WITNESS
SEE IT FILM IT CHANGE IT

Art. 50 - AI Act

Section 1. Questions in relation to Article 50(1) AI Act

Article 50(1) AI Act targets providers of interactive AI systems, notably systems that are intended to interact directly with natural persons. Providers should ensure that such systems are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system.
Recital 132 AI Act clarifies that when implementing the transparency obligation for interactive AI systems, the characteristics of natural persons belonging to vulnerable groups due to their age or disability should be taken into account to the extent the AI system is intended to interact with those groups. Article 50(5) AI Act furthermore requires that the information shall be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure. In addition, such information shall conform to the applicable accessibility requirements. Regarding the latter, recital 132 confirms that such information and notifications should be provided in accessible formats for persons with disabilities.
Article 50(1) AI Act exempts providers from this obligation if the interaction with the AI system can be considered obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use

Question 4. Are there aspects related to the scope or practical implementation of the transparency obligation for interactive AI systems under Article 50(1) for which you would seek further clarification?
(x) Yes
( ) No
Please, specify. 500 character(s) maximum
Further clarification is needed on the scope of “interactive” systems, especially for hybrid tools embedded in larger platforms or partially automated services. Guidance is also required on the form, timing, and visibility of disclosure to balance transparency with usability. Clear criteria on exemptions, proportionality of obligations, and consistency of enforcement across sectors and jurisdictions would help ensure practical, effective, and harmonized implementation. Any implementation of these disclosures must be privacy preserving.

Section 2. Questions in relation to Article 50(2) AI Act

Article 50(2) AI Act targets providers of AI systems, including general-purpose AI systems, capable of generating synthetic text, audio, image, and video content. Providers of such systems are required to employ technical solutions to ensure that the outputs of their systems are marked in a machine-readable format and enable detection that the content has been generated or manipulated by an AI system and not a human (see also recital 133).
AI systems that perform an assistive function for standard editing or that do not substantially alter the input data provided by the deployer or the semantics thereof are exempt from this obligation.

Furthermore, Article 50(2) AI Act does not apply if the generative AI system is authorised by law to detect, prevent, investigate, or prosecute criminal offences.

Question 5. Please provide practical examples of AI systems that generate synthetic text, audio, image, or video content as well as examples of systems for which there is doubt and you would seek clarification or consider them out of scope.

If you are aware of any AI systems that may fall under one or more of the exceptions of Article 50(2), such as AI systems that perform an assistive function for standard editing or that do not substantially alter the input data or the semantics thereof, or systems that can be authorised by law for law enforcement purposes, please include them in your response.

	Name and description of the system	**Is the AI system generating or manipulating synthetic audio, image, video or text content?**	**Motivate your answer, including whether and why the content should be considered synthetic**	**Does the AI system fall within one or more of the exceptions of Article 50(2)?**	**Motivate your answer and provide practical examples(s). For the law enforcement exemption provide the law that can authorise the use and describe if it includes any appropriate safeguards**
	Name/description	Select Yes No Not sure	Explain 500 character(s) maximum	Select Yes – the system performs primarily an assistive function for standard editing Yes – the system does not substantially alter the input data provided by the deployer or the semantics thereof Yes – the system is authorised by law to detect, prevent, investigate, or prosecute criminal offences No Not sure	Explain 500 character(s) maximum

				Depends on the circumstances and context	
	Multimodal systems that use image or video inputs to generate new media outputs (for example, Runway, Sora 2, Veo 3, and OmniHuman-type tools). These systems animate still images or expand short clips, blending authentic and synthetic elements.	Yes	Produces new frames or motion via AI on top of authentic inputs. The output mixes real pixels with AI-generated animation or new imagery.		During the 2025 protests in Turkey, real footage of a man in a Pikachu costume was transformed into AI-animated clips showing him chased by authorities. The lines between real and synthetic became blurred. Similar risks arise when archival stills are animated into events that never happened. Clear guidance is needed on disclosing the “recipe” of AI and human elements to preserve trust in what is real.
	Personalised AI feeds that blur the boundary between information and simulation	Yes	Personalised AI media or story-feed systems should be considered synthetic because they generate or assemble new audiovisual or textual outputs using generative models and algorithmic curation. They create bespoke,		**1. Personalised video generation and story feeds:** Sora type generative video pipelines can generate short video scenes from text prompts. When combined with a user’s interests or browsing data, they can produce

