Article 50 in practice: what changed and who needs to act?

SORAINEN

Ignas Sidaras, Krišjānis Knodze, Mihkel Miidla

Lithuania, Latvia, Estonia

02.09.2026

AI Act Article 50 (Transparency). What changed from August 2026?

Four key transparency obligations are now live

AI interaction disclosure (chatbots)

Machine-readable marking (most AI-generated content)

Deepfake labels

Public-interest text disclosure

Affects IT, customer support, marketing, communications and content creation.

Does not apply to natural persons using AI systems in the course of a purely personal, non-professional activity.

Non-compliance: fines up to EUR 15 million or 3% of worldwide annual turnover, whichever is higher.

AI interaction disclosure (chatbots)

SORAINEN

AI interaction disclosure (chatbots)

When	Users must be informed they are interacting with AI **before the interaction and throughout its duration**.
Who must ensure compliance	**Provider** (creator of AI chatbot) vs **Deployer** (user of AI chatbot)?
Four key elements must be present for a system to fall within the scope of Article 50(1)	• An AI system; • Intended to interact: exchange information or actions with a genuine conversational or responsive character; • Direct interaction: typically involves real-time or near real-time interaction; • Intended to interact directly with natural persons (including professionals, excluding AI Agents or other industrial systems).

SORAINEN

AI interaction disclosure (chatbots)

Examples of systems fall under the Article 50 (1):

- **AI-enabled voice assistants**, **chatbots** in various contexts (e.g. public service, customer support, e-commerce etc.),
- **AI hotline for incidents**, **fraud** or other **reporting**, **(humanoid) robots**, **AI companions**;
- **AI bots on social networks and media**, coding agents and **other AI agents capable of directly interacting with natural persons** (including in complex multi-agent architectures).

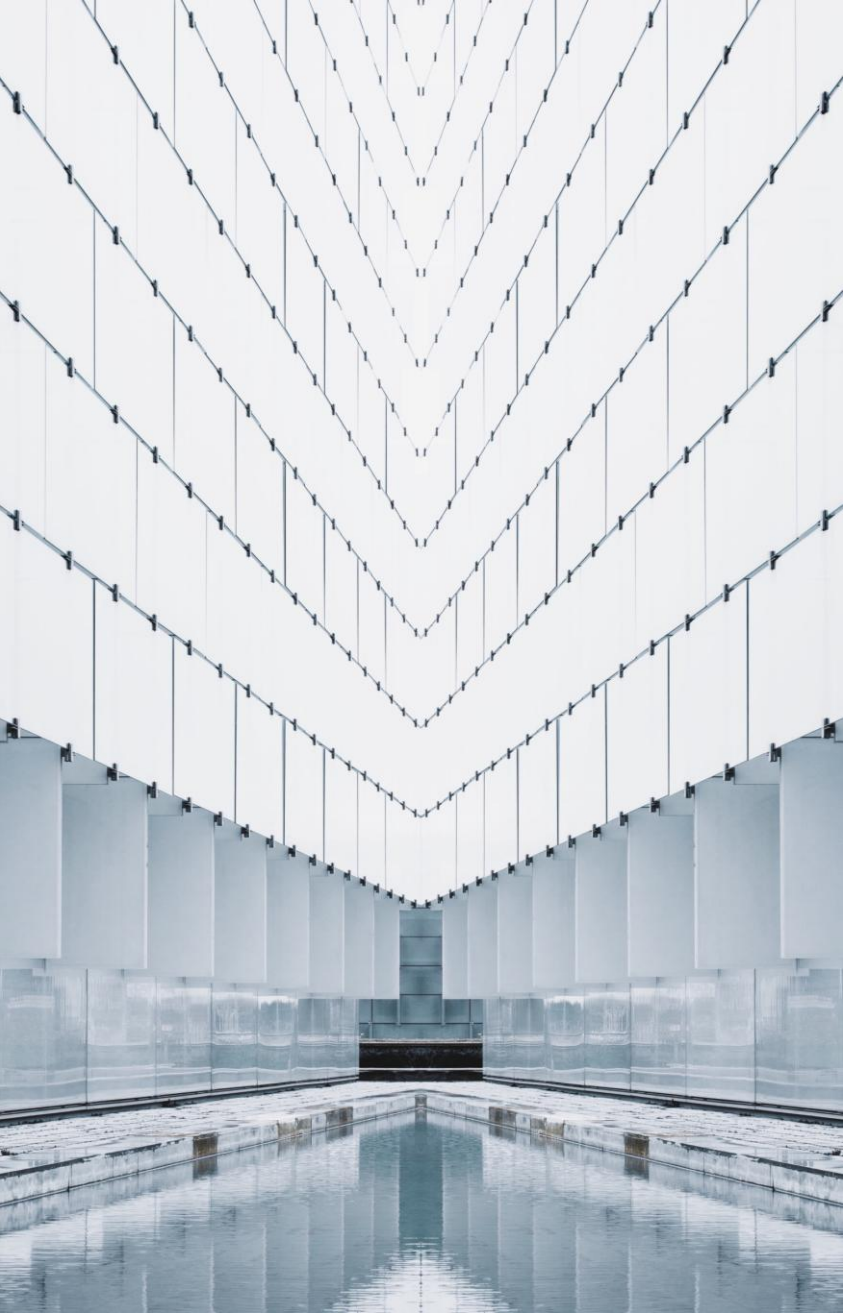
Examples of systems do not fall under the Article 50 (1):

