Code of Practice on Disinformation – Report of AI Forensics for the period 1 January–30 June 2025

Table of Content

Executive summary

Executive summary (max. 2 pages)

As the Code evolves and Signatories strengthen their collaboration within a shared framework, AI Forensics remains committed to its two core areas: algorithmic auditing and active participation in key working groups. In 2025, as the Code of Practice transitions to a Code of Conduct, we continue our engagement in the Generative AI and Elections Monitoring subgroups within the Crisis Response framework.

During the first half of 2025, we continued our research on the impact of emerging technologies on electoral integrity, we also looked into the algorithmic dynamics during the Polish Presidential elections.

We look forward to further collaboration with other Signatories, the European Commission, ERGA, and EDMO, reinforcing accountability and transparency in the digital ecosystem.

Guidelines for filling out the report

Reports are detailing how signatories have implemented their Commitments under the Code and signatories commit to provide regular reporting on Service Level Indicators (SLIs) and Qualitative Reporting Elements (QREs). The reports and data provided should allow for a thorough assessment of the extent of the implementation of the Code's Commitments and Measures by each signatory.

Reporting period

The reporting period to be covered in the reports is 6 months for signatories who are not offering very large online platform services. Signatories shall submit reports outlining policy updates and actions taken to implement the Commitments and Measures they signed up to under the Code. All data and policy updates should be reported for 12 months period from the submission of last reports.

Adjusting the reporting template

Signatories who are not offering very large online platform services can adapt the template to specific commitments and measures they subscribed to. This may include adapted wording for commitments, measures, QREs and SLIs. Relevant signatories will report only on commitments and measures they subscribed to and provide Member State-level data only if feasible.

Reporting per Service

When filling in a report for several services, use colour codes to clearly distinguish between services. At the beginning of the report, clarify what colour is used for which service.

Reporting in text form

Reporting in the form of written text is required for several parts of the report. Most of them are accompanied by a target character limit. Please stick to the target character limit as much as possible. We encourage you to use bullet points and short sentences. When providing information to the QRE, please make sure that your answer covers all the elements of the associated commitment and measure. Links should only be used to provide examples or to illustrate the point. They should not be used to replace explanations or to provide data in the forms. All relevant explanations and data must be included in the report directly, in written form.

Reporting SLIs and data

Reporting on SLIs requires quantitative information to be reported on in this harmonised reporting template.

- Where relevant and feasible, SLIs should be reported on per Member State.
- If no data is available on Member State level, SLIs might, instead, be exceptionally reported on per language. (NB that signatories agreed to revisit this issue after the first reporting, to ensure harmonised and meaningful reporting.)
- Please report data in the format provided by the harmonised reporting template, not through external links. Please use the Member State/language template provided in the harmonised reporting template. Where the table asks for "Other relevant metrics", please name the metric that you would like to report on in addition to the ones already provided. You may include more than the number of additional fields provided where necessary; in that case, please adjust the table as needed.
- Please contextualize all data as much as possible, i.e. include baseline quantitative information that will help contextualize the SLIs (e.g. number of pieces of content labelled out of what volume of content).
- If there are no relevant metrics to report on, please leave the respective columns blank.

Reporting on TTPs

If subscribed to Commitment 14, Integrity of Services, we ask you to report on each identified TTP individually. The number of identified TTPs may vary per service. Where more than one TTP are reported under the same action, clarify the reasoning in the methodology. Where input is not provided, keep the placeholder for the relevant TTP and explain reasons and planned remedial action. Additionally, as with all other SLIs, data can be provided per Member State for each individual TTP.

Missing Data

In case that at the time of reporting there is no data available yet, the data is insufficient, or the methodology is lacking, please outline in the dedicated field (i.e. in the field about further implementation measures planned) how this will be addressed over the upcoming six months, being as specific as possible.

Signatories are encouraged to provide insights about the data/numbers they provide by inserting possible explanations in the boxes of the template "*Methodology of data measurement & insights on data provided*". This should aim to explain the why of what is being reported, for instance - *Are there trends or curiosities that could*

require or use contextual explanation? What may be driving the change or the difference in the number? Please also indicate inconsistencies or gaps regarding methodology in the dedicated box.

Attachments

We ask you not to enclose any additional attachments to the harmonised reporting template.