			multimodal narratives shaped by prompts, user data and interaction history. Each user's feed is therefore a unique artefact of algorithmic generation rather than a neutral reflection of reality. Because these outputs merge authentic and synthetic elements and are often indistinguishable from real media, they should be classified as artificially generated or manipulated. Transparency is essential. Users who opt into such feeds or services should be clearly informed that the media they receive is curated and generated through AI systems, and may not represent unbiased or real-world context Personalised AI media or story feeds should be		personalised explainers or narratives. "AI Vibes" or "Reels remix" experiments on short-form platforms such as TikTok or YouTube Shorts are testing AI composed highlight feeds that remix user content with synthetic transitions, music and commentary. Runway or Pika include features that let users generate unique short films or advertisements with user specific prompts, effectively creating personalised narratives. **2. Synthetic news explainer or daily briefing feeds:** Some research projects and start-ups are building systems that summarise or restyle daily news into personalised video or audio explainers, for example AI anchors, synthetic presenters or voice cloned

			considered synthetic, as they generate or assemble new audiovisual or textual outputs using generative models and algorithmic curation. Each user's feed becomes a unique, algorithmically produced artefact rather than a neutral reflection of reality. Since such outputs mix authentic and synthetic elements and may resemble real media, they should be classed as artificially generated. Transparency and clear user disclosure are essential.		summaries. These operate in the text to speech and video for information space, where Article 50(4) will intersect with 50(2). **3. Interactive AI entertainment and virtual influencer ecosystems:** AI avatars and influencer platforms, for example Character.ai, Replika style or Meta's AI characters, generate customised storylines and media fragments tailored to each user's profile. Increasingly, this kind of AI slop feed mixes generated scenes, dialogue and synthetic commentary drawn from real world data.
	Voice cloning and AI dubbing systems used for localisation or accessibility	Yes	These systems generate synthetic speech in a target voice or language.		Clear standards are needed to preserve authenticity in AI assisted dubbing and translation, and to clarify when such outputs count as assistive editing versus synthetic

					generation (i.e. not to undermine real content being assisted by AI for localisation). Disclosure should indicate that AI assistance was used, without undermining the integrity of genuine content.

If you have more examples, please enter them in the section below, following the structure of question 5.

Article 50(2) AI Act specifies that the technical marking and detection solutions implemented by the provider should be effective, interoperable, robust and reliable as far as this is technically feasible taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards. Recital 133 AI Act gives examples of such marking techniques based on watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints, or a combination of such techniques. Furthermore, Recital 133 also clarifies that such techniques and methods can be implemented at the level of the AI system or at the level of the AI model, including general-purpose AI models generating content, thereby facilitating fulfilment of this obligation by the downstream provider of the AI system. Recital 133 also clarifies that the detection methods can be made accessible, as appropriate, to enable the public to effectively distinguish AI-generated content.

Question 6. Please provide examples of marking and detection solutions, including combinations of techniques, that can be employed to mark in a machine-readable format AI-generated or manipulated content and enable detection whether the content has been generated or manipulated by AI.

	Technology's name	**Type of solution, one or combination of multiple techniques**	**Application field per modality**	**Technology maturity**	**Link to the source (e.g. paper, journal)**	**Concise description of the technique and how it works, along with its specificities and potential limitations for modalities and costs of implementation if known**

1	C2PA	Watermarks Metadata Cryptographic methods Fingerprint	Multi-modal	Limited market adoption	https://c2pa.org/	C2PA is becoming central to machine-readable provenance, with much of its ecosystem hinging on the Conformance Program. While the standard is open, accessibility to the Program must be ensured. C2PA also needs to be shaped in a way that preserves privacy and avoids the usage of PII. Conformance administrators and implementers alike must adopt strong safeguards to protect privacy.
	TRIED Benchmark for post-hoc detection and evaluation	Detection and benchmarking framework combining forensic and model-based methods	Multi-modal	Early research	https://www.witness.org/ai-detection-global-benchmark-witness-2/	TRIED provides an open, rights-based benchmark for assessing detection tools used to identify AI-generated content. It evaluates accuracy, explainability and fairness. Complements marking standards such as C2PA by enabling transparent comparison of detection

						systems when provenance data is absent.

Question 7. For each of the solutions included in the previous question, please clarify whether there is relevant information that can help you competently assess their effectiveness, interoperability, robustness and reliability as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art. Please also assess to what extent the detection mechanisms are accessible and enable people exposed to the AI generated or manipulated content to identify its origin.

