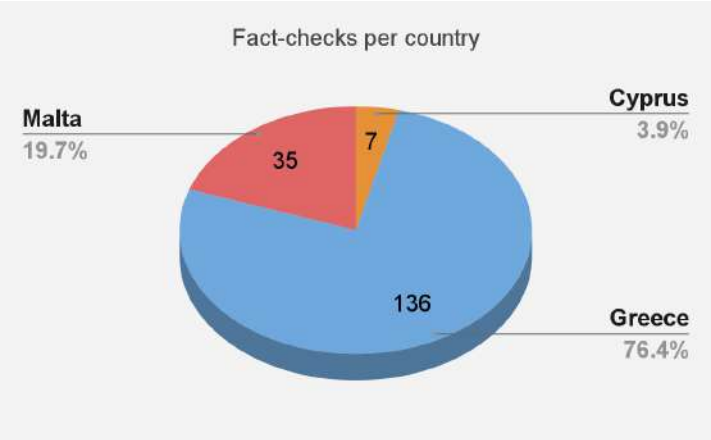
**How we derived if the content was labelled by the platform:**

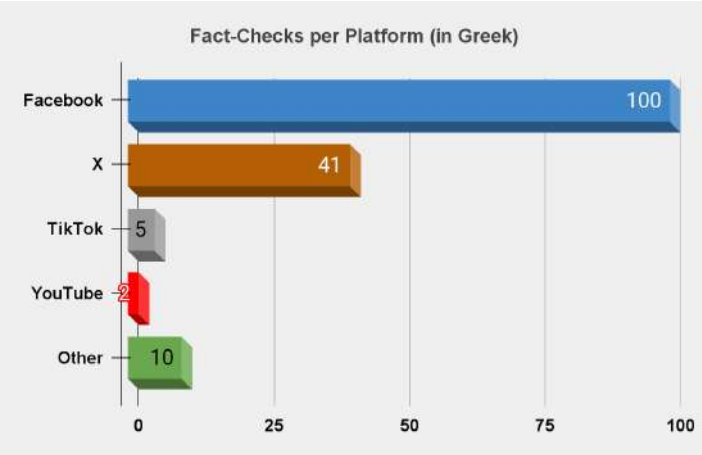
In cases of active content on platform (in cases where the fact-check articles contained links to the posts, etc.) we could verify if the content was treated with a label or not

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

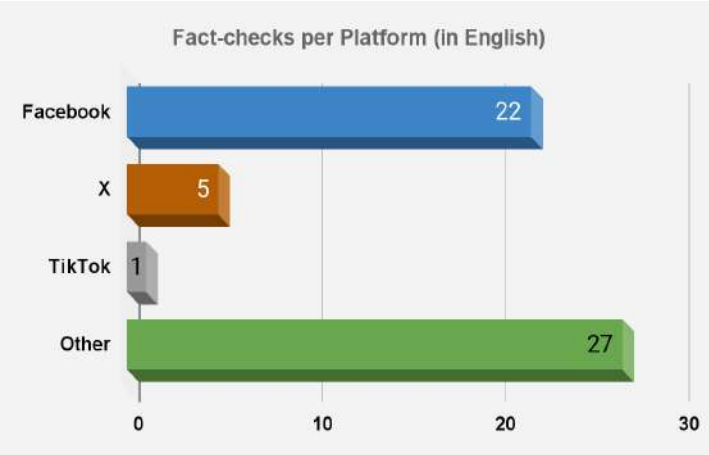
If active content was labelled and there was reference to the fact-checking organisation article for the specific topic.

| **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.** |

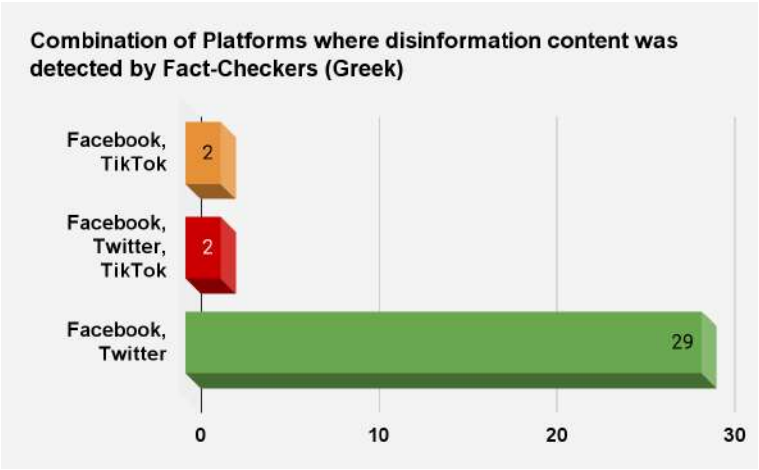


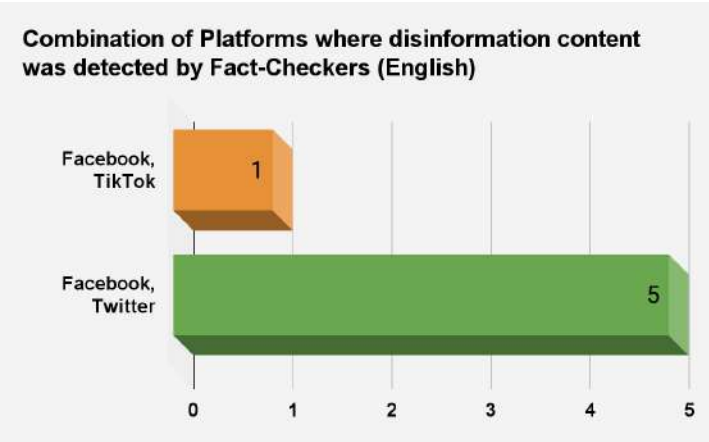



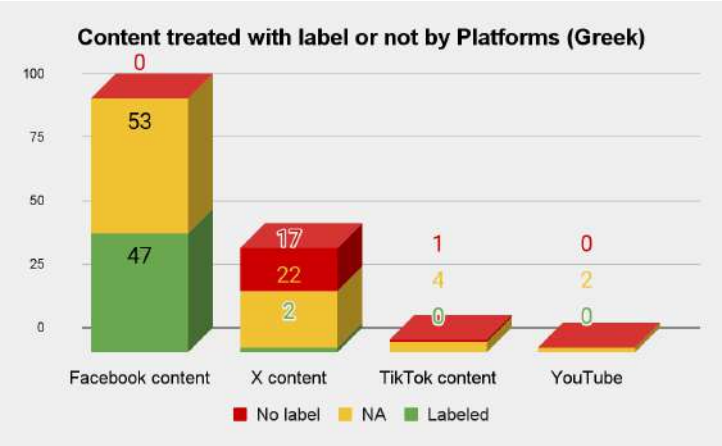

*__Other__ refers to online news outlets



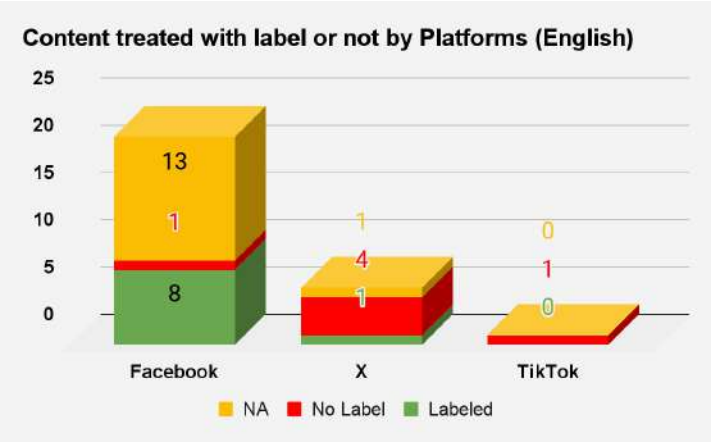

*__Other__ refers to online news outlets



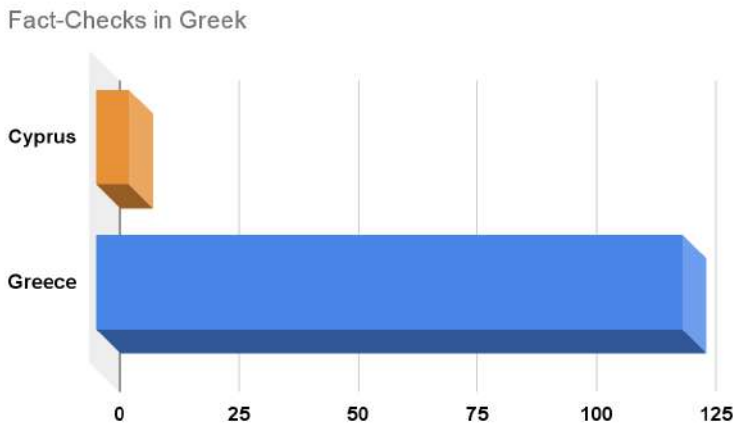






**NA:**




**NA:**



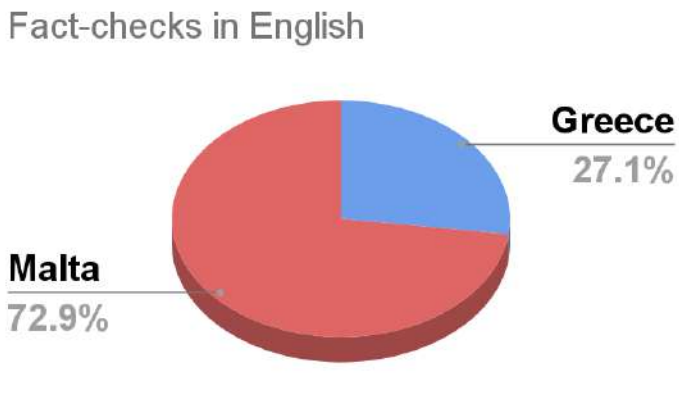




Table 30: Main findings from the MedDMO fact-checks analysis

## 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<br>**Greece:** legislative<br>**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.<br>**Greece:** Yes<br>**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.<br>**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.<br>**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.<br>**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.<br>**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.<br>**Cyprus:** Yes, see previous responses.<br>**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<br>**Cyprus:** Press Law does not deal with disinformation.<br>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]<br>(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<br>(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).<br>**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<br>**Cyprus:** No<br>**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<br>**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.<br>**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.<br>**Cyprus:** dissemination of fake news, false news, inaccurate publications, no definition is used, only the common terms.<br>**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<br>**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.

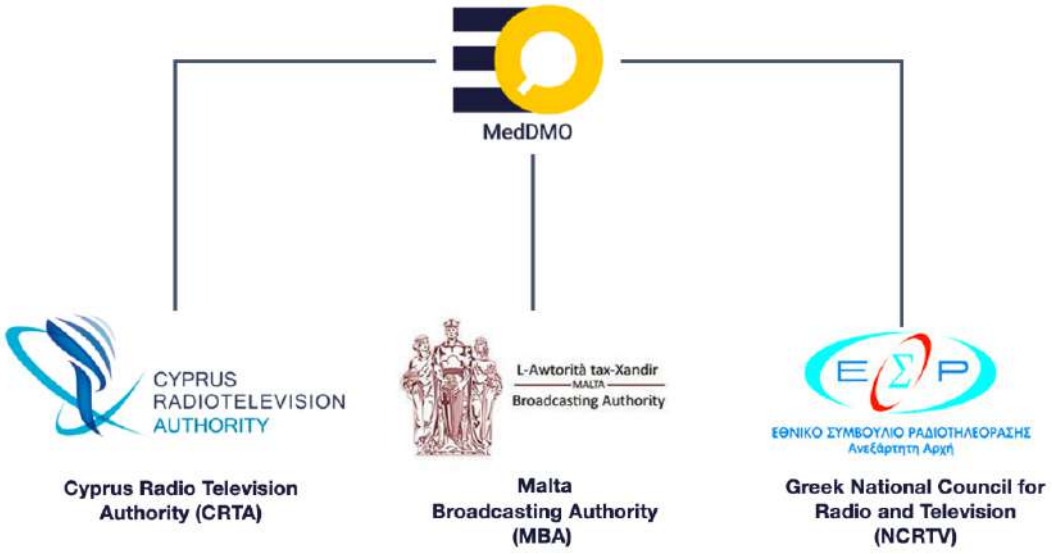



Figure 26: MedDMO collaboration with the national media authorities in Cyprus, Malta, and Greece

**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**<br>https://www.facebook.com/ads/library/ | Meta Content library and API<br>https://developers.facebook.com/docs/content-library-api/ |
| **TikTok** | **TikTok Commercial Content API**<br>https://developers.tiktok.com/products/commercial-content-api | **TikTok Research API**<br>https://developers.tiktok.com/products/research-api/ |
| **Google** | **Google Ads Transparency Centre**<br>https://adstransparency.google.com/political | **YouTube:** https://research.youtube/<br>**Google Trends:** https://trends.google.com/trends/<br>**Fact-check Explorer:**<br>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. The content of this document is © the author(s).
For further information, visit www.meddmo.eu

# Annex II: Mapping Disinformation Circulation in Cyprus, Greece, and Malta and Evaluating Platforms' Moderation Practices: Insights from MedDMO Fact-Checks Archive - Part A of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta)



MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu



Mediterranean Digital Media Observatory

# D5.1

## MedDMO Fact-Checks Archive Analysis for monitoring the media landscape in Cyprus, Greece, and Malta

## 2nd Report of platform practices and national authorities support 2024
## Part A

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

| | |
|---|---|
| **Deliverable title** | 2nd Report of platform practices and national authorities support –2024 – Part A |
| **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, Paschalia Spyridou, Michael Sirivianos |
| **Reviewer(s)** | CERTH, UCY |
| **EC Project Officer** | Mark Vella Muscat |


| | |
|---|---|
| **Abstract** | This is the MedDMO annual report detailing the results of monitoring online platform practices to support the national authorities of Cyprus, Greece, and Malta in combating active disinformation campaigns. This report presents findings from our analysis of the MedDMO fact-check archive (1 January–30 June 2024). We mapped the channels—social media platforms, local news outlets, and other online or offline channels—through which disinformation circulated in Cyprus, Greece, and Malta, and evaluated how each platform's moderation tools (warning labels, removals, context panels, and community annotations) were deployed in response. These insights will both inform our assessment of the platforms' Code of |



MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756.


The content of this document is © the author(s).


For further information, visit www.meddmo.eu

| | |
|---|---|
| | Practice on Disinformation reports and strengthen ongoing monitoring of the media landscape across all three countries. |
| **Keywords** | Monitoring Social media platforms, disinformation campaign, fact-checking, Code of Practice on Disinformation |

## Copyright

© Copyright 2022 MedDMO Consortium

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the MedDMO Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

## Revision History

| Version | Date | Modified By | Comments |
|---|---|---|---|
| V0.1 | 01/11/2024 | CUT | First Draft Table of Contents |
| V0.2 | 08/11/2024 | CUT | Final Table of Contents |
| V0.3 | 02/12/2024 | CUT | First Draft |
| V0.4 | 18/01/2025 | CUT | Second Draft |
| V0.5 | 25/02/2025 | CUT | Final Second Draft |
| V0.6 | 25/04/2025 | CUT | Final Second Draft Review |
| V0.7 | 05/05/2025 | AUTH, UCY | Review |
| V1.0 | 12/05/2025 | CUT | Final Document |

## Glossary

| Abbreviation | Meaning |
|---|---|
| CoP | Code of Practice on Disinformation |
| VLOPs | Very Large Online Platforms |
| VLOSEs | Very Large Online Search Engines |
| AFP | Agence France-Presse |
| EH | Ellinika Hoaxes |
| ToM | Times of Malta |
| FCC | Fact-Check Cyprus |
| UoM | University of Malta |

# Table of Contents


- ***Executive Summary*** ***6***

***1 Introduction*** ***7***

***2 A Methodology for Leveraging the MedDMO Fact-Check Archive in Platform's Practices and Media Landscape Monitoring*** ***8***

***3 Analysis of Greece and Cyprus Fact-Checks*** ***12***

***4 Analysis of Malta Fact-Checks*** ***22***

***5 Conclusions*** ***26***

***6 References*** ***28***

# Index of Figures

Figure 1: Examples of Meta applying labels to content with false information to empower users (left: Facebook, right: Instagram) 10
Figure 2: Examples of TikTok and X mechanisms to empower users. TikTok's label on unverified content on the left and X's Community Notes on the right 11
Figure 3: The proportion of Fact-Checks published by each of the three fact-checking organizations (Fact-Check Cyprus, Ellinika Hoaxes, and Agence France-Presse) covering Greece and Cyprus in our dataset for the period of January 1st to June 30th, 2024. 12
Figure 4: Fact-Check Topics for Greece and Cyprus for the period of this analysis 13
Figure 5: Platforms mentioned in Greece and Cyprus' Fact-Checks 14
Figure 6: Combination of Platforms mentions in Greece and Cyprus' Fact-Checks 14
Figure 7: Number of instances containing false claims per platform found in Greece and Cyprus' Fact-Checks 15
Figure 8: Meta (Facebook)'s cases with inconsistent labelling/moderation 18
Figure 9: Example of Facebook's inconsistency in labelling content with identical false claims 18
Figure 10: Word cloud of Greek and Cypriot Online News Outlets mentioned in the Fact-Checks for the period under analysis 19
Figure 11: Online news outlets in Greece and Cyprus, ranked by the number of times they were cited in fact-checks, with markers indicating whether their content underwent any moderation actions. 21
Figure 12: The proportion of Fact-Checks published by each of the two fact-checking organizations covering Malta in our dataset for the period of January 1st to June 30th, 2024. 22
Figure 13: Fact-Check Topics for Malta for the period of this analysis 23
Figure 14: Platforms mentioned in Malta's Fact-Checks 24
Figure 15: Number of instances containing false claims per platform found in Malta's Fact-Checks 24
Figure 16: Findings summary of Malta's Fact-Checks analysis 25

# Index of Tables

Table 1: Platforms' moderation applied to instances containing false claims 17

## • Executive Summary

This report examines over 1,200 instances of false or misleading claims circulating in Cyprus, Greece, and Malta between January and June 2024, as documented by MedDMO's fact-checking partners. It aims to measure how these flagged claims were treated by very large online platforms and by online news outlets that shared them. For each fact-checked item, we tracked platform responses—whether a warning label was applied, content removed, contextual information provided, or no visible action taken—as well as any subsequent corrections or clarifications issued by the news sites themselves. Our findings show significant variation in how platforms handle identical content and reveal that news outlets seldom follow up with corrections, leaving many debunked claims accessible without context. These patterns underscore gaps in the practical enforcement of the EU Code of Practice on Disinformation and inform MedDMO's ongoing Task 5.1 monitoring efforts.

# 1 Introduction

Disinformation monitoring requires not only the identification of false claims but also rigorous tracking of how those claims are managed once exposed. The EU Code of Practice on Disinformation (CoP) **[European Commission, 2022]** is a voluntary, self-regulatory framework through which major online platforms pledge to curb the spread of false or misleading content. Its core objectives are to empower users with clear labelling and context, boost the visibility of authoritative sources, remove or demote harmful disinformation, and ensure greater transparency and collaboration with independent fact-checkers. Under the EU Code of Practice on Disinformation, signatory platforms submit periodic self-assessment reports that outline the policies, technical mechanisms, and partnerships they have established to counter false content, from fact-check labels and removal workflows to demotion algorithms and contextual info panels. Each report also provides quantitative metrics (e.g., numbers of labelled posts, takedowns, demotions, and context exposures) for the covered period, helping stakeholders assess the real-world impact of these measures. To validate and complement those self-reports—and to monitor platform practices more broadly—we analysed the MedDMO fact-check archive (1 January–30 June 2024), tracing how false claims circulated across Cyprus, Greece, and Malta and evaluating exactly how each platform's moderation tools were applied to content flagged as disinformation. For the purposes of this report, we adopt the EU's original 2018 Code of Practice definition of "**Disinformation**" as follows: *"Verifiably false or misleading information which, cumulatively, (a) Is created, presented and disseminated for economic gain or to intentionally deceive the public; and (b) May cause public harm, intended as threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security."* **[European Commission 2018].**

MedDMO's fact-checking consortium publishes fact-checking articles for disinformation spreading in the three countries, Cyprus, Greece, and Malta. These fact-checking articles often mention where the false claims were spread (which platform, online news outlets, online or offline spaces) and give the links to the archived or original content carrying the false claim. In our effort to monitor the platforms and generally other channels that contribute to the false information circulation and their practices to combat disinformation and assess the platforms' Code of Practice on disinformation reports, we decided to analyse the information we can find from the MedDMO fact-checks archive. We have compiled a dataset of 1,400+ links to content with verified falsehoods circulating in Cyprus, Greece, and Malta. This report focuses exclusively on that dataset, mapping each debunked claim to the visible actions—or inactions—taken by four signatories of the EU Code of Practice (Meta, YouTube, TikTok, and X) and by online news outlets or other media used to circulate disinformation that often serve as "sources of information" that are mentioned on the social media posts for online audiences. By analysing which posts were labelled, removed, received additional context, or left unaltered, we provide an evidence-based view of platform compliance and editorial accountability.

The report is structured as follows. First, we outline the methodology used to analyse the MedDMO fact-check archive, detailing our inclusion criteria for fact-check articles, as well as platform and publisher tracking procedures. In Section 3, we aggregate and analyse fact-checks for Greece and Cyprus, treating them as a combined corpus due to extensive overlap in language and circulating narratives, highlighting how each platform and online news outlet responded to shared disinformation content. Section 4 then turns to Malta, presenting the treatment of locally false claims by the same set of online platforms, news outlets, and other channels where false claims were circulating. Finally, the report closes with a summary of our key findings and recommendations.

# 2 A Methodology for Leveraging the MedDMO Fact-Check Archive in Platform's Practices and Media Landscape Monitoring

The objective of this analysis is to examine the nature, dissemination, and platform-level response to disinformation identified by fact-checking organisations participating in the MedDMO initiative across three countries. Specifically, the study seeks to understand how false claims were distributed across different channels, including social media platforms, messaging services, online news outlets, and other media formats, as well as to assess the extent to which these platforms engaged in moderation practices—such as content labelling, removal, or referencing fact-checks—to mitigate the spread of false information.

To support this investigation, we collected and analysed fact-checks from the MedDMO Fact-Check archive[1], covering the period from January 1st to June 30th, 2024[2]. The selected fact-checks were reviewed to identify the type and topic of disinformation, the platform or channel through which it spread, and any moderation actions taken by platforms, where identifiable. By systematically examining these dimensions, the analysis aims to contribute insights into the evolving dynamics of disinformation in the Mediterranean region and the responsiveness of digital platforms in addressing it. The results of the analysis of MedDMO fact-checks for the same period in 2023 are available in MedDMO's 1st Report of platform practices and national authorities support 2023 (for Cyprus, Greece & Malta) **[Leonidou et al., 2023].**

MedDMO Fact-Check archive has fact-checks in Greek and English. The fact-checks in Greek refer to disinformation that spreads in Greece and/or Cyprus. At the same time, fact-checks in English may refer to Malta, Greece, and/or Cyprus.

From the **MedDMO Fact-Checks archive,** we have the following information:

1. Fact-check Title
2. Fact-check Article Link
3. Author (and corresponding MedDMO fact-checking organisation that conducted the fact-check)
4. Category of fact-checked information (i.e., Health, Politics, Technology, etc.)
5. Date of fact-checking article publication

Through manual examination of each fact-checking article, we extracted the following data:

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. **Sources** where false claims were detected by fact-checkers included online platforms, news outlets, websites, TV and radio broadcasts, public speeches, and other media formats. Website domains were recorded when applicable.
4. **Links** to the content that contains false claims (if available)
5. **Post-fact-checking treatment** of the content by the source (disinformation label, link to fact-checking article, availability of the content (if it was removed)

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

[2] *1st of January to 30th of June 2024 is the period covered in the Signatories of the Code of Practice on Disinformation reports No.4. that MedDMO will publish a report on the assessment of the reports of META, Google and TikTok soon.*

Following the manual examination of the fact-checks, we developed an extended version of the MedDMO Fact-Checks archive for the specified period. This enriched collection, referred to as the Fact-Checks Archive Extended Dataset, includes additional metadata and annotations derived from our analysis. For clarity, we will refer to this collection as "the dataset" throughout the remainder of this report.

**Identifying the Source of Disinformation**

To determine the medium through which disinformation was disseminated in each fact-checking article, we applied a manual review process based on the following criteria:

- Screenshots or embedded media within the article displaying the false claim on a specific platform (e.g., Facebook, TikTok).
- Direct links to the original or archived instances of the disinformation.
- Explicit references in the article text identify the platform or source of the claim.

In cases where a platform was mentioned or shown via a screenshot, we conducted targeted searches on the respective platform to retrieve the link to the content containing the false claim and to assess whether any moderation actions, such as labelling, removal, or fact-checking tags, had been applied. However, we did not carry out broader, non-specific searches for the claims across platforms, due to the time-intensive nature of such an approach. This methodology ensured that all attributions to platforms or sources were evidence-based and directly traceable to specific elements within the fact-checking articles.

**Determining Whether Content Was Labelled or Linked to Fact-Check by the Platform**

Each platform adopts a distinct strategy for addressing false information through a mix of labels, warning screens, removals, and contextual tools. In our analysis, we evaluated the platforms' practices to combat disinformation as reported by the platforms' CoP reports No.4. Meta's services (Facebook and Instagram) apply disinformation labels or warning screens to posts flagged by accredited fact-checking partners and link directly to the relevant fact-check articles for added context; any content judged to carry harmful disinformation, such as claims that could undermine democratic processes or public health, is removed outright. In January 2025, Meta announced that beginning in the United States, it would end its collaboration with third-party fact-checking organisations, who had been independently rating claim accuracy, enabling Meta to apply warning labels and demote misleading content, and would instead roll out a Community Notes system similar to X's approach **[Kaplan 2025, BBC 2025].**

Google's video platform, YouTube, takes a contrasting route by focusing almost exclusively on removing videos that violate its disinformation policies, which target content with demonstrably harmful false claims and do not affix "disinformation" labels to any remaining videos. TikTok likewise prioritises the removal of content that breaches its harmful-disinformation rules, but it also introduces an "Unverified content" label when moderator reviews or partner fact-checks are inconclusive; however, it does not apply explicit "false information" tags, choosing instead to pull down policy-violating material. Finally, X (formerly Twitter) relies on its community-driven context tool, Community Notes, to surface additional information: any user can contribute a note on a post they believe is misleading, and those notes become visible only when a diverse group of contributors agrees they are helpful. Unlike the other platforms, X neither removes content for false claims nor attaches "false information" labels; instead, it depends on community-sourced annotations to provide nuance and context.

To assess whether the identified content with false claims in our dataset had been labelled by the platform (see examples of platforms' techniques to label content in Figure 1 and Figure 2), we examined active and publicly accessible content whenever possible. This included cases where fact-checking articles provided direct links to the original content, or where related posts were identified through platform-specific search features. In these instances, we verified whether the content was accompanied by a disinformation label, a warning notice, or any form of platform intervention indicating the claim was false.

We also specifically checked whether the platform included a reference or direct link to a fact-checking article that debunked the claim, either as part of a moderation label or embedded within the content interface. In the case of online news articles or websites, we considered it a form of labelling when the article explicitly stated that a claim was false, such as including phrases like *"the statement that [X] is true is fake news"*—even if no technical or visual moderation label was present. These textual disclaimers were treated as an editorial form of disinformation recognition.

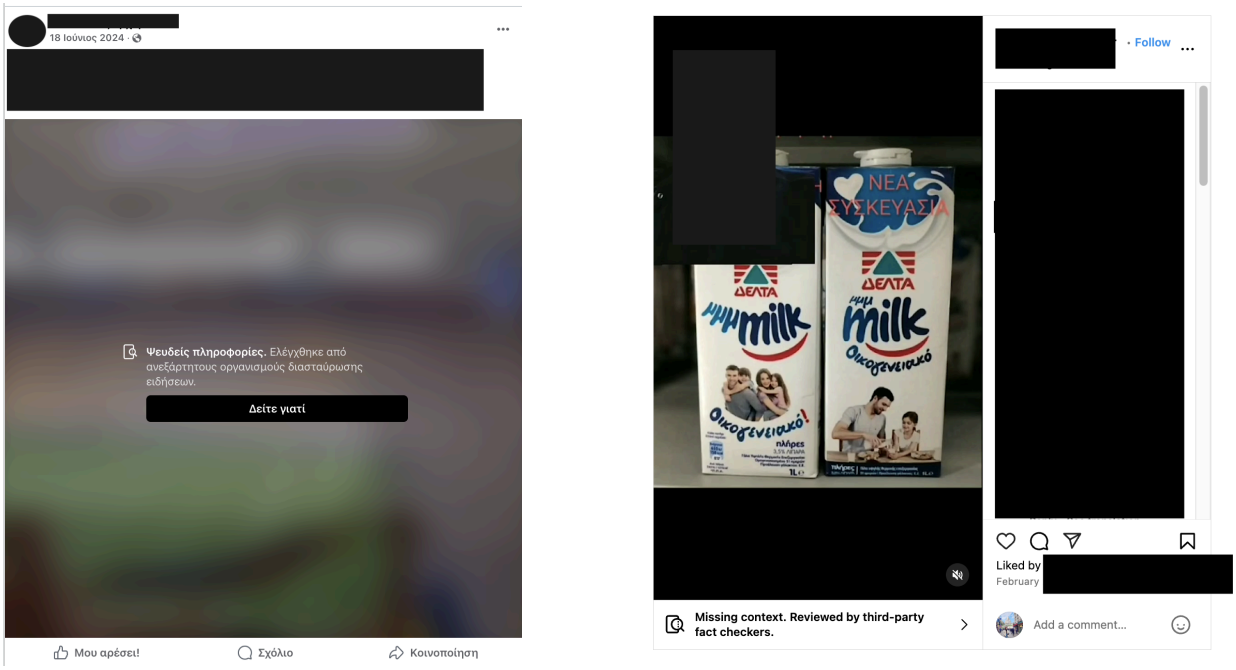



Figure 1: Examples of Meta applying labels to content with false information to empower users (left: Facebook, right: Instagram)

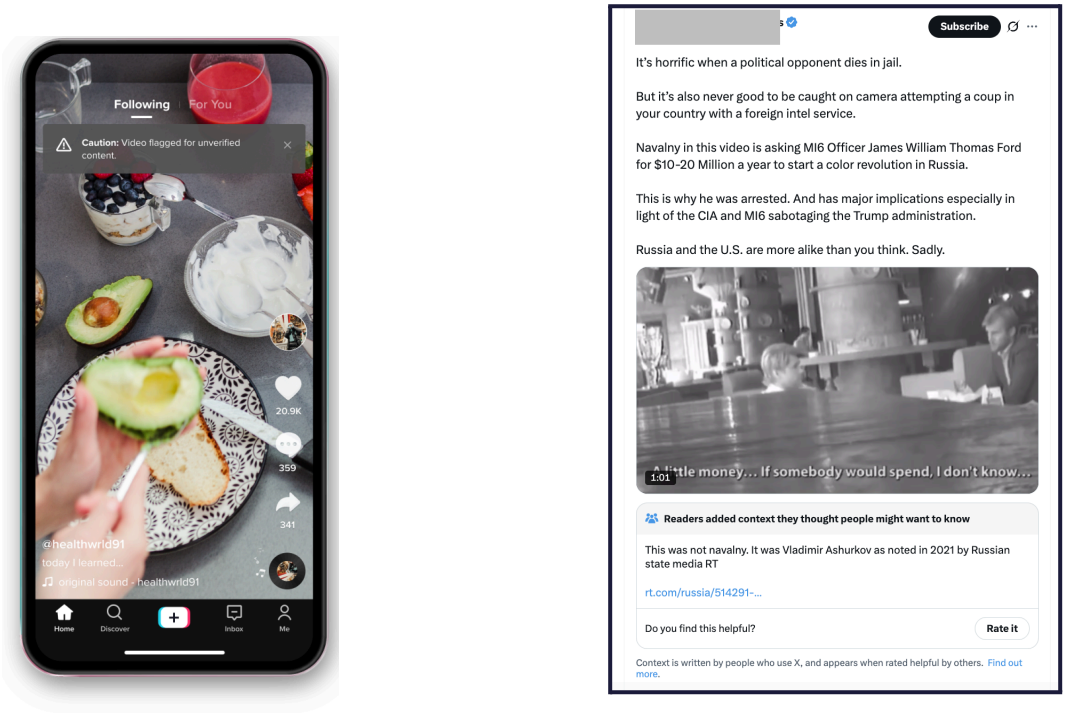



Figure 2: Examples of TikTok and X mechanisms to empower users. TikTok's label on unverified content[3] on the left and X's Community Notes on the right

**Determining Whether Content Was Removed**

In cases where the content was **no longer accessible on the platform**—as determined through direct or archived links provided in the fact-checking articles—we considered the content to have been **removed**, either by the platform, by the user, or as a result of **modified privacy settings** that made the post no longer publicly visible. This classification was based solely on the inaccessibility of the original content at the time of review.

**The extended MedDMO Fact-Checks archive Dataset:**

In our dataset, we analysed 150 fact-check articles covering Greece and Cyprus, 138 of them in Greek, 12 in English, and 34 fact-checking articles covering Malta (in English). There are **1226 links** to online content (content from social media platforms or news outlets' websites) that contain false information. Below we present the results of our analysis separately for Malta, and Greece, and Cyprus together.

[3] Image source: https://newsroom.tiktok.com/en-us/new-prompts-to-help-people-consider-before-they-share

# 3 Analysis of Greece and Cyprus Fact-Checks

As a result of the common language in Greece and Cyprus (Greek), it is very difficult to distinguish between false claims that are spread in the two countries. Usually, false claims spread in Greece also reach the Cypriots and vice versa. Due to this fact, we present the results of this analysis for Cyprus and Greece together.
Our dataset includes fact-checks coming from three different fact-checking organisations, Agence France-Presse (AFP), Ellinika Hoaxes (EH), and Fact-Check Cyprus (FCC), only from the period of January 1st to June 30th, 2024. In total, there are 150 fact-checks; the number of fact-checks per fact-checking organisation is depicted in Figure 3.

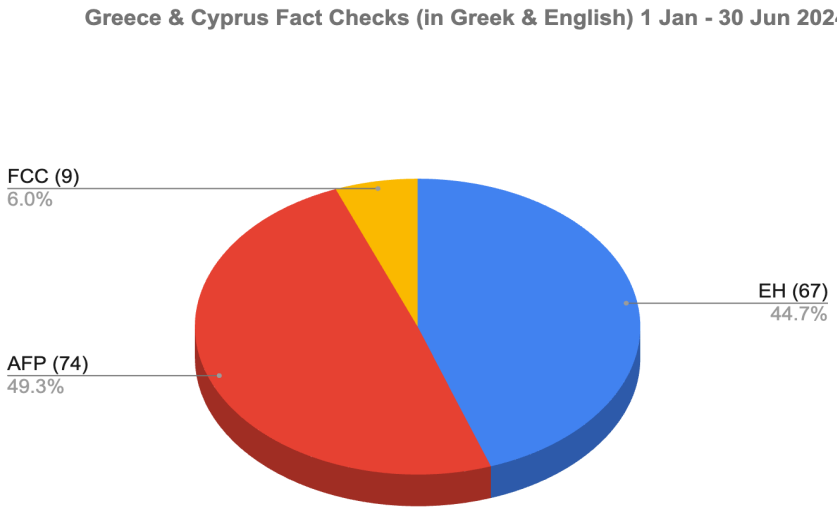



Figure 3: The proportion of Fact-Checks published by each of the three fact-checking organizations (Fact-Check Cyprus, Ellinika Hoaxes, and Agence France-Presse) covering Greece and Cyprus in our dataset for the period of January 1st to June 30th, 2024.

The fact-check articles' topics in our dataset are presented in Figure 4. It is important to note that the fact-checks can have more than one topic, i.e., Health, Covid-19. A great number of fact-checks are relevant to Society and Politics. Several claims related to COVID-19, and other health-related topics were spread during this period. Furthermore, false claims relevant to the Russia and Ukraine war, LGBTQ+, and migration are part of this dataset since those topics' enhanced interest during this period in Cyprus and Greece. Claims relevant to science, technology, environment, climate, and economy are also included.

