JUN. 26

INFORMATIVE NOTE

TECHNOLOGY, MEDIA AND TELECOMMUNICATIONS

The transparency of AI-generated content

Code of Conduct

I. Background

On 10 June 2026, the AI Office published the Code of Conduct on Transparency of AI-Generated Content. This was done to further elaborating on Article 50 of Regulation (EU) 2024/1689 (the AI Act). This article concerns the transparency obligations for providers and deployers of AI systems that generate or manipulate content. The Code is currently undergoing a fitness check by the European Commission and the AI Committee. It is expected to be supplemented by Commission guidelines on the scope of the transparency obligations set out in Article 50 of the AI Act. A public consultation on this topic ran until early July.

Advances in generative AI enable the creation of content that is so realistic it is difficult to identify as AI-generated. This raises concerns about the potential for misleading the public and impacting public opinion and citizens' ability to make informed decisions. To this end, paragraphs 4 and 5 of Article 50 of the AI Act require a statement to be made 'that the content has been artificially generated or manipulated'.

The Code may be adopted to support compliance with the AI Act. It provides measures to assist providers in flagging and detecting content created or modified by AI and deployers, users and those responsible labelling such content.

Adherence to this Code does not constitute proof of compliance with the obligations set out in Article 50 of the AI Act. Providers and deployers who decide to comply with these obligations by other means must demonstrate that their measures are adequate; this will be assessed by national market surveillance authorities.

Pedro Lomba
Mafalda Sequeira Roldão
Technology, Media and Telecommunications team

Advances in generative AI enable the creation of content that is so realistic it is difficult to identify as AI-generated.

II. Measures set out in the Code

1. Labelling and detection of content created or manipulated by AI (providers)

In accordance with Article 50(2) of the AI Act, providers must use machine-readable means to detect and label content created or modified by AI. Providers are defined as entities that develop and market generative AI systems capable of producing audio, images, video or text. The Code encourages organisations to develop AI labelling or detection solutions, even if they are not directly responsible for the labelling.

Signatories to the Code may adopt various types of labelling, such as:

- **Digitally signed and time-stamped metadata, which is secure and tamper-resistant.** The inclusion of other metadata is also recommended, provided it does not contain personal or commercial data.
- **An imperceptible watermark that is difficult to remove from the content.** It is recommended that the model incorporate watermarking.
- **Fingerprinting** and **logging** are presented as complementary measures. Fingerprinting is more suitable for audio and visual content, while logging is more suitable for text. On their own, they are not sufficient to meet the requirements of Article 50 (effectiveness, interoperability, robustness and reliability).

Using a single type of marking is considered proportionate and sufficient where: (i) a generative AI system is embedded in physical products in a technically controlled, closed environment of a predominantly educational nature; and (ii) effective technical measures are in place to prevent outputs from leaving the product's environment. Using a single type of marking is also considered proportionate and sufficient where free-form text is concerned, as it is not capable of carrying metadata.[1] In all other cases, multi-layer labelling is proposed to ensure that the outputs of the systems include at least two types of machine-readable labelling.

Signatories are further recommended to:

- Retain and not alter metadata tags created by other AI systems when such content is used as input and subsequently transformed into output by their own AI systems;
- Promote a ban on the intentional removal or tampering with metadata markings by users through acceptable use policies, terms and conditions, or other documentation accompanying the AI system.
- Where technically feasible, provide additional information on the origin of content throughout workflows, in particular.

  i) The name of the AI system

  ii) The name of the service provider

1 The Code recommends using watermarking for free-form text of more than 200 characters. It also suggests restricting access to the relevant detection tools to verified expert users.

iii) The timestamp of the creation or manipulation of the content

iv) The identifier and version of the underlying model

v) A record of the type of operation (e.g. removal of objects). If there are multiple operations, these should be identified within a single metadata tag

- A feature should be provided that allows a visible label to be applied directly when the output is generated. This will make it easier for deployers to comply with their obligation to identify deep fakes[2] and text created or manipulated by AI (Article 50(4) of the AI Act).[3]

- Promote AI literacy among staff in relevant roles to ensure compliance with Article 50(2) and (5) and Article 4 of the AI Act.

- Maintain a documented compliance process that describes how the various measures have been implemented to ensure compliance with Articles 5(2) and (5) of the AI Act, at a general level. This measure will be applied proportionately, taking into account the size and resources of the signatory, particularly in the case of SMEs and start-ups.

- Provide users with a tool enabling them to detect whether content has been created or manipulated by AI, regardless of the labelling type, free of charge as a rule. It is also a requirement to communicate detection results in a clear, comprehensible and accessible manner to the target audience.

  i) The Code allows for detection results to be exported with a digital signature including a hash, identifier and timestamp. Access to less reliable mechanisms (e.g. watermark detection in free-form text) and forensic detection mechanisms for unlabelled content (e.g. removed labels) will be restricted to specialists.

The Code allows for detection results to be exported with a digital signature.

The technical solutions used to detect content created or manipulated by AI must be:

- **Effective** - Solutions will be considered effective if individuals are able to access and understand the meaning of the detection results. As there is no specific quantitative metric, a user-centred assessment is required.

- **Reliable** - Reliability refers to the ability to correctly identify content that has been generated or manipulated by AI. The appropriate metrics (e.g. detection error rate) must be used and low error rates must be demonstrated across varied samples covering both in-house and third-party content.

- **Robust** - Solutions must maintain performance levels under different conditions, such as screenshots, resizing, rotation, translation, blurring of faces, removal of markings, copying or modification, and attempts to mask the origin of the content. Robustness will be assessed using the same metrics as for reliability.