- AI-enabled **traditional industrial robots** that operate in a closed industrial setting and not intended to interact with humans (e.g. electronic assembly robots),
- **AI-enabled recommender systems**, **spam filters**, **automated translation** or **transcription tools**
- **AI-enabled search and retrieval of information systems that do not generating or modify content**.

SORAINEN

Rules of a thumb for a sufficient AI interaction disclosure

- **Disclosure must be in-operation:** *do not burry it in T&Cs*
- **Know your audience:** *People's vulnerabilities due to age or disability must be considered whenever the AI is meant to interact with them or is likely to do so*
- **Disclosure must be clear and distinguishable:**
  - **Sufficient disclosure:** *"You are communicating with an automated system powered by artificial intelligence, not a human member of staff"*
  - **Insufficient disclosure:** *"Hi, I am your virtual assistant Sarah"*
- **A person must clearly understand when an interaction with AI begins and ends:** *For instance, when a staff member takes over the conversation, this must be clearly indicated (e.g., "Our staff member Megan has now joined the conversation")*



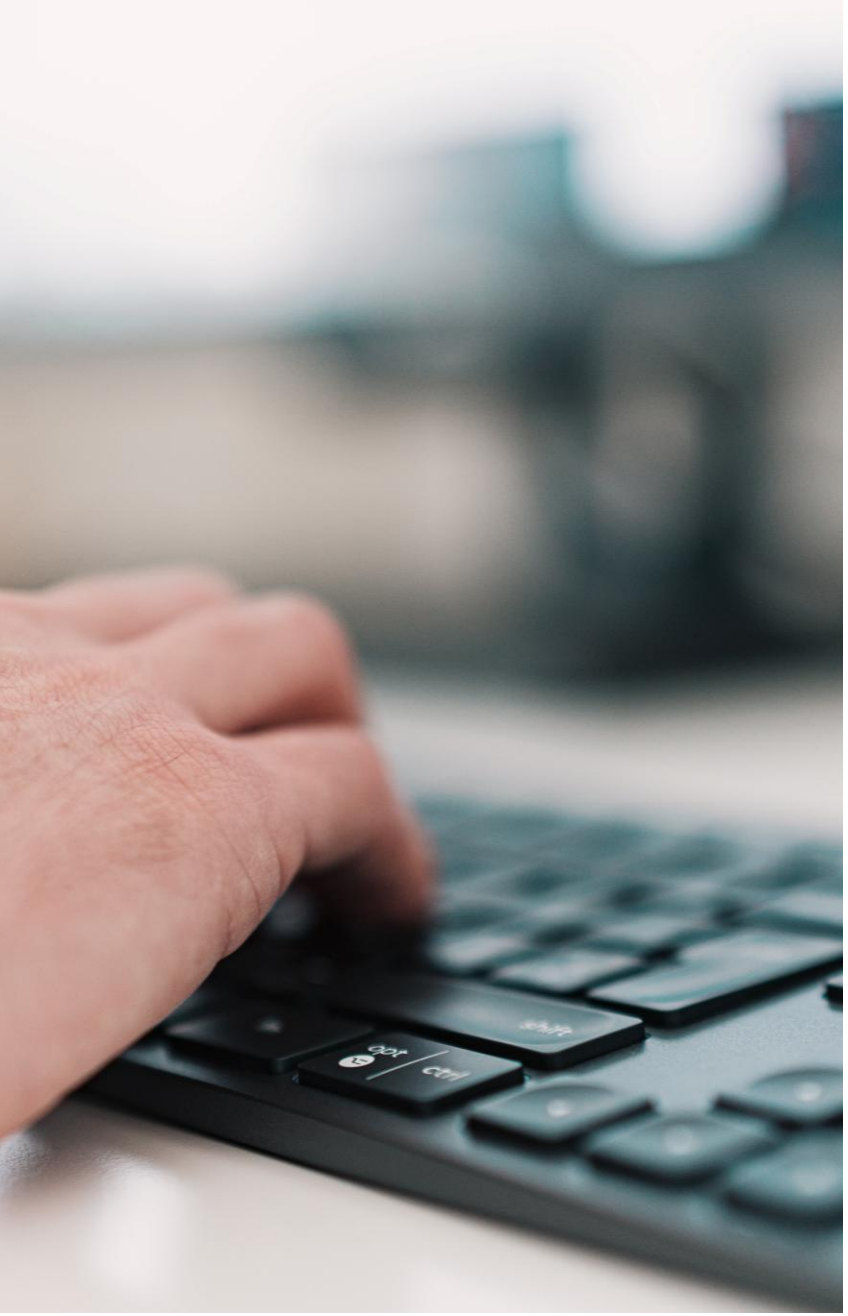
The “Obvious Interaction” Exception

- **Exception exists but is narrow** – interpret it restrictively
- To be able to rely on the exception, providers/deployers will need to assess and demonstrate:
  - the obvious artificial nature of the interaction to;
  - a natural person who must be **reasonably well-informed**, observant and circumspect, taking into account the circumstances and the context of use
- **Examples of “obvious”:** overtly robotic voice; cartoon robot avatar