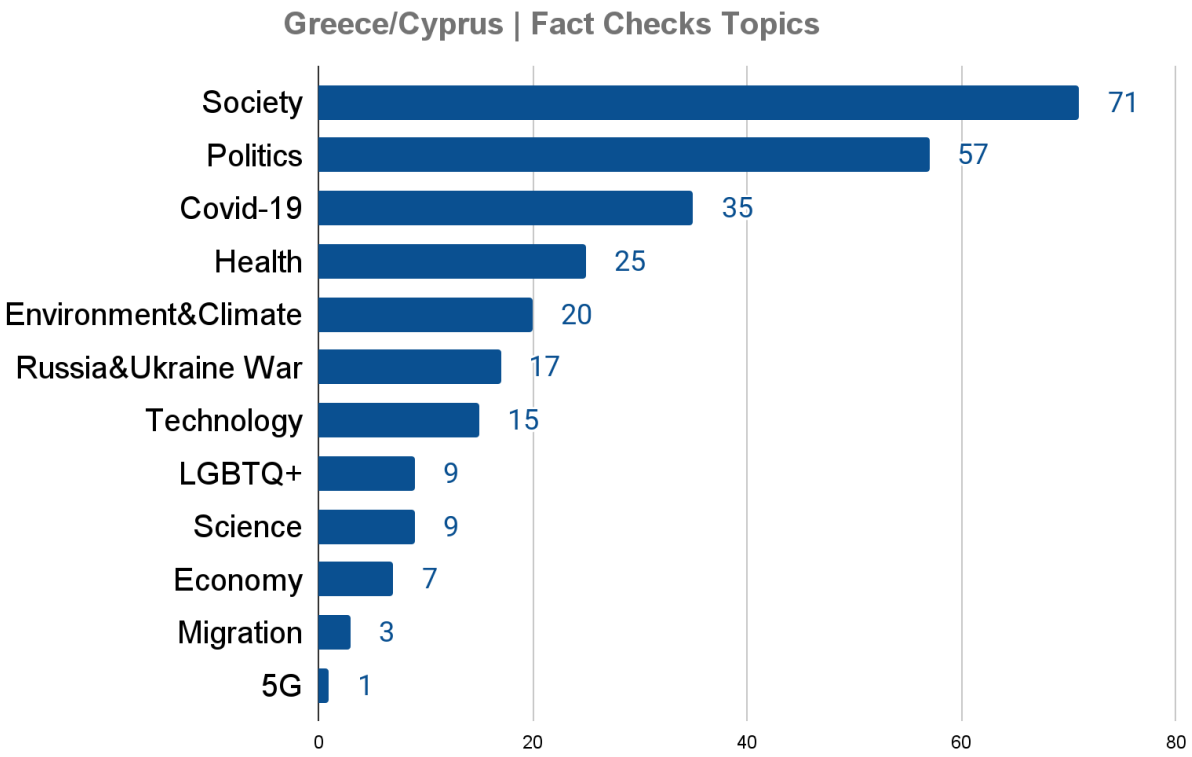



Figure 4: Fact-Check Topics for Greece and Cyprus for the period of this analysis

### False Claims in Cyprus/Greece and Social Media Platforms:

Fact-checks often include information, either in their text or accompanying visuals, indicating where the false claims were detected. Based on our analysis, we present in Figure 5 how many times each platform was referenced in fact-checks as a source where relevant false claims were spreading. A 91% of fact-checks in our dataset mentioned that the false claim was identified in Facebook, and 42% reported false claims found in X (prior Twitter). There are a few references to Instagram, YouTube, and TikTok. There is also a high percentage, 64%, of references to other online news outlets in Greece and Cyprus or platforms that we do not cover in our analysis (i.e., Telegram, LinkedIn, Threads) that are referred to as "Other" in Figure 5. Each fact-check may refer to multiple platforms spreading the false claims (see Figure 6). "Facebook and Other," and "Facebook, Other and X" are the two most frequent combinations of platforms.

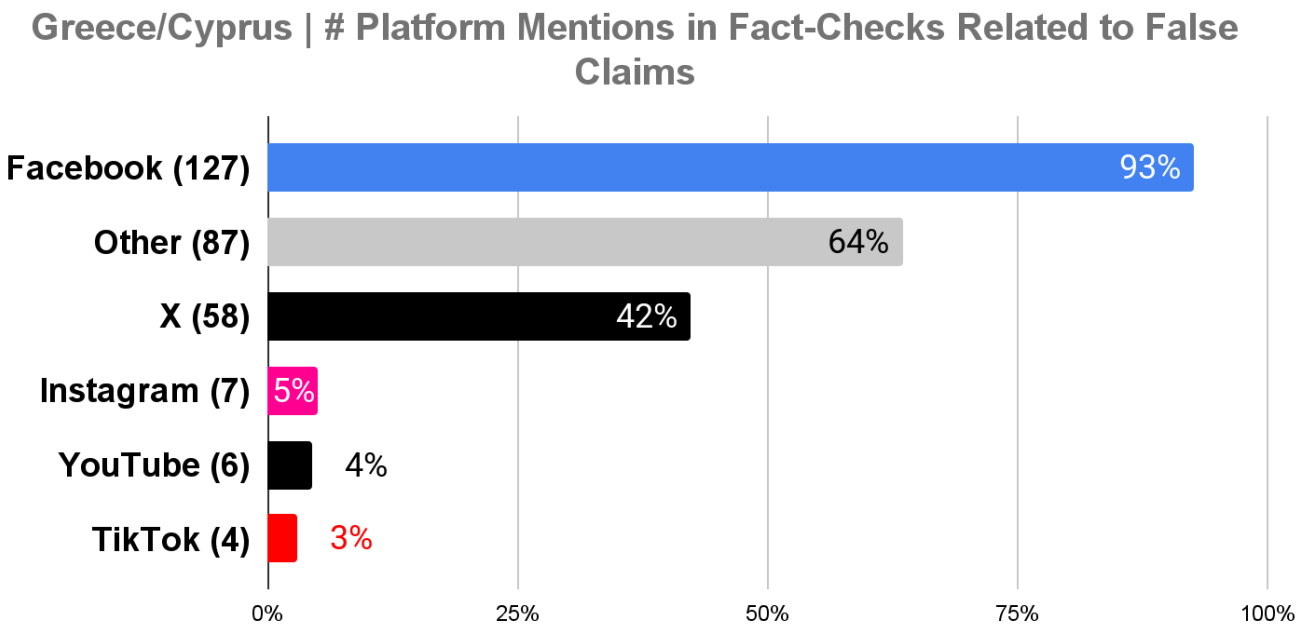



Figure 5: Platforms mentioned in Greece and Cyprus' Fact-Checks

*__Other__ refers to online news outlets, online platforms that are not covered in this survey, and other media

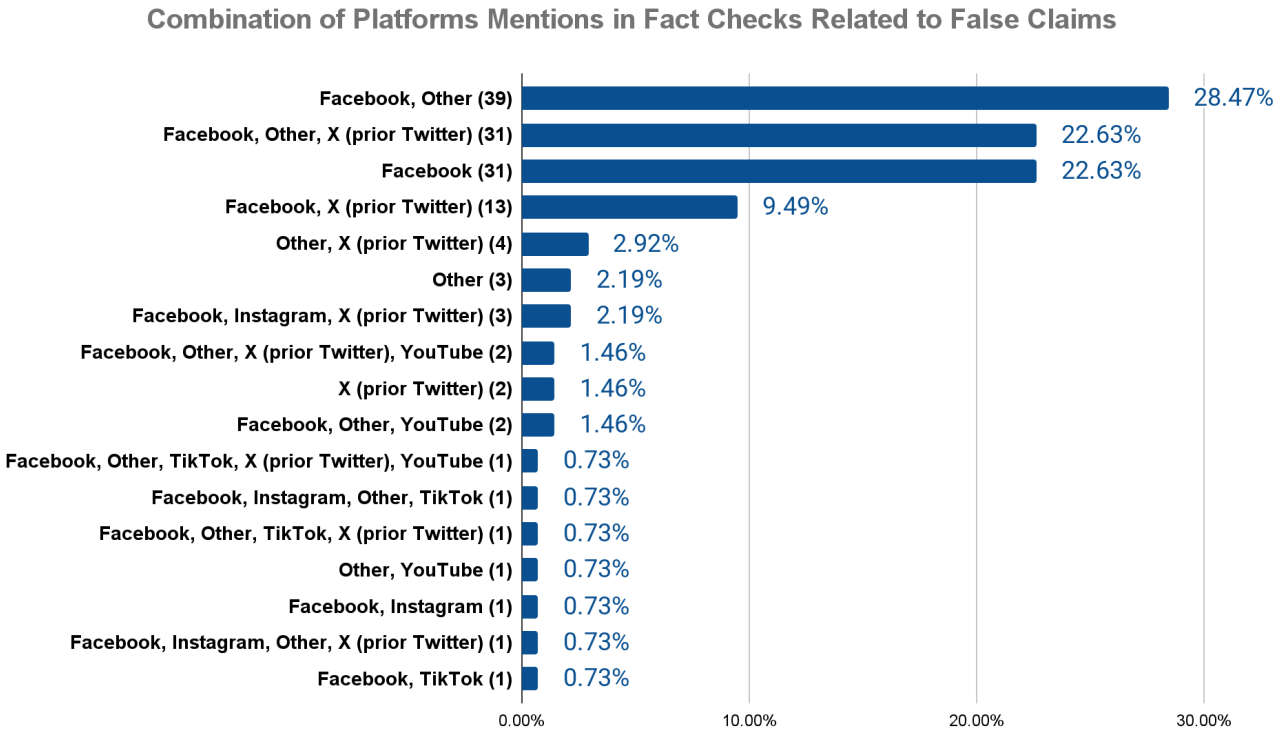



Figure 6: Combination of Platforms' mentions in Greece and Cyprus' Fact-Checks

In total, we found **1242** instances containing false claims from the fact-checks in our dataset. ~59% of these instances, 729 instances, come from Facebook. There are 80 instances from X, eight from Instagram, and fewer from other platforms such as YouTube and TikTok (see Figure 7).

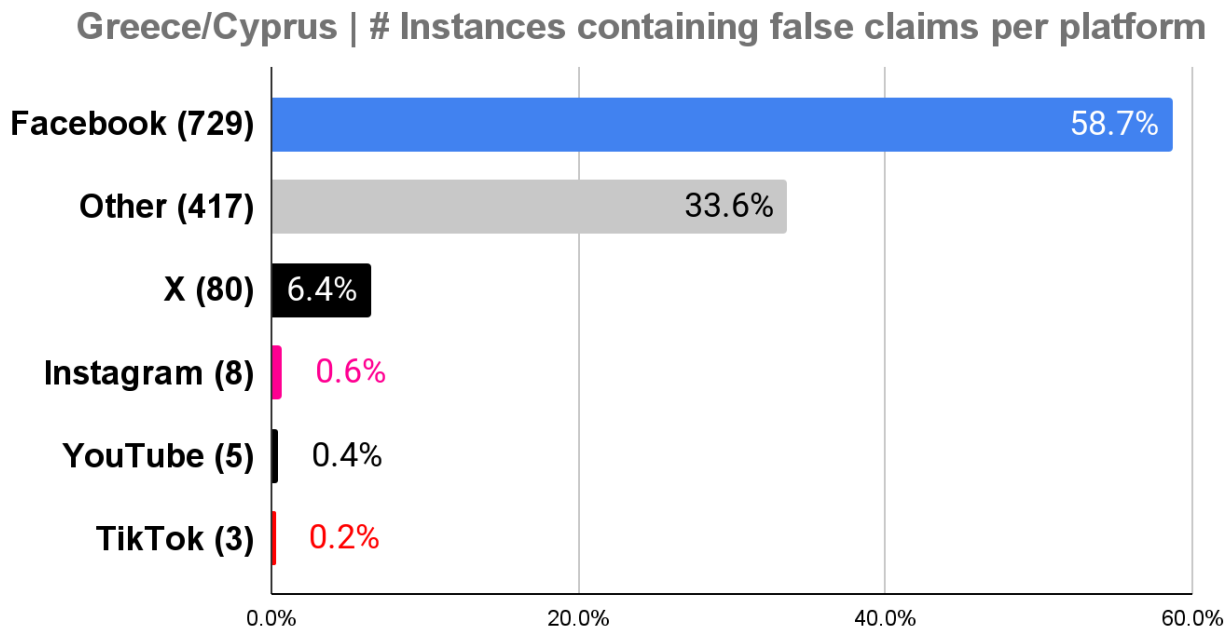



Figure 7: Number of instances containing false claims per platform found in Greece and Cyprus' Fact-Checks

Other than the aforementioned platforms, there were instances of content with false claims from Telegram (5), Threads (1), LinkedIn (1), and Rumble (1).

**Platforms Handling Content with False Claims**

Each entry of the dataset, i.e., instance containing a false claim, is categorised by whether it was labelled as disinformation, linked to a fact-check article, or no longer publicly accessible. Importantly, disappearance from public view ("Removed") may result from various user-side actions (e.g., post deletion, account suspension, or privacy setting changes) and should not always be interpreted as platform-driven moderation. See the summary of the results of this analysis in Table 1.

Facebook had the largest number of flagged posts, with 55.6% labelled as disinformation and linked to fact-checks. However, only 5.5% of the entries were no longer publicly accessible. Facebook (and Instagram) applies labels, giving context by adding a link to fact-checks, and downranking based on third-party fact-checking, rarely removing content unless it violates specific harm-related policies. This aligns with the observed pattern of high labelling but limited disappearance of posts. Instagram showed similar trends to Facebook, with 11.1% labelled and 33.3% of content no longer publicly available.

X (prior Twitter) displayed a minimal intervention pattern: only 4.2% of posts were “labelled”, and just 3.2% became inaccessible. While X maintains policies against civic disinformation, its enforcement relies primarily on user-driven Community Notes. So, labelled in the case of X’s instances means displaying a community note. In practice, few instances received labels or were removed, consistent with its reduced reliance on centralised moderation during this period.

TikTok had very few entries, and none of these three entries received a disinformation label or a link to a fact-checking article.

YouTube had five entries, none labelled, and all remaining visible. YouTube addresses disinformation primarily through policy-based removals (e.g., health or election disinformation) and non-targeted information panels. The lack of labelling and continued visibility of flagged content was expected since YouTube moderation policies are not focused on labelling and removing content with false claims.

Table 1: Platforms' moderation applied to instances containing false claims

| Platform (total number of instances) | # Instances (%) Labelled as False | # Instances (%) Links to Fact-Check | # Instances (%) Removed from/ Not Publicly Accessible Platform |
|---|---|---|---|
| Facebook (729) | Yes: 418 (57.3%)<br>No: 255 (35.0%)<br>NA: 56 (7.7%) | Yes: 418 (57.3%)<br>No: 255 (35.0%)<br>NA: 56 (7.7%) | Yes: 41 (5.6%)<br>No: 670 (91.9%)<br>NA: 18 (2.5%) |
| X* (80)<br>*Community notes | Yes: 4 (5.0%)<br>No: 66 (82.5%)<br>NA: 10 (12.5%) | Yes: 0 (0.0%)<br>No: 70 (87.5%)<br>NA: 10 (12.5%) | Yes: 3 (3.8%)<br>No: 70 (87.5%)<br>NA: 7 (8.8%) |
| Instagram (8) | Yes: 1 (12.5%)<br>No: 3 (37.5%)<br>NA: 4 (50.0%) | Yes: 1 (12.5%)<br>No: 3 (37.5%)<br>NA: 4 (50.0%) | Yes: 3 (37.5%)<br>No: 4 (50.0%)<br>NA: 1 (12.5%) |
| YouTube (5) | Yes: 0 (0%)<br>No: 5 (100%)<br>NA: 0 (0%) | Yes: 0 (0%)<br>No: 5 (100%)<br>NA: 0 (0%) | Yes: 0 (0%)<br>No: 5 (100%)<br>NA: 0 (0%) |
| TikTok (3) | Yes: 0 (0%)<br>No: 1 (33.3%)<br>NA: 2 (66.7%) | Yes: 0 (0%)<br>No: 1 (33.3%)<br>NA: 2 (66.7%) | Yes: 1 (33.3%)<br>No: 1 (33.3%)<br>NA: 1 (33.3%) |

**NA: cases where we lacked access to the content and therefore could not draw any conclusions.*

**The case of Facebook's false information labels and links to fact-checking articles:**

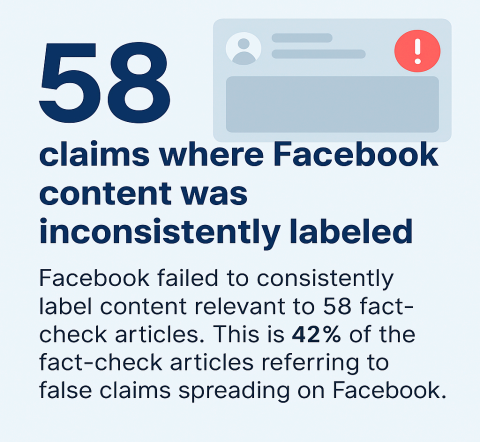



Figure 8: Meta (Facebook)'s cases with inconsistent labelling/moderation

From our analysis, 127 fact-check articles referenced false claims that had been disseminated via Facebook content. Among these, as explained in Figure 8, at least 58 fact-checks/claims were linked to multiple Facebook posts where the same false claim appeared. In those 58 cases, we found that labelling was applied inconsistently: some posts received a disinformation label while others did not, despite referring to the same debunked claim (see Figure 9). While we cannot confirm consistent labelling across all remaining fact-checks, these 58 clearly illustrate that Facebook's application of disinformation labels is not uniformly enforced, even within identical claim contexts.

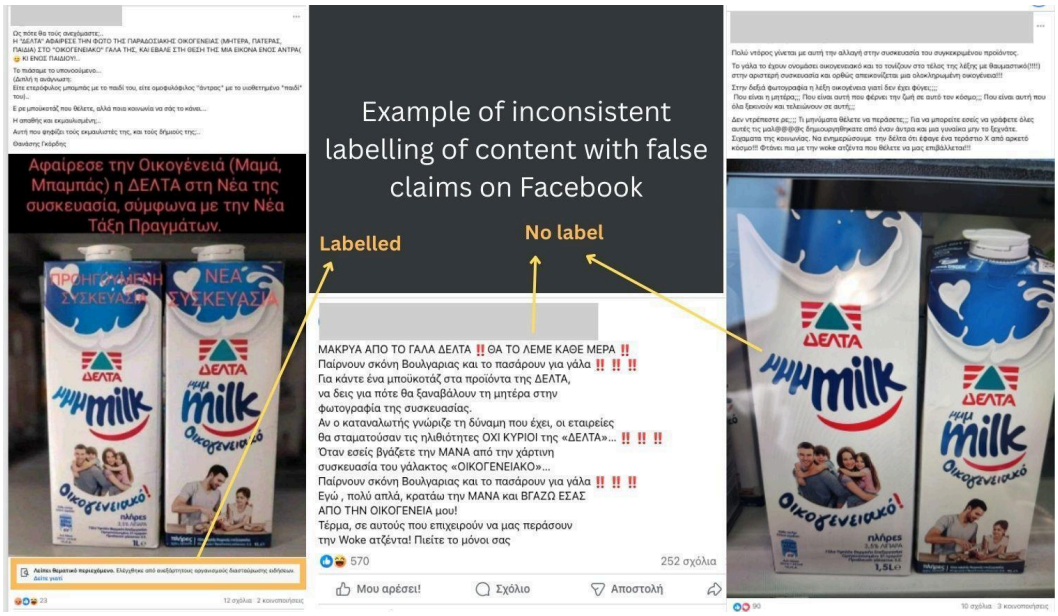



Figure 9: Example of Facebook's inconsistency in labelling content with identical false claims

**Greek News Websites that spread false claims through their published articles in the period of January-June 2024:**

In our dataset, there are 417 instances with the value “Other” as the platform that spread the false claim. These instances are coming from Greek news websites (406), other platforms that are not covered in our analysis (i.e., Telegram, Threads, LinkedIn), or do not mention a specific platform. From the analysis of the MedDMO Fact-Checks archive in the period of January-June 2024, we found from our data **235 unique Greek news websites/media** that published articles containing false claims[4]. Figure 10 shows the word cloud of the websites' names/domains and demonstrates the websites in our data that published articles with false claims; the font size of the names/domains increases if the websites publish a greater number of false claims.




[4] The websites listed have been identified in fact-checking articles as having published false or misleading information. Our listing is based solely on MedDMO fact-check sources and does not imply that these websites consistently or intentionally share false claims. This list is not exhaustive, and the presence or absence of any website does not indicate a blanket judgment about its overall credibility or intent. Our goal is to reflect documented cases of disinformation as flagged by fact-checkers — not to make final claims about the reliability or motives of these sources.

Figure 10: Word cloud[5] of Greek and Cypriot Online News Outlets mentioned in the Fact-Checks for the period under analysis

Only 20 out of the 406 articles (~5%) with false claims were corrected (i.e., altered the text or media, or mentioned the claim was false). More specifically, only five (~1.2%) mentioned that the claim was false, and only three (~0.75%) had a link to the fact-checking articles. In our analysis, we found 34 articles (16.5%) that we could not find online, and we assumed that they had been removed by the site admin. A ~84% (341) of the articles with false claims are still publicly available with no correction. For 12 instances, we were not able to reach the website (i.e., the website was down, the domain was for sale, we got the "forbidden" error response, etc.). In Figure 11, we list all the websites found in our dataset, categorised by the number of articles that contain false claims. The figure also illustrates whether a site was unreachable, corrected at least one article, or removed at least one article that contains disinformation.

[5] The Word Cloud was generated with the online tool from https://www.wordclouds.com/




Figure 11: Online news outlets in Greece and Cyprus, ranked by the number of times they were cited in fact-checks, with markers indicating whether their content underwent any moderation actions.

# 4 Analysis of Malta Fact-Checks

For the period of January 1st to June 30th, 2024, there are 34 fact-checks for Malta. In our analysis, we excluded one fact-check since it was a summary of false claims shared in 2023, which is out of the period this analysis focuses on. The fact-checks for Malta in MedDMO are produced by the Times of Malta (ToM) and the University of Malta (UoM). In the specific period under study, only one fact-check was produced by the UoM (see Figure 12). Times of Malta was the first Maltese organisation conducting fact-checks under the MedDMO project. The University of Malta later began producing such articles to support the Times' efforts. It is important to note that the two organisations do not have any collaboration with VLOPs and VLOSEs to fact-check claims that are distributed on their platforms. Meta, TikTok, and Google CoP reports (covering the period of January 1st to June 30th, 2024) do not explicitly mention any collaboration with any organisation covering Malta or the Maltese language.

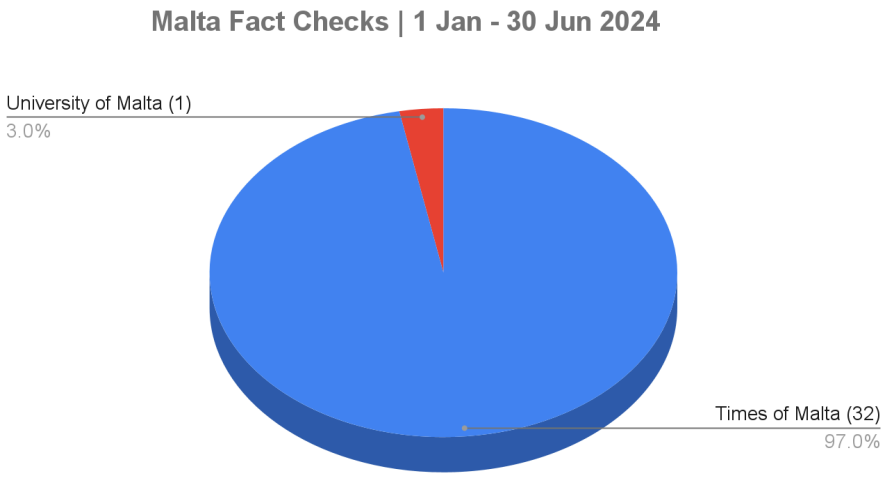



Figure 12: The proportion of Fact-Checks published by each of the two fact-checking organizations covering Malta in our dataset for the period of January 1st to June 30th, 2024.

The fact-check topics in our dataset are shown in Figure 13. As with the case of Greece and Cyprus, it is important to note that a single fact-check may be associated with more than one topic. A significant number of fact-checks are related to Politics, which emerges as the most prominent category. This is followed by Society and Economy, indicating a focus on political discourse, social issues, and financial matters during the period in question. A number of fact-checks also addressed online scams, AI-generated content, and manipulated media, reflecting growing concerns about digital disinformation and deceptive practices online.

Fewer fact-checks were observed on topics such as Technology, Health, Environment and Climate, and Migration, suggesting either lower disinformation activity in these areas or less fact-checking focus. Nonetheless, these topics are still part of the dataset and reflect the broader range of narratives addressed in Malta. Topics like migration and climate, while less frequent, are of regional interest and could become more prominent depending on the socio-political context.

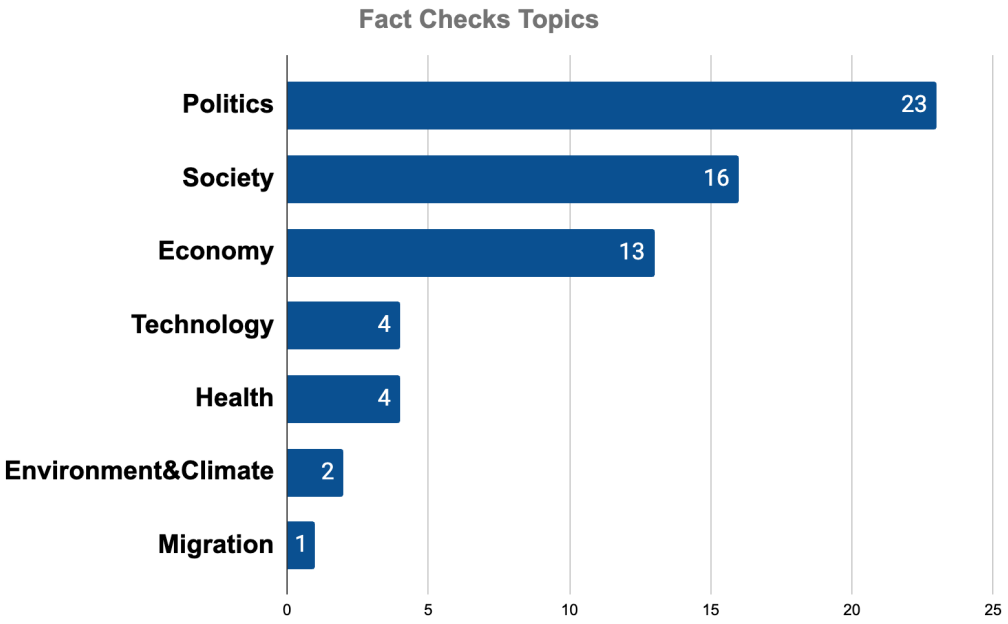



Figure 13: Fact-Check Topics for Malta for the period of this analysis

## False Claims circulating in Malta and Social Media Platforms:

Fact-checks often include references—either in the body of the text or in accompanying visuals—indicating where the false claims were initially identified. Figure 14 presents the number of times each platform was mentioned in Malta-based fact-checks as a source of disinformation. Facebook is the most frequently mentioned platform, appearing in 33.33% of the fact-checks.

There are also mentions of Instagram (6.06%), YouTube (3.03%), and TikTok (3.03%), though these are much less frequent. A significant portion of the fact-checks—66.67%—fall under the “Other” category. In the case of Malta, this includes a broad range of sources such as political speeches (e.g., by the Prime Minister or Opposition leader), WhatsApp conversations, online news outlets, public statements from organisations, or, in some cases, no specific source being identified in our analysis of the fact-checks.

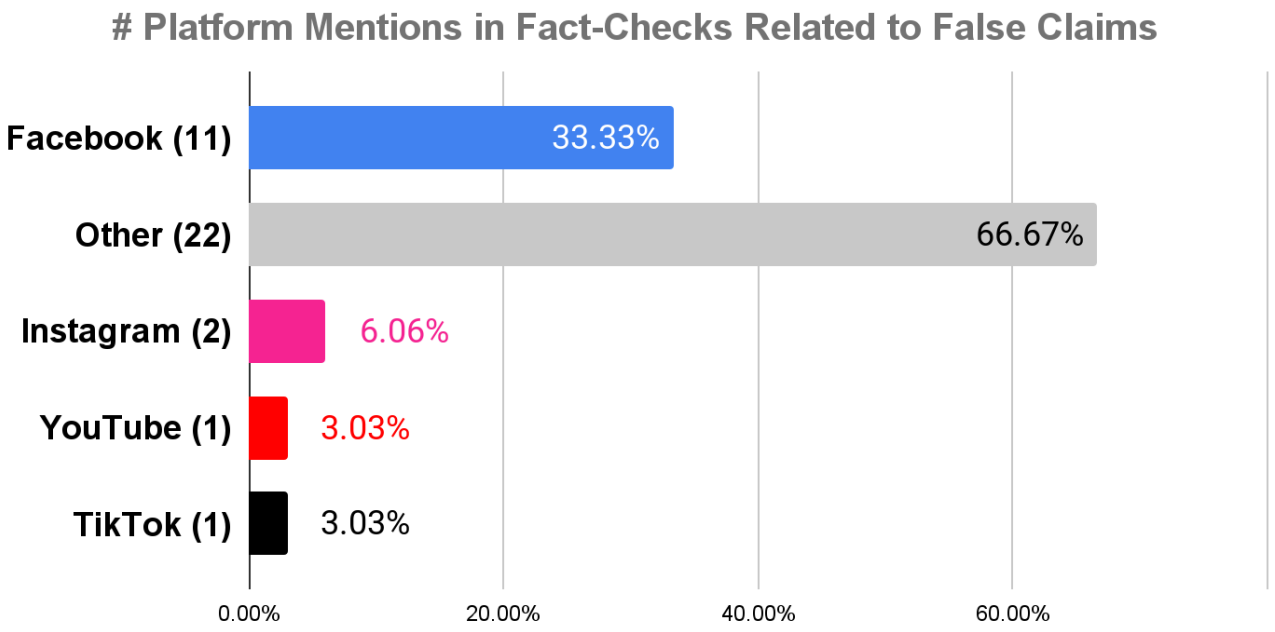



Figure 14: Platforms mentioned in Malta's Fact-Checks

In the Malta fact-checks included in our dataset for the specific period, direct links to online content such as social media posts or news articles were generally absent (see Figure 15). As a result, most platform mentions in the fact-checks did not include direct URLs to the original instances of disinformation, making it difficult to determine whether the content had been removed, labelled, or linked to a fact-checking article by the platforms themselves.

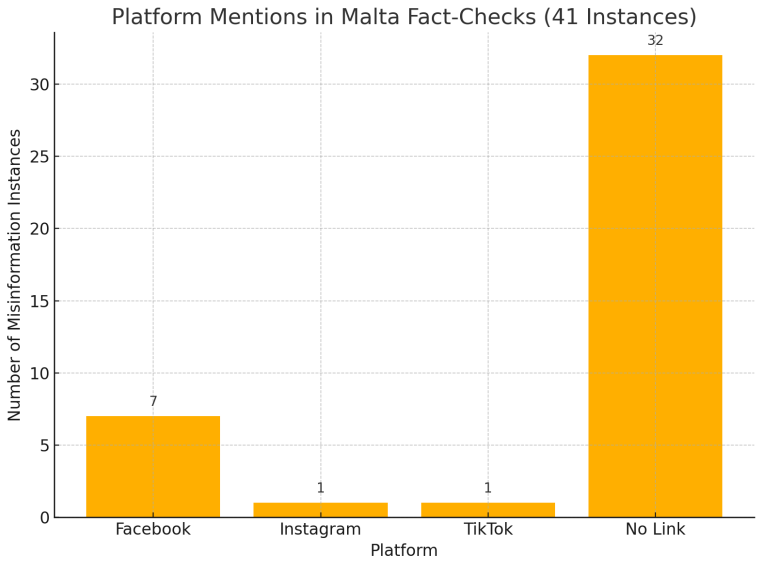



Figure 15: Number of instances containing false claims per platform found in Malta's Fact-Checks

Out of 41 distinct instances of disinformation, seven included Facebook links, while one was linked to Instagram and one to TikTok. Among these, just one post was flagged with a disinformation label and linked to a

fact-checking article, published by Open.online [Open Online 2024], a fact-checking organisation based outside of Malta. This article addressed a recurring false claim that had circulated across multiple countries. Similarly, the instances of false claims identified on TikTok and Instagram did not receive any form of platform labelling, removal, or links to fact-checking resources.

Despite the limited platform intervention, the false claims observed in Malta exhibited high levels of user engagement, including large numbers of likes, comments, shares, and views. This underscores the broad reach of such content and highlights the urgent need for social media platforms to actively label and contextualise disinformation to help users identify false narratives.

Within the MedDMO framework, fact-checking activities are carried out by the Times of Malta and the University of Malta, focusing on debunking false claims circulating locally. However, these organisations are not part of any third-party fact-checking programmes, such as those operated by Meta or TikTok. This means that platforms do not incorporate their findings into official disinformation response mechanisms, leaving content targeting the Maltese public largely unmoderated by platform-based fact-checking systems.




Figure 16: Findings summary of Malta's Fact-Checks analysis

The analysis of fact-checks in Malta reveals a limited but telling snapshot of the disinformation landscape. Most fact-checks focused on political, societal, and economic issues, with Facebook being the most frequently referenced platform, followed by Instagram, TikTok, and YouTube. However, the majority of false claims lacked

direct links, making it difficult to assess platform responses. Crucially, no systematic labelling or moderation of false content was observed, even in posts that achieved high levels of user engagement, highlighting a gap in platform accountability. Additionally, AI-generated content and online scams were part of the disinformation ecosystem.

Despite the efforts of local actors such as the Times of Malta and the University of Malta, there is no formal collaboration between these organisations and major platforms like Meta or TikTok. This absence significantly limits the visibility and impact of local fact-checking in platform-level content moderation.

These findings underline the importance of greater platform transparency, enhanced labelling practices, and inclusion of local fact-checkers in global disinformation response frameworks.

# 5 Conclusions

This report examines how over 1,200 online content that carried false claims identified in 150 fact-checks from the MedDMO's fact-checking partners between January and June 2024 were treated by social platforms and online news outlets, and we have the following findings.

In the contexts of Cyprus and Greece, Meta indeed applies disinformation labels and provides direct links to fact-checking articles, but we observed significant inconsistency in their deployment. Identical posts carrying the same false claim sometimes bore a warning banner and reference link, while others did not, suggesting a flaw or gap in Meta's labelling system (and raising the question of whether this inconsistency is specific to Greek-language content). On X, Community Notes were used to flag only a small fraction of misleading posts, leaving most false claims unannotated. Instagram, YouTube, and TikTok each yielded too few flagged instances in our dataset to support firm conclusions about their disinformation interventions. Finally, among the leading online news outlets in Greece and Cyprus where these false claims appeared, fewer than one in five articles were ever corrected, "labelled" as fake, or taken down—meaning that debunked content remains live without context.

In Malta, although our fact-check sample was small, false claims recurred across Facebook, Instagram, YouTube, and TikTok. The majority of the checks focused on politics, society, or the economy, often tracing falsehoods back to speeches by the Prime Minister or Opposition leader, forwarded WhatsApp messages, online news outlets, official statements from organisations, or—in some cases—no identifiable origin. We observed only a single instance of a Meta-warning label—on a claim already debunked internationally by a non-local fact-checking organisation. There are no systematic removals or labels on Maltese-language content across the platforms. This can be justified since there are no known collaborations between Malta's fact-checking organisations and online social media platforms. Fact-checking in Malta was recently established, driven by the Times of Malta and later the University of Malta under MedDMO, and remains disconnected from platform moderation channels, so local debunkings rarely feed into in-platform remediation or labelling. These few traced posts or online content nonetheless drew high user engagement, underscoring the absence of any in-platform context tools for Maltese audiences.

Together, these findings reveal critical gaps in both platform moderation and publisher accountability. Inconsistent labelling, sparse takedowns, and near absence of post-publication corrections allow debunked content to persist unchallenged. Strengthening collaboration between MedDMO fact-checkers, national

authorities, and Very Large Online Platforms—and advocating for uniform labelling, timely removals, and broader language coverage—will be central to achieving the objectives of MedDMO Task 5.1 and promoting digital information resilience in Cyprus, Greece, and Malta. Moreover, this methodology leverages the MedDMO fact-check archive as a "ground truth": recent fact-checked entries can be regularly sampled to verify whether each was labelled, contextualised, demoted, or removed on the platform where it appeared, and to track over time any improvements—or backsliding—in inconsistent labelling or delayed takedowns. By combining automated data collection with targeted manual audits, Task 5.1 can deliver robust, nuanced compliance assessments that drive continuous platform accountability. Coupled with reinforced regulatory oversight—especially through alignment with the EU's Digital Services Act—these efforts can foster more consistent platform compliance and meaningful publisher corrections. Ultimately, by institutionalising these evidence-based practices within national media frameworks, Cyprus, Greece, and Malta can safeguard democratic discourse and restore public trust in the online information environment.