Technology's name	**Assessment of effectiveness based on the grade below: - N/A (Not known/not sure) 1. Very low 2. Low 3. Moderate 4. High 5. Very high**	**Assessment of interoperability based on the grade below: - N/A (Not known/not sure) 1. Very low 2. Low 3. Moderate 4. High 5. Very high**	**Assessment of robustness based on the grade below: - N/A (Not known/not sure) 1. Very low 2. Low 3. Moderate 4. High 5. Very high**	**Assessment of reliability based on the grade below: - N/A (Not known/not sure) 1. Very low 2. Low 3. Moderate 4. High 5. Very high**	**Assessment of transparency and accessibility to people based on the grade below: - N/A (Not known/not sure) 1. Very low 2. Low 3. Moderate 4. High 5. Very high**	**Motivate your answer, including by providing sources, further information and evidence that supports the assessment**
C2PA	Effective - 5	Interoperable - 4	Robust - 4	Reliable - 4	Transparent and accessible - 4	Explain 750 character(s) maximum Interoperability and access depends on effective access to the Conformance Program and financial and procedural capacity to implement the standard. Barriers must be reduced. Interoperability may (or should) be limited if privacy is not guaranteed. Effective transparency also depends on UX,

						which is vague in legislation.
TRIED	Effective- 4	Interoperable - 3	Robust - 4	Reliable - 4	Transparent and accessible - 5	TRIED provides a framework to evaluate detection tools used to identify AI-generated or manipulated content. Its effectiveness lies in benchmarking accuracy, bias and explainability rather than performing detection itself. Interoperability is moderate; the framework can integrate diverse detectors. Robustness and reliability depend on dataset quality and transparent methodology. Accessibility is high because the benchmark is open and rights-based, promoting equitable evaluation of detection systems where provenance data is absent.

Question 8: Are you aware of technical standards or ongoing standardisation activities relevant in the context of the obligation for generative AI systems in Article 50(2) AI Act?

( ) No

(x) Yes

Please, specify and provide links and further information. 500 character(s) maximum

Apart from CEN-CENELEC's AI Act work, relevant standardisation includes **C2PA** for watermarking and labelling AI content, **JPEG Trust** for authenticity tagging, **ISO 22144** for content provenance, and **AMAS** (including ITU, ISO, ISE) coordinating watermarking and trust metadata standards across media. These efforts support Article 50(2) obligations on transparency and traceability for generative AI systems.

Question 9. Are there any other aspects related to the scope or the practical implementation of the transparency obligation for generative AI systems under Article 50(2) for which you would seek clarification?

( ) No

(x) Yes
Please, specify. 500 character(s) maximum
Clarification is needed on how the CoP will address the following issues (a) protect privacy and personal identifiable information (b) ensure systems are globally accessible/interoperable to providers, deployers, and users alike; (c) avoid weaponization by non-EU governments; (d) protect free expression and satire; (e) ensure technologies remain transparent and understandable to consumers; and (f) ensure equitable access to classifier tools like SynthID for a broad, diverse range of intermediaries.

Section 4. Questions in relation to Article 50(4) AI Act*

Article 50(4), subparagraph 1, AI Act requires deployers of AI systems generating or manipulating image, audio or video content constituting a deep fake to disclose that the content has been artificially generated or manipulated. The definition of a deep fake is provided in Article 3(60) AI Act, which defines 'deep fakes' as 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 as authentic or truthful.
If the deep fake content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency requirement is limited to the disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work. The transparency obligation in Article 50(4), subparagraph 1, does not apply if the AI system is authorised by law to detect, prevent, investigate, or prosecute criminal offences.

Question 14. Please provide practical examples of AI-generated or manipulated content for which you would seek clarification regarding its classification as a 'deep Fake'. (400 character(s) maximum)
Clarification is needed on 1) AI re-enactments of public figures used to undermine democracy, 2) non-consensual intimate imagery, 3) quasi-factual "AI slop" mixing real and synthetic content, 4) personalised deepfakes such as Sora2-type timelines, 5) AI-generated animations of historical photos, and 6) AI-dubbed or translated media.