2 Definition 60 of the AI Act: 'deep fake' means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful.

3 This measure does not affect the liability of deployers, who are still required to clearly and conspicuously label the content in question in accordance with Article 50(4) and (5) of the AI Act.

- **Interoperable** - Solutions must function seamlessly across systems, suppliers and contexts to enable detection regardless of the technique used. In the absence of established standards, a phased implementation approach will be adopted, with interoperable watermarking solutions expected **by 2 February 2027**.

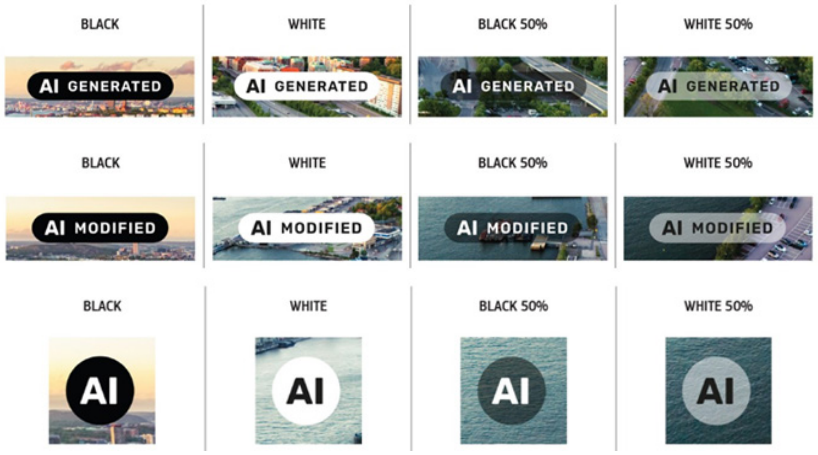
Prior to placing a generative AI system on the market or making it available, signatories may also use alternative marking and detection techniques. However, they must demonstrate compliance to the competent authorities based on recognised assessment methods and benchmarks. Recognised methods and benchmarks are to be established, particularly those to be approved by the AI Office. Until then, signatories must test and report on the performance of their solutions using internal benchmarks and industry best practice. They may also involve independent experts in the testing process, particularly for assessing resilience to attacks (red teaming), or make use of AI regulatory testing environments (regulatory sandboxes) under Article 57 of the AI Act.

2. Labelling of deepfakes and text created or manipulated by AI (deployers)

Under Article 50(4) and (5) of the AI Act, deployers of generative AI systems are required to label any content created or manipulated using AI. The Code provides guidelines on: i) image, audio or video content that constitutes a **deepfake**; and ii) **text intended to inform the public on matters of public interest** that has been created or manipulated using AI **without undergoing human review or editorial control**.

The Code encourages compliance with the following measures:

- Consistently and effectively disclosing the artificial origin of deepfakes or text using EU iconography or an equivalent label that complies with the design specifications.



- Display the acronym 'AI' visibly and directly within the content for a sufficient duration to be noticed, preferably accompanied by information on whether it was generated or modified using AI.
- Use an audio disclaimer at the start of the content and repeat it periodically throughout if visual disclosure is not possible.
- Ensure accessibility for all, particularly people with special needs, in accordance with the two applicable pieces of EU legislation. The first is Directive (EU) 2019/882 on accessibility requirements for products and services, implemented in Portugal by Decree-Law 82/2022 of 6 December. The second is Directive (EU) 2016/2102 on accessibility requirements for the websites and mobile applications of public bodies, implemented in Portugal by Decree-Law 83/2018 of 19 October. Accessibility for all must be ensured using audio disclaimers, tactile solutions or alternatives to visual content, high-contrast iconography, and content that can be detected by assistive technologies.
- Contribute to the creation of a task force to develop EU iconography by:

  i) improving design and usability

  ii) creating interactive solutions

  iii) harmonising dissemination practices

  iv) sharing sector-specific best practices

- Put in place appropriate internal processes, including a documented compliance process. Keep documentation ready to share with the relevant authorities, including examples of the solutions adopted and how they are disclosed.
- Establish mechanisms for reporting labelling errors and swiftly rectifying non-compliance.
- Train staff in relevant roles on this obligation and the implementation of designs and specific positioning requirements.
- Implement internal policies to ensure human review of content and public identification of the entity responsible for editorial content. Media service providers who are signatories under Article 2(2) of Regulation (EU) 2024/1083, and who are already subject to editorial obligations, may use the exception in the second subparagraph of Article 50(4). They apply their existing procedures for review, editorial control and professional standards to do this.

Implement internal policies to ensure human review of content and public identification of the entity responsible for editorial content.

PL
MJ

III. Conclusion and next steps

The Code of Conduct is a significant step towards implementing the AI Act. It is expected to guide the standardisation of market supervision by competent authorities across the EU. The Code promotes the consistent, practical and proportionate application of the transparency obligations set out in the AI Act. However, it does not replace the Act itself or the Commission's guidelines on Article 50. Instead, it provides signatories with a practical, EU-wide recognised framework for demonstrating compliance with these obligations.

The measures set out in the Code are voluntary. However, the underlying obligations (Article 50 of the AI Act) will take effect on 2 August 2026.

In view of the above, it is recommended that providers undertake the following:

- Adoption of technical solutions (metadata, watermarking);
- Review of acceptable use policies/terms and conditions to prevent the removal of labels both internally and by users;
- Preparation of evidence of compliance (audit trail).

Deployers should:

- Label content with EU iconography when publishing deepfakes or texts intended to inform the public on matters of public interest that have been created or altered by AI systems.
- Adapt the icons to comply with current accessibility rules.
- Ensure that internal editorial policies, where they exist, comply with these measures.
- Implement mechanisms for reporting labelling errors and for rapid rectification in cases of non-compliance.
- Preparation of evidence of compliance (audit trail). ■