# 6 References

**[European Commission 2022]** European Commission (2022). Strengthened Code of Practice on Disinformation. https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

**[European Commission 2018]** European Commission (2018). Code of Practice on Disinformation. https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation

**[Leonidou et al. 2023]** Pantelitsa Leonidou, Nikos Salamanos, Michael Sirivianos, Loukia Taxitari, Socrates Ioakeim, Andreas Phylactou (2023), 1st Report of platform practices and national authorities support 2023 (for Cyprus, Greece & Malta). https://meddmo.eu/wp-content/uploads/2024/05/FINAL_1stReport_Monitoring_the_platform_practices_compressed.pdf

**[Kaplan 2025]** Kaplan, J. (2025, January 7). More speech and fewer mistakes. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

**[BBC 2025]** BBC News. (2025, May 4). TikTok adds ‘unverified content’ warning to videos. https://www.bbc.com/news/articles/cly74mpy8klo

**[Open Online 2024]** Fact-checking Team. (2024, January 5). Attenzione alla truffa dei fantomatici bagagli smarriti in aeroporto “venduti a due euro”. Open Online. https://www.open.online/2024/01/05/truffa-bagagli-smarriti-aeroporto-venduti-due-euro-fc/


Mediterranean Digital
Media Observatory

# Annex III: MedDMO report on monitoring Meta, TikTok, and Google's practices to combat disinformation and supporting national media authorities 2024 (for Cyprus, Greece & Malta) - Part B of the 2nd Report of platform practices and national authorities support 2024 (for Cyprus, Greece & Malta)

MedDMO is a Digital Europe SME Support Action project co-financed by the EC under Grant Agreement with project ID: 101083756. The content of this document is © the author(s). For further information, visit www.meddmo.eu




# D5.1

## MedDMO report on monitoring Meta, TikTok, and Google's practices to combat disinformation and supporting national authorities 2024 in Cyprus, Greece, and Malta

## 2nd Report of platform practices and national authorities support 2024 Part B

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

| | |
|---|---|
| **Deliverable title** | 2nd Report of platform practices and national authorities support –2024 |
| **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, Theodoros Christophides, Andreas Phylactou, Socratis Ioakim, Iacovos Mylona, Neophytos Lophitis, Athanasios Sitista, Loukia Taxitari, Kostantinos Papadamou, Pashalia Spyridou, Michael Sirivianos |
| **Reviewer(s)** | CERTH, UCY |
| **EC Project Officer** | Mark Vella Muskat |

| | |
|---|---|
| **Abstract** | This is the MedDMO annual report detailing the results of monitoring the online platform practices in support of the national authorities of Cyprus, Greece and Malta in combating active disinformation campaigns. In addition, the report highlights progress in the implementation of the Code of Practice on Disinformation by online platforms. |
| **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. The content of this document is © the author(s).
For further information, visit www.meddmo.eu

## Copyright

© Copyright 2022 MedDMO Consortium

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the MedDMO Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.
All rights reserved.

## Revision History

| VERSION | DATE | MODIFIED BY | COMMENTS |
|---|---|---|---|
| V0.1 | 01/10/2024 | CUT | First Draft Table of Content |
| V0.2 | 01/12/2024 | CUT | Final Table of Content |
| V0.3 | 01/02/2025 | CUT | First Draft |
| V0.4 | 20/05/2025 | CUT | Second Draft |
| V0.5 | 10/06/2025 | CUT | Final Second Draft |
| V0.6 | 15/06/2025 | CUT | Final Second Draft Review |
| V0.7 | 23/06/2025 | CUT | Review |
| V1.0 | 25/06/2025 | 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 |
| DSA | Digital Service Act |
| ICPSR | Inter-university Consortium for Political and Social Research |
| CIB | Coordinated Inauthentic Behaviour |
| 3PFC | Third Party Fact-Checking Program |
| EFCSN | European Fact-Checking Standards Network |
| IIB | Independent Intermediary Body |
| CASD | Centre D'Accès Sécurisé Aux Données |
| WHO | World Health Organisation |
| EMIF | European Media and Information Fund |
| VCE | Virtual Compute Environment |
| CRTA | Cyprus Radio Television Authority |
| MBA | Broadcasting Authority of Malta |
| NCRTV | National Council for Radio and Television |
| DSC | Digital Service Coordinator |
| MCA | Malta Communications Authority |
| EETT | Hellenic Telecommunications and Post Commission |

# Table of Contents

Revision History ..... 4
Glossary ..... 4
Index of Figures ..... 6
Index of Tables ..... 7
Executive Summary ..... 8
1 Introduction ..... 9
2 Methodology for Monitoring the Platforms' Practices ..... 11
3 CoP Signatories Reports Assessment Results ..... 15
3.1 Summary of Results ..... 15
3.2 Recommendations for Platforms and National Experts in Cyprus, Malta, and Greece to Strengthen the Fight Against Disinformation ..... 25
3.3 Meta (Facebook and Instagram) ..... 28
3.3.1 Meta: Empowering Users ..... 28
3.3.2 Meta: Empowering the Research Community ..... 34
3.3.3 Meta: Empowering the fact-checking community ..... 38
3.4 Google (Search, YouTube) ..... 42
3.4.1 Google: Empowering Users ..... 42
3.4.2 Google: Empowering the Research Community ..... 53
3.4.3 Google: Empowering the fact-checking community ..... 59
3.5 TikTok ..... 64
3.5.1 TikTok: Empowering the users ..... 64
3.5.2 TikTok: Empowering the Research Community ..... 82
3.5.3 TikTok: Empowering the fact-checking community ..... 87
4 MedDMO Fact-Checking Partners Collaboration with VLOPs ..... 93
5 Supporting the National Authorities ..... 98
6 Conclusion ..... 100
7 References ..... 101
Annex I: CoP Signatories Reports Assessment Scores and Missing SLIs. ..... 103
Annex II: Questionnaire for MedDMO fact-checking organizations partners ..... 107

# Index of Figures

Figure 1: Average scores for each CoP pillar and each platform ..... 15
Figure 2: Average scores for each platform under Empowering users pillar ..... 16
Figure 3: Average scores for each platform under Empowering the Research Community pillar ..... 18
Figure 4: Average scores of each platform under Empowering the Fact-Checking Community pillar ..... 22
Figure 5: Facebook Warning Screens ..... 29
Figure 6: Fact-checking rating label in Greek on a Facebook post and dialog window with link to the fact-checking article. ..... 31
Figure 7: Fact checking rating label in English on a Facebook post and dialog window with link to the fact-checking article. ..... 32
Figure 8: Warning message to prevent user from sharing the content in Greek and English. ..... 32
Figure 9: Google Search “About This Result” Feature (Greek) ..... 43
Figure 10: Google Search “About This Result” Feature (Maltese) ..... 43
Figure 11: Google Search "About this image" feature ..... 44
Figure 12: Google Search "About this image" feature (left: Greek and right: Maltese) ..... 44
Figure 13: YouTube’s Climate Change Information Panel example when language is set to Greek. ..... 45
Figure 14: YouTube: Information Panels Inconsistency ..... 46
Figure 15: Hit Pause Campaign Screenshot in Greek and English ..... 47
Figure 16: Fact Check Rich Snippet in Google Search Results (Greek). ..... 50
Figure 17: Fact Check Rich Snippet in Google Search Results (Maltese). ..... 50
Figure 18: Fact Check Rich Snippet in Google Search Results (English). ..... 50
Figure 19: Google's Fact-Check Explorer ..... 51
Figure 20: Search by Image Feature in Fact Check Explorer ..... 51
Figure 21: TikTok Climate Change Search Intervention [TikTok, 2024] ..... 65
Figure 22: Climate Change Intervention unavailable. ..... 66
Figure 23: Holocaust Search Intervention Inconsistency ..... 66
Figure 24: Monkeypox Search Intervention Availabilty and Inconsistency ..... 67
Figure 26: TikTok Video Notice Tags [TikTok, 2024] ..... 67
Figure 27: TikTok Video Notice Tag Language Issues with Redirecting Information Pages ..... 68
Figure 28: TikTok Video Notice Tag for Monkeypox Inconsistency ..... 69
Figure 30: TikTok AI-generated content label availability ..... 70
Figure 31: MedDMO collaboration with the national media authorities ..... 98

# Index of Tables

Table 1: CoP Commitments covered by this analysis ....13
Table 2: CoP Commitment-level assessment scaling system ....14
Table 3: Available resources for Researchers found in the Platforms Reports ....19
Table 4: Available resources for Fact-Checkers found in the Platforms Report ....24
Table 5: Meta's reported quantitative information for SLI 21.1.2, page 82. ....33
Table 6: Meta's reported quantitative information for SLI 31.1.3....40
Table 7: TikTok's reported quantitative information for SLI 17.1.1....74
Table 8: TikTok's reported quantitative information for SLI 21.1.1....81
Table 9: TikTok's reported quantitative information for SLI 21.1.2....81
Table 10: TikTok's reported quantitative information for SLIs 31.1.1 – 3 ....91
Table 11:MedDMO Fact-Checking Partners Collaboration with Meta ....93
Table 12: MedDMO Fact-Checking Partners Collaboration with TikTok ....96
Table 13: MedDMO Fact-Checking Partners Collaboration with Google and other platforms ....97
Table 14: Meta's CoP Report No.4 Summary of Assessment Results....103
Table 15: Google's CoP Report No.4 Summary of Assessment Results ....103

# 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 2024 (based on their CoP reports No.4. of September 2024). The report also outlines broader 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, disinformation proliferates through social media channels and websites, particularly during election periods and amidst significant news events such as the war in Ukraine. Greece experiences disinformation crises during natural disasters and in the context of political decisions, with actors ranging from government mechanisms to far-right movements disseminating false information. In Malta, state-sanctioned 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 disinformation 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 disinformation.

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

# 1 Introduction

The global challenge of disinformation remains highly prevalent, significantly affecting societies, politics, and public discourse worldwide. Between January and June 2024, the ongoing advancement of digital platforms and social media intensified the spread of false or misleading information, posing serious threats to information integrity, shaping public opinion, eroding trust in democratic institutions, and influencing political outcomes.

During the first half of 2024, both Greece and Cyprus faced similar disinformation challenges, prominently featuring disinformation related to COVID-19 vaccines, including unfounded health claims and conspiracy theories about vaccine safety. Politically motivated disinformation was widespread, employing manipulated or misrepresented visuals falsely linked to recent events such as natural disasters and geopolitical conflicts. Additionally, the Israel-Hamas conflict became a frequent subject of misleading narratives, increasing polarization and exacerbating social tensions in both countries. Often, identical false claims circulated simultaneously in Cyprus and Greece. Cyprus specifically encountered intensified disinformation during electoral periods and critical geopolitical developments, including the ongoing impact of the war in Ukraine. Misleading narratives proliferated on Cypriot social media platforms, news websites, and traditional media, often involving anonymous online accounts, politically aligned groups, or foreign entities aiming to influence public perception and destabilize trust in Cypriot institutions.

In Malta, the first half of 2024 witnessed significant disinformation around allegations of financial corruption, economic mismanagement, and misuse of public funds. There was notable usage of deepfake technology to spread falsehoods targeting political figures. Environmental disinformation also circulated widely, influencing public perception of urban planning and environmental policy. Furthermore, Malta faced extensive migration-related disinformation, with exaggerated reports of migrant numbers and impacts fuelling societal tensions. Online scams were also prevalent, exploiting public trust and causing economic vulnerabilities.

This report is the second part of MedDMO's monitoring of platform practices in Cyprus, Greece, and Malta for the first half of 2024. It presents a qualitative assessment of how Meta, Google, and TikTok implemented the Code of Practice on Disinformation—specifically focusing on the commitments under Empowering Users, Empowering the Research Community, and Empowering the Fact-Checking Community—as outlined in their CoP Signatory Report No. 4 (January–June 2024).

The analysis builds on findings from the first part of the monitoring round **[Leonidou et al., 2025],** which mapped the appearance and handling of false claims across platforms and media based on the MedDMO Fact-Checks archive. Results and findings further complement this after applying the EDMO methodology to assess the VLOPSEs September 2024 CoP Signatories report **[Botan et al., 2025]**.

While Greece, Cyprus, and Malta face distinct yet overlapping disinformation challenges, this study underscores the importance of tailored, country-specific strategies to improve digital literacy, strengthen fact-checking efforts, and enhance democratic resilience. As part of the MedDMO project, our aim is not only to share the findings of our monitoring work with the broader public, but also to inform and support other MedDMO activities—such as collaboration with national broadcasting authorities in the three countries. To that end, the report concludes by highlighting our ongoing initiatives related to disinformation, in partnership with the Cyprus

Radio Television Authority, the Greek National Council for Radio and Television, and the Broadcasting Authority of Malta.

The report is structured as follows: Section 2 presents the methodology used to evaluate platform practices, drawing from the Code of Practice Signatory reports. Section 3 begins with a summary of key findings for Meta (Facebook and Instagram), Google (Search and YouTube), and TikTok, followed by recommendations for how platforms and media experts can collaborate more effectively to combat disinformation in Cyprus, Greece, and Malta. It then provides a detailed analysis of the reported information under the commitments assessed within each of the three pillars: Pillar V (Empowering Users), Pillar VI (Empowering the Research Community), and Pillar VII (Empowering the Fact-Checking Community). Section 4 outlines platforms' fact-checking partnerships involving MedDMO partners such as Agence France-Presse and Ellinika Hoaxes. Finally, Section 5 highlights key MedDMO activities supporting the national authorities in the three countries in their efforts to tackle disinformation.

# 2 Methodology for Monitoring the Platforms' Practices

**Code of Practice on Disinformation:** The EU Code of Practice on Disinformation (CoP) **[European Commission, 2022]** is a voluntary, self-regulatory framework through which major online platforms pledge to curb the spread of false or misleading content. Its core objectives are to empower users with clear labelling and context, boost the visibility of authoritative sources, remove or demote harmful disinformation, and ensure greater transparency and collaboration with independent fact-checkers. Signatory platforms are required to submit periodic self-assessment reports outlining the policies, mechanisms, and partnerships in place to tackle disinformation, ranging from labelling systems and removal workflows to demotion algorithms and information panels. These reports also include quantitative metrics such as the number of labelled or removed posts, content demoted in feeds, or instances where users were exposed to contextual information. These metrics allow stakeholders to assess the platforms' implementation effectiveness across different regions and time periods. As of 13 February 2025, the European Commission and the European Board for Digital Services formally endorsed the integration of the Code of Practice into the Digital Services Act (DSA), marking its transition from a voluntary code to a co-regulatory instrument for VLOPs and VLOSEs (Code of Conduct). This transition, which takes effect in July 2025[1], reinforces the need for structured and transparent monitoring practices to ensure compliance and assess effectiveness.

In this report, we monitor the practices of online platforms in addressing disinformation by evaluating their self-reported actions submitted as part of the Code of Practice on Disinformation (CoP) Signatories Reports. Specifically, we examine the qualitative and quantitative data disclosed by Meta, Google, and TikTok under the September 2024 CoP reporting cycle, focusing on how these platforms claim to implement their commitments in Cyprus, Greece, and Malta. By systematically reviewing the official submissions made under the CoP framework, we aim to assess the clarity, completeness, and country-specific applicability of their policies, tools, partnerships, and interventions related to disinformation.

In the framework of the MedDMO project, we adopt the EU's original 2018 Code of Practice **definition of "Disinformation"** as "**Verifiably false or misleading information which, cumulatively, (a) is created, presented and disseminated for economic gain or to intentionally deceive the public; and (b) may cause public harm, intended as threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security"** **[European Commission, 2018]**.

This report is part of the second round of the Code of Practice (CoP) report analysis within the MedDMO project. The first report, available on the MedDMO website **[Leonidou et al., 2023]**, covered platform practices from January 1st to June 30th, 2023. In this second round, we build on our earlier monitoring efforts **[Leonidou et al., 2025]**, which used the MedDMO Fact-Check Archive[2] to assess platform responses to debunked disinformation claims across Cyprus, Greece, and Malta from January 1st to June 30th, 2024. That analysis examined whether platforms applied moderation actions such as labelling, removal, or contextual warnings to flagged content.

[1] https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation
[2] https://meddmo.eu/fact-checking/archives/

The current report shifts focus to evaluate how Meta (Facebook and Instagram), Google (Search and YouTube), and TikTok implemented their CoP commitments during the same period. Our assessment draws primarily from the CoP Signatories' Reports submitted in September 2024, complemented by our own investigations. It concentrates on three key pillars: empowering users, supporting the research community, and strengthening the fact-checking ecosystem (see the specific CoP Commitments we covered, outlined in Table 1). We assess both qualitative reporting elements (QREs) and quantitative and implementation data (SLIs) specific to Cyprus, Greece, and Malta. The evaluation also considers the transparency of platform actions and the accessibility of tools and processes for users, researchers, and fact-checkers.

Our assessment is aligned with the methodology proposed by **the EDMO Subgroup on CoP Report Analysis in September 2024** **[Botan et al., 2025].** More specifically, three reviewers/researchers reviewed the Signatories' reports and responded to the questions of the proposed methodology (Sections A and B). Afterwards, platforms' reported actions are evaluated using a structured scale (see Table 2), considering their coherence, transparency, and specificity in relation to each country's context.

In terms of context, Cyprus, Greece, and Malta share similar challenges related to disinformation but also present distinct media and political environments requiring tailored responses. The 2023 Media & News Survey **[European Commission, 2023]** showed that social media platforms dominate news consumption in all three countries. In Cyprus, 70% of respondents use social media for news, with Facebook, Instagram, YouTube, and WhatsApp being the top platforms. In Greece, 55% of the population uses social media, and 57% rely on online news sites, with Facebook and YouTube again ranking highest. Malta follows a similar pattern, with 70% of the population using social media, particularly Facebook and WhatsApp. Across all three countries, influencer-generated content also plays a growing role in how users engage with news—69% in Cyprus, 63% in Greece, and 61% in Malta use social media for news, with substantial portions preferring content from influencers.

While TikTok was not explicitly cited in the survey, it has gained relevance through increasing engagement levels, particularly among younger users. During the 2024 Romanian presidential election, TikTok disclosed the existence of 27,000 fake accounts involved in a coordinated disinformation campaign promoting a far-right candidate **[BBC News, 2025a]**. In Germany, a Global Witness investigation **[Global Witness 2025]** ahead of the federal election revealed that TikTok's recommendation system disproportionately promoted far-right AfD content to politically neutral new users. These examples underline the importance of rigorous monitoring and greater transparency from platforms, especially in how their algorithms surface political content.

Through this report, we aim to assess how well Meta, Google, and TikTok are addressing disinformation in Cyprus, Greece, and Malta. We explore their existing tools, their integration with local contexts and languages, their support for researchers and fact-checkers, and the extent to which their measures are transparent and effective. Ultimately, our goal is to support ongoing efforts in enhancing media literacy, platform accountability, and democratic resilience across the three countries.

Table 1: CoP Commitments covered by this analysis

| CoP Pillars | CoP Commitment | Issue covered by the measure |
|---|---|---|
| V Empowering Users | Commitment 17 | Media Literacy |
| | Commitment 21 | Better equipping users to identify disinformation |
| VI Empowering the Research Community | Commitment 26 | The provided access to platforms' data for researchers |
| | Commitment 27 | The provided access to data necessary to vetted researchers to undertake research on Disinformation |
| | Commitment 28 | Support good faith research into Disinformation |
| VII Empowering the Fact-Check Community | Commitment 30 | Cooperation with EU fact-checking community |
| | Commitment 31 | Integration of Fact-checkers' work in platforms |
| | Commitment 32 | The provided access to information for fact-checkers |

Table 2: CoP Commitment-level assessment scaling system

| **1 to 5 Rating Scale Guidelines** | |
|---|---|
| **Score** | **Interpretation** |
| **1** | **Very Poor:** The tools, activities, or partnerships had little to no impact, were poorly implemented, or had significant issues that prevented them from achieving their objectives. |
| **2** | **Poor:** The tools, activities or partnerships had limited impact, with several deficiencies in implementation or design that reduced their effectiveness. |
| **3** | **Fair:** The tools, activities or partnerships had a moderate impact, achieving some of their objectives but with noticeable areas for improvement. |
| **4** | **Good:** The tools, activities or partnerships were effective, achieving most of their intended outcomes, with only minor areas that could be enhanced. |
| **5** | **Excellent:** The tools, activities or partnerships were highly effective, fully achieving their objectives and demonstrating best practices with no significant shortcomings. |
| **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. |

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

# 3 CoP Signatories Reports Assessment Results

## 3.1 Summary of Results

In this section, we present the overall findings of the analysis. The evaluation of the CoP reports of the Signatories is based on the reported information by each signatory, looking at the completeness and clarity of the reported elements. In cases where it is feasible, we investigate the reported elements, looking for evidence of the reported data, 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), as well as the overall implementation of the reported policies and practices (QREs) by the platforms across the three countries.

A detailed presentation of the evaluation scores—broken down by platform and commitment—along with a table indicating whether relevant SLIs were reported in the Signatories report, is provided in Annex I: CoP Signatories Reports Assessment Scores and Missing SLIs.

**CoP Signatories Reports Assessment Scores and Key Findings per pillar:**

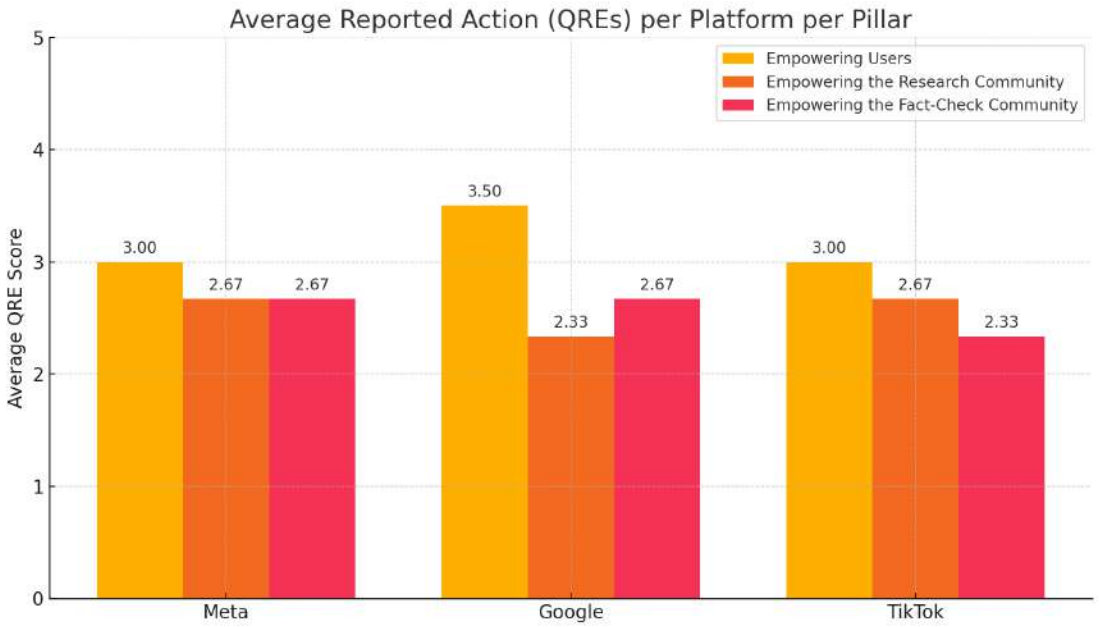



Figure 1: Average scores for each CoP pillar and each platform

Figure 1 illustrates the average scores for each Code of Practice (CoP) pillar for each platform. Among the three assessed pillars, "Empowering the Users" received the highest average scores, ranging from 3 to 3.5 out of 5, across Meta, Google, and TikTok. This suggests that platforms were generally more precise and more complete

in reporting their tools, activities, and partnerships aimed at empowering users with media literacy and verification capabilities.

In contrast, the "Empowering the Research Community" and "Empowering the Fact-Checking Community" pillars received lower scores. These lower ratings primarily stem from insufficient or vague responses in the qualitative reporting elements (QREs). Platforms often repeated the same information across different QREs or failed to provide concrete, detailed descriptions of their activities.

TikTok's report stood out for its clarity and structure, including illustrative screenshots of tools and features, which helped evaluators navigate the submission more easily. However, this clarity in presentation should not be interpreted as indicating a higher overall level of action or effectiveness in addressing disinformation compared to Meta and Google.

***Empowering the Users***

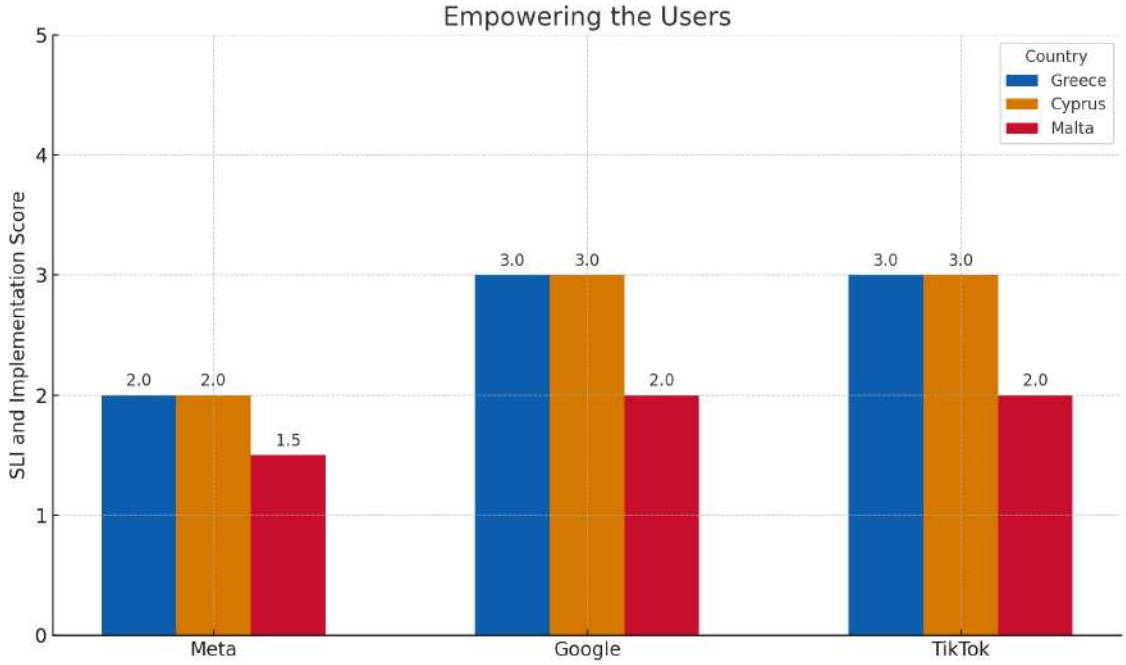



Figure 2: Average scores for each platform under Empowering users pillar

Figure 2 illustrates the assessment scores of the efforts of Meta, Google, and TikTok towards empowering users to identify disinformation across Greece, Cyprus, and Malta. More specifically:

**Lower Scores for Meta - Limited Efforts and Inconsistent Labelling:** Meta received the lowest score among the three platforms for empowering users. The platform reported fewer relevant efforts during the evaluation period and showed weak engagement in localized media literacy initiatives. Our analysis of debunked Facebook content [Leonidou et al., 2025] revealed inconsistent application of false information labels. Even when posts featured identical text or images related to the same false claim, some were labelled while others were not. This inconsistency suggests that Meta's automated labelling system may not reliably flag misleading content, limiting users' ability to make informed decisions.

**Higher Scores for Google and TikTok - More Documented Tools and Local Language Support:** In contrast, Google and TikTok received higher scores for reporting more tools, campaigns, and events aimed at empowering users. Both platforms made some of their resources available in the Greek language, which contributed to improved accessibility for users in Greece and Cyprus. However, the level of detail and localization of these efforts varied, with generic campaigns being more prominent than country-specific ones.

**Absence of Local Media Literacy Events Across All Platforms:** None of the three platforms reported organizing localized media literacy events specifically tailored to Greece, Cyprus, or Malta. This gap highlights a broader issue: a lack of structured collaboration with national media literacy experts and institutions. While global initiatives like Google's HitPause Campaign are technically accessible, they are not adapted to the unique media environments of these countries, reducing their potential impact.

**Language Barriers in Malta:** Lack of Maltese Language Integration: A significant challenge identified in Malta is the limited support for the Maltese language in media literacy tools and campaigns. The platforms primarily rely on English-language materials, which may limit accessibility for Maltese users. Additionally, none of the platforms engaged in fact-checking or media literacy activities specifically in Malta, which may have contributed to Malta's lower scores.

**Limited Collaborations with Local Experts and Fact-Checkers:** Across all three countries, platform partnerships with local experts i.e., journalists, educators, and fact-checkers are either minimal or absent. This hinders the design and adaptation of tools or campaigns to address national disinformation challenges. Malta is particularly affected, as the lack of engagement with fact-checking bodies prevents the development of user-focused features.

**TikTok Shortcomings: Inconsistent Tags and Limited Search Functionality:** Although TikTok scored well in clarity and structure of reporting, it has notable limitations in implementation. Our review found that TikTok's Video Notice Tags were not consistently applied, and its search interventions were unreliable, especially for Greek language queries or those with minor typographical errors. Furthermore, TikTok's "Unverified Content" label does not apply to Maltese content, as there is no local fact-checking partner for the country.

**YouTube Information Panels: Greek Support Available, but Consistency Concerns Remain:** YouTube's Information Panels are available in Greek, which is beneficial for users in Greece and Cyprus. However, the Maltese language is not supported. Users in Malta receive these panels in English by default. During testing, we identified a case where an information panel failed to appear when the interface language was set to Greek but did appear when set to English (UK) (see Figure 14). This inconsistency raises concerns about the reliability of the panels and whether language settings may hinder access to credible contextual information, particularly for non-English-speaking users.

**Meta and TikTok Lack Fact-Checking Partnerships in Malta:** Neither Meta nor TikTok has established fact-checking collaborations in Malta. As a result, disinformation circulating locally is unlikely to be flagged or labelled by platform mechanisms. This further reduces user awareness and weakens the platforms' ability to empower users in the region effectively.

***Empowering the Research Community***

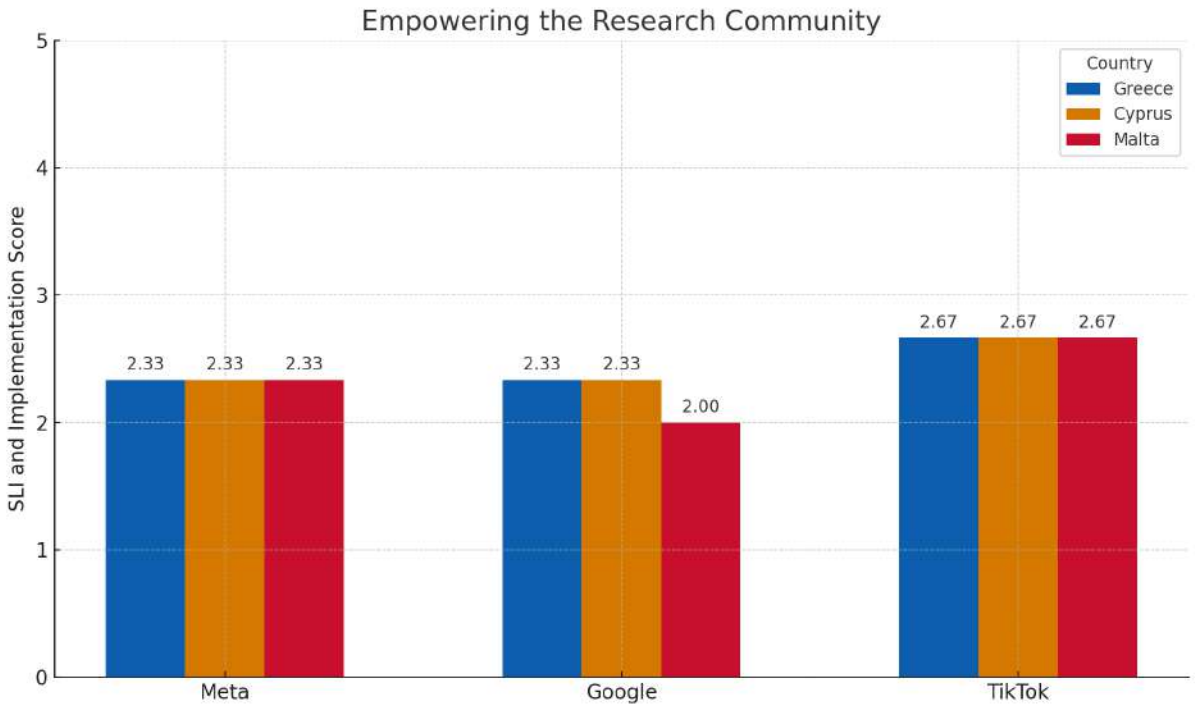



Figure 3: Average scores for each platform under Empowering the Research Community pillar

Figure 3 illustrates the assessment of the efforts of Meta, Google, and TikTok in empowering researchers to combat disinformation across Greece, Cyprus, and Malta. More specifically:

**Consistent but Modest Platform Performance in Supporting Researchers:** All three platforms—Meta, Google, and TikTok—received similar, moderate scores for their efforts to empower the research community across Cyprus, Greece, and Malta. While each platform reports the existence of publicly accessible tools and APIs for disinformation-related research, the actual usage or engagement from these countries is extremely limited, and no meaningful metrics were provided to assess their effectiveness.

**Limited Google Research Tools Functionality for Malta:** Google received a slightly lower score compared to Meta and TikTok, primarily due to limited tool functionality for Malta. For instance, Google's Fact Check Explorer does not index fact-check articles relevant to Malta, and Google Trends lacks country-specific features, such as omitting Malta from the "Trending Now" location filter. These limitations reduce research visibility into Maltese disinformation trends.

**Promoting the Tools to Researchers:** Although platforms did fulfil their obligation to list data access tools and repositories for researchers, these tools appear to be under-promoted and underutilized. To improve impact, platforms must actively promote, localize, and provide guidance on how to use these tools within each country's academic and research communities.

**Disinformation Labels from APIs:** Crucially, researchers also need access to deeper layers of platform data, such as whether content has been labelled, modified, or removed under disinformation policies. Improved

transparency and access to such data via APIs is essential to enable high-quality, independent research on disinformation dynamics within and across the three countries.

Table 3 lists several resources (tools and data repositories) available to researchers as mentioned in their CoP reports No.4.