- **Allowed only:** where there is almost no doubt for the specific audience (example: *Interactive AI systems intended only to be used by properly trained health professionals to support medical diagnosis and suggest related treatments.*)
- **Common mistake:** assuming *“everyone knows it's a bot”* – this does not satisfy Article 50(1)

SORAINEN

Machine-readable marking

SORAINEN



Machine-readable marking

- **Obligation falls on providers**
- **Covers:** nearly ALL AI-generated image, audio, video, and text
- **Two linked duties:** mark outputs AND make them detectable
- **Quality standard:** effective, interoperable, robust, and reliable
- **Exceptions**: standard editing or non-substantial changes, very short text, source code and machine-to-machine outputs
- **Grace period:** systems placed on market before 2 Aug 2026 have until **2 December 2026** to comply with marking.

SORAINEN

How marking works: the two main approaches

1. Invisible Statistical Watermarks (e.g., Google SynthID)

- Embeds imperceptible signals at point of generation
- Survives cropping (up to 75%), JPEG compression, colour adjustments
- Covers image, audio, video, and text outputs

2. Cryptographic Provenance Signing (C2PA / Content Credentials)

- Attaches a signed chain-of-custody record to the file
- Records: creator, tool, AI model, timestamp, full edit history
- Adopted by Adobe, Microsoft, OpenAI, Google, BBC, AP, Reuters

Key limitation:

- C2PA is stripped by screenshots and most social media upload pipelines
- Watermarks survive distribution but carry only a lookup identifier
- **Best practice:** combine both methods

Photo & image example

Google SynthID example:

SynthID adds an invisible digital watermark to an AI-generated image.

The watermark doesn't change the image quality.

It's added the moment content is created, and designed to stand up to modifications like cropping, adding filters, changing frame rates, or lossy compression.