Question 16. If you are aware of any examples of disclosure practices that can be employed with deep fake content to duly disclose the artificially generated or manipulated origin of such content to natural persons exposed thereto, please provide them in your response. For each disclosure practice, determine whether the type and the content of the disclosure practice is appropriate for clearly, distinguishably and accessibly informing natural persons about the artificially generated or manipulated origin of the content and the timing of the notification. In cases where the disclosure practice is used on deep fake content that forms part of an evidently creative, satirical, artistic, fictional or analogous work or programme, determine whether the disclosure is done in an appropriate manner that does not hamper the display or enjoyment of the work.

	Description of the disclosure practice	**Specify to which type of deep fake it applies**	**Determine whether the type and the content of the disclosure practice is**	**Motivate your answer and, where available, provide practical examples with links and further information**	**For deep fake content part of evidently creative satirical, artistic,**	**Motivate your answer and provide practical example(s)**

			appropriate for clearly, distinguishably and accessibly informing natural persons		**fictional or analogous works, determine whether the disclosure does not hamper the display or enjoyment of the work**	
1	Describe	Select Audio Image Video	Select Appropriate Not appropriate	Explain 500 character(s) maximum	Select Hampers the display or enjoyment Does not hamper the display or enjoyment Not sure	500 character(s) maximum
	Striscia la Notizia, Italian TV Show, low-fidelity face swap video with a Striscia watermark and a clear text-based disclaimer	Video	Appropriate	In September 2019, the well-known Italian satirical TV show Striscia la Notizia posted a low-fidelity face-swap video of former prime minister Matteo Renzi sitting at a desk insulting his then coalition partner Matteo Salvini with exaggerated hand gestures on social media. The content had a Striscia watermark and a clear text-based disclaimer, but some viewers believed the video was genuine. Link: https://x.com/Striscia/status/1176191956558462976	Does not hamper the display or enjoyment	This example demonstrates a context shift: Once any piece of media, even labeled and watermarked, is distributed across politicized and closed social media groups, its creators lose control of how it is framed, interpreted, and shared.

	Welcome to Chechnya	Video	Appropriate	In the documentary *Welcome to Chechnya*, vulnerable interviewees were digitally disguised with the help of inventive synthetic media tools like those used to create deepfakes. In addition, subtle halos appeared around their faces, a clue for viewers that the images they were watching had been manipulated, and that these subjects were taking an immense risk in sharing their stories. Link: https://www.welcometochechnya.com/	Does not hamper the display or enjoyment	When a disclosure is baked into the media itself, it can't be removed, and it can actually be used as a tool to push audiences to understand how a piece of media was created and why.
	Kendrick Lamar's 2022 music video *"The Heart Part 5"*	Video	Appropriate	Kendrick Lamar's 2022 music video, "The Heart Part 5," the directors used deepfake technology to transform Lamar's face into both deceased and living celebrities such as Will Smith, O. J. Simpson, and Kobe Bryant. This use of technology is written directly into the lyrics of the song and choreography, like when Lamar uses his hand to swipe over his face, clearly indicating a deepfake edit. Link: https://www.youtube.com/watch?v=uAPUkgeiFVY	Does not hamper the display or enjoyment	The resulting video is a meta-commentary on deepfakes themselves. The video transcends easy answers and give audiences space to interpret for themselves.

	Republican National Committee's ad against Biden	Video	Appropriate	The Republican National Committee put out a video advertisement against Biden, which featured a small disclaimer in the top left of the frame: "Built entirely with AI imagery." Link: https://arstechnica.com/tech-policy/2023/04/gop-releases-100-ai-generated-ad-to-fearmonger-over-bidens-reelection-bid/	Does not hamper the display or enjoyment	Critics questioned the diminished size of the disclaimer and suggested its limited value, particularly because the ad marks the first substantive use of AI in political attack advertising. As AI-generated media become more mainstream, many have argued that text-based labels, captions, and watermarks are crucial for transparency.
	Alt-Text as Poetry	image	Appropriate	In the art project Alt-Text as Poetry, audiences are encouraged to draft alt-text descriptions of images for visually impaired audiences that are poetic rather than perfunctory. Link: https://alt-text-as-poetry.net/		Just like artistic disclosures, alt-text helps explain—or disclose—contextual information, ideally in a creative way. The artists explain that they approach access "generously, centering disability culture, rather than focusing on compliance."

Question 18. Please provide practical examples of AI-generated or manipulated textual content for which you would seek clarification regarding whether or not such content is published with the purpose of informing the public about matters of public interest, or whether or not such content has undergone human review or editorial control. 400 character(s) maximum.