Table 3: Available resources for Researchers found in the Platforms Reports

| **Tool / Data Repositories for researchers** | **Description** |
|---|---|
| **Meta** | |
| Meta Content Library and API | Enables researchers to access public content from Facebook and Instagram. Researchers can apply through the Inter-university Consortium for Political and Social Research (ICPSR). Includes search, filtering, and download functions with access thresholds. |
| Meta Ads Library | Publicly searchable database of all active ads across Meta platforms, including issue, electoral, and political ads. Offers ad creative, spend, targeting data, and impressions. |
| Ad Targeting Dataset | Political and social issue ad targeting data since August 2020. |
| URL Shares Dataset | Aggregated engagement data for links. |
| Influence Operations Research Archive | Data on Coordinated Inauthentic Behaviour (CIB) networks. |
| Data for Good | Dashboards with aggregated insights. |
| **Google** | |
| Google Researcher Program | Grants are available to EU researchers (including vetted non-academics) for limited metadata access across Search and YouTube to study public data. A dedicated team reviews applications; successful applicants receive API credentials and usage guidance. |

| | |
|---|---|
| YouTube Researcher Program | Provides qualifying academic/non-profit researchers worldwide with expanded access to anonymized, global YouTube video metadata (titles, descriptions, views, likes, comments, channel information, and search results) via a Data API. Includes formal application and support. |
| Google Trends | Publicly accessible tool showing aggregated, anonymized search-interest data over time and by geography. Researchers can analyze search trends for keywords, topics, and regions to study information demand and potential disinformation spikes. |
| Fact Check Explorer | Index of ClaimReview-tagged fact-checked articles from certified publishers worldwide. Researchers can search by keyword, URL, or date to retrieve verified claims and metadata (claim origin, publisher, verdict). |
| Claim Search API | Programmable interface allowing researchers to retrieve ClaimReview fact-check results in JSON format. Useful for bulk analysis of claim metadata, publisher details, and review dates. |
| Google Ads | A publicly accessible database that provides transparency into political and issue-based advertising across Google platforms. It includes detailed information about ad content, spend, impressions, advertiser identity, geographic targeting, and more. Researchers and the public can search and explore ads to better understand how different topics are promoted and targeted across regions. |
| Search Researcher Result API | Beta API that returns anonymized, aggregated search results (SERP snippets, ranking signals) to approved researchers. Enables analysis of how certain queries produce different result types, without exposing individual user data. |
| Lumen Database (Harvard Berkman) | Independent repository of legal takedown and content-removal notices voluntarily shared by companies (including Google). Researchers can query millions of notices to study patterns in online content availability and removal requests. |
| AMMeBa Dataset | Publicly released dataset of 135,838 ClaimReview fact-checks annotated for media modality (image, video, AI-generated, etc.). Enables researchers to study historical and evolving trends in media-based disinformation from 1995 onward. |

| **TikTok** | |
|---|---|
| Research API | Programmatic access to *public* TikTok content and account metadata—including comments, captions, subtitles, counts (views/shares/likes), and follower/following lists—available to qualified non-profit academic researchers in major regions. |
| Virtual Compute Environment | Secure, sandboxed environment for non-academic, not-for-profit EU researchers to query and analyse TikTok's public data under strict privacy and security controls. |
| Commercial Content API | API providing detailed advertising data—creative assets, run dates, targeting criteria, impressions, spend ranges, and disapprovals—for ads served in the EU; access via approved TikTok for Developers accounts. |
| Commercial Content Library | Publicly searchable repository of all paid ads (and tagged commercial-nature posts) run in the EEA, UK, and Switzerland, with creatives, metadata (targeting parameters, impressions), and one-year retention after last view. |
| TikTok Transparency Centre | Houses TikTok's core moderation and policy reports for EU audiences: COPD (biannual), Community Guidelines Enforcement (quarterly), and DSA (biannual) reporting with downloadable aggregated data. |

***Empowering the Fact-Checking Community***

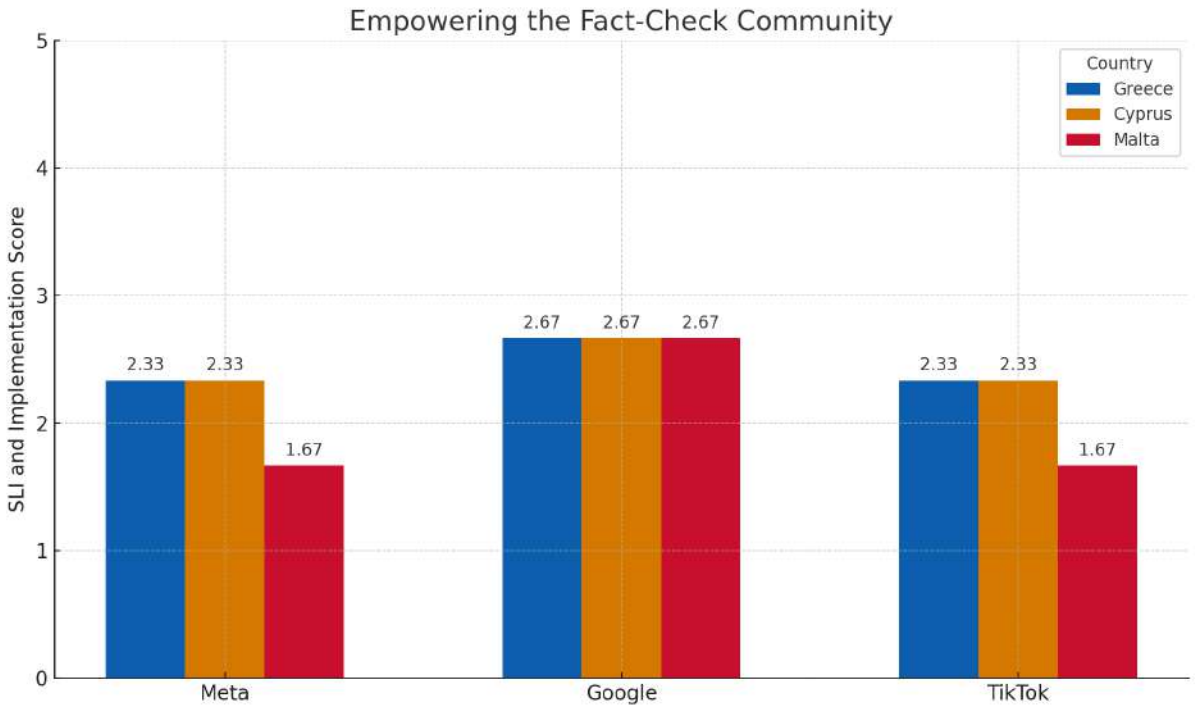



Figure 4: Average scores of each platform under Empowering the Fact-Checking Community pillar

Figure 4 illustrates the assessment scores of the reported efforts of Meta, Google, and TikTok in empowering the fact-checking community across Greece, Cyprus, and Malta. The scores reflect platform practices in terms of operational partnerships, transparency of tools, integration of fact-checks, and regional funding distribution. Key observations include:

**Uneven Integration and Coverage Gaps**: Meta, Google, and TikTok exhibit moderate engagement in Greece and Cyprus, but a significantly weaker presence in Malta. The absence of partnerships with Maltese fact-checkers, combined with limited or no support for the Maltese language, has resulted in serious coverage gaps that reduce the platforms' ability to address disinformation effectively at the national level.

**Differences in Fact-Checking Integration Mechanisms**: Each platform adopts a different model for integrating fact-check outcomes. Meta uses fact-checkers' ratings to label content as "false information" and links to the relevant fact-check articles directly. TikTok does not apply fact-check labels directly; instead, it routes fact-checkers' assessments to internal moderators, who then decide whether to remove or demote the content—making the integration less direct and opaque. In contrast, Google does not label content as true or false. Instead, it integrates fact-checking outcomes by surfacing fact-checking articles in its search results through the ClaimReview markup and by displaying information panels alongside content to provide contextual information.

**Lack of Methodological and Impact Transparency**: All three platforms report ongoing collaboration with fact-checking partners (primarily through IFCN), yet none of them provide clear methodologies or performance assessments for evaluating the effectiveness of fact-checking integrations. TikTok and Meta do not disclose usage metrics for their fact-checking dashboards or internal tools. Google also lacks follow-up metrics for its

ClaimReview markup and does not provide post-placement impact data (e.g., click-throughs, or engagement resulting from fact-check appearances).

**Minimal Country-Specific Disaggregation of Impact Data**: TikTok shares country-level data on videos assessed and removed due to fact-checker input. However, the numbers are very low. Meta reports the number of articles used to support content ratings and shares cancelled after users saw false information warnings. Google provides fact-checking impression metrics and article counts by language and country in its Fact Check Explorer. Yet, across all three platforms, this data lacks sufficient granularity or context to meaningfully assess impact at the national level.

**Generic or Inaccessible Fact-Checker Tools**: Google promotes general tools such as Search Console and YouTube Studio, which fact-checkers can use to assess site traffic and video performance. However, these tools are not tailored for fact-checking workflows and are available to all users. Google's only tool explicitly designed for fact-checking—the ClaimReview markup—also lacks a feedback mechanism. Meta and TikTok reference internal dashboards, but do not offer public details on their structure, usage, or effectiveness.

**Funding Gaps and Opaque Support Structures**: Google indirectly supports fact-checkers through funding bodies such as IFCN, EFCSN, and EMIF. In 2024, only fact-checking projects in Cyprus and Greece received limited grants. Malta received no support under the most recent IFCN funding phases. Meta does not disclose country-level funding details or outline how resource allocation reflects local disinformation risks or fact-checking capacity.

**In summary**, while the platforms maintain some form of collaboration with fact-checkers and promote select tools, they fall short in offering country-specific impact data, tool transparency, and effective integration frameworks—particularly in Malta. Clearer data on tool performance, regional investment, and integration outcomes is essential to ensure that fact-checking communities across all three countries are fully empowered in accordance with the Code of Practice on Disinformation.

Table 4 lists several tools available to support fact-checking organizations in their efforts. Only Google reported such tools, while Meta and TikTok only mentioned the dashboards available for fact-checking organizations they partner with.

Table 4: Available resources for Fact-Checkers found in the Platforms Report

| **Tool for fact - checkers** | **Description** |
|---|---|
| Google Search Console | Monitors a site's presence in Google Search, reporting on impressions, clicks, queries, and country-specific views so fact-checkers can track their web performance. |
| YouTube Studio | Central dashboard for channel management and analytics, showing fact-check video views, watch time, traffic sources, and search terms that lead viewers to content. |
| ClaimReview HTML Markup | A schema.org format that fact-checkers embed in their articles; allows Google to index and display structured fact-check metadata in Search and YouTube. |
| YouTube Create | A mobile app for rapid video production—offering templates, filters, music, and editing controls—so organizations (including fact-checkers) can quickly create and publish short-form debunk videos. |

## 3.2 Recommendations for Platforms and National Experts in Cyprus, Malta, and Greece to Strengthen the Fight Against Disinformation

In this section, we present our recommendations to enhance the coverage and effectiveness of platform practices—both existing and future—in addressing disinformation in Cyprus, Greece, and Malta. The suggestions are organized according to the three pillars of the Code of Practice: Empowering Users, Empowering the Research Community, and Empowering the Fact-Checking Community. This structure allows for targeted, actionable insights tailored to each area of intervention.

***Empowering Users***

**Develop Localized Tools and Language-Specific Resources:** Platforms should ensure that all media literacy tools—such as disinformation warnings, fact-check labels, context panels, and in-app educational features—are available in the national and regional languages used in Cyprus, Greece, and Malta. This includes full support for Greek, Cypriot Greek, and Maltese. Localized content should also reflect context-specific disinformation narratives, such as those tied to regional elections, migration, or public health debates. Without language and cultural adaptation, such tools risk being underused or misunderstood by local users.

**Improve Visibility and Discoverability of Media Literacy Tools:** Many users remain unaware that platforms offer tools such as fact-check notices, reverse image search guidance, or credibility signals. **Platforms should integrate clear prompts, banners, and tutorials** directly into the user interface, especially around high-risk events like elections. These features should be visible in news feeds, search results, and trending topics. Additionally, **periodic awareness campaigns**—using pop-ups or short explainer videos in local languages—can enhance user familiarity with available verification tools.

**Encourage Stakeholder Feedback and Co-Design of MIL Tools:** Platforms should actively engage national experts in media literacy (e.g., educators, journalists, NGOs, university researchers) from Cyprus, Greece, and Malta in the design, testing, and refinement of their media literacy and fact-checking features. This can be achieved through local focus groups, usability studies, or consultative forums. Tools co-designed with local stakeholders are more likely to address cultural norms, digital habits, and the specific vulnerabilities of each country's information environment.

**Ensure Consistent Labelling of False Content Across Languages and Regions:** One of the critical issues observed is the uneven application of fact-check labels, particularly for Greek-language posts across different platforms and regions. Platforms must implement cross-regional consistency so that if a claim has been debunked (e.g., by a Greek fact-checker), identical posts in Cyprus or in Cypriot Greek also trigger the same label. This consistency ensures that all users are not confused and that all audiences receive timely and accurate correction notices.

**Provide Local-Language Context Panels and Correction Prompts:** When disinformation is flagged, platforms should not only label it but also display short summaries of why the content is misleading, link to relevant fact-checks, and do so in the local language. These context panels should appear prominently (e.g., above the post or in the comment feed), helping users understand the basis of the correction rather than simply seeing a warning badge. This enhances user trust and increases the likelihood of behaviour change.

**Broaden Community-Driven Annotation Programs:** Platforms that use systems like Community Notes (on X/Twitter) should expand them to support local languages and regional contributors. This includes recruiting Greek- and Maltese-speaking users to participate in community-based verification and incentivizing diverse participation. Expanding community moderation gives local users a greater sense of agency in combating disinformation and can help flag emerging false narratives before they scale.

**Implement Safe Design Features to Limit Spread of Disinformation:** Platforms should adopt "friction" features—such as pop-up warnings before resharing flagged content, slowing down virality of unverified posts, or hiding likes on content that has been debunked. These interventions reduce impulse-sharing of falsehoods while preserving transparency. Platforms should also down-rank content that has been fact-checked as false, so it appears less frequently in feeds and recommendation systems, especially during periods of heightened risk (e.g., elections or health crises).

**Boost Digital Literacy and Awareness Campaigns:** To maximise reach, platforms should collaborate with schools, media outlets, and civil society organisations to co-develop media literacy campaigns tailored to Cyprus, Greece, and Malta. These could include educational videos, classroom toolkits, or influencer-led explainers promoting critical thinking and digital verification techniques. Platforms should fund or co-host such initiatives in both online and offline settings, especially targeting younger and first-time voters.

***Empowering Researchers***

**Provide Access to Data for Independent Research:** Enable vetted researchers and EDMO hubs (such as MedDMO) to access granular platform data relevant to disinformation in each country. This includes sharing anonymized data on content views, shares, takedowns, and labelling actions in Cyprus, Greece, and Malta. Such access is a precondition for effective monitoring of the local disinformation ecosystem, allowing researchers to identify trends and gaps in platform responses.

**Include Small Languages in Transparency Reporting:** In transparency reports and archives, break down information by country and language so that data from Malta and Cyprus is not hidden in aggregate. Platforms should report country-specific metrics – for instance, the number of fact-checked posts, false content removals, and user warnings issued in Greek and Maltese – to facilitate accountability. This granular transparency enables researchers and policymakers to gauge coverage and effectiveness of platform measures at the national level.

**Improve Researcher API Tools and Support:** Enhance existing APIs and data-sharing tools to ensure they cover content in local languages (Greek, Maltese) and are easy for research teams to use. Platforms should streamline data request processes under the Digital Services Act's researcher access provisions and provide documentation or liaisons for researchers in Cyprus, Greece, and Malta. By lowering technical barriers, platforms empower the research community to study disinformation dynamics and the impact of interventions in these markets.

**Collaborate on Joint Studies and Risk Assessments:** Work directly with academic institutions and observatories in the region to investigate platform-specific disinformation risks. For example, platforms can invite local researchers to co-design studies or share insights when performing risk assessments for upcoming elections in Greece or Cyprus. This cooperation ensures that platform policies are informed by on-the-ground expertise and evolving disinformation tactics in each country.

***Empowering the Fact-Checking Community***

**Integrate Local Fact-Checkers into Platform Programs:** Formally include credible national fact-checking organizations in Cyprus, Greece, and (especially) Malta into the platforms' third-party fact-checking partnerships or equivalent collaboration frameworks. In Malta (where no such partnerships exist yet), platforms should onboard local fact-checkers (e.g. Times of Malta, University of Malta team) so that their debunks feed into in-app warnings and reduced content reach. Likewise, any Greek or Cypriot fact-checking outlets not already in programs should be added. Bringing local fact-checkers into the fold **widens fact-checking coverage across the EU** and ensures even smaller communities are protected.

**Increase Visibility of Local Fact-Check Results:** Adjust algorithms and notification systems to boost the prominence of local fact-checks on the platforms. When local fact-checkers debunk a false story targeting Greek or Cypriot audiences, platforms should swiftly label the content and display the fact-check link or summary to all users who have seen or shared it. This practice amplifies trustworthy local sources and helps correct falsehoods before they spread further.

**Support Capacity-Building for Fact-Checking Organizations:** Provide direct support to strengthen fact-checking efforts in these countries. This could include offering training on platform policies and tools, providing grants or advertising credits for public-awareness campaigns by fact-checkers, and sharing relevant data or tools (like claim detection algorithms) with them. By investing in the local fact-checking community, platforms improve the **effectiveness** of moderation – verified false claims get addressed faster and more reliably.

**Facilitate Cross-Border Fact-Checking Collaboration:** Encourage collaboration between fact-checkers in Greece, Cyprus, and Malta as well as those in other EU countries. Platforms can host regular roundtables or an online portal for fact-checkers to exchange information on emerging disinformation themes (for instance, similar false narratives that appear in Greek and then in Maltese contexts). Such networks, supported by platform data where possible, ensure that fact-checking coverage is coordinated and no country is left behind when dealing with region-wide disinformation campaigns.

Next, we present a platform-by-platform analysis, highlighting key reported information and evaluating each platform's practices. Our assessment focuses on tools, activities, and partnerships that platforms report under their disinformation mitigation efforts. Where applicable, we illustrate these tools with examples and share findings from our own investigations—especially in cases where we observed strong performance or notable gaps.

The evaluation includes both general commentary on the quality and clarity of the reported elements, as well as specific insights into their relevance for the three countries under review: Cyprus, Greece, and Malta. For each commitment, we include an evaluation matrix that displays the scores assigned to the Qualitative Reporting Elements (QREs) and the per-country scores for Service Level Indicators (SLIs) and implementation. Each section concludes with a summary of the reported content we assessed, followed by a discussion of the key observations and shortcomings that influenced our scoring.

## 3.3 Meta (Facebook and Instagram)

Our analysis of Meta's practices is based on the information provided in Meta's Code of Practice Signatory Report No. 4, submitted in September 2024, covering the period from January 1st to June 30th, 2024 **[Meta, 2024]**. It is important to note that in January 2025, Meta announced a significant policy change: it would begin phasing out its collaboration with third-party fact-checking organisations—initially in the United States—who had been independently assessing the accuracy of claims to enable Meta to apply warning labels and demote misleading content. Instead, Meta plans to implement a Community Notes system, resembling the one used by X (formerly Twitter) **[Kaplan, 2025][BBC, 2025b].** The CoP report assessed in this study reflects the period prior to this announcement and does not include any information on Meta's new direction.

### 3.3.1 Meta: Empowering Users

| **Pillar-V: Empowering Users**<br>***Commitment 17: Enhancing Media Literacy,***<br>*page 66-71* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 17** | 3 | 2 | 2 | 2 |

**Meta Media Literacy Tools**

Meta has reported several features aimed at improving user awareness and promoting media literacy across its platforms:

- **Warning Screens** on Facebook and Instagram for content containing sensitive material such as violent or graphic imagery, bullying or harassment, nudity, or references to suicide or suicide attempts.
- **Verified Badges** for public figures and notable accounts on Facebook and Instagram to help users identify authentic profiles and reduce impersonation and scams.
- **Notification Screens for Outdated Articles** on Facebook, which appear when users attempt to share news content that is over 90 days old. These notifications aim to provide users with contextual awareness. Notably, official health content from government agencies or recognized global organizations is exempt to avoid obstructing the flow of verified information.

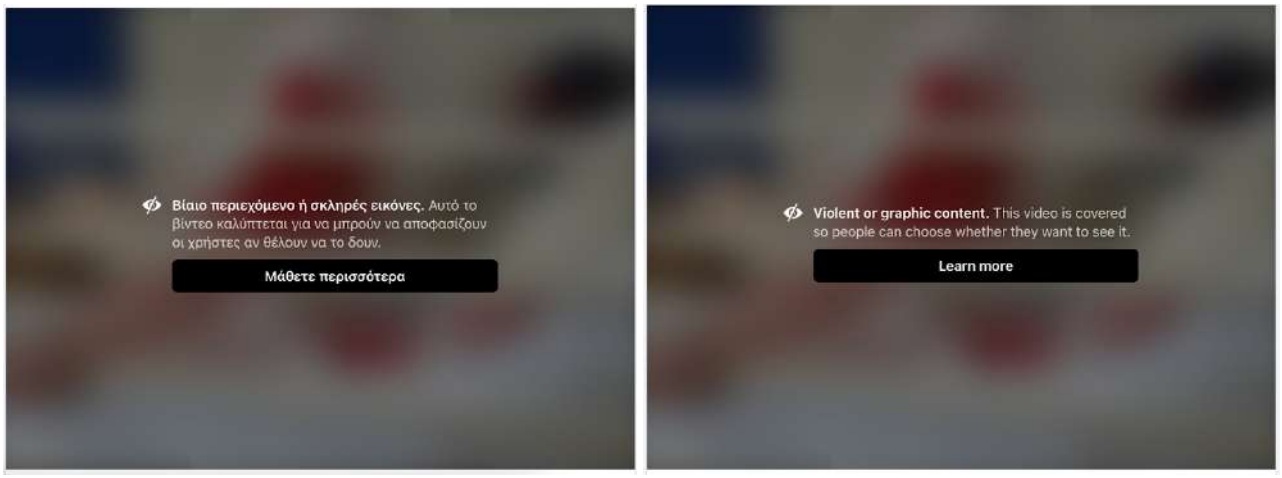



Figure 5: Facebook Warning Screens

**Meta Media Literacy Activities**

During the reporting period, Meta promoted a range of initiatives to support informed civic engagement and counter disinformation across Europe:

- European Parliament Elections 2024
  - Ran awareness campaigns in Germany, Poland, Slovakia, Lithuania, and Romania, promoting Facebook and Instagram tools related to electoral integrity.
  - Promoted the #BeElectionSmart website by EDMO and included it on Meta's dedicated election hub[3].
  - Provided ad credit support to ERGA (European Regulators Group for Audiovisual Media Services) for its anti-disinformation campaign.
- Country-Specific Efforts
  - Germany: Supported the Federal Returning Officer's "get-out-the-vote" campaign.
  - France: Partnered with AFP Fact Check and participated in the disinformation awareness campaign led by Génération Numérique. Meta also ran its own election awareness campaign in the country.
- Thematic and Regional Campaigns
  - EFCSN Collaboration: Supported workshops and educational efforts focused on AI-generated disinformation.
  - Youth Awareness Campaigns: Conducted campaigns in Spain, Ireland, and the Netherlands to promote well-being tools for young users.
  - European Disability Forum: Delivered a webinar focused on online campaigning techniques for disability rights organizations.
  - CEE Digital Awareness Day: Co-organized a regional event featuring media literacy experts from 17 Central and Eastern European countries.

[3] https://www.facebook.com/government-nonprofits/eu

**Meta's Partnerships with Media Literacy Experts**

Meta reported collaborations with NGOs, academics, and fact-checking organizations to enhance media literacy and combat disinformation, including:

- Ongoing work with **AFP Fact Check** and **Génération Numérique** in France.
- Training and capacity-building initiatives with the **European Fact-Checking Standards Network (EFCSN)**.
- Youth-centered campaigns and digital well-being initiatives in various EU member states.
- Participation in regionally coordinated events to address local disinformation narratives and improve digital resilience.

These activities were designed to reinforce the transparency and reliability of online information, particularly in politically sensitive periods such as elections.

**Major Comments:**

- Warning screens on sensitive content are translated into Greek but not Maltese when users set the platform language to Greek or Maltese, respectively (see
- Figure 5 in which on the left the language setting is in Greek and on the right it is in English). These screens are not applied to content containing disinformation, misinformation, or misleading claims.

- No data is provided on the number of accounts with verified badges at the member-state level. There is also no metric reported to assess how verified badges affect user interactions or trust.

- No quantitative information is reported under SLI17.1.1. This means there is no data available at either the national or EU level on the impact of Meta's media literacy tools—such as total impressions or user engagement —as recommended in the Code of Practice.

- Media literacy campaigns listed in Meta's report were not tailored to Cyprus, Greece, or Malta. There are no metrics on reach or engagement specific to these countries.

- Meta reported providing ad credit support to ERGA for an EU election awareness campaign but gave no implementation details. National broadcasting authorities in Cyprus and Greece confirmed the action, including that the campaign was translated into Greek.

- Meta's reporting on partnerships with experts is general and lacks granularity by member state. It is unclear whether any partnerships exist in Cyprus, Greece, or Malta.

| Pillar-V: Empowering Users<br>*Commitment 21: Better Equipping Users to Identify Disinformation,*<br>*page 80-84* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 21** | 3 | 2 | 2 | 1 |

**Meta Third Party Fact-Checking Program (3PFC):**

Meta collaborates with over 45 independent third-party fact-checkers certified by the International Fact-Checking Network (IFCN) and the European Fact-Checking Standards Network (EFCSN) across Europe. In the EU, Meta works with 29 partners covering 23 languages in 26 countries. Their fact-checking efforts have a global impact, as false-rated posts receive demotions, warnings, and notifications worldwide.

Globally, Meta's 3PFC includes nearly 100 organizations working in over 60 languages to combat misinformation. Fact-checkers operate independently, reviewing content and assigning **ratings** such as **False, Altered, Partly False, Missing Context, Satire, and True**. Meta then labels, limits distribution, and sanctions repeat offenders based on these ratings. Meta also offers the link to the fact-checking article related to the fact-checking label on the post. If a user is about to share content that has received the fact-checking label, they receive a warning or reminder that the content contains false information, in order to prevent the spread of disinformation.

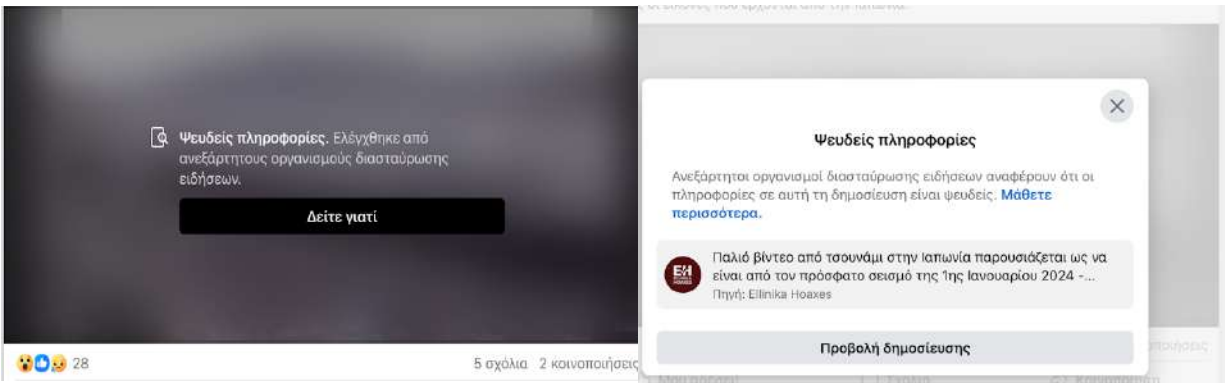



Figure 6: Fact-checking rating label in Greek on a Facebook post and dialog window with link to the fact-checking article.

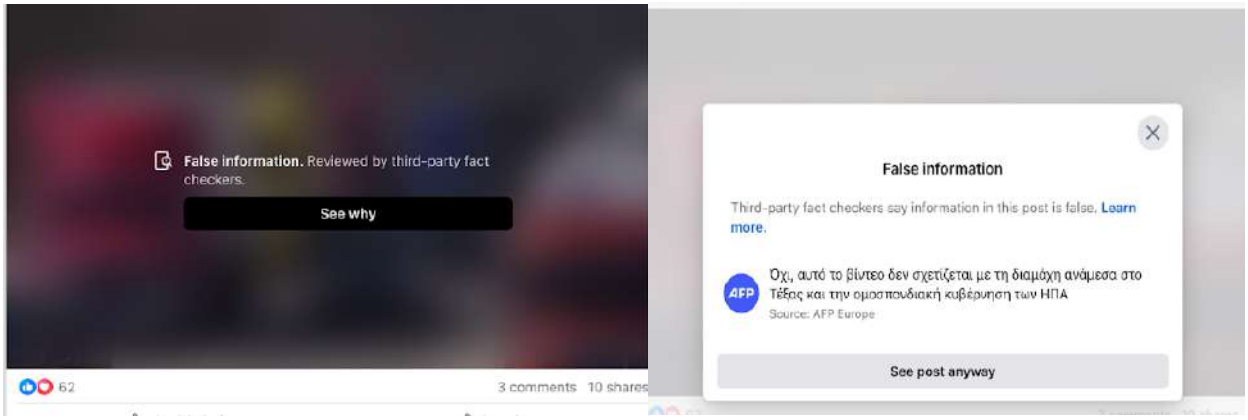



Figure 7: Fact checking rating label in English on a Facebook post and dialog window with link to the fact-checking article.

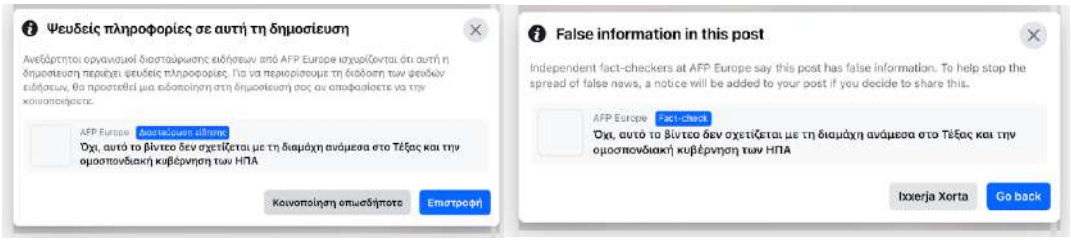



Figure 8: Warning message to prevent user from sharing the content in Greek and English.

**Major Comments:**

- Meta does not have partnerships with any fact-checking organizations covering Malta or the Maltese language. In contrast, Cyprus is covered by AFP, and Greece is covered by both AFP and Ellinika Hoaxes—both of which are partners in the MedDMO project and members of Meta's Third-Party Fact-Checking Program (3PFC).
- Fact-check rating labels are currently available in Greek (see Figure 6 and Figure 8) but not in Maltese (see Figure 7 andFigure 8). When users select Maltese as their language, Meta prompts them to add a secondary language to access features that are not supported.
- An analysis of the Greek MedDMO's fact-check archive and Facebook post searches related to false claims—specifically those verified by AFP and Ellinika Hoaxes—reveals inconsistencies in Meta's labelling system **[Leonidou et al., 2025].** While some posts receive fact-check labels, others presenting the same false claims remain unlabelled. This inconsistency suggests that Meta's algorithmic application of fact-check labels is not systematic, even when dealing with content reviewed by verified 3PFC partners. These gaps should be addressed to improve Meta's disinformation response.
- According to metrics reported under SLI21.1.2 (see table below), Facebook consistently displayed more fact-check articles and generated more treated content views than Instagram across all countries, suggesting either higher misinformation volumes or stronger enforcement on Facebook.

- Reshare prevention rates on Facebook were highest in Malta (57%), followed by Cyprus (49%) and Greece (47%). On Instagram, Greece led with 48%, followed by Cyprus (44%) and Malta (40%).
- Meta does not provide a ratio comparing fact-check impressions to the original reach of the false content, making it difficult to evaluate how many users saw corrective information.
- The platform also does not disclose how many users were penalized as repeat misinformation offenders or provide a breakdown of fact-check labels by category (e.g., False, Altered, Partly False, Missing Context, Satire, True), limiting transparency around enforcement practices.
- Global metrics under SLI21.1.2, which were included in previous reports, were not reported by Meta in this reporting round.
- Meta did not report any internal studies or evaluations assessing the effectiveness or accuracy of its labelling and rating systems.

Table 5: Meta's reported quantitative information for SLI 21.1.2, page 82.

| SLI 21.1.2 | Number of Articles written by third party fact checkers to justify rating | | Content viewed on Facebook and treated with fact checks, due to a falsity assessment by third party fact | | % of reshares attempted that were not completed on treated content | |
|---|---|---|---|---|---|---|
| | Instagram | Facebook | Instagram | Facebook | Instagram | Facebook |
| Cyprus | over 8,000 | over 30,000 | over 35,000 | over 260,000 | 44% | 49% |
| Greece | over 12,000 | over 46,000 | over 96,000 | over 1,200,000 | 48% | 47% |
| Malta | over 5,100 | over 23,000 | over 15,000 | over 110,000 | 40% | 57% |
| Total EU | over 39,000 | over 150,000 | over 990,000 | over 30,000,000 | 43% | 46% |

### 3.3.2 Meta: Empowering the Research Community

| Pillar-VI: Empowering the Research Community,<br>*Commitment 26: Empowering the Research Community,*<br>*pages 96 - 100* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 26** | 3 | 2 | 2 | 2 |

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

**Meta Content Library and API[4]:**

- Rolled out in June 2023.
- Includes data: for Instagram– accounts, posts and comments, for Facebook – Pages, posts, comments, profiles, groups and events.
- 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[5]** 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[6]** 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[7] 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.

[4] https://developers.facebook.com/docs/content-library-and-api
[5] https://developers.facebook.com/docs/content-library-api/get-help
[6] https://developers.facebook.com/docs/content-library-api/get-access
[7] https://transparency.meta.com/researchtools/ad-library-tools

- 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

**Major Comments:**

- Meta's Content Library and API are well-documented and accessible to researchers in the EU through the SOMAR platform[8].
- Meta reports global usage metrics for the Meta Content Library, it does not provide country-level uptake data, leaving a gap in understanding local research engagement.
- Meta states its commitment to improving the Content Library and has introduced new data fields, including access to comments and account profile information.
- However, the Content Library and API still lacks key data fields that would be especially valuable for disinformation research—such as endpoints related to fact-checking ratings, links to fact-checking articles, or labelling decisions.
- The Meta Content Library limits access to public posts on Facebook and Instagram based on follower and page size thresholds[9]—specifically, only pages with more than 15,000 followers, and personal or creator accounts with over 25,000 followers are included in the downloadable subset. While public content from accounts with at least 1,000 followers is available for viewing, these higher thresholds for data export significantly restrict researchers' ability to analyse disinformation dynamics in smaller countries like Cyprus and Malta, where such large-scale accounts are comparatively rare.

[8] https://icpsr.atlassian.net/servicedesk/customer/portal/53
[9] https://developers.facebook.com/docs/content-library-and-api/appendix/data-dictionary

| Pillar-VI: Empowering the Research Community,<br>***Commitment 27: Governance Structure for Access to Data for Research Purposes Requiring Additional Scrutiny,***<br>*pages 100 - 103* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 27** | 3 | 3 | 3 | 3 |

**Data Sharing Programs:**

Meta has been engaged in the EDMO working group on Platform-to-Researcher data sharing since 2019. They participated in a pilot program with EDMO researchers, testing the platform-to-researcher data sharing process. Meta's reported information highlights its active participation in the EDMO working group and support for the creation of an **Independent Intermediary Body (IIB)**[10]. As part of this pilot, they shared data with vetted researchers via Centre D'Accès Sécurisé Aux Données (CASD), a France-based third-party secure data access center, ensuring a privacy-preserving approach.

Meta also mentions engaging with ICPSR (Inter-university Consortium for Political and Social Research) as another third-party organization involved in vetting researchers and hosting access to datasets. Specifically, ICPSR is currently vetting researchers and providing access to datasets related to:

- **The US 2020 election**[11]
- T**he Meta Content Library and API**.

They have also expanded their Influence Operations (IO) Archive dataset since 2021, which provides access to over 100 removed networks related to Coordinated Inauthentic Behavior (CIB[12]), allowing researchers to conduct both qualitative and quantitative analyses.

[10] An Independent Intermediary Body (IIB) is a third-party organization designed to facilitate the sharing of data between digital platforms and vetted researchers while ensuring compliance with privacy laws, such as the GDPR. The IIB acts as a neutral intermediary to oversee and secure the data-sharing process, ensuring that personal data is handled appropriately and that the research conducted aligns with legal and ethical standards. It helps build trust and transparency between platforms, researchers, and the public.

[11] https://www.icpsr.umich.edu/web/about/cms/5024

[12] https://about.fb.com/news/tag/coordinated-inauthentic-behavior/

**Major Comments:**

- There is no clear disclosure of funding for the Independent Intermediary Body, and while Meta references ongoing engagement, the timeline for establishing this third-party entity lacks the specificity expected under Measure 27.1. Providing more concrete details would strengthen the credibility of the report.
- Meta's response would also benefit from additional clarity regarding the number of research teams participating in the pilot programs, along with any preliminary findings or insights. It remains unclear whether these pilots include access to content that has been removed from the platform, as stipulated in Measure 27.4.
- Overall, the reported information is overly generic and difficult to evaluate. To our knowledge, no research institutes in Cyprus, Malta, or Greece have been granted access to platform data as vetted researchers under these pilot initiatives.

| **Pillar-VI: Empowering the Research Community,** *Commitment 28: Cooperation with Researchers,* *pages 103 - 105* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 28** | 2 | 2 | 2 | 2 |

**META Research Support & Engagement:**

Meta has a dedicated team to facilitate research and engage with the research community. Information on available research tools is provided through the Transparency Center[13].

**Meta provides access (only to researchers) to the following datasets:**

1. Meta Content Library & API – Public posts, pages, groups, and events.
2. Ad Targeting Dataset – Political and social issue ad targeting data since August 2020.
3. URL Shares Dataset – Aggregated engagement data for links.
4. Influence Operations Research Archive – Data on Coordinated Inauthentic Behavior (CIB) networks.
5. Data for Good – Dashboards with aggregated insights.

[13] https://transparency.meta.com/researchtools/

**Major Comments:**

- META outlines the available research tools but does not specify how they engage with researchers beyond providing tools.
- Meta does **not** disclose specific financial allocations for disinformation research under its cooperation with the European research community.
- There is no information to assess the reported information at a member-state level.

### 3.3.3 Meta: Empowering the fact-checking community

| ***Pillar-VII: Empowering the fact-checking community*** <br> ***Commitment 30: Cooperation with the fact-checking community,*** <br> *page 108-113* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 30** | 3 | 2 | 2 | 1 |

As previously mentioned in Commitment 21, Meta has agreements with independent fact-checking organizations that follow the International Fact-Checking Network (IFCN) Code of Principles. From 2024, fact-checkers certified by the European Fact-Checking Standards Network (EFCSN) may also be onboarded.

The list of fact-checking partners covering specific countries or languages during the particular period is available in Meta's Report No. 4, on pages 111-112 **[Meta, 2024]**.