Crisis and elections reporting template

Relevant signatories are asked to provide proportionate and appropriate information and data during a period of crisis and during an election. Reporting is a part of a special chapter at the end of the harmonised reporting template and should follow the guidelines:

- The reporting of signatories' actions should be as specific to the particular crisis or election reported on as possible. To this extent, the rows on "Specific Action[s]" should be filled in with actions that are either put in place specifically for a particular event (for example a media literacy campaign on disinformation related to the Ukraine war, an information panel for the elections), or to explain in more detail how an action that forms part of the service's general approach to implementing the Code is implemented in the specific context of the crisis or election reported on (for example, what types of narratives in a particular election/crisis would fall into scope of a particular policy of the service, what forms of advertising are ineligible).
- Regarding elections, signatories are expected to provide specific information on their **experience with the RRS for FR and RO elections**. This can be included in the first two rows ("Threats observed…" / "Mitigations in place …"). In addition, **regardless of the RRS activation, signatories should report on relevant actions in place for elections at national level** (parliamentary/presidential) in EU Member States during the reporting period – specifying the country(ies) and election(s).
- Signatories who are not offering very large online platform services and who follow the invitation to report on their specific actions for a particular election or crisis may adapt the reporting template as follows:
    - They may remove the "Policies and Terms and Conditions" section of the template, or use it to report on any important changes in their internal rules applicable to a particular election or crisis (for example, a change in editorial guidelines for fact-checkers specific to the particular election or crisis)
    - They may remove any Chapter Section of the Reporting Template (Scrutiny of Ads Placement, Political Advertising, Integrity of Services etc.) that is not relevant to their activities
- The harmonised reporting template should be filled in by adding additional rows for each item reported on. This means that rather than combined/bulk reporting such as "Depending on severity of violation, we demote or remove content based on policies X, Y, Z", there should be individual rows stating for example "Under Policy X, content is demoted or removed based on severity", "Under Policy Y, content […]" etc.
- The rows should be colour-coded to indicate which service is being reported on, using the same colour code as for the overall harmonised reporting template.

Reporting should be brief and to the point, with a suggested character limit entry of 2000 characters.

Uploading data to the Transparency Centre

The reports should be submitted to the Commission in the form of the pdf via e-mail to the address CNECT COP TASK FORCE CNECT-COP-TASK-FORCE@ec.europa.eu within the agreed deadline. Signatories will upload all data from the harmonised reporting template to the Transparency Centre, allowing easy data access and filtering within the agreed deadline. It is the responsibility of the signatories to ensure that the uploading takes place and is executed on time. Signatories are also responsible to ensure that the Transparency Centre is operational and functional by the time of the reports' submission that the data from the reports are uploaded and made accessible in the Transparency Centre within the above deadline, and that users are able to read, search, filer and download data as needed in a user-friendly way and format.

V. Empowering Users

Commitment 17

In light of the European Commission's initiatives in the area of media literacy, including the new Digital Education Action Plan, Relevant Signatories commit to continue and strengthen their efforts in the area of media literacy and critical thinking, also with the aim to include vulnerable groups. [change wording if adapted]

Measure 17.1	Relevant Signatories will design and implement or continue to maintain tools to improve media literacy and critical thinking, for instance by empowering users with context on the content visible on services or with guidance on how to evaluate online content.				
QRE 17.1.1 [insert wording if adapted]	AI Forensics contributes to improve media literacy and critical thinking by publishing its work and disseminating it among media, stake-holders and decision makers. Furthermore, AI Forensics participates in conferences, meetings and events that serve as platforms to inform the larger public on the importance of algorithmic auditing, accountability and transparency.				
SLI 17.1.1 - actions enforcing policies above [change wording if adapted]	Methodology of data measurement [suggested character limit: 500 characters]				
	Total count of the tool's impressions	Interactions/ engagement with the tool	Other relevant metrics	Other relevant metrics	
Data					
Measure 17.2	Relevant Signatories will develop, promote and/or support or continue to run activities to improve media literacy and critical thinking such as campaigns to raise awareness about Disinformation, as well as the TTPs that are being used by malicious actors, among the general public across the European Union, also considering the involvement of vulnerable communities.				
QRE 17.2.1 [insert wording if adapted]	AI Forensics is dedicated to increasing critical thinking among users and helping them restore their self-agency. Our innovative data-driven methodology provides journalists, researchers, and policymakers with timely evidence of systematic violations of users' interests and digital rights, particularly for minority groups and communities that are often overlooked in the design of technology. We believe that consistent and coordinated scrutiny is the path to restoring the balance of power between big tech platforms and its users. Therefore, AI Forensics will continue producing research and investigations that are fulfilling this aim.				
SLI 17.2.1 - actions enforcing policies above [change wording if adapted]	Methodology of data measurement [suggested character limit: 500 characters]				
	Nr of media literacy/ awareness raising activities organised/ participated in	Reach of campaigns	Nr of participants	Nr of interactions with online assets	Nr of participants (etc)
Data					