AI-manipulated content impersonating politicians, AI-drafted campaign statements, AI-crafted narratives targeting minorities and women, and personalised creations and curation of content should be addressed in the CoP. It's also important to clarify how the CoP covers content with little or no human editorial control, how disclosure rules apply, and what information will be shared with the public.

Question 20. Are there any other aspects related to the scope or the practical implementation of the transparency requirement for deployers of AI systems that generate deep fakes and text publications on matters of public interest under Article 50(4) for which you would seek clarification?
( ) No
(x) Yes
Please, specify. 500 character(s) maximum
(a) how to address satire, parody, artistic works; (b) identification of AI-generated segments within mixed real + synthetic content (c) ensuring transparency/labels maintained when shared across platforms; (d) proportionality of obligations for small vs. large deployers; (e) treatment of open-source tools; (f) personalized, limited reach content and likeness usage with consent, e.g. in Sora2 (g) what qualifies as "editorial control" (h) how disclosures can be presented consistently

Section 5. Other horizontal questions in relation to the implementation of Article 50 AI Act

This section covers a set of questions that relate to horizontal issues regarding Article 50 AI Act. First, it addresses the requirements from Article 50 (5) AI Act which apply horizontally when providing the information as required by paragraphs 1-4 of Article 50. *Second, the section addresses Article 50(6) and the interplay between the AI Act's requirements from Article 50(1)-(5) and other transparency obligations of the AI Act or other Union or national legislation. Finally, it also asks for recommendations and good practices for the Code of Practice.*
Article 50(5) AI Act requires that the information to be provided under the various transparency requirements from Article 50 shall be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure. In addition, such information shall conform to the applicable accessibility requirements. Regarding the latter, recital 132 confirms that such information and notifications should be provided in accessible formats for persons with disabilities.

Question 21. *Are there aspects related to the AI Act's horizontal requirements in Article 50(5), including their interplay with the requirements in Article 50(1)–(4), for which you would seek clarification?*
( ) No
(x) Yes
Please, specify. 500 character(s) maximum
We would be interested in clarifications regarding how the horizontal requirements apply to providers or distributors of open AI weights and parameters that may subsequently be used to generate synthetic audio, image, video or text content, especially as it concerns provisions pertaining to 50(2). Added to that, more clarification is needed on deployers and providers will collaborate to maintain marking across distribution, modification and re-use/remix.

Question 22. Are there any further aspects related to the transparency obligations under Article 50(1)-(5) for which you would seek clarification regarding their interplay with other obligations in the AI Act?
( ) No
(x) Yes
Please, specify which aspects require clarification and point to specific provisions in the AI Act. (500 character(s) maximum)
Further clarification on how the transparency duties from article 50 interact with risk management (Art. 9), data governance (Art. 10), and information to users (Art. 13), would be welcome. Added to that, when

applying to high-risk models, the CoP needs to provide further clarification on which are the applicable rules and what's the exact timeline for the obligations to go into force. .

Question 23. Are there any further aspects related to the transparency obligations under Article 50(1)-(5) for which you would seek clarification regarding their interplay with obligations in other Union or national legislation (e.g. data protection regulation such as Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016 /680, Regulation (EU) 2024/900 on the transparency and targeting of political advertising or Regulation (EU) 2022/2065 on a Single Market For Digital Services)?
( ) No
(x) Yes

Please, specify which specific aspects require clarification and point to specific provisions in Union or national legislation. 500 character(s) maximum
Further clarification on how the Article 50 obligations interplays with the transparency obligations under the GDPR and Digital Services Act, as well as the ones applicable to political ads would be welcome. Specifically, guidance is needed on reconciling disclosure duties, user consent, automated decision explanations, and cross-border enforcement to understand what are the concrete obligations for AI deployers.

Question 24. Are there any recommendations or good practices you would like to share as input for the Code of Practice to operationalise the implementation of the transparency obligations regarding interactive and generative AI systems?
( ) No
(x) Yes
Please, specify. 750 character(s) maximum

WITNESS recommends leveraging existing frameworks and research to help operationalise transparency obligations for interactive and generative AI systems. Some of the emerging examples we would like to highlight are the (a) C2PA Harm Assessments, a framework by the C2PA (Coalition for Content Provenance and Authenticity) that provides practical guidance for identifying and mitigating risks associated with provenance technologies; and (b) WITNESS' most recent report on C2PA and Human Rights, which outlines how content provenance and authenticity standards intersect with human rights considerations, offering actionable insights for transparency and accountability.