Meta reported that they have a team dedicated to working with fact-checkers, gathering feedback, and improving processes. Regular reviews and consultations are mentioned, but without details on frequency or impact.

Meta engages with EDMO regional hubs but does not outline specific collaboration outcomes.

**Major Comments:**

- META does not fully provide details on expected outcomes and the impact of fact-checking.
- META lacks transparency regarding per-country funding and the effectiveness of fact-checking.
- Meta mentions regular review and feedback of the fact-checking process and the cooperation with the fact-checking organizations, but there are no results or findings of these reviews and feedback.
- Meta lists the fact-checking organizations covering EU countries and languages; however, the lack of disclosed funding details and the number of fact-checkers per organization involved in Meta's rating makes it unclear how the funding and agreements effectively meet the needs of each country.
- The Meta 3PFC program covers Cyprus and Greece (Greek language) through their collaboration with AFP and Ellinika Hoaxes, but not Malta (Maltese language)

| ***Pillar-VII: Empowering the fact-checking community*** <br> ***Commitment 31: Use and integration of fact-checking*** <br> *page 114-119* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 31** | 3 | 3 | 3 | 2 |

Meta's actions towards Commitment 31 focus on integrating independent fact-checkers, labeling misinformation, and assessing the impact of fact-checking efforts across its platforms:

Meta's collaboration with fact-checkers leads to labelling content as "false", "altered", or "missing context", reducing visibility and issuing warnings before users share it, along with links to fact-check articles.

In January–June 2024, Meta (in the EU) labelled more than 30 million Facebook posts and more than 990,000 Instagram posts, relying on more than 150,000 fact-checking articles. Additionally, 46% of Facebook and 43% of Instagram users cancel reshares after seeing these warnings.

Meta funds fact-checking initiatives but does not detail contributions to a centralized fact-check repository or shared technological solutions.

**Major Comments:**

- Meta incorporates fact-checkers' ratings by labelling content and linking to relevant fact-checking articles, demonstrating a degree of integration of fact-check outcomes into its services. However, no fact-checking organizations currently cover Malta or the Maltese language, limiting the effectiveness of these efforts for Maltese users.
- Despite referencing ongoing research, Meta has not provided a transparent methodology for evaluating the effectiveness of its fact-checking integration. The platform reports general figures, such as average monthly active users across the EU (see Table 6), but these do not offer meaningful insight into the actual impact of fact-checking measures.
- Additionally, Meta has not disclosed any information regarding the development or existence of a dedicated fact-checks repository, further limiting transparency and evaluation of its approach.

Table 6: Meta's reported quantitative information for SLI 31.1.3

| SLI 31.1.3 | Facebook | Instagram |
|---|---|---|
| | Average of monthly active users on Facebook in the European Union between 1/01/2024 and 30/06/2024 | Average of monthly active users on Instagram in the European Union between 1/01/2023 and 30/06/2023 |
| **Total Global** | 260.7 million | 264.3 million |

| *Pillar-VII: Empowering the fact-checking community*<br>*Commitment 32: Fact-checkers access to relevant information*<br>*page 119-121* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 32** | 2 | 2 | 2 | 2 |

Meta's third-party fact-checking partners have access to a **dashboard** Meta built in 2016. According to Meta, this dashboard surfaces potential misinformation for fact-checkers and provides "data points" to help them prioritize which content to review.

Meta reported using feedback from the fact-checking community and similarity-detection signals to highlight likely misinformation for fact-checkers. They may also send content to fact-checkers if Meta "becomes aware" content may be false.

Meta indicates it will continue to work with EDMO (European Digital Media Observatory) and with an "elected body representative of independent European fact-checking organisations," as referenced in Commitment 32. They also claim to be "constantly working to further strengthen" relationships with fact-checkers.

In preparation for the **2024 European Parliamentary Elections**, Meta launched a refresher training series for fact-checking partners. This included training on:

- Reviewing generative AI content.
- Understanding the different labels for misinformation.
- Using the "Newswhip" dashboard to track trending topics.

**Major Comments:**

- Meta refers to ongoing improvements to its fact-checking dashboard and "constant work" to enhance collaboration with fact-checkers. Still, it does not provide specific timelines, implementation milestones, or a roadmap outlining when new tools or functionalities will be introduced.
- Although fact-checker feedback is acknowledged, Meta does not explain how this input is collected, processed, or integrated into its systems for iterative improvement.
- While all fact-checking organizations participating in the 3PFC reportedly have access to the dashboard, Meta provides no data on actual usage, such as how frequently it is accessed, the volume of content reviewed, or metrics indicating success in identifying or addressing misinformation.
- Meta also mentions plans to collaborate with EDMO but offers no concrete details on how or when it intends to align with any future joint frameworks for data access by fact-checkers, only a general indication that such alignment is intended.

## 3.4 Google (Search, YouTube)

Our analysis on Google's practices is based on the information provided in **Google's Code of Practice Report, delivered in September 2024, No.4** [Google, 2024], covering the period 1st of January 2024 to 30th June 2024.

### 3.4.1 Google: Empowering Users

| **Pillar-V: Empowering Users**<br>*Commitment 17:* ***Enhancing Media Literacy,***<br>*page 123-136* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 17** | 4 | 3 | 3 | 2 |

**Media Literacy Tools:**

Google Search has deployed the **"About This Result"** feature that appears as three vertical dots next to a search result on Google (whether on desktop or mobile), an information panel appears providing useful context about the result. This panel includes a brief description of the source—typically from Wikipedia or the site itself—to help users understand who is behind the content. It also explains how the result was generated, such as through keyword matches, links from other websites, or relevance to the user's region. Additionally, it indicates whether the website uses a secure HTTPS connection and offers feedback options allowing users to report results they find unhelpful or misleading. Additionally, there's a "More about this page" button that links to a dedicated page with deeper insights into the source's web presence, publication history, and content credibility, further supporting users in making informed decisions.

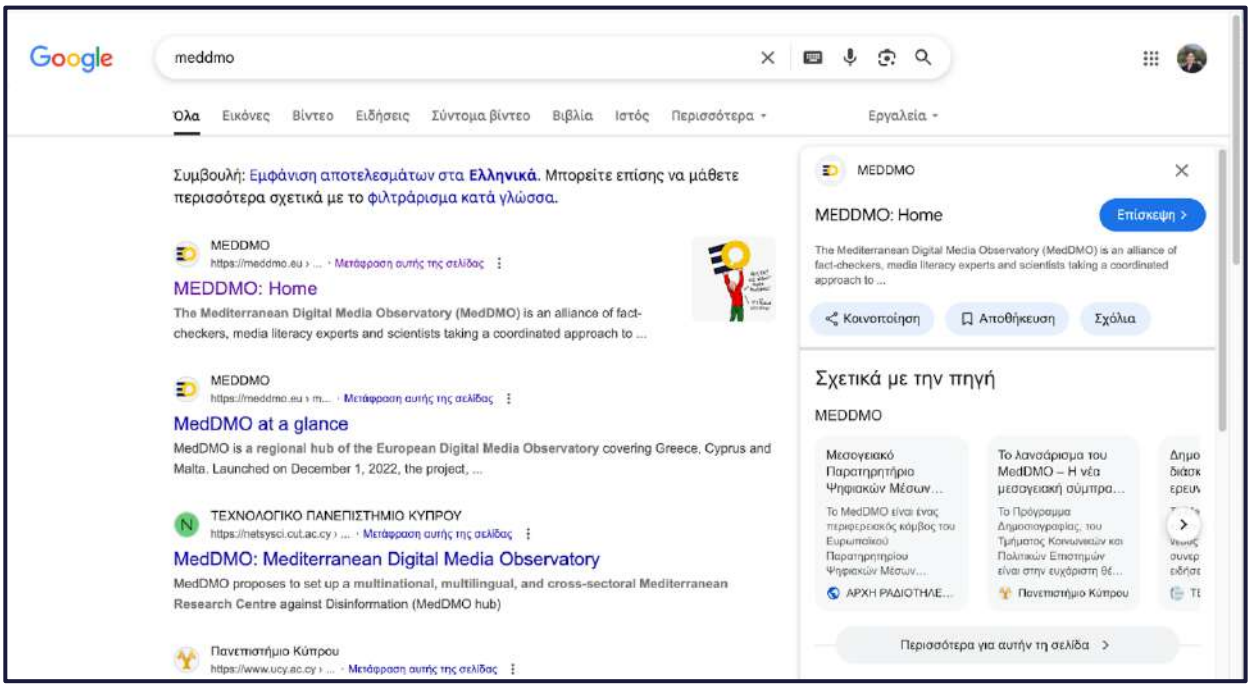



Figure 9: Google Search "About This Result" Feature (Greek)

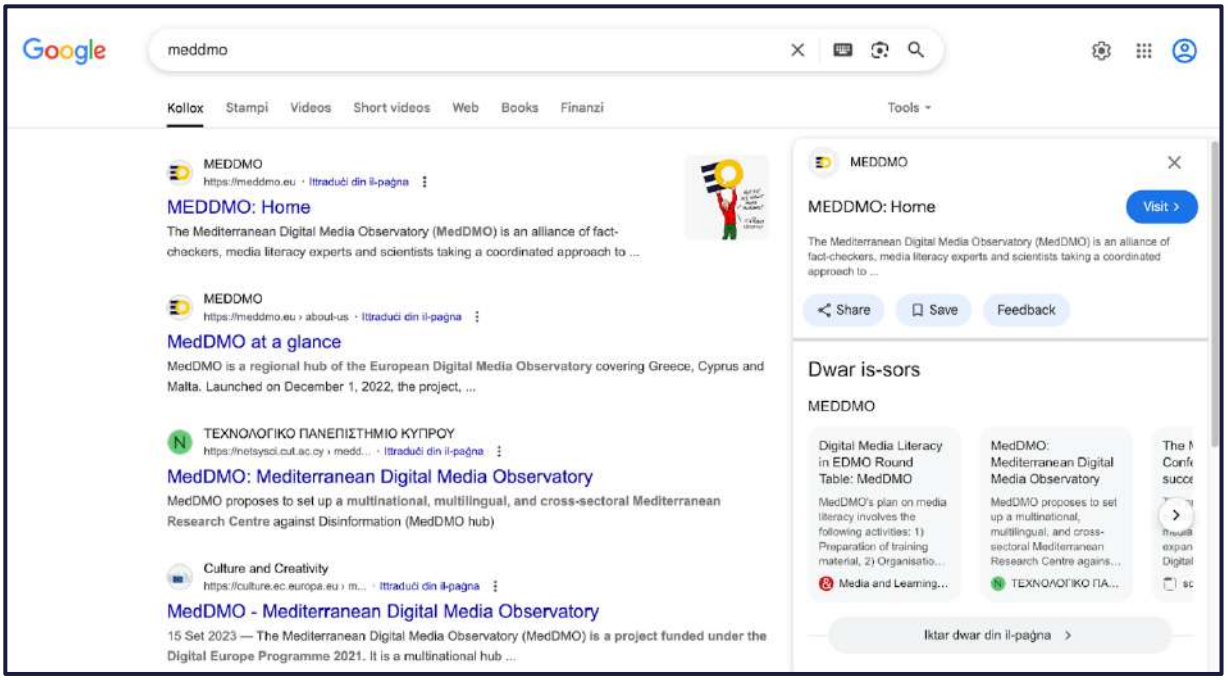



Figure 10: Google Search "About This Result" Feature (Maltese)

Google has expanded its **"About This Image"** tool to 40 additional languages, helping users evaluate online images more easily. "About this image" feature appears as a small clickable section next to or below the image in the info panel. It gives the image context like 1) When the image was first indexed, 2) Where it has appeared online, and 3) Whether it was used in fact-checking articles.

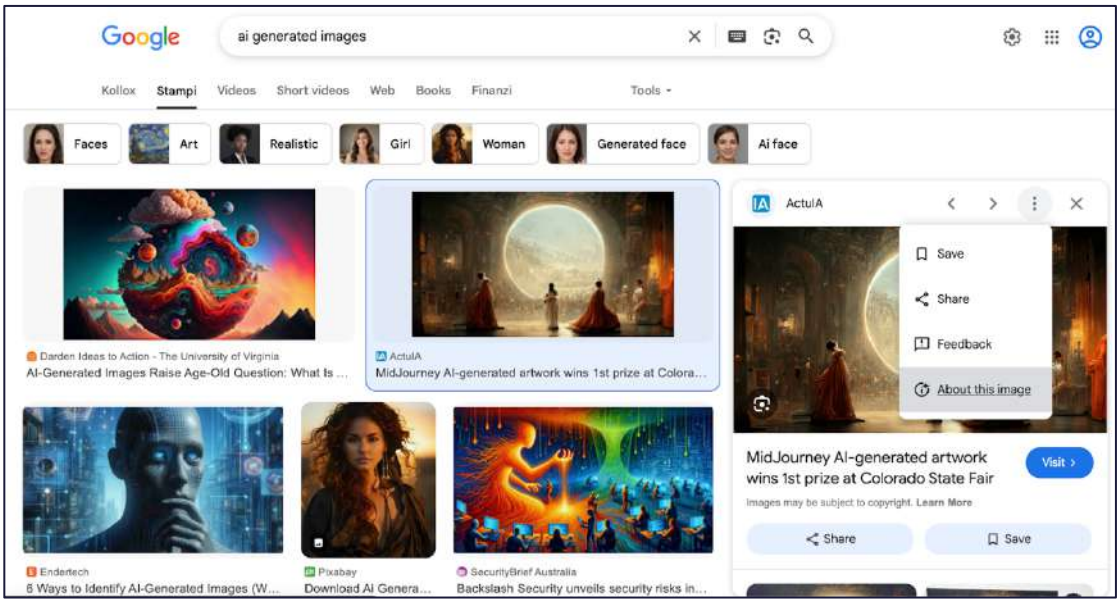


Figure 11: Google Search "About this image" feature

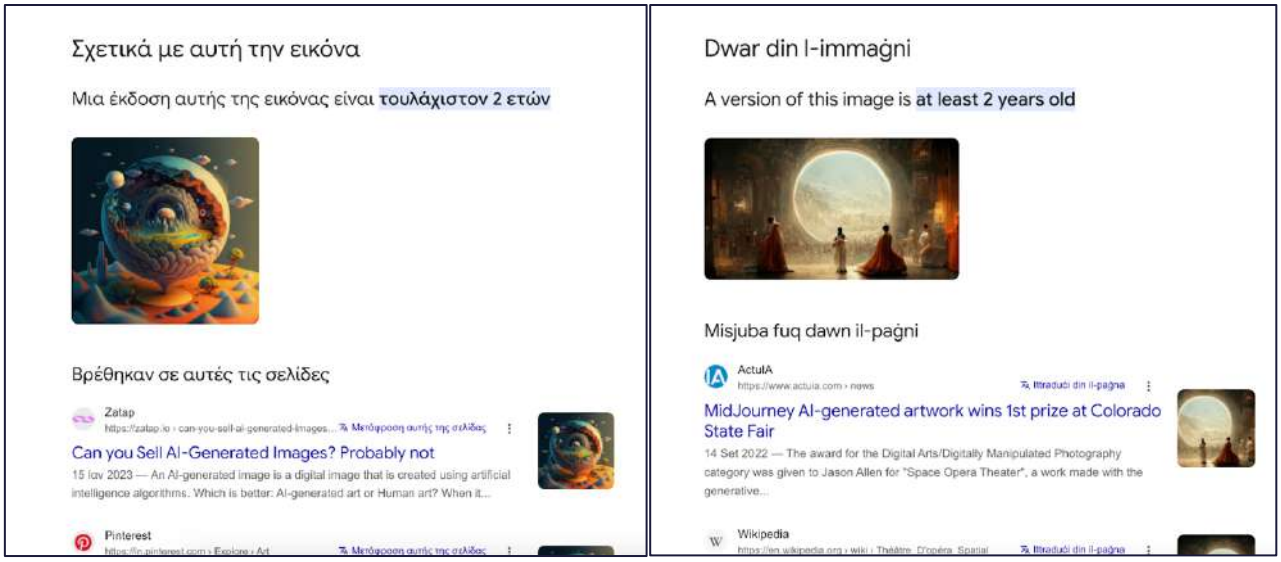


Figure 12: Google Search "About this image" feature (left: Greek and right: Maltese)

Google introduced **SynthID text watermarking** as part of its Responsible Generative AI Toolkit. If an AI-generated image created with Google tools contains the SynthID text watermark, the "About This Image" section will include a new field to indicate this, with a text "Made with Google AI". In that way, users will know if an image is authentic or AI-generated.

Google Search's **Content Advisory Notices**[14] are designed to alert users when reliable information is limited or when people are searching for a topic faster than accurate information is available.YouTube uses **Information Panels** to provide viewers with additional context on topics prone to misinformation, such as COVID-19, climate change, and elections. These panels appear alongside videos and in search results, offering concise information sourced from authoritative third-party organizations like the World Health Organization (WHO), Wikipedia, and Encyclopedia Britannica. The aim is to help users make informed decisions by presenting reliable information directly within the YouTube platform.

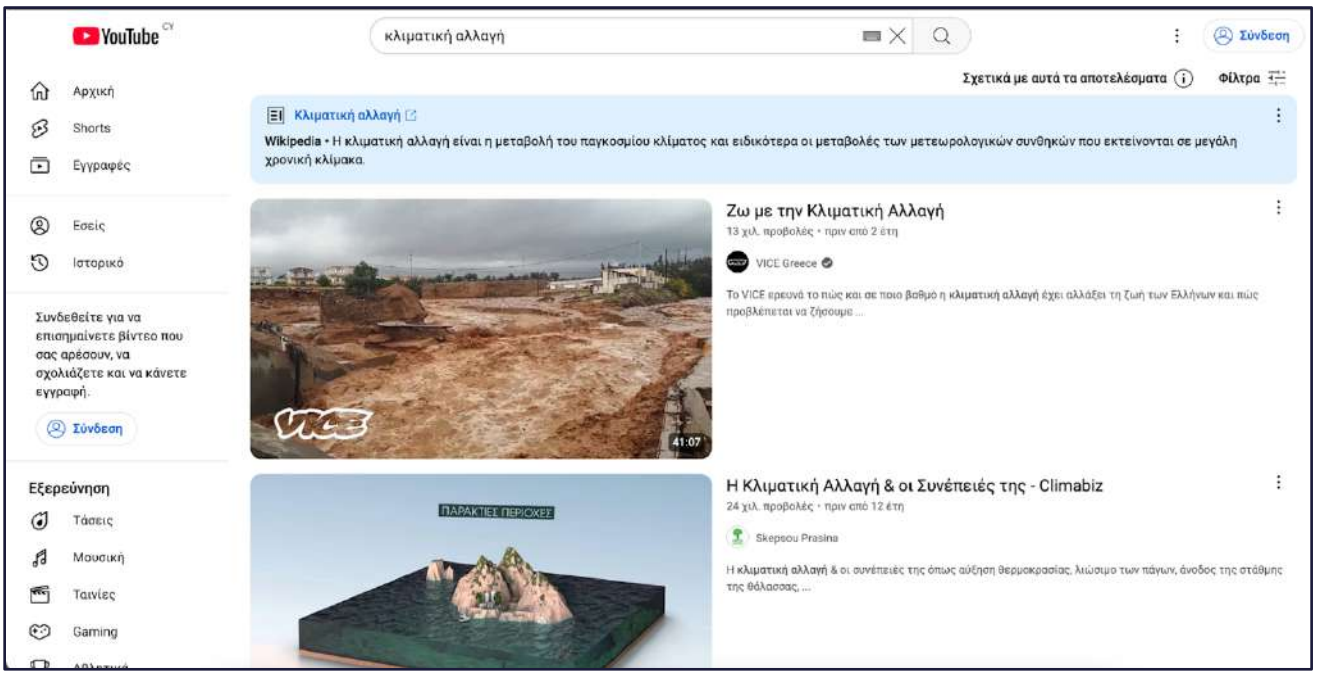


Figure 13: YouTube's Climate Change Information Panel example when language is set to Greek.

[14] https://blog.google/products/search/google-search-new-fact-check-features-europe/

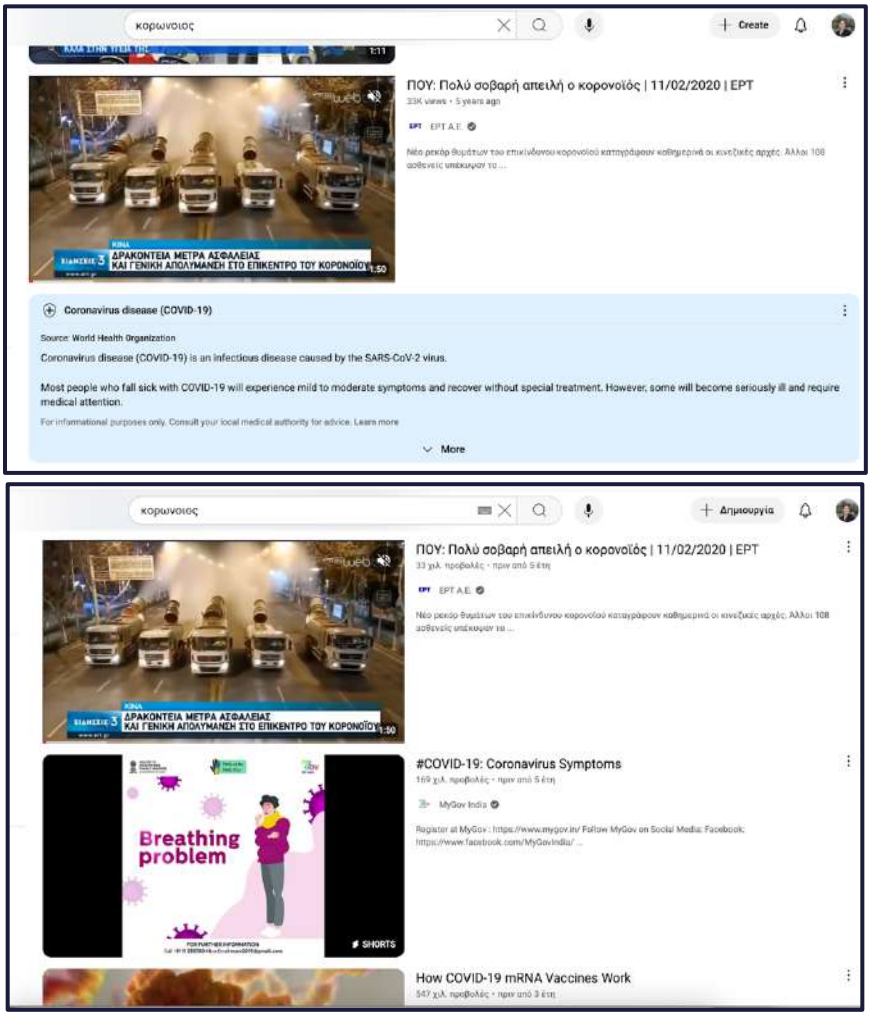



Figure 14: YouTube: Information Panels Inconsistency

**Campaigns**

Hit Pause Campaign[15] , a global media literacy initiative featuring a series of animated videos, was expanded to all EU Member States to teach critical thinking about misinformation. YouTube collaborated with the National Association for Media Literacy Education (NAMLE) to develop its "Hit Pause" campaign.

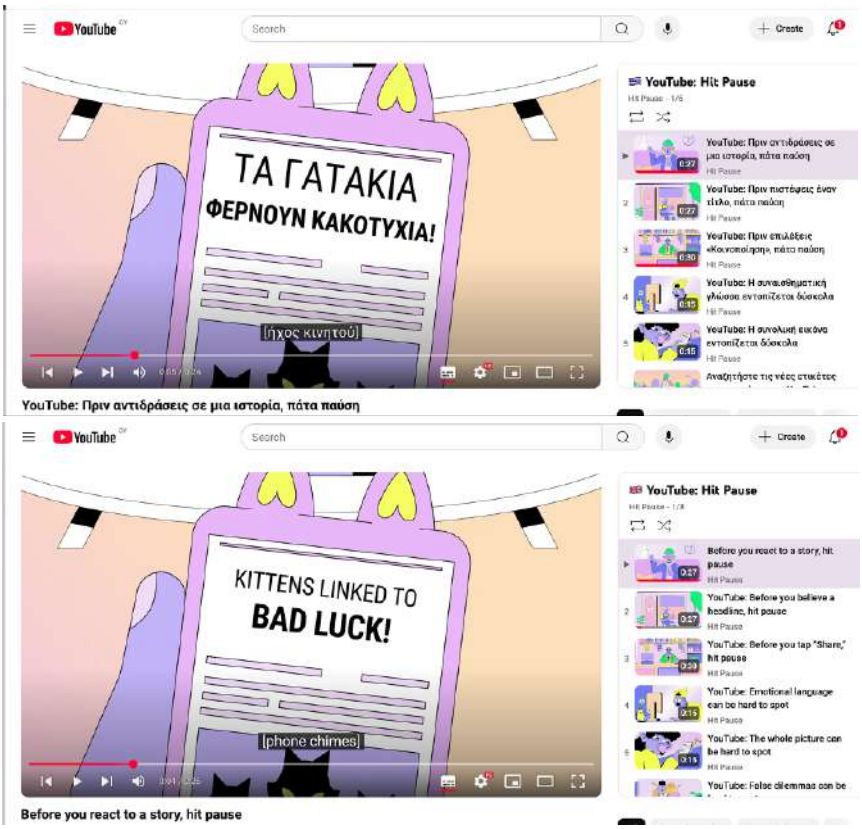



Figure 15: Hit Pause Campaign Screenshot in Greek and English

**Research and Grants**

Google through its Jigsaw unit, conducts research on information interventions and publishes its results regularly. In H1 2024, they published two research works relevant to misinformation:

- January 2024 – Research by Jigsaw and collaborators **[Maertens et al., 2025]** shows that inoculation interventions (text or video) can reduce misinformation susceptibility for up to a month, and their effectiveness can be extended with memory-based boosters.
- May 2024 – Ethnographic research with Jigsaw **[Hassoun et al., 2024]** reveals that GenAI enables everyday users to creatively remix and produce misinformation, calling for a shift in focus from passive consumers to active "bricoleurs."

[15] https://www.youtube.com/hitpause

Google granted funds to two organizations for media literacy activities in H1 2024:

- Barnardos ($1M) [16]
- ThinkYoung ($1M) [17]

Google partnered with Public Libraries 2030[18] to implement the "Super Searchers" training program. However, this program did not run in EU countries for first half of 2024.

**Major Comments:**

- Google has developed and deployed useful features to give more context to users about their search results.
- The **'About This Result'** feature is now available in the Greek (see Figure 9) and Maltese (see Figure 10) languages, following its expansion to 40 new languages in December 2023. This feature enables users to access information about the source and other contextual details of the search results information. Google reported that the "Source" section of this feature was viewed 291,106 times in Cyprus, 2,779,146 times in Greece, and 184,810 in Malta, the "Your Search and these results" was viewed 1,111,636 times in Cyprus, 11,819,482 times in Greece, and 662,658 in Malta, and the "Personalization" section was viewed 1,104,946 times in Cyprus, 11,769,060 times in Greece, and 657,346 in Malta.
- The **'About This Image'** feature is available in the Greek language and partially available in the Maltese language (see Figure 12). English is used to complement the text or feature that is not supported by Google in the Maltese language. This is a helpful tool, that provides context for users regarding images. Google reported that this feature was viewed 101,432 times in Cyprus, 639,346 times in Greece, and 57,930 times in Malta.
- Although we were unable to confirm whether the **Content Advisory Notices** feature is fully available in the Greek and Maltese languages, Google's reported data indicates that users in Cyprus, Greece, and Malta were exposed to these notices during H1 2024. Specifically, the impression proportions for low relevance results were relatively similar across the three countries (Cyprus: 0.183%, Greece: 0.168%, Malta: 0.173%), suggesting that users in all three regions encountered searches with limited high-quality information at comparable rates. In contrast, the impression proportions for rapidly changing results and potentially unreliable sets of results were significantly lower (Cyprus: 0.0006% and 0.0001088%, Greece: 0.00009% and 0.0000109%, Malta:

[16] https://www.barnardos.ie/google-org-extends-support-to-barnardos-to-help-make-the-internet-safer-for-children/
[17] https://blog.google/around-the-globe/google-europe/fighting-misinformation-online-elections/
[18] https://publiclibraries2030.eu/who-we-are/

0% and 0.0000238%), which may indicate either fewer trending or emerging topics in search or a more limited deployment of this specific advisory category in these regions.

- **YouTube Information Panels** are also available in the Greek language, while the Maltese language is not supported. Users in Malta receive the information panels in English by default. We detected an instance where the information panel is not displayed when the language is set to Greek (see Figure 14 on the bottom) but it is displayed when the language is set to English (UK) (see Figure 14 on the top). This instance indicates that the method to display information panels to users is unclear and inconsistent. Google reported the number of impressions of information panels (excluding fact-checks panels, crisis resource panels and non-COVID medical panels). The information panels were viewed 5,389,481 times in Cyprus, 42,428,213 times in Greece, and 2,653,667 times in Malta.

- YouTube's **"Hit Pause" campaign.** Greece had 6,663,804 received impressions from this campaign; Cyprus had 353,502 impressions and Malta had 182,871 impressions. These numbers indicate that the campaign has reached users in the three target countries. The campaign is also translated into the Greek language, while we did not find it translated into Maltese. Maltese users can find the campaign in English language.

- There is no mention of localized events, training sessions, or campaigns specifically targeting Greek, Cypriot, or Maltese audiences.

- There are no known Google partnerships with Media Literacy Experts in the three countries.

| **Pillar-V: Empowering Users**<br>*Commitment 21:* ***Better Equipping Users to Identify Disinformation,***<br>*page 155-165* | | | | |
|---|---|---|---|---|
| Evaluation with scale 1-5 | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 21** | 3 | 3 | 3 | 2 |

Google does not directly label or flag content in Search or YouTube using disinformation tags based on ratings from third-party fact-checking organisations, as Meta does, nor does it explicitly collaborate with fact-checkers to apply such labels.

However, Google does incorporate fact-checks into Search through features such as **Fact Check Rich Snippets**, which rely on the ClaimReview markup implemented by publishers. In the first half of 2024, Google Search displayed **298,382 fact-checked articles globally** using these rich snippets.




Figure 16: Fact Check Rich Snippet in Google Search Results (Greek).




Figure 17: Fact Check Rich Snippet in Google Search Results (Maltese).




Figure 18: Fact Check Rich Snippet in Google Search Results (English).

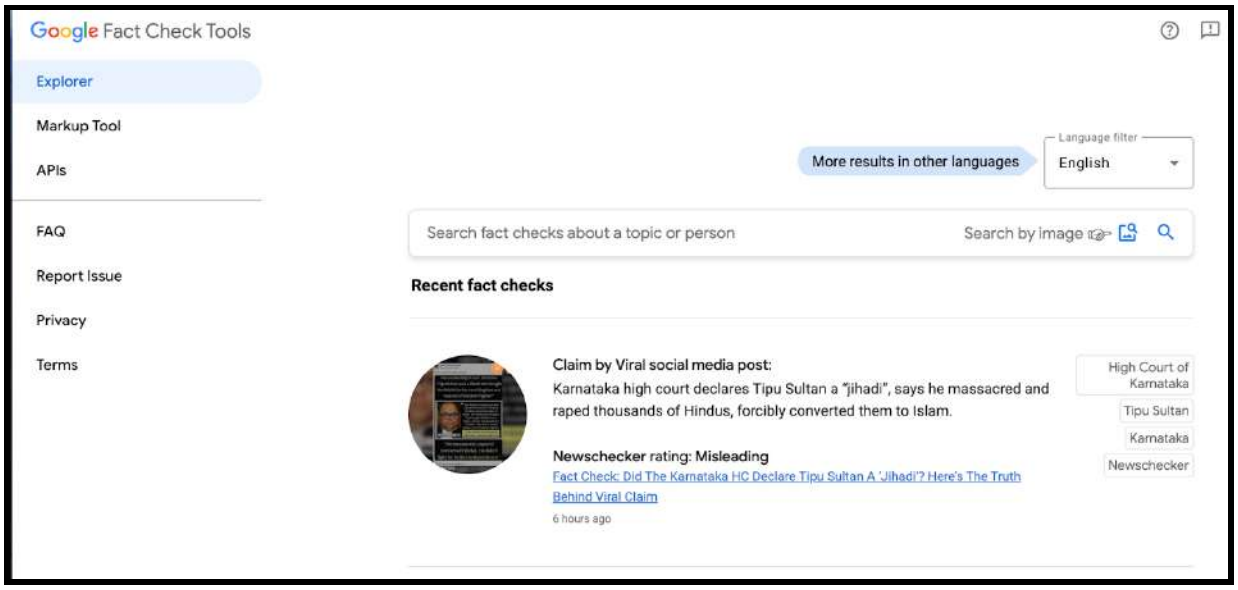



Figure 19: Google's Fact-Check Explorer

The Google Fact Check Explorer (see Figure 19) and the Claim Search API allow users and journalists to explore verified claims and images.

**Regarding Fact-Checking Tools, Google highlighted two tools (in beta testing at the time of reporting):**

1) **Image Search in Fact Check Explorer** (see Figure 20): Users can search for existing fact-checks by image. Deployed during the EU elections through the **Elections24Check** platform.

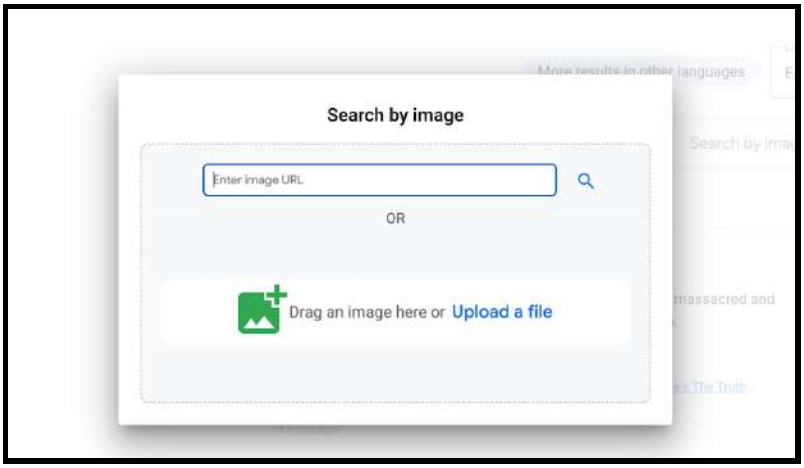



Figure 20: Search by Image Feature in Fact Check Explorer

2) **Image Timeline Tool**: Provides contextual timelines for images.

**YouTube Information Panels** were expanded to include fact-check articles in video searches relying on the ClaimReview tagging system. YouTube's algorithm elevates authoritative sources and content using ClaimReview tagging.

Fact-checking organizations are supported in developing **short- and long-form video content on YouTube**. **YouTube Studio** analytics and **Creator Support** help fact-checkers grow their reach.

**Research and Data Contributions**

Google Research, in collaboration with four fact-checking organizations—Maldita (Spain), Full Fact (UK), Duke Reporters Lab (USA), and Factly (India)—conducted a two-year study **[Dufour et al., 2024]** analysing online media-based misinformation using human raters and a large sample of 135,838 fact-checks tagged with ClaimReview metadata. The study, which spans fact-checks dating back to 1995, found that around 80% of recent misinformation claims involve media content. Historically, images were the most common format associated with misinformation, but by 2022, videos had become the dominant medium, accounting for over 60% of fact-checked media claims. Furthermore, AI-generated content, which was virtually absent before 2023, saw a significant rise in fact-checks from the spring of that year onward. As part of this research, Google released the **Annotated Misinformation, Media-Based (AMMeBa)**[19] to support further study and evaluation of misinformation mitigation methods, marking the first large-scale dataset of its kind to track the types and modalities of online misinformation.

**Major Comments:**

- The integration of Fact Check Rich Snippets into Search results is available in the Greek and English languages. When the browser language is set to Maltese, the feature is not available, not even with a fallback language (see
- Figure 17). However, the label design does not capture the user's attention. The results with the fact-check label are almost indistinguishable from the other results. This is something that Google should try to improve. The impression number there cannot be assumed to mean that the users saw the fact-check label or that it was just there in the search results; the user may not have even seen it.
- In the first half of 2024, Google reported the usage of fact-check labelled results in Google Search across Cyprus, Greece, and Malta, demonstrating the visibility and reach of its disinformation countermeasures in these countries. According to the data, users in Greece were shown fact-check rich snippets 2,488,609 times, in Cyprus, the number of impressions reached 267,880, while in Malta, the figure was 125,929.

- Regarding the Fact Check Explorer tool, which indexes articles reviewed by fact-checking organizations using structured metadata, the data shows a modest but consistent increase in Greek-language fact-checking activity during the first half of 2024. Specifically, the number of Greek-language fact-checked articles rose from 1,772 at the beginning of the period to 1,898 by the end, reflecting ongoing contributions from Greek-speaking fact-checkers to the global

[19] https://www.kaggle.com/datasets/googleai/in-the-wild-misinformation-media

verification ecosystem. In contrast, Maltese-language content was not listed, pointing to the limited visibility or integration of fact-checking in Malta within Google's tools.