Measure 17.3	For both of the above Measures, and in order to build on the expertise of media literacy experts in the design, implementation, and impact measurement of tools, relevant Signatories will partner or consult with media literacy experts in the EU, including for instance the Commission's Media Literacy Expert Group, ERGA's Media Literacy Action Group, EDMO, its country specific branches, or relevant Member State universities or organisations that have relevant expertise.
QRE 17.3.1 [insert wording if adapted]	Outline relevant actions [suggested character limit: 2000 characters]

VI. Empowering the research community	
Commitment 28 Relevant Signatories commit to support good faith research into Disinformation that involves their services. [change wording if adapted]	
Measure 28.1	Relevant Signatories will ensure they have the appropriate human resources in place in order to facilitate research, and should set-up and maintain an open dialogue with researchers to keep track of the types of data that are likely to be in demand for research and to help researchers find relevant contact points in their organisations.
QRE 28.1.1 [insert wording if adapted]	AI Forensics continues to strengthen its multidisciplinary, socio-technical research approach, with a dedicated team of 17 members. We maintain a collaborative model, working closely with civil society, media partners, and academic institutions to produce independent audits of platforms' systemic risks. In the first half of 2025, our research addressed a broad range of platform accountability gaps: • Meta's Failing Ad Moderation: Health Scams Targeting EU Users We published the first large-scale audit of Meta's advertising ecosystem under the DSA, revealing that the platform approved over 46,000 deceptive health-related ads—many involving deepfakes or impersonations—that reached 292 million impressions across the EU. This investigation underscored Meta's failures in compliance with DSA Article 34, including inadequate risk mitigation, weak advertiser verification, and insufficient transparency. • TikTok's Polish Election Labels: Only Sometimes, and Only for Some We examined TikTok's inconsistent labelling practices during Poland's 2025 Presidential Election, finding systemic shortcomings such as the exclusion of over 20 million diaspora users from election context labels and widespread failures to annotate posts spreading fraud allegations or featuring AI-generated imagery. The investigation contributed to regional debate on electoral integrity and was featured by Demagog Poland.

	• **TikTok's Research API: Problems Without Explanations** Our audit of TikTok's Research API revealed major transparency gaps, including missing metadata for high-profile content, persistent exclusions of creator accounts, and inaccessible ads. By testing 260,000 TikTok URLs via data donation and web scraping, we demonstrated that the API falls short of enabling meaningful oversight under the DSA. Findings were amplified by Tech Policy Press and civil society coalitions, and we launched a public monitoring dashboard to track inaccessible content in real time. • **When Personal Becomes Profitable: Sensitive Targeting on X** We investigated X's advertising system and found widespread targeting based on sensitive personal data categories, including political opinions, sexual orientation, religious beliefs, health conditions, and ethnic origin. Major advertisers such as TotalEnergies, Dell Technologies, and the Saudi Public Investment Fund used or excluded targeting terms tied to protected characteristics. As part of this work, we launched **HaveIBeenTargeted.online**, a public tool that enables users to see whether they may have been profiled by such ads. Together, these investigations illustrate our role in enabling independent scrutiny of platform infrastructures across elections, advertising, and transparency mechanisms. Our work featured in major outlets such as *Le Figaro* and *Le Monde*, and informed policy debates at the EU Digital Policy Summit in Gdańsk.
Measure 28.2	Relevant Signatories will be transparent on the data types they currently make available to researchers across Europe.
QRE 28.2.1 [insert wording if adapted]	AI Forensics releases their investigations and its methodology openly in their reports.
Measure 28.3	Relevant Signatories will not prohibit or discourage genuinely and demonstratively public interest good faith research into Disinformation on their platforms, and will not take adversarial action against researcher users or accounts that undertake or participate in good-faith research into Disinformation.
QRE 28.3.1 [insert wording if adapted]	AI Forensics works towards the goal of promoting good faith research.