- This absence is likely due to several factors. Fact-checking activity in Malta is still relatively recent, and the organizations involved are not currently certified by the International Fact-Checking Network (IFCN) or the European Fact-Checking Standards Network (EFCSN)—both of which are widely recognized by platforms like Google. Furthermore, fact-checks produced in Malta are primarily published in English, rather than Maltese, and may not use the ClaimReview markup required for inclusion in tools like Fact Check Explorer and for generating rich snippets in Search. As a result, while some fact-checking work is taking place in Malta, it may not yet be technically or institutionally connected to Google's structured fact-checking ecosystem, highlighting a gap in localized, language-specific disinformation resources for Maltese users. The same applies to the fact-checks published by the Fact-Check Cyprus.
- We could not verify if the information panels with links to fact-checking articles are available in Greek or Maltese. YouTube did not provide SLIs for the impressions of the information panels with third-party fact checking information (neither at the EU level nor at the Member State level).
- There is no research or evaluation of mechanisms in place targeting the audience in Cyprus, Greece and Malta to assess the effectiveness of these tools. Google reported a research study published, but it was not specifically an evaluation study of the developed features to equip users to identify disinformation.

### 3.4.2 Google: Empowering the Research Community

| **Pillar-VI: Empowering the Research Community,** *Commitment 26: Empowering the Research Community* *pages 184 - 195* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 26** | 4 | 4 | 4 | 3 |

**Google provides the following tools to researchers:**

1. **Google Trends**[20] – Offers access to anonymized and aggregated search request data from Google Search and YouTube. It allows researchers to track interest in topics globally or at a more localized level. Google Trends is publicly available, and it does not have an API integration.
2. **Google Fact Check Explorer**[21] – A tool that enables researchers to explore fact-checked articles. It is integrated with the **Fact Check Claim Search API**[22], which allows developers to query verified fact-checking data.
3. **Google Researcher Program**[23] – Grants eligible EU researchers access to publicly available data from Google Search and YouTube. Researchers receive limited metadata scraping permissions to analyse publicly available data.
4. **YouTube Researcher Program**[24] – Provides expanded access to global video metadata from YouTube via a Data API[25]. Eligible academic researchers affiliated with accredited institutions can access data including video titles, descriptions, view counts, likes, comments, and channel metadata.

The Google and YouTube Researcher Program have a three-step application process to review the applicants' eligibility to access the data. The application for Google is available here, and for YouTube here.

**Major Comments:**

- Google offers several tools that are useful for disinformation research. Google Trends helps detect trending search topics in specific locations. Fact Check Explorer can support the labelling of data and analysis of disinformation themes over time. Publicly available data from Google Search and YouTube can assist in studies on false claims detection, user interactions, and the broader spread of misleading content.
- However, Google Trends does not offer an API, which limits its use for large-scale research. It does allow users to download search results as CSV files. Notably, the "Trending Now" tab excludes Malta from the location filter, with only Greece and Cyprus available. The "Explore" tab is accessible for all three countries. Since Google Trends only provides a sample of actual queries, there is no information on whether this sample is representative of search behaviour in Greece, Cyprus, or Malta. Google reported 13,353 users in Cyprus, 36,501 users in Greece, and 6,011 users in Malta used Google Trends researching Google Search. Additionally, Google reported 667 users in Cyprus, 1,184 users in Greece, and 6 users in Malta used Google Trends researching YouTube. While these figures suggest some level of engagement, it is unclear whether the usage was for research purposes.

[20] https://trends.google.com/trends/
[21] https://toolbox.google.com/factcheck/about#fce
[22] https://toolbox.google.com/factcheck/apis
[23] https://transparency.google/tools-programs/researchers/
[24] https://research.youtube/
[25] https://developers.google.com/youtube/v3/docs/search/list#videoPaidProductPlacement

- Google Fact Check Explorer is available to researchers in all three countries. However, fact-checks from Malta and Cyprus are not indexed in the repository. Google reported 93 users in Cyprus, 446 users in Greece, and 58 users in Malta accessed the tool during the reporting period. These numbers indicate relatively low engagement. Fact-Check Explorer can also be accessed via the Google FactCheck Claim Search API, but Google did not provide user metrics by Member State for this API.
- During the reporting period, no researchers from Cyprus, Greece, or Malta applied for or were granted access to the YouTube Research Program. Similarly, Google did not report the number of applications or granted accesses to the broader Google Research Program.

| **Pillar-VI: Empowering the Research Community,** ***Commitment 27: Governance Structure for Access to Data for Research Purposes Requiring Additional Scrutiny,*** *pages 195 - 198* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 27** | 1 | 1 | 1 | 1 |

Regarding the development of an independent third-party intermediary body for data access, as envisioned in Commitment 27, Google confirms its participation in the European Digital Media Observatory (EDMO) Working Group, which is focused on shaping the body's governance, functions, staffing, and budget. Despite this involvement, the intermediary body has not yet been established.

For Search, Google reports that it is exploring options for pilot programs aimed at sharing data with vetted researchers to support investigations into misinformation and disinformation. However, no specific details have been disclosed regarding participating teams, the scope of research topics, or concrete progress on these pilot efforts.

In the case of YouTube, Google maintains the existing YouTube Researcher Program that offers expanded access to global video metadata through a Data API. Google did not report any new pilot initiatives or expansions during the current reporting period.

**Major Comments:**

- The qualitative information submitted by Google is described as limited and largely repetitive of earlier disclosures. While it confirms ongoing engagement with EU institutions and researchers, it does not present substantive new insights or progress updates, and thus only partially fulfils the transparency objectives of Commitment 27.
- Google has not provided verifiable milestones or tangible evidence of progress beyond affirming its continued support and collaboration with the EDMO process.
- In summary, while Google remains formally engaged with relevant initiatives and institutions, its reporting under Commitment 27 reflects ongoing intentions rather than demonstrable new actions or outcomes.

| **Pillar-VI: Empowering the Research Community,** *Commitment 28: Cooperation with Researchers,* *pages 198 - 202* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 28** | 2 | 2 | 2 | 2 |

**Google Support for Research Programmes:**

**Google Researcher Programme**

Google reports that there are dedicated teams for the Google Researcher Programme. These teams manage access applications for researchers and evaluate updates and developments for the Programme. Information to support researchers is available in the Help Center Support of the Search Researcher Result API[26] guidelines and the Google Transparency Center[27].

**YouTube Researcher Programme**

[26] https://support.google.com/websearch/answer/13856826?visit_id=638332129583125760-4204336310&p=searchresearcher&rd=1
[27] https://transparency.google/

YouTube's program staff manage access applications, provide technical support to grantees, and solicit researcher feedback to shape future platform enhancements. Researchers accepted into the program can reach out to a dedicated e-mail alias and access YouTube API code samples on GitHub[28]. Additionally, researchers can access the API documentation[29] to review the data they can access. Moreover, YouTube's Product and Policy teams maintain ongoing dialogues with independent researchers who seek to understand platform mechanics or receive input on research design.

**Workshops for researchers:**

Google hosted two workshops for researchers throughout H1 2024:

- January 1st to June 30th, 2024: Google hosted 25 researchers, experts in Trust & Safety-related research areas. Bringing together scholars working on misinformation, child safety, violent extremism, privacy, and at-risk user issues. These gatherings aimed to foster interdisciplinary collaboration and share insights across siloed research fields.
- A workshop with 30 attendees, including academics at the Trust & Safety Forum in France, on the topic of Trust & Safety by Design frameworks and implementation designs, including misinformation.

**Partnerships:**

**Google has a partnership with Lumen[30],** an independent research project managed by the Berkman Klein Centre for Internet & Society at Harvard Law School. The Lumen database holds online content takedowns after a report or complaint. Reports or complaints on Google products are shared with Lumen for publication. The Lumen repository can support researchers from academia or industry in the topic of online content availability.

**Research Grants:**

Google also maintained its **Trust & Safety Research Awards[31]**, offering grants to researchers investigating priority topics such as scams, generative AI, and misinformation. In June 2024, Google launched its **Academic Research Awards[32] (GARA) program**, which solicits proposals across a broad range of computing and technology subjects, including those relevant to Trust & Safety.

In 2021, Google contributed €25 million to the **European Media and Information Fund** (EMIF), which is administered by the European University Institute and Calouste Gulbenkian Foundation and supports media literacy, fact-checking, and disinformation research across Europe. Although EDMO advises EMIF's grant-making process, Google explicitly states it does not participate in selecting or evaluating individual EMIF applications.

[28] https://github.com/youtube/api-samples
[29] https://developers.google.com/youtube/v3/getting-started/?target=_blank
[30] https://lumendatabase.org/
[31] https://research.google/programs-and-events/trust-safety-research/
[32] https://research.google/programs-and-events/google-academic-research-awards/

**Major Comments:**

- Google reported its programs, resources, and partnerships to support researchers. They provide some participation figures (e.g., workshop attendees, application volumes).
- However, more information is missing on the dedicated teams supporting researchers, i.e., the number of Google employees responsible for supporting the research programs.
- At the same time, there is room for improvement in quantifying program outputs, publishing grant budgets and selection criteria. Addressing these gaps would enhance both the completeness and verifiability of Google's commitment to empowering the European research community in the fight against disinformation.
- Based on the reported information, it is not clear if researchers from the three countries were benefiting from Google's researchers support measures.

### 3.4.3 Google: Empowering the fact-checking community

| *Pillar-VII: Empowering the fact-checking community*<br>***Commitment 30: Cooperation with the fact-checking community,***<br>*page 207-217* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 30** | 3 | 3 | 3 | 3 |

**Google supports fact-checking professionals through the following fundings/partnerships:**

- **International Fact-Checking Network (IFCN) Global Fact Check Fund ($13.2 M USD total):** Grants to support 135 + fact-checking organizations worldwide. In early 2024, Phase 3 (Engage) awarded $100 K each to six EEA teams (Croatia, France, Italy, Poland, Spain), and Phase 4 (Build) awarded $25 K each to eight EEA teams (including Cyprus and Greece).
- **European Fact-Checking Standards Network (EFCSN) / Elections24Check (€1.5 M EUR):** Funds a coalition of 40 + news and fact-checking organizations across 36 countries to verify European Parliamentary election claims, publish an open disinformation database, and provide beta Fact Check Explorer features (e.g., image search, provenance tools).
- **European Media and Information Fund (EMIF, €25 M EUR 2021–26):** Awards grants to at least 87 media literacy, misinformation, and fact-checking projects in 25 countries. EDMO serves as a scientific advisor; Google does not select grantees.

**Google resources/tools to support fact-checking organizations:**

- **Google News Initiative Training:** Delivers digital verification workshops to journalists; since 2015, 145 000+ European journalists trained and 400,000+ visits to free online curricula.
- **YouTube Partner Program (YPP)[33]:** The YPP lets eligible creators—such as certified fact-checking organizations—earn ad revenue from their videos. To join, a channel must have 1,000 subscribers, 4,000 watch hours in the past year, and must follow YouTube's guidelines. Once accepted, creators gain access to ad monetization, advanced analytics, and direct support from YouTube's Creator Support team. They also benefit from features like longer upload limits and improved visibility in search and recommendations, helping trusted fact-checkers reach a wider audience. YPP enables 11 EU-based,

[33] https://support.google.com/youtube/answer/72851?hl

IFCN-certified fact-checkers (Belgium, France, Germany, Italy, Portugal, Spain) to monetize fact-check videos, receive Creator Support, and access guidance on best practices.

- **ClaimReview Markup (schema.org):** A structured data format that fact-checkers embed in their articles. When implemented, Google can automatically index, and display verified fact-checks in Search and YouTube.

**Major Comments:**

- Google does not directly fund fact-checking organizations to label content on their platforms. However, Google grants funds to organizations such as IFCN, EFCN, and EMIL that are responsible for deciding on how to distribute funds to fact-checking organizations. Fact-checking organizations receive fundings from IFCN and EFCN. Google highlights that the decision to grant funds to organizations by reviewing applications is not their responsibility.
- Under IFCN's Global Fact Check Fund (supported by Google's $13.2 M commitment), Fact Check Cyprus (Cyprus)[34] and Greece Fact Check (Greece)[35] selected in Phase 4 (Build) for a $25K funding. Greece Fact Check received also a $5K
- No fact-checking organization in Malta received IFCN funding in either Phase 3 or Phase 4.
- In IFCN Phase 3 (Engage), no grantees came from Cyprus, Greece, or Malta.
- EMIL funded 5 research projects[36] in Greece in 2024. No research projects in Cyprus and Malta received funding from EMIL in 2024.
- We are not aware of fact-checking organizations in Cyprus, Greece and Malta that use the YouTube Partner Program (YPP). However, for recently established fact-checking organizations the metrics that make a channel eligible for YPP may be too hard to reach.
- The report notes regular meetings with IFCN and EFCSN, as well as collaboration with EU-based fact-checking organisations. However, it does not specify details regarding the exact meeting schedules or the impact of this collaboration on Google's mechanism to combat disinformation or support fact-checking organisations.

[34] https://factcheckcyprus.org/%cf%87%cf%81%ce%b7%ce%bc%ce%b1%cf%84%ce%bf%ce%b4%cf%8c%cf%84%ce%b7%cf%83%ce%b7-%ce%ba%ce%b1%ce%b9-%ce%b1%ce%bd%ce%b5%ce%be%ce%b1%cf%81%cf%84%ce%b7%cf%83%ce%af%ce%b1/

[35] https://www.factchecker.gr/xrimatodotisi-anexartisia/

[36] https://gulbenkian.pt/emifund/projects/?filter%5Btax%5D%5Bfcg_proj_priority%5D%5B%5D=&filter%5Btax%5D%5Bfcg_project_funding%5D%5B%5D=&filter%5Btax%5D%5Bfcg_project_year%5D%5B%5D=67&filter%5Btax%5D%5Bfcg_proj_country%5D%5B%5D=85&filter%5Btax%5D%5Bfcg_project_size%5D%5B%5D=&filter%5Btax%5D%5Bfcg_project_status%5D%5B%5D=

| *Pillar-VII: Empowering the fact-checking community*<br>*Commitment 31: use and integration of fact-checking*<br>*page 218-221* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 31** | 3 | 3 | 3 | 3 |

**Integration of fact-checking work on Google platforms:**

Google does not take any action to label content on their platform as false information. Google released the **ClaimReview HTML markup,** which enables fact-checkers to tag their content for better discoverability. The markup is used in the Search Results and Information Panels to help users with giving context.

YouTube encourages fact-checking organizations to produce fact-checks in a video format, allowing them to promote more fact-checking content on their platforms. They offer the **YouTube Create**[37] mobile app, which helps users edit videos on their mobile devices, simplifying the process of producing high-quality videos. This tool is available only in specific countries in EU[38].

**Major Comments:**

- Google has made efforts to support fact-checkers and incorporate their work into its services. However, it falls short in providing detailed quantitative data or conducting impact assessments. The platform's reporting lacks disaggregated metrics by country and platform, limiting the evaluation of its effectiveness in specific contexts.
- YouTube's initiative to encourage fact-checkers to create video-based content is timely and appropriate, especially given the popularity of short-form videos across age groups. However, the YouTube Create app—intended to support such content creation—is not currently available in Cyprus, Greece, or Malta.
- Google does not offer dedicated tools for fact-checkers beyond the ClaimReview Markup tool. Fact-checkers lack streamlined access to content metadata or additional contextual information that would aid in verifying claims and improving the efficiency of the fact-checking process.

[37] https://support.google.com/youtube/thread/235901340/
[38] https://support.google.com/youtube/answer/13952912

| *Pillar-VII: Empowering the fact-checking community*<br>*Commitment 32: fact-checkers access to relevant information*<br>*page 221-226* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 32** | 2 | 2 | 2 | 2 |

**Google Tools for fact checkers to quantify the impact of their work:**

**Search Console**[39]: a free toolkit that helps website owners, including fact-checking organizations, to track and improve their visibility in Google Search. Its key feature—the Search Performance report—reveals how often your pages appear (impressions), which queries trigger them, click-through rates, and country-specific views across the EU. By using these insights, fact-checkers can see which claims attract the most attention, identify search terms to target, and spot any dips in performance to troubleshoot issues, ensuring their corrections reach users effectively.

**YouTube Studio** offers to YouTube creators, including fact-checking organizations, a centralized dashboard to manage and grow their channels while monitoring video performance. Using the Channel Analytics panel, teams can track key metrics—total views, watch time, traffic sources, and the specific search terms that drive viewers to their fact-check videos—enabling them to understand which topics resonate and where to focus outreach. Additionally, YouTube's Help Center documents "information panels," text-based overlays that can provide viewers with context or source links during playback. Together, these tools enable fact-checkers to fine-tune their content strategy, engage their audience more effectively, and ensure accurate information reaches the widest possible audience.

**Google Communication channels with the fact-checking community:**

- Google reports that they are in regular contact with the International Fact-Checking Network (IFCN) and the European Fact-Checking Standards Network (EFCSN) to discuss collaborations and directions to support the fact-checking community.
- Fact-checking organizations that participated in the Elections24Check[40] have access to new beta features in Fact-Checker Explorer.
- Representatives from Google and YouTube participated and hosted events related to misinformation.
- YouTube, through its Creator Support teams, is available to support the fact-checking organizations that participate in the YouTube Partner Program (YPP).

[39] https://support.google.com/webmasters/answer/9128668?hl=en&ref_topic=9128571
[40] https://elections24.efcsn.com/

**Major Comments:**

- Google offers generic tools such as Search Console for websites and YouTube Studio for video channels, which fact-checking organizations can use to monitor metrics like impressions, clicks, watch time, traffic sources, and search queries. While these tools provide general visibility into content performance, they are not specifically designed to support fact-checking workflows.

- Despite the availability of these dashboards, Google does not report any usage statistics or engagement metrics specific to fact-checkers using Search Console. Since access to the tool is available to all webmasters without requiring identification as a fact-checking entity, it is impossible for organizations to evaluate how their fact-check pages perform relative to other types of content.

- A similar limitation exists for YouTube Studio. Although fact-checkers can monitor analytics such as views, watch time, and referral sources, Google's reporting does not include data on the extent of usage by fact-checking channels or the effectiveness of these features in promoting viewer engagement with fact-check content.

- While Google supports the use of the ClaimReview markup to make fact-checks visible in Search and YouTube, it provides no follow-up reporting or performance feedback. This absence of a feedback loop means fact-checking organizations have no visibility into whether applying ClaimReview markup improves content visibility or user engagement.

- Google claims that fact-checking content appears in search results and recommendation systems but does not provide granular data on how often this occurs, which queries trigger fact-check panels, or what user actions follow these exposures. Without such metrics, it is difficult to assess the actual impact of fact-check integration.

- Finally, although the report references ongoing collaboration with the fact-checking community—including through partnerships with IFCN and EFCSN—it lacks information on the frequency, scale, or effectiveness of these engagements. There are no details about the number of trainings, workshops, or meetings held, nor their outcomes, making it challenging to evaluate the true reach and value of Google's fact-checking support efforts.

## 3.5 TikTok

Our analysis on TikTok's practices is based on the information provided in **TikTok's Code of Practice Report, delivered in September 2024, No.4** [TikTok, 2024] covering the period 1st of January 2024 to 30th June 2024.

### 3.5.1 TikTok: Empowering the users

| **Pillar-V: Empowering Users**<br>***Commitment 17: Enhancing Media Literacy,***<br>*page 101-150* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 17** | 3 | 3 | 3 | 2 |

TikTok employs a multi-faceted approach to enhance media literacy and counter misinformation. Their efforts can be categorised into three main areas: tools, activities, and partnerships.

**Tools:**

- TikTok implemented **in-app intervention tools** that flag content related to specific topics (e.g., Holocaust, Israel-Hamas conflict, climate change, and EU Parliament elections) and redirect users to reliable sources.
- **Video notice tags** (see
- Figure 25) provide educational prompts for specific hashtags, directing users to in-app guides or third-party information.
- TikTok applies a **state-controlled media label**[41] to content created by accounts that is subject to control or influence by government. The user can see more information on why they see the label in an **in-app page**[42]. They strengthened policies regarding state-affiliated media, prohibiting such accounts from targeting audiences outside their home countries. In the EU, Iceland, and Liechtenstein, TikTok reported that it restricted access from accounts i.e., RT- Russia Today UK-Germany-France-Spanish, Sputnik,Rossiya RTR / RTR Planeta, etc. See the full list in **[TikTok, 2024],** pages 108-109**.**
- A new **AI-generated content label** encourages creators to disclose content that is either fully or significantly AI-generated.

[41] https://www.tiktok.com/transparency/en-us/state-affiliated-media/
[42] https://www.tiktok.com/tns-inapp/pages/state-affiliated-media

- TikTok applies an "**Unverified content" label** to content that fact-checking partners have assessed but cannot verify as accurate to encourage users not to share this content. If users continue to share this content, it becomes ineligible for the TikTok recommendation algorithm. Additionally, the creator of the content is notified and provided with information on why the content receive the unverified label.
- TikTok offers a range of **dedicated information resources**—both online and within its app—to help users access accurate, up-to-date content from trusted authorities. Depending on the topic and country, users may be directed to external sites (for example, a national electoral commission), an in-app information center (such as for the Ukraine war), or specialized pages in TikTok's Safety Center or Transparency Center.
- The **Safety Center**[43] explains TikTok's **policies on harmful misinformation**[44] (including COVID-19[45] and election integrity[46]), online challenges[47], and guidelines for sharing content related to tragic events[48], while its **"safety partners" page**[49] highlights collaborations with experts and NGOs in the context of building a safer platform.
- **TikTok Transparency Center**[50] available in a number of EU languages, hosts transparency reports and information on TikTok commitments to ensure the integrity of the platform.
- **TikTok Newsroom**[51] is available for users to be up-to-date with the latest TikTok updates and efforts. Users can select their country to get the most relevant information. The country selection list includes also EU as an option.

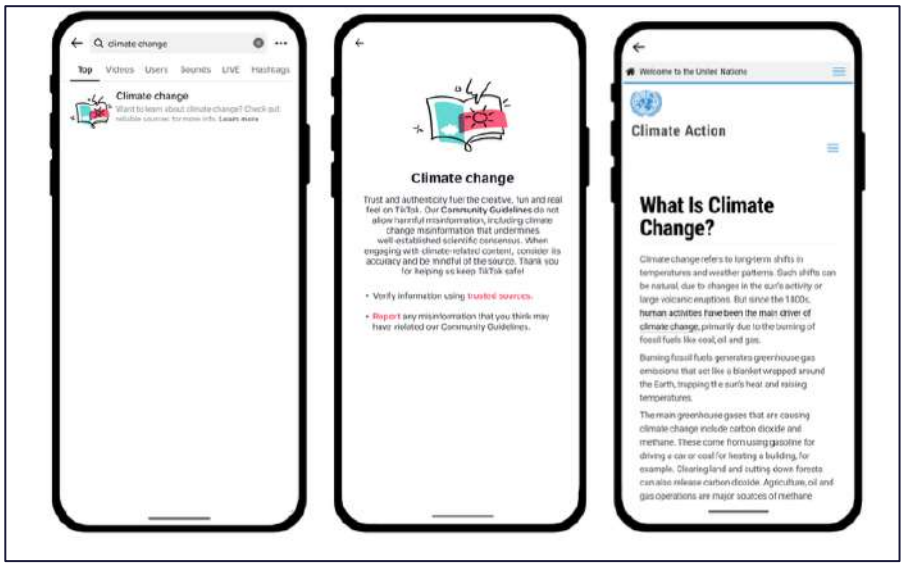



Figure 21: TikTok Climate Change Search Intervention [TikTok, 2024]

[43] https://www.tiktok.com/safety/el-gr/
[44] https://www.tiktok.com/safety/en-gb/harmful-misinformation-guide
[45] https://www.tiktok.com/safety/en/covid-19/
[46] https://www.tiktok.com/safety/en/election-integrity/
[47] https://www.tiktok.com/safety/en-us/online-challenges
[48] https://www.tiktok.com/safety/en/tragic-events-support/
[49] https://www.tiktok.com/safety/en-us/safety-partners/
[50] https://www.tiktok.com/transparency/en-us/
[51] https://newsroom.tiktok.com/en-gb

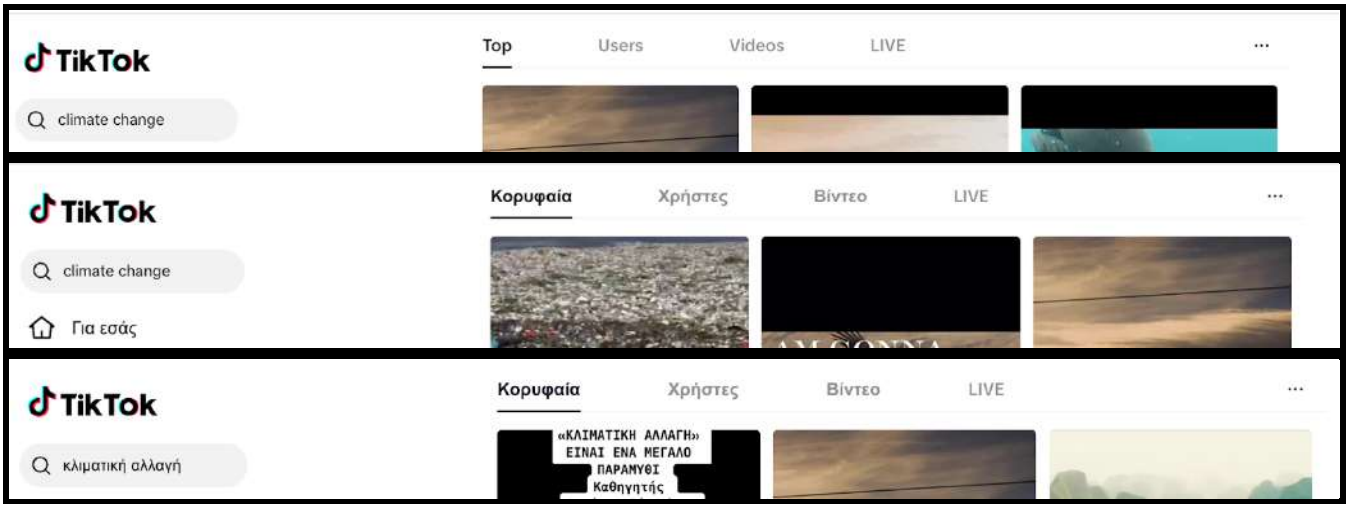



Figure 22: Climate Change Intervention unavailable.

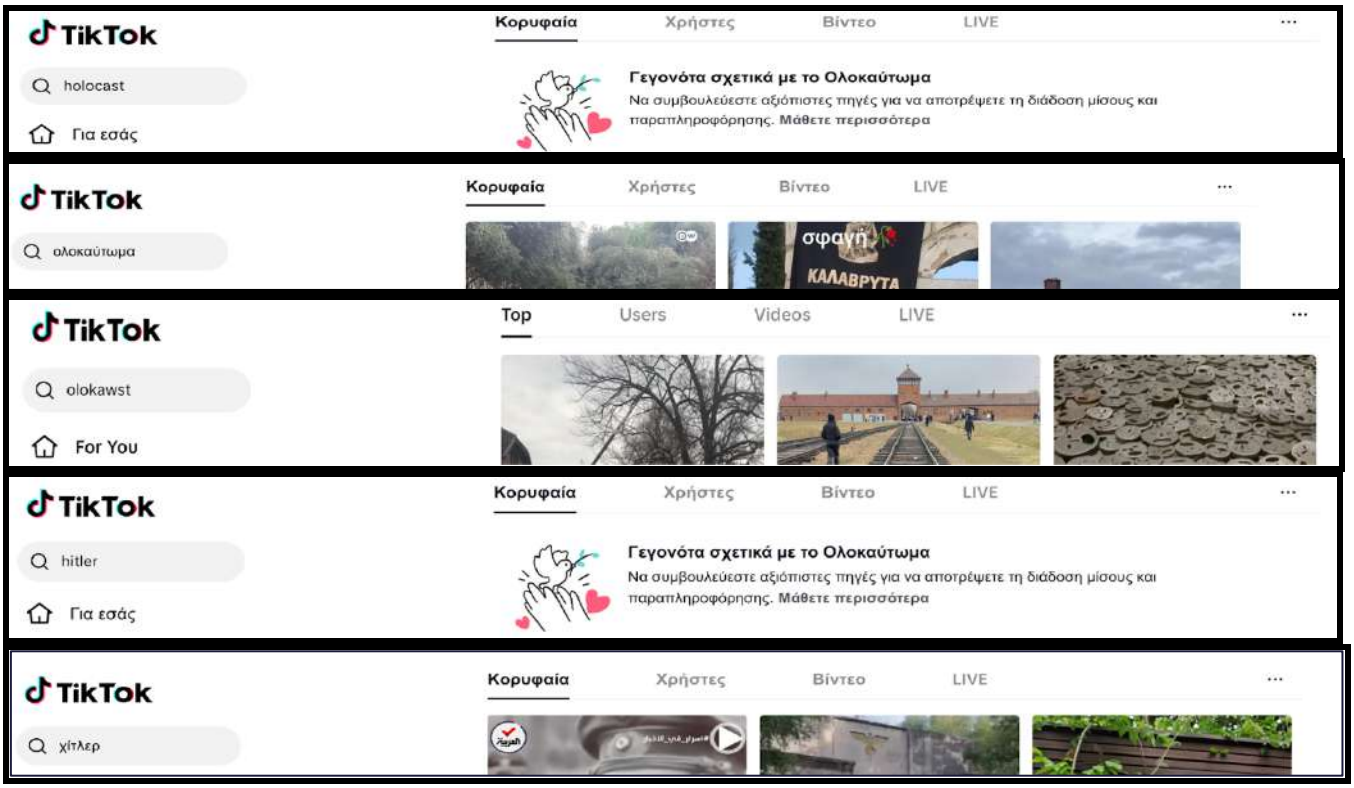



Figure 23: Holocaust Search Intervention Inconsistency.

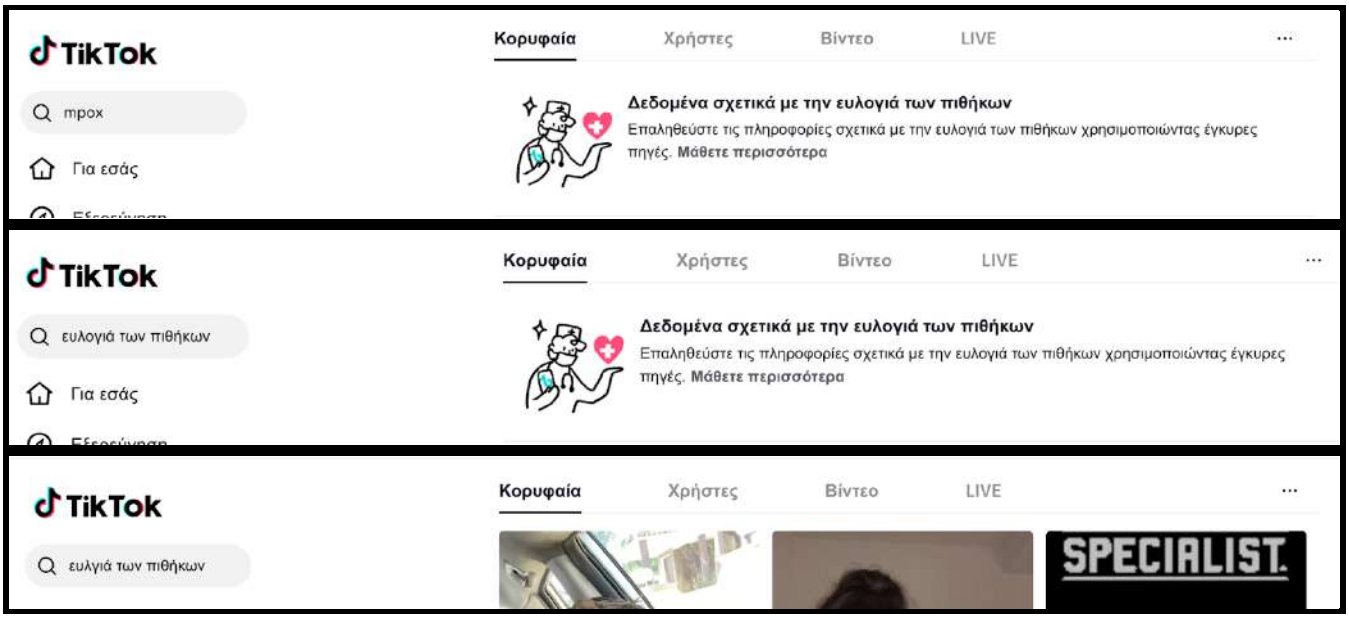



Figure 24: Monkeypox Search Intervention Availabilty and Inconsistency

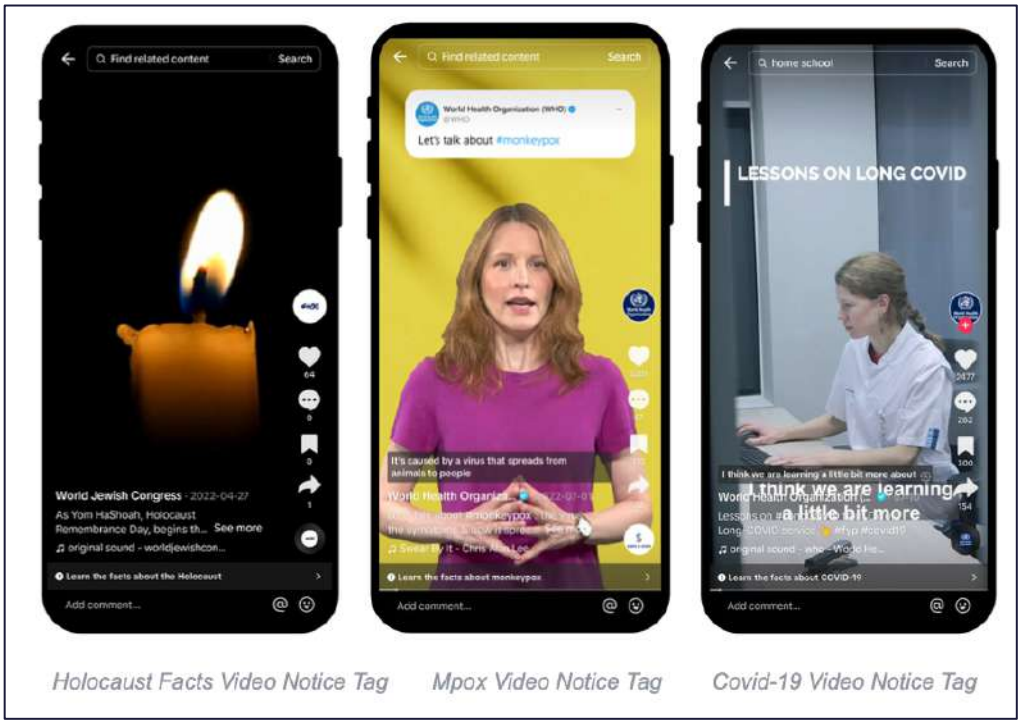



Figure 25: TikTok Video Notice Tags [TikTok, 2024]

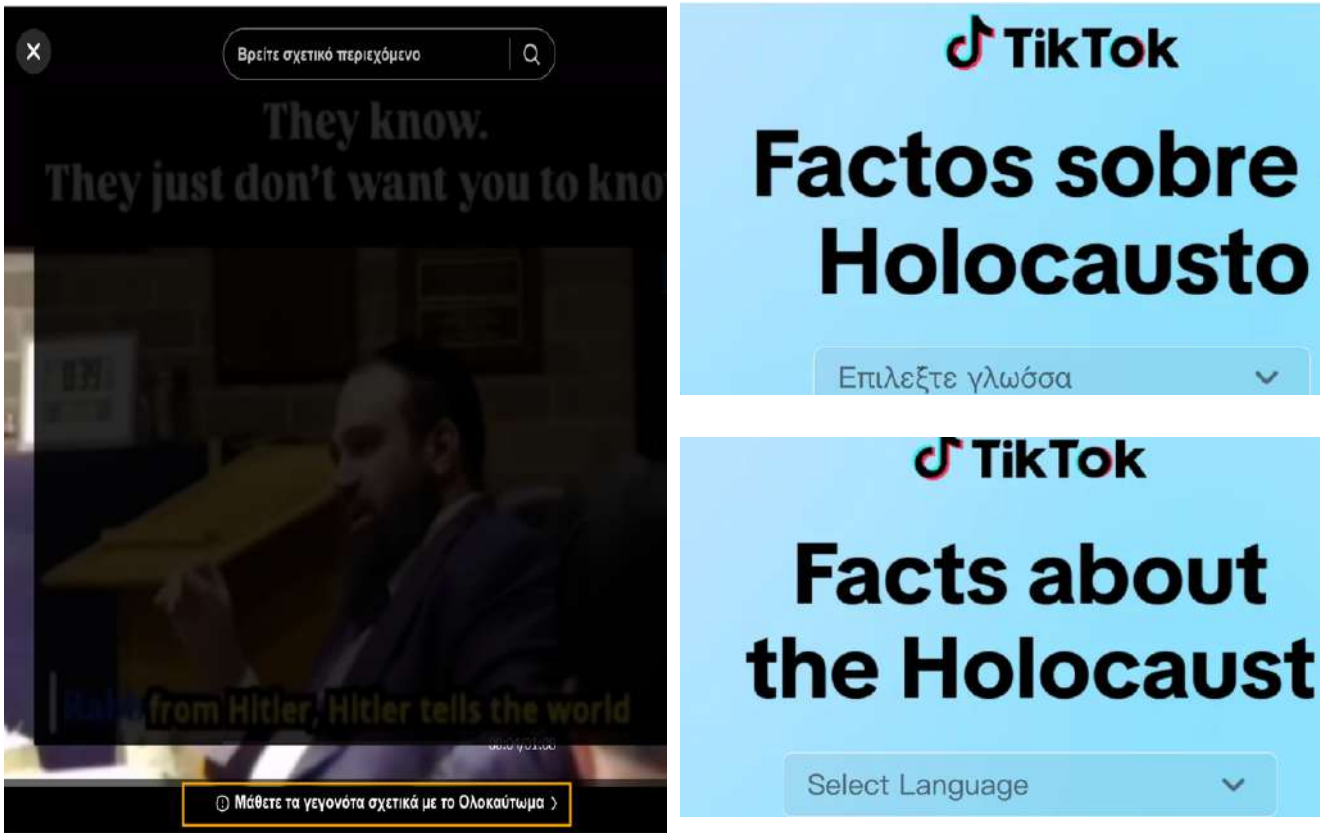



Figure 26: TikTok Video Notice Tag Language Issues with Redirecting Information Pages

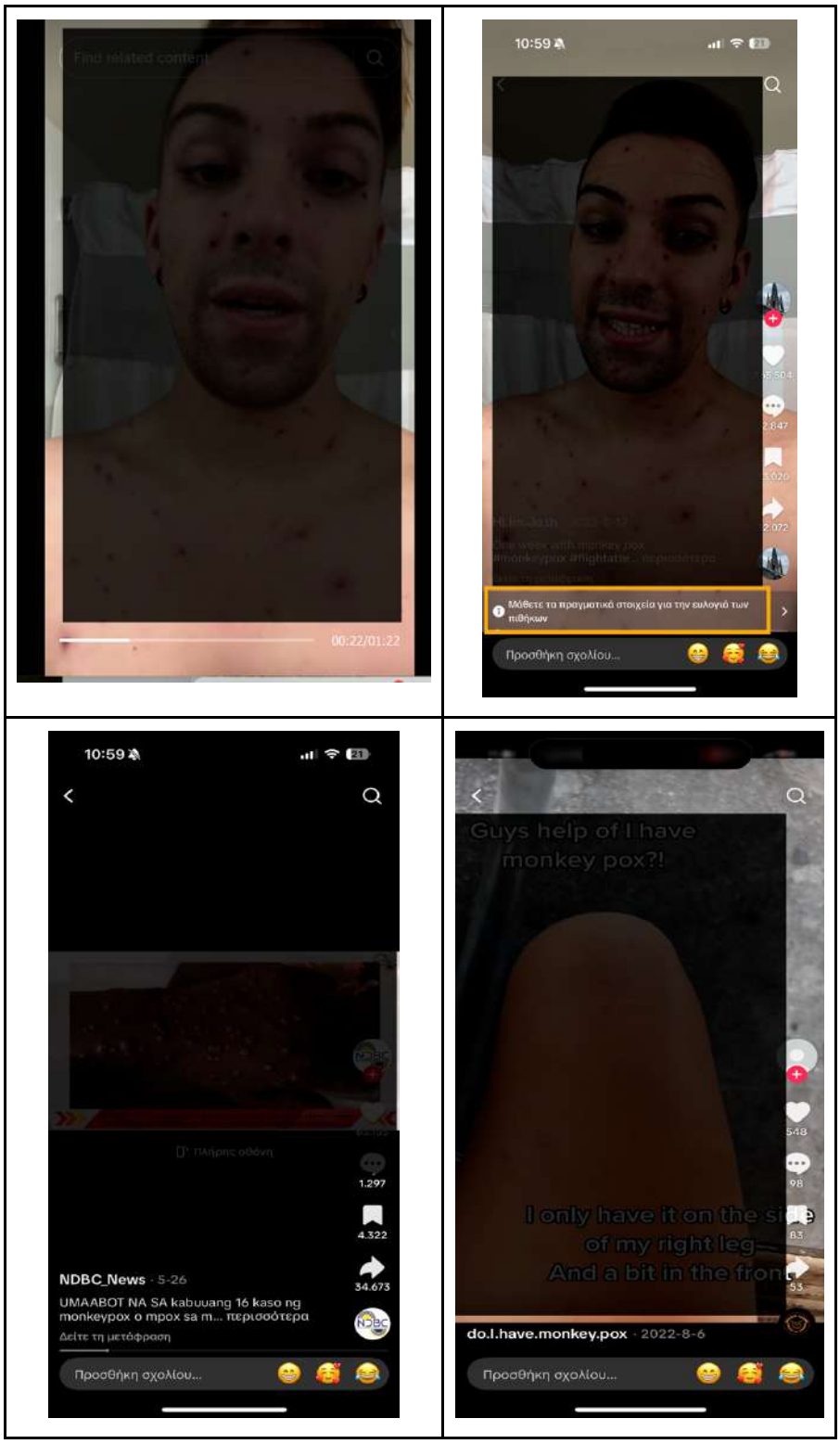



Figure 27: TikTok Video Notice Tag for Monkeypox Inconsistency

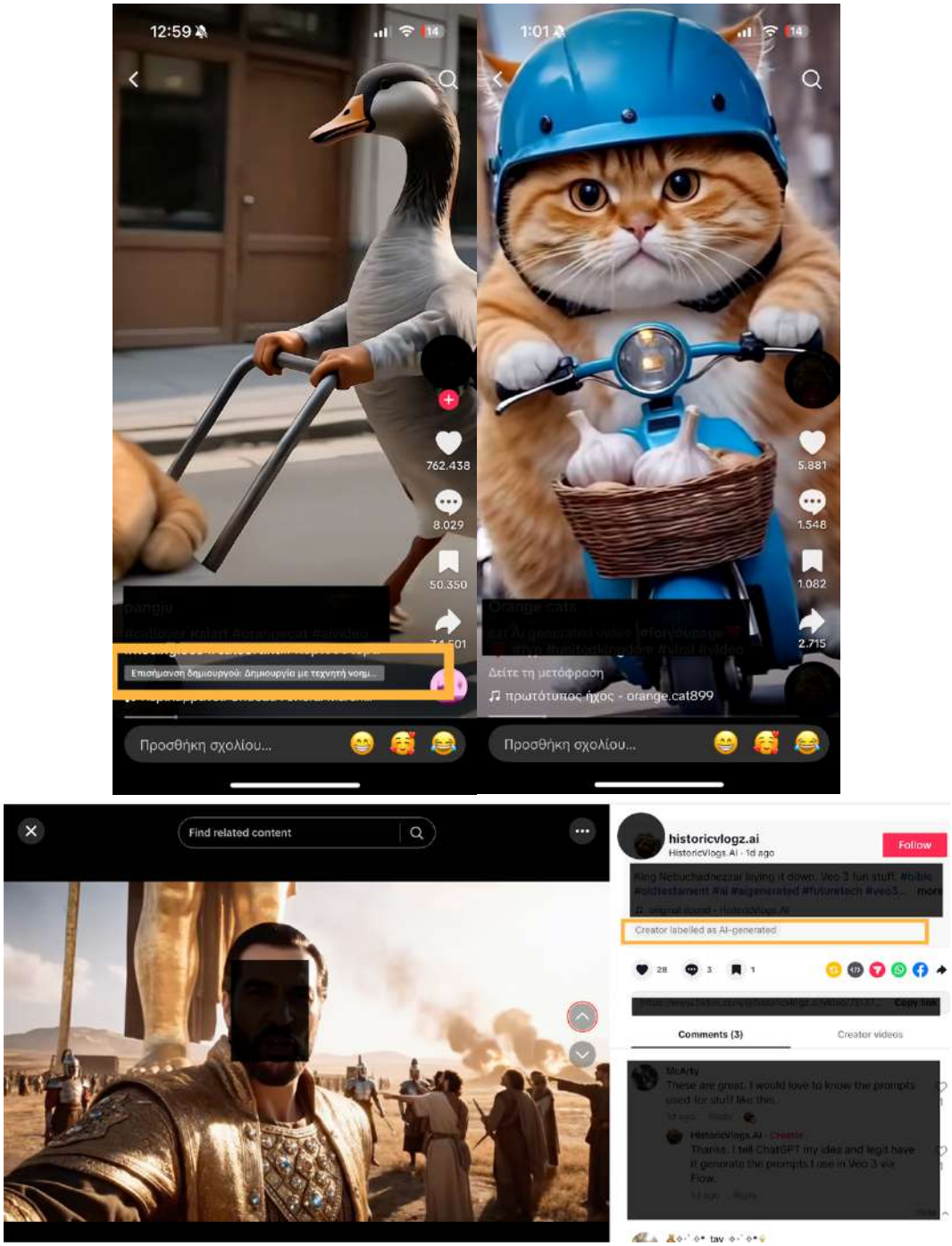



Figure 28: TikTok AI-generated content label availability

**Activities:**

- **Media literacy campaigns** were launched across all 27 EU member states, focusing on the 2024 EU Parliamentary election[52]. Within other country-specific campaigns, TikTok launched an Election Center[53] for the French Parliamentary Election 2024, providing users with up-to-date information relevant to the elections. There was a dedicated section for combating false information including videos created in partnership with the fact-checking organization Agence France-Presse (AFP).
- A **climate change search intervention tool** (see Figure 21) was introduced in 23 official EU languages, directing users to authoritative information on the subject.
- **Media Literacy campaigns** for **War in Ukraine** – available only for 14 countries: Austria, Bulgaria, Czech Republic, Croatia, Estonia, Germany, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia, Ukraine. – and **Israel-Hamas conflict** through Search Interventions when users search for relevant keywords.
- **TikTok's Election Speaker Series** collaborated with external experts to promote election integrity, including a partnership with Agence France-Presse (AFP) for Greece and Cyprus.
- TikTok rolled out critical thinking skills campaigns in the Netherlands, and Denmark.

**Partnerships:**

- TikTok partner with external experts[54] around the world for a range of topics, including media literacy.
- TikTok reported their partnership with organisations (i.e., Agence France-Presse (AFP), Deutsche Presse-Agentur (dpa), Demagog.pl, Demagog.cz, Facta, Faktograf, Logically Facts, Newtral, Poligrafo, Delfi.lt, The Journal, Nieuwscheckers, Funky Citizens, DigiQ, Ostro.) with expertise in media literacy for promoting the election integrity on TikTok, including campaigns for the 2024 EU Parliamentary Election, and the Election Speaker Series.
- TikTok reported collaboration with media literacy experts in Austria, Germany, Slovakia, etc. for the War in Ukraine-related campaigns and with WHO Tech Taskforce, and other EU fact-checking organizations (including Agence France Presse) for Covid-19 related misinformation to improve the in-app intervention tools.

[52] https://newsroom.tiktok.com/en-eu/our-work-to-prepare-for-the-2024-european-elections
[53] https://activity-i18n.tiktok.com/magic/eco/runtime/release/666aaeaeb991e002dd1e6a59?isFromGCP=1&appType=muse&magic_page_no=1&lang=en®ion=Default&use_spark=1
[54] https://www.tiktok.com/safety/en-gb/safety-partners/

**Major Comments**:

- The Maltese language is generally not supported by TikTok.

- The Search Intervention Tools and Video Notice Tags features are available in the Greek language as well. However, these features appear to be inconsistent in many ways:
    - Minor typos in search query results in the search intervention not to appear while still fetch related videos. For example,
    - Figure 24 shows that Monkeypox Search Intervention fails to display when there are minor typos in the search terms, even if relevant videos are returned. Similarly, the Israel–Hamas Search Intervention (in **Error! Reference source not found.**) is not activated for queries with minor typos are in the search terms—even when relevant videos are retrieved—highlighting a lack of linguistic and input flexibility.
    - Using keywords related to a topic again may not bring the search intervention section (see
    - Figure 22, and **Error! Reference source not found.**). For example, the Climate Change Intervention not available when searching in Greek and English, and the language is set to Greek and English respectively (see
    - Figure 22).
    - Using the Greek or Maltese language in search query may not display the search intervention feature. For example, the Holocaust Search Intervention (in
    - Figure 23) appears when the language settings are set to Greek and the search query is in English. However, it does not appear when the query is entered in Greek or Maltese, even if the language settings are set to Greek or English, respectively. Similarly, the Israel–Hamas Search Intervention (in **Error! Reference source not found.**) appears for English-language keywords when the language settings are set to either English or Greek. However, it does not activate for queries using Greek-language keywords.
    - The redirection page when clicking on the Video Notice Tag may not be available for some languages. For example, in Figure 26 there is a Video Notice Tag in Greek for Holocaust-related content. When the user clicks on the tag, they are redirected to the information page shown at the top right. However, when the interface language is set to Greek, the linked page is incorrectly displayed in Spanish, as Greek is not available in the language menu. The information page is accessible in multiple other languages, including English (bottom right).
    - Same videos on browser and in mobile app receive or not receive the video notice tag. In Figure 27, the Video Notice Tag for Monkeypox is visible when accessed via the TikTok mobile app (top right), and at the same the same video appears without the notice tag when viewed on a web browser (top left), indicating inconsistency in the display of the tag across devices.
    - Some videos receive the Video Notice Tag, while others on the same topic do not, even though the tag feature is available for the specific topic (i.e., monkeypox) (see Figure 27).

The tag is available in Greek, as shown in the top right of the figure. However, it is not consistently applied to all videos related to the topic (see examples at the bottom).

- The **Search intervention tool for the Ukraine War** is not available for the three countries. There is no reasoning behind who gets the tools and who does not.
- **The State-controlled media label** is now available in Cyprus, Greece, and Malta[55]. However, we did not manage to find accounts with the label in the three countries.
- We did not find examples of the **'Unverified Content Labels'** to examine if they are accessible to the audience of the three countries.
- The **AI-generated content label** is also available in Greek language. However, there is still AI-generated content that does not receive a label (see Figure 28). The label is available in the mobile application (top left of the figure) and on the browser view of TikTok as shown in the bottom of the figure. We found also examples of AI-generated content with no label (see top right of the figure). We assume those are the cases when the users do not self-label their content, and the TikTok detection mechanism fail to detect those videos and label them accordingly.
- Some of the **Safety Center** pages and the dedicated information on misinformation-related topics are also available in Greek, i.e., Harmful Misinformation[56], Online Challenges[57], Tragic Events[58].
- The **TikTok's Transparency Center** is not available in the Greek or Maltese languages.
- **The TikTok's Newsroom** offers Greece (in Greek) and EU (in English) as options in the country list to customize the view for the blogs' page. Cyprus and Malta are not options on the list.
- TikTok reported quantitative information for the reach and user engagement of their tools at a Member State level. See below the figures presenting the metrics for each tool in the three countries. The numbers show that the Video Notice Tag feature reaches more users, looking at the impressions and the number of clicks than the Search Intervention feature in most of the topics. This was expected as users mostly consume content on TikTok through their For You Feed rather than searching for specific content. For Video Notice Tags it would be beneficial to better understand the effectiveness of the number of users who interacted (i.e., share) with the content after receiving a tag. The report lacks quantitative information on the Unverified Content label and the AI-generated label. Additionally, TikTok could provide more engagement metrics for the

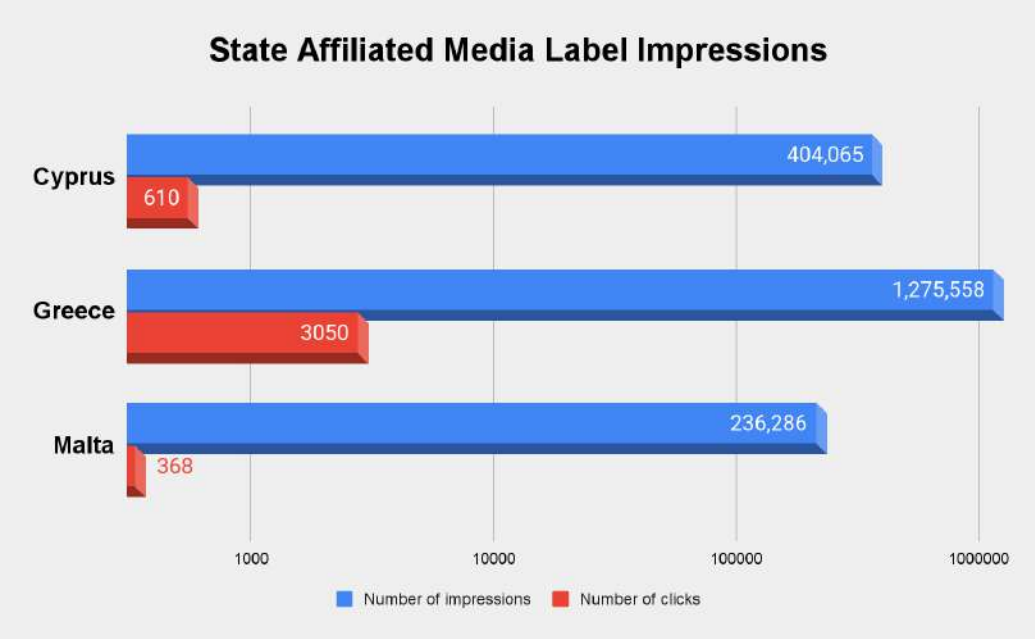
[55] https://www.tiktok.com/transparency/en-us/state-affiliated-media/
[56] https://www.tiktok.com/safety/el-gr/harmful-misinformation-guide
[57] https://www.tiktok.com/safety/el-gr/online-challenges
[58] https://www.tiktok.com/safety/el-gr/tragic-events-support

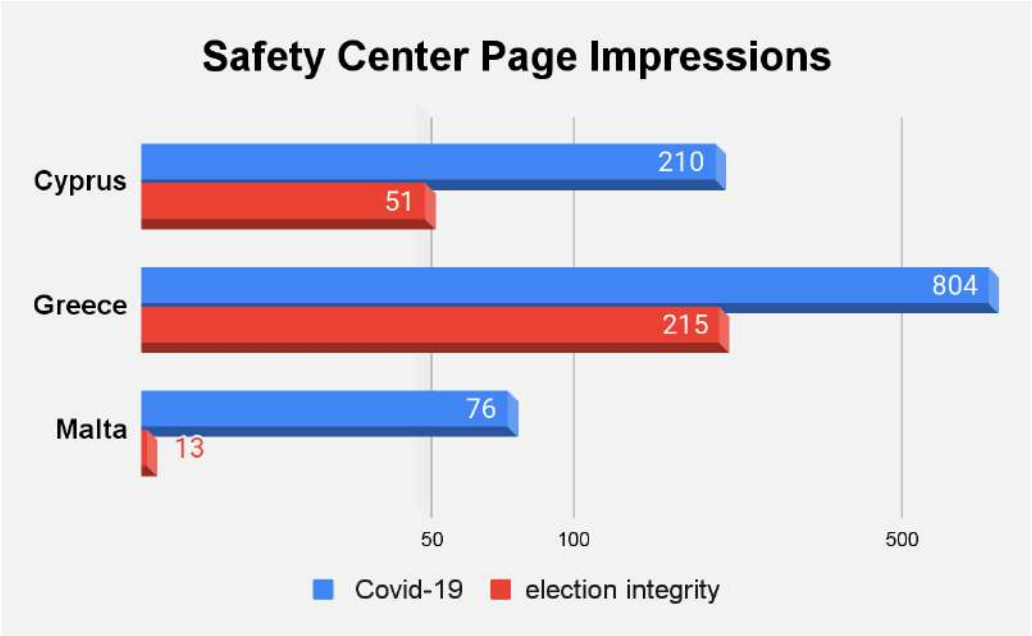
Transparency Page, misinformation related-specific Safety Pages, and Newsroom. No metrics are available for the search interventions for Israel-Hamas conflict or the War in Ukraine.

- Initiatives to enhance media literacy in Greece, Cyprus and Malta are limited (especially in Malta). Almost all activities and tools used in the three countries are part of broader initiatives undertaken by TikTok and concern major and often "controversial" topics (e.g. elections, climate crisis). Campaigns or programmes addressing the needs and challenges of the three countries individually were not materialised.

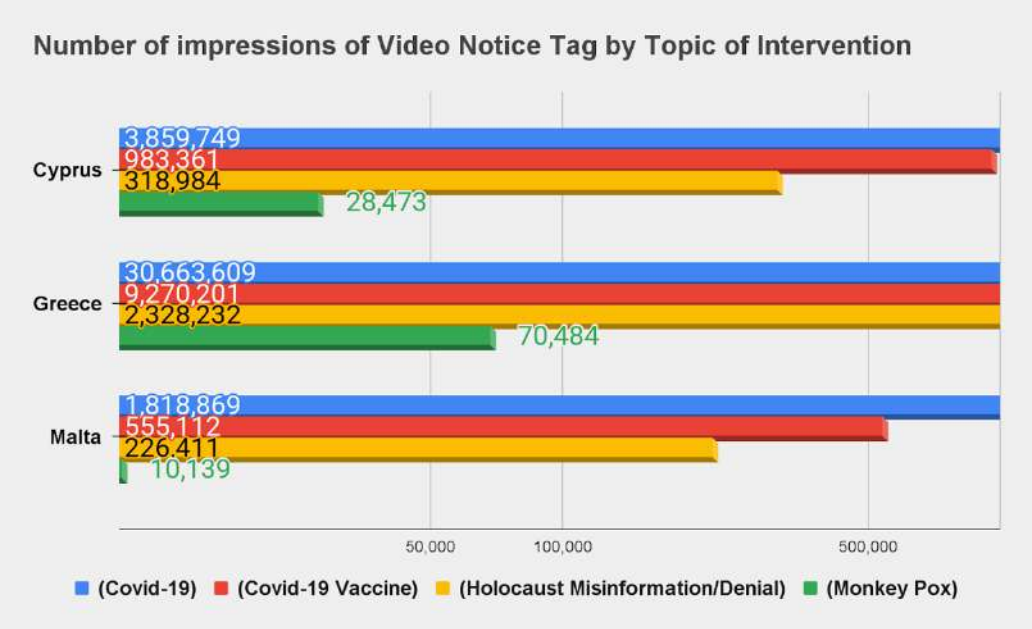
The TikTok reported information for SLI 17.1.1 can be found in the TikTok CoP 2024 report [TikTok, 2024], pages 112-137, next we present the reported numbers in figures:

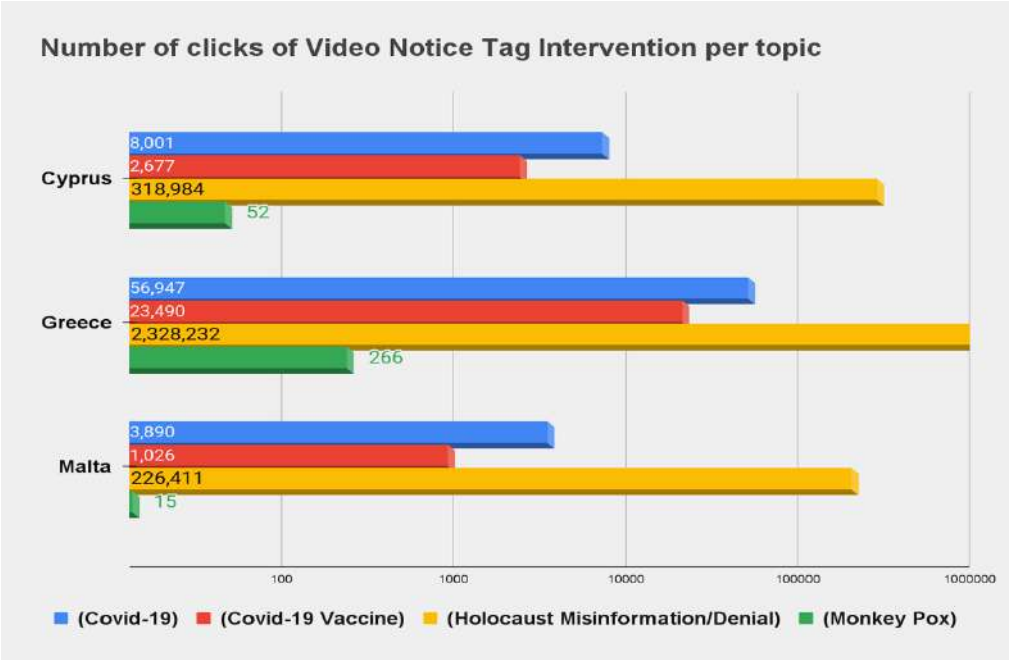
Table 7: TikTok's reported quantitative information for SLI 17.1.1

| State Affiliated Media Label |
| --- |
|  |
| Safety Centre Page Impressions |

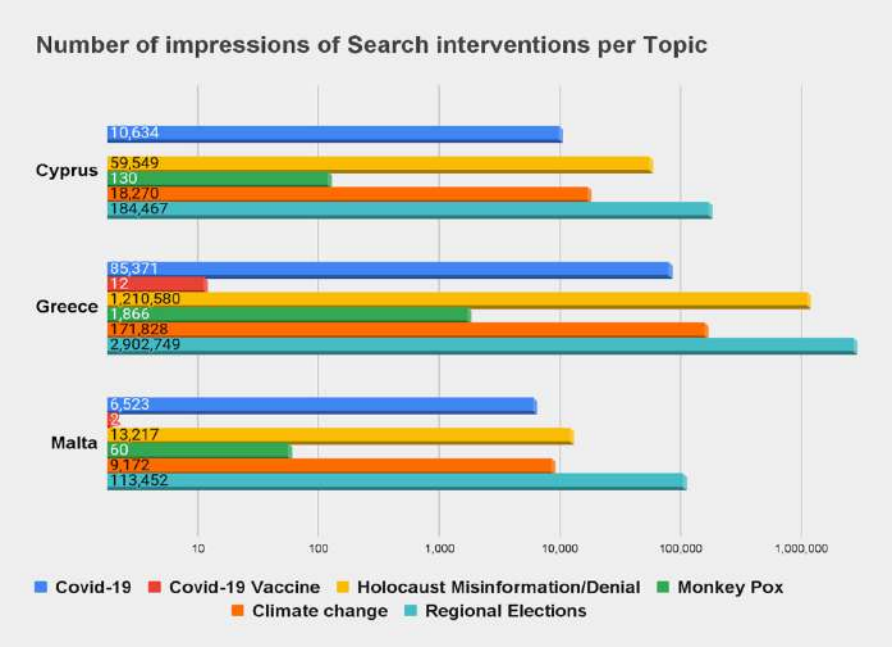


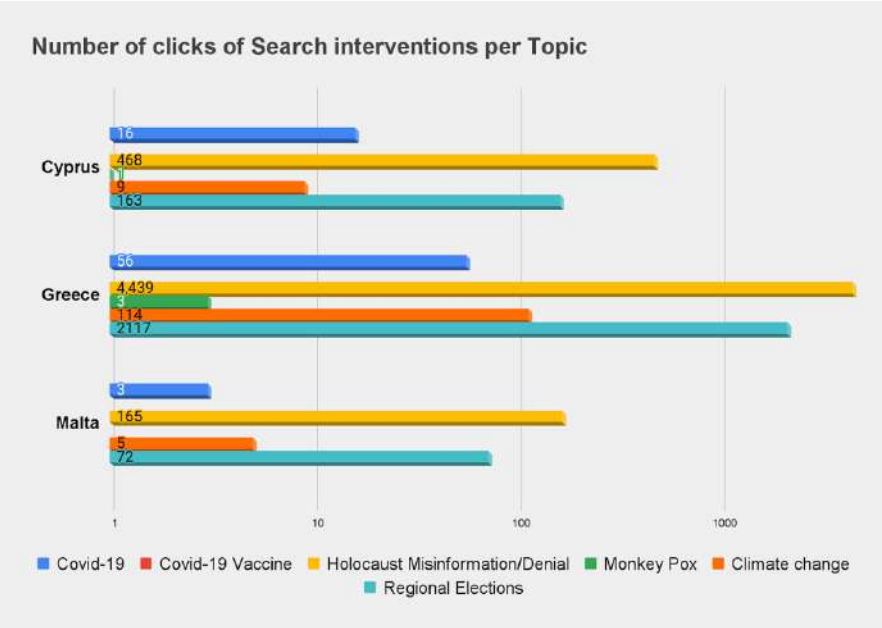

**Video Notice tags intervention**






**Search Intervention**




Number of clicks of Search interventions per Topic
Cyprus
16
468
1
9
163
Greece
56
4,439
3
114
2117
Malta
3
165
5
72
1
10
100
1000
Covid-19
Covid-19 Vaccine
Holocaust Misinformation/Denial
Monkey Pox
Climate change
Regional Elections

| Pillar-V: Empowering Users<br>*Commitment 21: Better Equipping Users to Identify Disinformation,*<br>*page 174-187* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 21** | 3 | 3 | 3 | 3 |

TikTok has made some efforts to ensure transparency and combat misinformation, particularly by partnering with independent fact-checkers and using tools to notify users when content is flagged. Key aspects of their approach include:

- **Independent Fact-Checking**: Enforcement of misinformation policies through 12 IFCN accredited fact-checking partners in Europe who provide fact-checking coverage in 23 official EEA languages, including at least one official language of each EU Member States. Fact-checking organizations that collaborate with TikTok review and assess content, while TikTok's misinformation moderators determine on actions such as removing or demoting content that violates its guidelines. When the fact-checking of a claim is inconclusive then, labels such as "**unverified content**" may be applied. Non-violating but potentially harmful content has its visibility reduced on the For You feed. Users receive inbox notifications when action is taken on flagged content, and creators are also notified when their videos are marked as unsubstantiated

- **User awareness on fact-checking partnerships:** Users can learn about fact-checking partnerships and labels through the Safety Center & Transparency Center in more than 25 languages.

- **In-App Tools**: These tools provide information on specific topics like **COVID-19**, **election integrity**, and **climate change** to improve user awareness.

- **New AI-generated label:** Building on a new AI-generated label for creators to disclose content that is completely AI-generated or significantly edited by AI.

**Partnerships with Researchers to design and evaluate TikTok Tools to empower users to identify misinformation:**

- **Massachusetts Institute of Technology (MIT) & University of Regina:** Consulted on a study to measure how showing users accurate information after they'd seen a misinformation claim affects their behaviour.
- **Irrational Labs (behavioural scientists):**

  - Partnered in 2021 to design and test the "unverified content" prompts, which reduced shares of labelled videos by 24% and likes by 7% **[Irrational Labs, 2021].**
  - Worked on developing and user-testing the "state-controlled media" label (across English, Spanish, Arabic), finding "[country] state-controlled media" to be the clearest and most accurate phrasing.
- **IFCN-accredited fact-checkers (EU):** Help assess the accuracy of content; where fact-checks are inconclusive (e.g. during unfolding events), videos receive the "unverified content" label.
- **Media experts, political scientists, academics, and representatives from international organisations & civil society:** Consulted across North/South America, Africa, Europe, Middle East, Asia, and Australia to inform TikTok's state-affiliated media policy.
- **TikTok Safety Advisory Councils:** Engaged in advance of launching the AI-generated content label to advise on policy and user experience.
- **Industry experts, including Dr David G. Rand (MIT):** Drew on Dr Rand's research into how viewers perceive AI-disclosure labels to guide the design of TikTok's "AI-generated content" label.

**Major Comments:**

- TikTok's efforts represent a step toward improved media literacy and content moderation, but there are significant gaps in their reporting and data-sharing practices.
- TikTok listed their partnerships with experts on consultation to design or evaluate their tools. However, there is no detailed information on the partnerships and the findings of the studies TikTok used for their design or evaluation. Important to mention that there are no studies to evaluate the actual impact of these warnings and interventions.
- TikTok reported that the fact-checking organizations review content, but the decision is made by their content moderators. It is not clearly described how the moderators take the decisions to remove or demote content.
- TikTok provided quantitative data such as the share cancel rate for content that received the unverified content label, and the share of content removals under the misinformation policy, etc. (see Tables below).
- Malta sees the biggest effect, with 30.2% of users stopping a share when warned, followed closely by the EU average (29.7%). Cyprus trails at 25.0% and Greece at 27.0%, suggesting room to optimize the pop-up's impact there.
- Cyprus and Greece have roughly a 30% share of all removals under this policy, indicating a significant volume of flagged content. Malta's comparatively low 12.1% could reflect fewer overall

cases, different content patterns, or lower moderation coverage of content from Maltese creators.

- TikTok takes down proactively the vast majority of flagged misinformation before any user report- with average 97% in the EU. Proactive removals share in Malta and Cyprus exceeds 95%, while in Greece it is 87%.
- 63% (Cyprus), 64% (Greece) and 75% (Malta) of videos judged to violate policy are removed before garnering any views.
- Most removals happen within 24 hours of posting, with Cyprus leading at 84.5%, Greece at 80.8% and Malta at 75%.
- TikTok also reports the number of contents receiving the 'unverified label', ranging from 1 to 2K for the three countries.
- The data provided by TikTok shows that the unverified content labels have a significant impact on users sharing the content, and at the same time that the share of TikTok content removals under the misinformation policy holds a considerable share of the content removals, and they also do that timely before reaching the users. However, we are not able to independently assess these metrics.

Table 8: TikTok's reported quantitative information for SLI 21.1.1

| TikTok | | | | | |
|---|---|---|---|---|---|
| **SLI 21.1.1** | **Methodology of data measurement:** 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 the total removals of each policy. 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 |
| Cyprus | 25% | 31.9% | 95.7% | 62.9% | 84.5% |
| Greece | 27% | 29.5% | 87.1% | 64% | 80.8% |
| Malta | 30.2% | 12.1% | 98.6% | 75% | 75% |
| Total EU | 29.7% | 26.6% | 97.2% | 80% | 83.7 |

Table 9: TikTok's reported quantitative information for SLI 21.1.2

| 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 | 1.033 | 25 |
| Greece | 1.874 | 27 |
| Malta | 1.101 | 32.5 |
| Total EU | 82.147 | 29.7 |

### 3.5.2 TikTok: Empowering the Research Community

| Pillar-VI: Empowering the Research Community,<br>*Commitment 26: Empowering the Research Community*<br>*pages 208 - 215* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5 based on EDMO methodology** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 26** | 3 | 3 | 3 | 3 |

**TikTok Access to data for researchers:**

**Research API**[59]: Provides programmatic access to *public* TikTok content and account metadata, including comments, captions, subtitles, counts of comments/likes/shares/favourites, follower/following lists, etc. Initially launched for U.S. academic researchers; now open to qualified non-profit academics in the EEA, UK and Switzerland.

**Virtual Compute Environment (VCE):** A secure, sandboxed environment for qualifying academic, not-for-profit researchers in the EU. There is a two-stage process to analyse public data from TikTok: 1) Test Stage: SDK queries return random samples (up to 5,000 records per day). 2) Execution Stage: Researchers submit analysis scripts against the full public dataset; TikTok reviews the scripts outputs to ensure only aggregated data is extracted. If TikTok approves the output, the researcher will receive a link to download the results.

**Commercial Content API**[60]**:** As mandated by the DSA, provides ads and advertiser metadata (creative, run dates, targeting parameters, reach, etc.). Access by approved TikTok for Developers account holders worldwide, with EU data currently included.

**Commercial Content Library**[61]**:** Publicly searchable repository of all paid ads (and tagged commercial-nature content) served in the EEA, UK and Switzerland. Includes creatives, metadata (targeting criteria, impressions, dates), and remains available for one year after last view.

**TikTok Transparency Centre:** The following reports are available in the Transparency Center:

**1) COPD Transparency Reports**[62] (biannual): granular take-down and moderation data for EU/EEA.

[59] https://developers.tiktok.com/products/research-api/
[60] https://developers.tiktok.com/products/commercial-content-api
[61] https://library.tiktok.com/ads
[62] https://www.tiktok.com/transparency/en/copd-eu/

**2) Community Guidelines Enforcement Reports**[63] (quarterly): proactive removals and account actions since 2019, with downloadable aggregated datasets.

**3) DSA Transparency Reports**[64] (biannual): Detailed EU content-moderation metrics under the Digital Services Act. Expanded downloadable data (e.g., removal counts by policy category for top-50 markets, live-stream enforcement, view-counts prior to removal).