VI. Empowering the research community	
Commitment 29 Relevant Signatories commit to conduct research based on transparent methodology and ethical standards, as well as to share datasets, research findings and methodologies with relevant audiences. [change wording if adapted]	
Measure 29.1	Relevant Signatories will use transparent methodologies and ethical standards to conduct research activities that track and analyse influence operations, and the spread of Disinformation. They will share datasets, research findings and methodologies with members of the Task-force including EDMO, ERGA, and other Signatories and ultimately with the broader public

QRE 29.1.1 [insert wording if adapted]	All the research produced by AI Forensics is publicly available for anyone to access.
Data	

VIII. Transparency Centre	
Commitment 34 To ensure transparency and accountability around the implementation of this Code, Relevant Signatories commit to set up and maintain a publicly available common Transparency Centre website. [change wording if adapted]	
Measure 34.1	
Measure 34.2	
Measure 34.3	
Measure 34.4	
Measure 34.5	

IX. Permanent Task-Force	
Commitment 37 Signatories commit to participate in the permanent Task-force. The Task-force includes the Signatories of the Code and representatives from EDMO and ERGA. It is chaired by the European Commission, and includes representatives of the European External Action Service (EEAS). The Task-force can also invite relevant experts as observers to support its work. Decisions of the Task-force are made by consensus. [change wording if adapted]	
Measure 37.1	
Measure 37.2	
Measure 37.3	
Measure 37.4	
Measure 37.5	
Measure 37.6	
QRE 37.6.1 [insert wording if adapted]	AI Forensics is an active participant in the Monitoring and Reporting Subgroup, in the AI Generative Subgroup as well as in the Crisis Response Subgroup (where we participate in both in the Elections Steering Committee).

X. Monitoring of Code	
Commitment 38 The Signatories commit to dedicate adequate financial and human resources and put in place appropriate internal processes to ensure the implementation of their commitments under the Code. [change wording if adapted]	
Measure 38.1	
QRE 38.1.1 [insert wording if adapted]	AI Forensics has two representatives, directly involved in the work of the Subgroup and Working group; we are part of ensuring full compliance with relevant Commitments taken under the Code.

X. Monitoring of Code
Commitment 39 Signatories commit to provide to the European Commission, within 1 month after the end of the implementation period (6 months after this Code's signature) the baseline reports as set out in the Preamble. [change wording if adapted]

X. Monitoring of Code	
Commitment 40 Signatories commit to provide regular reporting on Service Level Indicators (SLIs) and Qualitative Reporting Elements (QREs). The reports and data provided should allow for a thorough assessment of the extent of the implementation of the Code's Commitments and Measures by each Signatory, service and at Member State level. [change wording if adapted]	
Measure 40.1	
Measure 40.2	
Measure 40.3	
Measure 40.4	
Measure 40.5	
Measure 40.6	

X. Monitoring of Code

Commitment 42
Relevant Signatories commit to provide, in special situations like elections or crisis, upon request of the European Commission, proportionate and appropriate information and data, including ad-hoc specific reports and specific chapters within the regular monitoring, in accordance with the rapid response system established by the Taskforce. [change wording if adapted]

X. Monitoring of Code
Commitment 43
Relevant Signatories commit to provide, in special situations like elections or crisis, upon request of the European Commission, proportionate and appropriate information and data, including ad-hoc specific reports and specific chapters within the regular monitoring, in accordance with the rapid response system established by the Taskforce. [change wording if adapted]

Reporting on the service's response during a period of crisis

Reporting on the service's response during a crisis		
[Name of crisis]		
Threats observed or anticipated at time of reporting: [suggested character limit 2000 characters].		
Mitigations in place at time of reporting: [suggested character limit: 2000 characters].		
[Note: Signatories are requested to provide information relevant to their particular response to the threats and challenges they observed on their service(s). They ensure that the information below provides an accurate and complete report of their relevant actions. As operational responses to crisis/election situations can vary from service to service, an absence of information should not be considered a priori a shortfall in the way a particular service has responded. Impact metrics are accurate to the best of signatories' abilities to measure them].		
Policies and Terms and Conditions		
Outline any changes to your policies		
Policy	**Changes (such as newly introduced policies, edits, adaptation in scope or implementation)**	**Rationale**
Scrutiny of Ads Placements		
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.		
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention	

	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Political Advertising	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Integrity of Services	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention

	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Empowering Users	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Empowering the Research Community	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Empowering the Fact-Checking Community	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention

	Indication of impact (at beginning of action: expected impact) including relevant metrics when available

Reporting on the service's response during an election

Reporting on the signatory's response during an election

Polish Presidential elections

Threats observed during the electoral period: [suggested character limit 2000 characters].