**Major Comments:**

- TikTok provides a relatively comprehensive set of tools, documentation, and processes to support access to non-personal data for disinformation research.
- TikTok has documented definitions for public data (Research API Codebook[65] and the Commercial API Data structures[66]) and the application processes for accessing these datasets.
- Researchers in Cyprus, Greece, and Malta have access to the reports. Additionally, researchers can freely access the Commercial Library. However, for accessing the Research API, Commercial API, and VCE, researchers must complete an application form which will be later reviewed by TikTok based on a number of criteria.
- TikTok reported the numbers of applications received for researchers' access to the Researcher API and Commercial API. There were no applications from researchers in Cyprus, and Malta for both the APIs. In Greece, two applications were submitted for the Research API, and only one got accepted.
- Looking at the codebook of the Research API, it is not clear if the researchers can get information if a video received the 'unverified content label', or the 'state-affiliated label', nor if the video was demoted from the recommendation algorithm.
- There are no metrics on the impressions, or downloads of the reports TikTok mentioned.
- Overall, TikTok has shown a commitment to supporting academic research and has taken steps in this direction. While improvements have been noted, further efforts are necessary to provide more targeted, complete, and meaningful data as well as information to verify the data provided by the platform.

63 https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2021-2/
64 https://www.tiktok.com/transparency/en/dsa-transparency/
65 https://developers.tiktok.com/doc/research-api-codebook?enter_method=left_navigation
66 https://developers.tiktok.com/doc/commercial-content-api-get-ad-details?enter_method=left_navigation

| Pillar-VI: Empowering the Research Community,<br>***Commitment 27: Governance Structure for Access to Data for Research Purposes Requiring Additional Scrutiny,***<br>*pages 215 - 217* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 27** | 3 | 3 | 3 | 3 |

**TikTok Pilot Programs for data sharing:**

TikTok completed the data access pilot with EDMO, which trialled the process for sharing data with vetted researchers designated under the DSA.

TikTok refined their standard operating procedure for vetted researcher access to ensure compliance with the provisions of the Delegated Act on Data Access for Research.

TikTok reported that they actively participate in the EDMO working group for the creation of the Independent Intermediary Body (IIB) to support research on digital platforms.

**Major Comments:**

- Completing the EDMO pilot is a concrete milestone that demonstrates TikTok's early technical and procedural readiness for DSA-mandated data-sharing. Beyond noting completion, TikTok does not provide metrics or case studies—e.g., the number of researchers participated, the types of queries run, or lessons learned—that would illustrate real-world usability.
- TikTok's involvement in the IIB working group shows strategic commitment to shaping the intermediary's design. The report lacks specifics on TikTok's contributions—such as proposed governance models or procedural recommendations—that would evidence substantive influence on the body's emerging framework.
- TikTok reported refining internal processes to signal compliance readiness. Without details on what was changed (e.g., data-handling safeguards, researcher vetting criteria), it's difficult to assess whether these refinements materially enhance transparency or researcher protections.
- TikTok's narrative confirms formal participation and completion of key tasks—but offers limited new insight into outcomes, researcher experiences, or quantitative achievements.

| Pillar-VI: Empowering the Research Community,<br>*Commitment 28: Cooperation with Researchers,*<br>*pages 217 - 224* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 28** | 3 | 3 | 3 | 3 |

TikTok outlines a broad and growing set of initiatives to support independent research and transparency, particularly in relation to misinformation and disinformation. Key elements include:

- **Research Tools and Data Access**: TikTok has expanded its Research API to provide access to public content and account data and introduced the VCE (Verified Research Environment) for qualifying non-academic, not-for-profit researchers in the EU. These tools ensure privacy and security while enabling meaningful analysis of platform dynamics.

- TikTok also maintains open access to its Commercial Content API and Commercial Content Library, which include ad metadata, targeting information, and metrics on ad impressions.

- **Transparency and Accountability**: TikTok operates four Transparency and Accountability Centers (TACs) in Dublin, Los Angeles, Singapore, and Washington, DC. These facilities host stakeholders and allow them to observe TikTok's moderation and platform safety processes.

- **Collaborations with Expert Councils**: TikTok engages with nine Safety Advisory Councils, including the European Safety Advisory **Council** and the Youth Advisory Council, to incorporate expert and civil society input into policy design and feature development—such as the AI-generated content label.

- **Academic and Civil Society Engagement**: TikTok collaborates with external experts and academics (e.g., MIT's Dr. David Rand) and consults with researchers on topics such as election misinformation and AI content disclosure. They held expert briefings (e.g. the Election Speaker Series) with fact-checking organizations across the EU in the lead-up to the 2024 European Parliamentary Elections.

- **Event Participation and Outreach**: TikTok actively participates in research events, including GlobalFact **11**, hosted webinars with EDMO, and other digital rights and tech summits. They estimate reaching ~135 research stakeholders during the reporting period through such engagements.

- **EDMO and DSA Engagement**: TikTok completed a pilot program with EDMO to trial data sharing with vetted researchers under the Digital Services Act (DSA)**.** They expressed a willingness to continue collaborating with EDMO and the forthcoming Independent Intermediary Body (IIB)**.**

- **Commitment to Continuous Improvement**: TikTok states that it continuously refines its tools, data-sharing processes, and transparency mechanisms, and will report further developments in future Code of Practice on Disinformation (CoP) submissions.

**Major Comments:**

- TikTok reported having multiple dedicated teams (e.g., product, policy, data science, legal, outreach) to support research engagement. They mentioned hosting Transparency and Accountability Centers (TACs) and working with several advisory councils. However, no specific reference is made to dedicated points of contact or resources in Cyprus, Greece, or Malta, and no country-specific activities or research collaborations were reported.

- TikTok described multiple tools available to researchers, including the Research API, the new VCE (for vetted non-academic researchers), and the Commercial Content API. While these tools offer access to public account and content data, the platform did not provide metrics on usage or uptake by researchers from Greece, Cyprus, or Malta. Importantly, it is unclear how accessible these tools are for smaller, less-resourced research institutions in these Member States.

- TikTok did not disclose any concrete figures or breakdown of financial or technical resources allocated specifically for disinformation research. Statements on resource allocation are vague, and there is no evidence of country-level funding support or investment in localized research.

- TikTok noted the completion of a data-sharing pilot with EDMO, which is a positive step. However, there is no follow-up on resource allocation, researcher uptake, or whether any researchers from Cyprus, Greece, or Malta participated in or benefited from this pilot.

### 3.5.3 TikTok: Empowering the fact-checking community

| ***Pillar-VII: Empowering the fact-checking community***<br>*Commitment 30: cooperation with the fact-checking community,*<br>*page 226-234* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 30** | 3 | 3 | 3 | 2 |

**TikTok and Fact-checking organizations:**

TikTok works with 19 IFCN-accredited fact-checking organisations globally, covering 23 official languages in the European Economic Area (EEA): Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Norwegian, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish, Swedish. In Europe, TikTok collaborates with 12 IFCN-accrediated fact-checking organizations: Agence France-Presse (AFP) –covering Greece and Cyprus, dpa Deutsche Presse-Agentur, Demagog, Facta, Faktograf (onboarded in 2024), Lead Stories, Logically Facts, Newtral, Poligrafo, Reuters, Science Feedback, Teyit.

During high-risks events, TikTok makes temporary agreements with fact-checking partners to cover additional languages. For example, during the 2024 EU Elections, TikTok also covered fact-checking in the Maltese language.

TikTok Fact-checking Partners receive videos flagged either algorithmically or via user reports and evaluate each piece against TikTok's classification schema (e.g., misinformation, unverified, partly false). Fact-checkers assign a rating and, where appropriate, submit written reports on emerging disinformation trends. TikTok moderators then use these independent assessments to remove content, prevent it from being recommended, or apply an "unverified content" label.

All fact-checking partners sign from TikTok's master services agreement template, which defines:

- Scope of services (content review, rating, trend reporting)
- Performance expectations and service levels
- Language coverage commitments and ad-hoc project support
- Reporting obligations and optional proactive flagging
- Fees, term length, and renewal processes

**Media Literacy Campaigns**: TikTok also reported a number of collaborations with its fact-checking partners for media literacy campaigns (e.g., 2024 EU Parliamentary Election, 2024 French Parliamentary Election, 2024 Slovak Presidential Election, etc.).

**Major Comments:**

- Overall, TikTok shows a strong commitment on collaboration with fact-checkers, as evidenced by its partnerships with multiple IFCN-accredited organisations and the development of advanced tools to detect misinformation.
- There is no fact-checking organization covering Malta, or the Maltese language. Agence France Presse (AFP) covers Cyprus and Greece, and the Greek language.
- Regional adaptations, such as temporary Maltese language support during EU 2024 elections, reflect TikTok's responsiveness to important local events.
- However, the lack of detailed reporting on resource allocation, evaluation frequency, and the depth of collaborations limits the ability to assess the program's actual impact.
- Additionally, the frequency and method of reviews with fact-checkers are not described in detail, though TikTok claims to have regular meetings and ongoing dialogue with partners.
- Emphasis on fair compensation and transparent processes for fact-checking partners is a positive step. The absence of specific quantitative data reduces transparency, especially regarding how compensation is distributed, and which countries receive the most support.

| *Pillar-VII: Empowering the fact-checking community*<br>*Commitment 31: use and integration of fact-checking*<br>*page 235-243* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 31** | 2 | 2 | 2 | 1 |

**Fact-checking work integration on TikTok:**

The fact-checking partners collaborate with the platform in three primary ways:

1. Fact-checking specific videos flagged by moderators.
2. Contributing to a global database of fact-checked claims.
3. Participating in a proactive detection program to flag new and evolving claims.

Fact-checkers offer advisory input only, while TikTok's moderation team makes the final decisions on content. If a fact-check is inconclusive, TikTok informs viewers with a banner ("Unverified content label") suggesting reduced sharing. Such videos may be excluded from recommendations in the "For You" feed.

- Major Comments:

- TikTok states that moderators use fact-checkers' ratings to determine whether to remove content or label it as "Unverified Content." However, the exact decision-making process remains unclear. There is no transparency on how fact-checker inputs are weighed or translated into moderation actions, leaving a gap in understanding the real influence of third-party verification on enforcement decisions.

- During the reporting period, TikTok submitted just 17 videos from Cyprus, 215 from Greece, and a single video from Malta to fact-checking partners. These figures are exceptionally low—particularly in Malta and Cyprus—and raise questions about the coverage and prioritisation of content in these markets. The platform appears to be significantly underutilising its partnerships with fact-checkers in these countries.

- Following fact-checking assessments, TikTok moderators removed only 2 videos from Cyprus, 31 from Greece, and none from Malta. These numbers are disproportionate when compared to the total volume of misinformation likely circulating on the platform, suggesting that either the

threshold for removal is high or that many potentially harmful posts are not reaching fact-checkers at all.

- TikTok also reported removals based on internal moderation triggers—such as policy guidelines, known misinformation trends, and a knowledge-based repository. These resulted in 329 removals from Cyprus, 2,294 from Greece, and 72 from Malta, none of which underwent fact-checking assessment. The significant gap between removals via internal systems versus fact-checked content calls into question the role and value of the fact-checking process in TikTok's overall moderation pipeline.

- There is no baseline or reference data provided, making it impossible to fully contextualise or assess the meaning of these figures. Without publicly available historical data or benchmarks, the numbers—however precise—remain difficult to interpret.

- All reported data originate from TikTok's internal systems, with no external audits or third-party verification mechanisms in place. Aside from informal input from fact-checking partners, there is no local or EU-level infrastructure to corroborate the statistics. This raises doubts about the reliability of the reporting.

- The sharp contrast between videos removed after fact-checking assessments and those removed purely through policy enforcement is striking across all three countries. It suggests either limited integration of fact-checking workflows or inefficiencies in flagging high-risk content to external verifiers.

- TikTok claims to be developing a repository of fact-checked claims to support moderator decisions but provides no technical details or scope of the initiative. The absence of information on how this repository functions, what content it includes, or how often it is updated makes it impossible to assess its effectiveness.

- Despite repeated references to fact-checking partnerships, TikTok offers little insight into the actual contributions of fact-checking organisations. There is no data on verification turnaround times, feedback loops, or the types of claims reviewed, leaving their operational role poorly defined.

- TikTok should substantially improve transparency by disclosing how fact-checking decisions are made, how content is routed for review, and the criteria moderators use in applying or dismissing fact-checker recommendations. The absence of such contextual detail significantly undermines the credibility of the reported metrics and hampers meaningful evaluation of their effectiveness.

Table 10: TikTok's reported quantitative information for SLIs 31.1.1 – 3

| 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 due to 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 pol,icy as all enforcement is based on the policy guidelines, known misinformation trends and a knowledge-based repository. | | **(SLI 31.1.3)** Methodology of data measurement: The metric we have provided demonstrates the % of videos that have been removed as a result of the fact-checking assessment, in comparison to the total number of videos removed due to violating 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** | **17** | **2** | **329** | **0.6%** |
| **Greece** | **215** | **31** | **2294** | **1.4%** |
| **Malta** | **1** | **0** | **72** | **0** |
| **Total EU** | **8.729** | **1.661** | **262.652** | **0.6%** |

| *Pillar-VII: Empowering the fact-checking community*<br>*Commitment 32: fact-checkers access to relevant information*<br>*page 244-245* | | | | |
|---|---|---|---|---|
| **Evaluation with scale 1-5** | | | | |
| | **Evaluation of Reported Actions (QREs)** | **Evaluation of Implementation (SLIs)** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Commitment 32** | 2 | 2 | 2 | 2 |

**TikTok's tools for their fact-checking partners**:

**Dashboard:** TikTok provides fact-checkers with access to a dashboard where they can review flagged content. The dashboard provides some quantitative information, including the number of videos queued for assessment, the time taken for review, and the ratings applied by fact-checkers in their previous assessments.

**Major Comments:**

- TikTok does not report any additional interfaces or automated tools beyond the dashboard for fact-checkers.
- No details are available on the dashboard or user engagement metrics.
- TikTok does not specify the communication channels or frequency of exchanges with the fact-checking community.
- Without objective evidence, it is challenging to validate TikTok's claims or assess their compliance with relevant commitments.

# 4 MedDMO Fact-Checking Partners Collaboration with VLOPs

The MedDMO consortium includes three fact-checking organisations certified by the International Fact-Checking Network (IFCN): Agence France-Presse (AFP), Ellinika Hoaxes (EH), and the recently established Fact-Check Cyprus. In parallel, the Times of Malta and the University of Malta carry out fact-checking activities focusing on misinformation trends within Malta. All fact-checks related to Malta are accessible via the MedDMO website.

Two of our IFCN-certified partners, Ellinika Hoaxes and Agence France-Presse, maintain ongoing collaborations with certain Very Large Online Platforms (VLOPs) in the fight against disinformation. These collaborations include participation in VLOP third-party fact-checking programmes, engagement in media literacy campaigns, and the application of platform-provided tools to label or contextualise false or misleading content.

To gain further insight into the nature of these collaborations, we invited both organisations to respond to an open-ended questionnaire detailing their experiences working with the platforms. The questionnaire is available in **Annex II**. Next, we present the main findings of the questionnaire.

Table 11:MedDMO Fact-Checking Partners Collaboration with Meta

| **Collaboration with Meta: Third Party Fact-Checking Program (3PFC)** | | |
|---|---|---|
| | **EH** | **AFP** |
| **Collaboration:** | Ellinika Hoaxes has been collaborating with Meta as part of the company's third-party fact-checking program since May 2019. EH fact-checks content on Facebook, Instagram, and, since 2024, on Threads as well. | AFP has collaborated with Meta in its Third-Party Fact-Checking (3PFC) program since 2017. The collaboration extends to fact-checking in 26 languages worldwide, covering various platforms including Facebook and Instagram. |
| **Fact-Checked Content:** | Under the 3PFC program, EH can fact-check content on Facebook, Instagram, and Threads, including text-only posts, images, videos, reels, and sponsored content. | 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 four stages: Detection, Evaluation, Refutation through investigation, and Ratings Submission. Check-worthy claims are first | AFP's fact-checking process involves manually monitoring social networks, Facebook queues, and WhatsApp queries. The process includes evaluating the fact-checkability and potential |

| | | |
|---|---|---|
| | detected, primarily through manual monitoring or shared with EH by our audience. Subsequently, they are evaluated based on their virality and potential harm. Claims deemed worthy are investigated, and fact-checks are written and published. Ratings on the original content spreading the claim are submitted through the platform's provided system. | 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:** | Figures specifically detailing the impact of EH's work, as provided by the platform, are limited and we are unable to share them. Metrics on the overall impact of the 3PFC program are publicly shared by the company under its obligations to the DSA and CoP. | Meta publicly shares information about its collaboration with fact-checking partners, including AFP in its reports to the DSA, and the Code of Practice. Metrics regarding the impact of fact-checking programs are provided, including 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 55 fact-checking articles per month, but the exact number of ratings applied in content by Meta 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, although specific figures are considered confidential and not disclosed. AFP seeks access to the internal archives of platforms relevant to their fact-checking work, 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 **10 editorial members** contributing through their work 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 12: MedDMO Fact-Checking Partners Collaboration with TikTok

| Collaboration with TikTok: Fact-Checking Program | |
|---|---|
| | **AFP** |
| **Collaboration:** | AFP has collaborated 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, assessing potential harm, and virality. 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 its Code of Practice and DSA reports regarding the impact of 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 platforms relevant to their fact-checking work, 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[67]. 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. |

[67] https://activity.tiktok.com/magic/eco/runtime/release/64400a0478c79d0360a77740?appType=tiktok&magic_page_no=1

Table 13: MedDMO Fact-Checking Partners Collaboration with Google and other platforms

| **Other Collaborations** | | |
|---|---|---|
| | **EH** | **AFP** |
| **Google** | Ellinika Hoaxes does not have an active collaboration with Google. Like most fact-checking organizations globally, EH provides some data from its fact-checks through ClaimReview. This data is integrated into Google's search results and Fact Check Explorer. No compensation is provided for this data. | 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[68], English[69], Spanish[70], and Portuguese[71]. AFP also create tips and techniques videos in this context (French[72], English[73], Spanish[74]). |
| **Other platforms** | Ellinika Hoaxes doesn't have any other active collaboration with platforms beyond META. Yet we often fact-check content spreading on platforms like X, Telegram, and YouTube. Ellinika Hoaxes emphasises the importance of various platforms engaging more with fact-checking initiatives for an enhanced impact. | **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. |

[68] https://fr.digitalcourses.afp.com/?_gl=1*x...
[69] https://digitalcourses.afp.com/
[70] https://es.digitalcourses.afp.com/?_gl=1*78krj...
[71] https://br.digitalcourses.afp.com/?_gl=1*1lb6m...
[72] https://www.youtube.com/playlist?list=PLo9T0OZu4qjk7MVqK7VxTFoiI5LTM4FYF
[73] https://www.youtube.com/playlist?list=PLo9T0OZu4qjk7MVqK7VxTFoiI5LTM4FYF
[74] https://www.youtube.com/playlist?list=PL3oLC6iSclxCCkNtfzRpxw9Fh0hsKzDe4

# 5 Supporting the National Authorities

As part of the MedDMO project, our mission is to assist media authorities in Cyprus, Malta, and Greece in addressing the complex challenge of disinformation.

To support this goal, we have actively engaged with the respective regulatory bodies—the Cyprus Radio Television Authority (CRTA), the Broadcasting Authority of Malta (MBA), and the National Council for Radio and Television (NCRTV) in Greece—with the aim of building collaborative partnerships. These relationships are designed to enhance information sharing, coordination, and strategic responses to disinformation threats across the Mediterranean region.

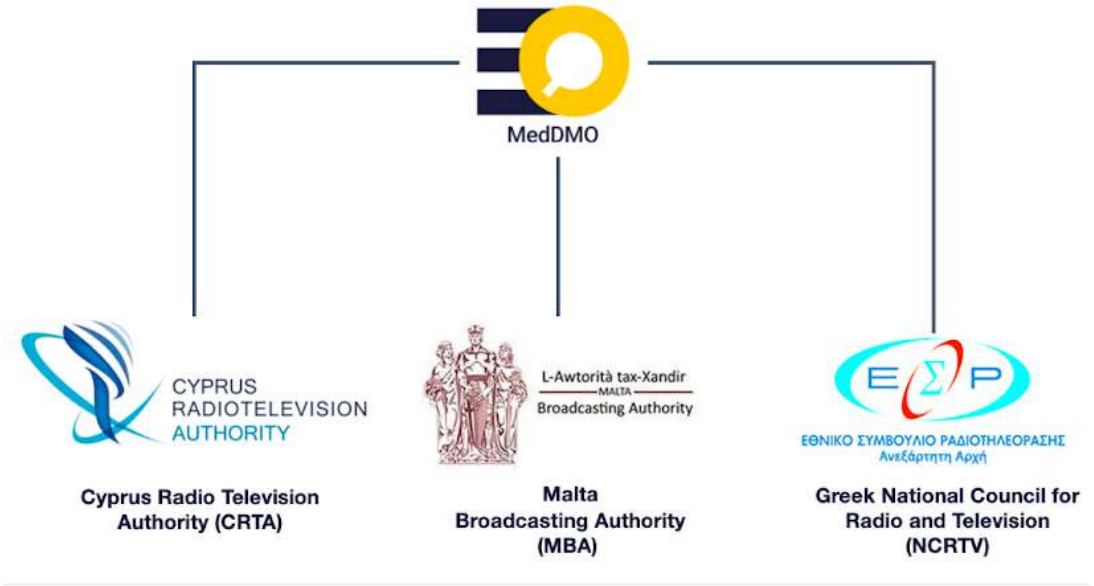



Figure 29: MedDMO collaboration with the national media authorities

**Key Areas of Collaboration Between National Media Authorities and MedDMO:**

The following initiatives illustrate the ongoing support and collaborative efforts with each of the three authorities in 2024:

**The Case of Cyprus**

Towards supporting Cyprus Radio Television Authority[75]:

- Cyprus Radio Television Authority (CRTA) members visited CUT and participated in a Fact-checking process Seminar (05/07/2024). A discussion followed for further collaboration between MedDMO and CRTA, mainly in media literacy activities.

[75] https://crta.org.cy/en/

- CUT visited CRTA premises (29/11/2024) and conducted a Fact-checking workshop for the staff of CRTA. This activity was a follow up activity after the CRTA visit to CUT, they requested to repeat the workshop for all the relevant staff of CRTA in their premises for their convenience to participate. The workshop ended with a Q&A session. The participants were very engaged in the workshop and the case studies of specific false claims spreaded in Cyprus. CRTA expressed their interest in co-organizing similar workshops for professional journalists in Cyprus.

- CRTA suggested the creation of a disinformation awareness video that could be disseminated widely through exploiting the authority's channels in TV and Radio productions in Cyprus. MedDMO partners are working towards the creation of a series of videos.

**The Case of Greece**

Towards supporting the National Council for Radio and Television (NCRTV)[76]:

- On October 18, 2024, the National Council for Radio and Television (ESR), the Journalists' Union of Athens Daily Newspapers (ESIEA), and MedDMO successfully hosted a seminar on disinformation and modern propaganda at ESIEA in Athens. Experts discussed the challenges of combating fake news, verification tools, and the role of journalism in promoting media literacy. The event attracted journalists, media professionals, and students, concluding with an open discussion on strengthening fact-checking and transparency in the digital era. The president of the CRTA board participated in the event co-organized by MedDMO and NCRTV, showcasing the strong connection between the two countries and their broadcasting authorities. Additionally, members of the CRTA were able to remotely attend the event.

**The Case of Malta**

Towards supporting the Broadcasting Authority of Malta (MBA)[77]:

- Organization of the “Webinar for Journalists in Malta” (19/06/2024). Members from the MedDMO Fact-checking organizations (ToM, AFP, EH, FCC) discussed the fact-checking procedure for investigating interesting misinformation cases. Our coordinator Nikos Sarris introduced MedDMO to the Maltese Journalists.

In addition to our direct collaboration with each regulatory authority, we actively share MedDMO project outcomes—such as reports, policy findings, and educational materials—that are relevant to their work. We remain open to requests for support and aim to respond to their evolving needs. At the same time, we value the information and insights they share with us, which contribute to our reporting and analysis. We also regularly invite them to MedDMO-organised events and other initiatives we believe align with their mandate, fostering a continuous and mutually beneficial exchange.

[76] https://www.esr.gr/information/
[77] https://ba.org.mt/

**Collaboration with the Digital Service Coordinators in Cyprus, Greece and Malta**

Under the Digital Services Act (DSA), each EU Member State appoints a Digital Service Coordinator (DSC)[78] tasked with ensuring compliance, coordinating enforcement, and fostering a safer online environment. In this context, MedDMO has established direct communication with the DSC in Cyprus (CRTA) and recently reached out to the DSCs of Malta and Greece for organising an introductory meeting. The corresponding DSC in Malta is the Malta Communications Authority (MCA)[79] and in Greece is the Hellenic Telecommunications and Post Commission (EETT)[80]. During this session, the DSCs were introduced to MedDMO and EDMO initiatives, and possibilities for further collaboration were discussed. As a direct outcome, the Maltese DSC (MCA) invited a MedDMO representative to participate in an upcoming event, further strengthening these cooperative efforts.

# 6 Conclusion

This report highlights that while Meta, Google, and TikTok have taken important steps to counter disinformation in Cyprus, Greece, and Malta, their efforts remain uneven and, in many cases, insufficiently tailored to local needs. Key challenges include limited support for the Maltese language, weak integration of local fact-checkers, underutilized research tools, and a general lack of transparency and impact metrics across all three countries.

Despite these shortcomings, the platforms' continued participation in the Code of Practice represents a valuable framework for accountability and improvement. The availability of tools, partnerships with fact-checkers, and emerging transparency mechanisms offer a foundation on which to build.

Looking ahead, there is clear potential for progress. By enhancing localization, ensuring equitable access to tools and data, and strengthening collaboration with national experts, platforms can significantly improve their disinformation responses in the region.

As part of the MedDMO project, we will continue to monitor the practices of these platforms and offer evidence-based recommendations. Our goal remains to support media experts and national authorities in Cyprus, Greece, and Malta in building stronger, more resilient information ecosystems.

[78] https://digital-strategy.ec.europa.eu/en/policies/dsa-dscs
[79] https://www.mca.org.mt/initiatives/dsa
[80] https://www.eett.gr/

# 7 References

**[European Commission, 2022]** European Commission (2022). Strengthened Code of Practice on Disinformation. https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

**[European Commission, 2018]** European Commission (2018). Code of Practice on Disinformation. https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation

**[BBC News, 2025a]** The TikTokers accused of triggering an election scandal, 30 April 2025. https://www.bbc.com/articles/cqx41x3gn5zo

**[BBC News, 2025b]** Facebook and Instagram get rid of fact checkers, 7 January 2025. https://www.bbc.com/news/articles/cly74mpy8klo

**[Kaplan, 2025]** Kaplan, J. (2025, January 7). More speech and fewer mistakes. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

**[Global Witness, 2025**] TikTok and X recommend pro-AfD content to non-partisan users ahead of the German elections, 20 February 2025. https://globalwitness.org/en/campaigns/digital-threats/tiktok-and-x-recommend-pro-afd-content-to-non-partisan-users-ahead-of-the-german-elections/

**[European Commission, 2023]** European Commision,. Flash Eurobarometer: Media & News Survey 2023. https://europa.eu/eurobarometer/surveys/detail/3153

**[Meta, 2024]** Meta. Code of Practice on Disinformation – Report No. 4 (September 2024). https://disinfocode.eu/reports?signatory=&type=&year=2024

**[Google, 2024]** Google. Code of Practice on Disinformation – Report No. 4 (September 2024). https://disinfocode.eu/reports?signatory=&type=&year=2024

**[TikTok, 2024]** TikTok. Code of Practice on Disinformation – Report No. 4 (September 2024). https://disinfocode.eu/reports?signatory=&type=&year=2024

**[Leonidou et al., 2023]** Pantelitsa Leonidou, Nikos Salamanos, Michael Sirivianos, Loukia Taxitari, Socrates Ioakeim, Andreas Phylactou (2023), 1st Report of platform practices and national authorities support 2023 (for Cyprus, Greece & Malta). https://meddmo.eu/wp-content/uploads/2024/05/FINAL_1stReport_Monitoring_the_platform_p ractices_compressed.pdf

**[Leonidou et al., 2025]** Pantelitsa Leonidou, Nikos Salamanos, Paschalia Spyridou, Michael Sirivianos (2025). MedDMO Fact-Checks Archive Analysis for monitoring the media landscapes in Cyprus, Greece, and Malta https://meddmo.eu/mapping-disinformation-circulation-in-cyprus-greece-and-malta-and-evaluating-platforms-moderation-practices-insights-from-meddmo-fact-checks-archive-january-june-2024/

**[Botan et al., 2025**] Madalina Botan, Trisha Meyer (2025). Implementing the EU Code of Practice on Disinformation: An Evaluation of VLOPSE Compliance and Effectiveness (Jan–Jun 2024). https://edmo.eu/wp-content/uploads/2025/06/EDMO-Report-%E2%80%93-Implementing-the-EU-Code-of-Practice-on-Disinformation.pdf

**[Irrational Labs, 2021]** Irrational Labs (2021). How Behavioural Science Reduced the Spread of Misinformation on TikTok. https://irrationallabs.com/case-studies/tiktok-how-behavioral-science-reduced-the-spread-of-misinformation/

**[Hassoun et al., 2024]** Amelia Hassoun, Ariel Abonizio, Beth Goldberg, Katy Osborn, Cameron Wu (2024). The Influencer Next Door: How Misinformation Creators Use GenAI. 10.48550/arXiv.2405.13554.

**[Maertens et al., 2025]** Rakoen Maertens, Jon Roozenbeek, Jon Simons, Stephan Lewandowsky, Vanessa Maturo, Beth Goldberg, Rachel Xu, and Sander van der Linden (2025). Psychological Booster Shots Targeting Memory Increase Long-term Resistance Against Misinformation. PsyArXiv. March 11. doi:10.31234/osf.io/6r9as_v3.

**[Dufour et al., 2024]** Nicholas Dufour, Arkanath Pathak, Pouya Samangouei, Nikki Hariri, Shashi Deshetti, Andrew Dudfield, Christopher Guess, Pablo Hernández Escayola, Bobby Tran, Mevan Babakar, Christoph Bregler (2024). AMMeBa: A Large-Scale Survey and Dataset of Media-Based Misinformation In-The-Wild. https://arxiv.org/abs/2405.11697

# Annex I: CoP Signatories Reports Assessment Scores and Missing SLIs.

Table 14: Meta's CoP Report No.4 Summary of Assessment Results

**Meta**

Scores: 1 (“very poor”), 2 (“poor”), 3 (“fair”), 4 (good), 5 (excellent), n/a (“not applicable”)

| | Evaluation of Reported Actions/Policies (QREs) | Evaluation of Quantitative data (SLIs) & Implementation | | |
|---|---|---|---|---|
| | | Greece | Cyprus | Malta |
| **Pillar-V: Empowering Users** | | | | |
| **Commitment 17** | 3 | 2 | 2 | 2 |
| **Commitment 21** | 3 | 2 | 2 | 1 |
| **Average** | **3** | **2** | **2** | **1.5** |
| **Pillar-VI: Empowering the Research Community** | | | | |
| **Commitment 26** | **3** | **2** | **2** | **2** |
| **Commitment 27** | **3** | **3** | **3** | **3** |
| **Commitment 28** | **2** | **2** | **2** | **2** |
| **Average** | **2.7** | **2.3** | **2.3** | **2.3** |
| **Pillar-VII: Empowering the Fact-Check Community** | | | | |
| **Commitment 30** | 3 | 2 | 2 | 1 |
| **Commitment 31** | 3 | 3 | 3 | 2 |
| **Commitment 32** | 2 | 2 | 2 | 2 |
| **Average** | **2.7** | **2.3** | **2.3** | **1.7** |

Table 15: Google's CoP Report No.4 Summary of Assessment Results

**Google**

Scores: 1 (“very poor”), 2 (“poor”), 3 (“fair”), 4 (good), 5 (excellent), n/a (“not applicable”)

| | Evaluation of Reported Actions/Policies (QREs) | Evaluation of Quantitative data (SLIs) & implementation | | |
|---|---|---|---|---|
| | | Greece | Cyprus | Malta |
| **Pillar-V: Empowering Users** | | | | |
| **Commitment 17** | 4 | 3 | 3 | 2 |

| | | | | |
|---|---|---|---|---|
| **Commitment 21** | 3 | 3 | 3 | 2 |
| **Average** | **3.5** | **3** | **3** | **2** |
| **Pillar-VI: Empowering the Research Community** | | | | |
| **Commitment 26** | **4** | **4** | **4** | **3** |
| **Commitment 27** | **1** | **1** | **1** | **1** |
| **Commitment 28** | **2** | **2** | **2** | **2** |
| **Average** | **2.3** | **2.3** | **2.3** | **2** |
| **Pillar-VII: Empowering the Fact-Check Community** | | | | |
| **Commitment 30** | 3 | 3 | 3 | 3 |
| **Commitment 31** | 3 | 3 | 3 | 3 |
| **Commitment 32** | 2 | 2 | 2 | 2 |
| **Average** | **2.6** | **2.7** | **2.7** | **2.7** |

Table 16: TikTok's CoP Report No.4 Summary of Assessment Results

| **TikTok**<br>Scores: 1 (“very poor”), 2 (“poor”), 3 (“fair”), 4 (“good”), 5 (“excellent”), n/a (“not applicable”) | | | | |
|---|---|---|---|---|
| | **Evaluation of Reported Actions/Policies (QREs)** | **Evaluation of Quantitative data (SLIs) & implementation** | | |
| | | **Greece** | **Cyprus** | **Malta** |
| **Pillar-V: Empowering Users** | | | | |
| **Commitment 17** | 3 | 3 | 3 | 2 |
| **Commitment 21** | 3 | 3 | 3 | 2 |
| **Average** | **3** | **3** | **3** | 2 |
| **Pillar-VI: Empowering the Research Community** | | | | |
| **Commitment 26** | 3 | 3 | 3 | 3 |
| **Commitment 27** | 2 | 2 | 2 | 2 |
| **Commitment 28** | 3 | 3 | 3 | 3 |
| **Average** | 2.7 | 2.7 | 2.7 | 2.7 |
| **Pillar-VII: Empowering the Fact-Check Community** | | | | |
| **Commitment 30** | 3 | 3 | 3 | 2 |
| **Commitment 31** | 2 | 2 | 2 | 1 |
| **Commitment 32** | 2 | 2 | 2 | 2 |
| **Average** | **2.3** | **2.3** | **2.3** | **1.6** |

Table 17: Missing Service Level Indicators (SLIs) information in Signatories Reports

**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** | **Meta Instagram** | **Google Search** | **Google YouTube** | **TikTok** |
|---|---|---|---|---|---|---|
| **17.1** | **17.1.1** | **x** | **x** | ✓ | ✓ | ✓ |

| | | | | | | |
|---|---|---|---|---|---|---|
| **17.2** | **17.2.1** | ✓ | ✓ | **x** | ✓ | ✓ |
| **21.1** | **21.1.1** | ✓ **mentioned as same with SLI 21.1.2** | ✓ **mentioned as same with SLI 21.1.2** | ✓ | **x** | ✓ |
| | **21.1.2** | ✓ | ✓ | **N/A** | **N/A** | ✓ |
| **26.1** | **26.1.1** | ✓ | ✓ | ✓ | ✓ | **x** |
| **26.2** | **26.2.1** | | | | | ✓ |
| **27.3** | **27.3.1** | N/A | N/A | **x** | **x** | N/A |
| **30.1** | **30.1.1.** | ✓ **(reference to QRE 30.1.2)** | ✓ **(reference to QRE 30.1.2)** | **N/A** | **N/A** | ✓ |
| **31.1** | **31.1.1** | ✓ | ✓ | ✓ **(reference to SLI 21.1.1)** | **x** | ✓ |
| | **31.1.2** | ✓ | ✓ | **N/As** | **N/A** | ✓ |
| | **31.1.3** | ✓ | ✓ | **x** | **x** | ✓ |
| **32.1** | **32.1.1** | ✓ **(reference to QRE 30.1.2)** | ✓ **(reference to QRE 30.1.2)** | **N/A** | **x** | N/A |
| **Total Missing SLIs information** | | **1** | **1** | **3** | **5** | **1** |

# Annex II: Questionnaire for MedDMO fact-checking organizations partners

**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

4. 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.

5. 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.
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 15.

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

7. 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)

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

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

10. 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:.....

11. Did you 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.''***

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

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

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

15. Please also elaborate on your collaborations with other platforms (i.e., Google, TikTok)



# 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. The content of this document is © the author(s).
For further information, visit www.meddmo.eu