AI Forensics has been actively involved in election monitoring in 2025, with the following report published as its outcome:

TikTok's Polish Elections Labels: Only Sometimes, And Only For Some
AI Forensics investigated TikTok's election information labeling practices during the 2025 Polish Presidential Elections. Our research revealed that TikTok failed to meet European Commission guidelines for safeguarding electoral processes. Labels designed to provide users with reliable election information were applied inconsistently, leaving significant gaps in coverage. Moreover, TikTok restricted label visibility to users physically located in Poland, excluding the Polish diaspora of more than 20 million people worldwide.

Main threats:

- **Inconsistent application of election labels**: The majority of relevant political content lacked the mandated labels.
- **Geographic discrimination**: Labels were only visible in Poland, excluding millions of diaspora voters abroad.
- **Amplification of fraud allegations**: 23 videos spreading claims of rigged elections reached 4.5+ million views, with nearly 80% missing the required election labels.
- **AI-generated disinformation**: Generative AI imagery appeared in multiple posts without disclosure or detection.
- **Undermining trust in elections**: Content calling for the cancellation of results or discouraging participation circulated widely, with little platform intervention.

Our findings highlight TikTok's inconsistent enforcement of its own policies and inadequate protection of democratic processes, particularly disadvantaging diaspora communities who rely heavily on online information sources.

TikTok still appears unable to effectively flag the majority of election-related content and content undermining the electoral process. In our dataset of 23 videos spreading allegations of fraud or discouraging participation, 78% were not labelled at all. In the five days before the second round of the elections, 57% of monitored election-related posts remained unlabelled.

While TikTok did demonstrate relatively high responsiveness to moderation — OSCE reported over 80% of flagged content was addressed — its proactive labelling mechanisms remain insufficient. According to both TikTok's own commitments on election integrity and the European Commission's guidance on mitigating systemic risks, the platform should:

1. **Expand coverage of election labels** so that the vast majority of election-related posts are annotated.

2. **Maintain responsiveness to user-flagged content**, while scaling proactive labelling and potential deplatforming of content that undermines electoral trust.

[Note: Signatories are requested to provide information relevant to their particular response to the threats and challenges they observed on their service(s). They ensure that the information below provides an accurate and complete report of their relevant actions. As operational responses to crisis/election situations can vary from service to service, an absence of information should not be considered a priori a shortfall in the way a particular service has responded. Impact metrics are accurate to the best of signatories' abilities to measure them].

Policies and Terms and Conditions

Outline any changes to your policies

Policy	**Changes (such as newly introduced policies, edits, adaptation in scope or implementation)**	**Rationale**
		Our findings underline that TikTok's election labels were only visible to users physically located in Poland, omitting the estimated 20 million Polish diaspora and citizens voting abroad. Election labelling should not be IP-dependent but universally applied across all locations. Additionally, across 23 TikTok posts in our sample, four employed **generative AI imagery**, yet none were labelled as such — despite TikTok's terms requiring users to disclose AI-generated content and TikTok's stated capacity to automatically detect it. This gap demonstrates that existing rules on synthetic media disclosure are not effectively implemented in practice. As highlighted in our 2024 investigation into AI-generated imagery in French political campaigns, the use of synthetic content in election contexts is a growing phenomenon. Platforms must adopt stricter rules and ensure consistent enforcement to safeguard electoral integrity.

Scrutiny of Ads Placements

Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.

Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact including relevant metrics when available
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact including relevant metrics when available
Political Advertising	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	
	Indication of impact including relevant metrics when available
Integrity of Services	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	

Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention The inconsistencies identified — IP-based labelling exclusions, poor coverage of election-related posts, and undetected generative AI — represent systemic service-level risks. Platforms must ensure equal protections across geographies, enforce synthetic media labelling, and scale proactive safeguards in line with EU standards for electoral processes.
	Indication of impact including relevant metrics when available
Empowering Users	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact including relevant metrics when available
Empowering the Research Community	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention
	Indication of impact including relevant metrics when available
Empowering the Fact-Checking Community	

Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention Our report on TikTok´s "Others Searched for" Feature suggests several solutions to address the threats: **Fact-Checking and Flagging of Sensitive Content**: Implementing robust fact-checking mechanisms that flag potentially misleading or biased search suggestions would help young users navigate political content more responsibly.
	Indication of impact including relevant metrics when available