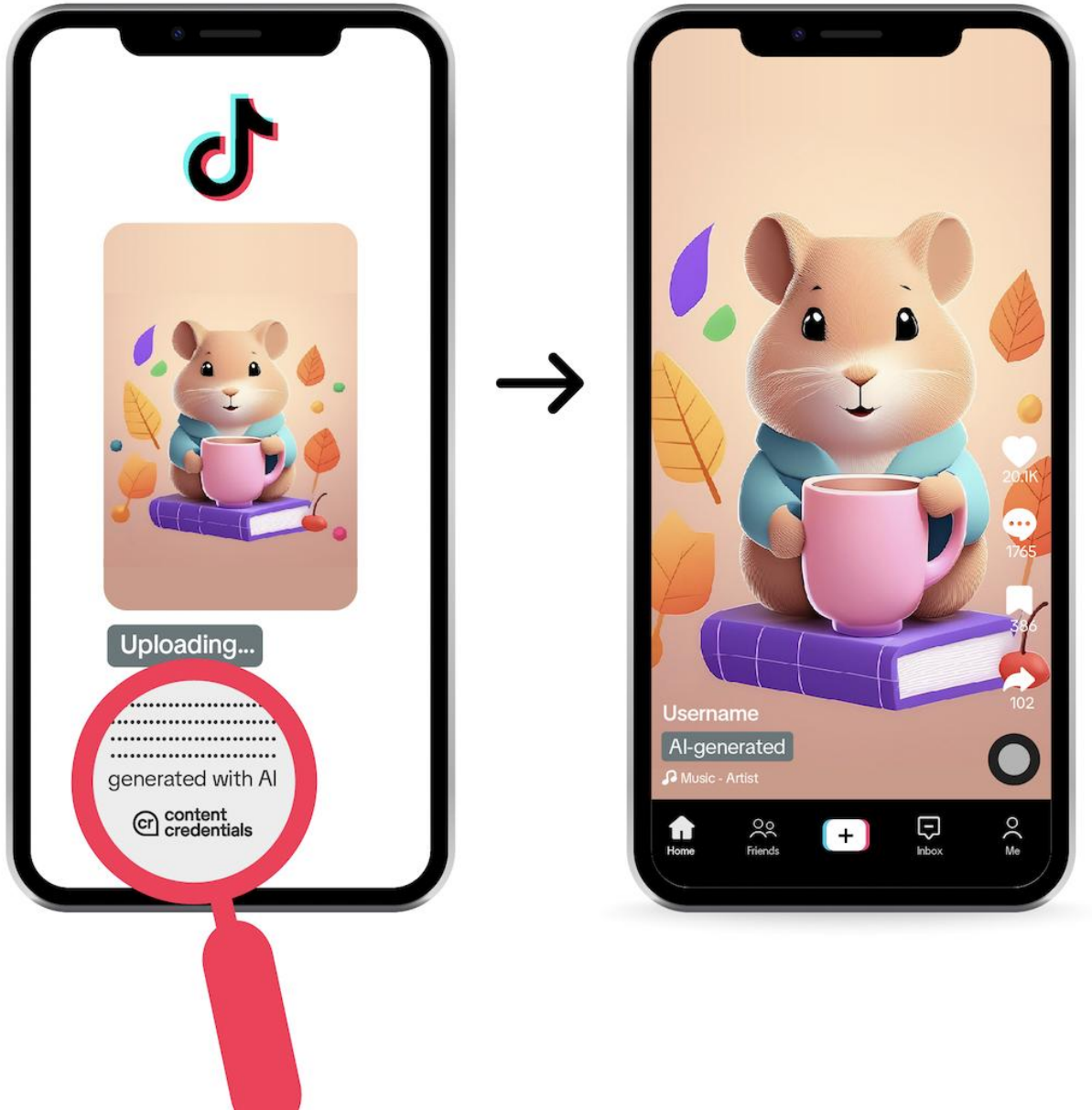
SORAINEN

Video & audio examples

TikTok example:

TikTok has already tagged more than 1.3 billion videos as AI-generated, and more than half of all videos on the platform now use some form of AI.

The crackdown on unlabeled AI content has jumped sharply, up 340% from 2024, with over 51,000 synthetic posts removed in the second half of 2025.



SORAINEN

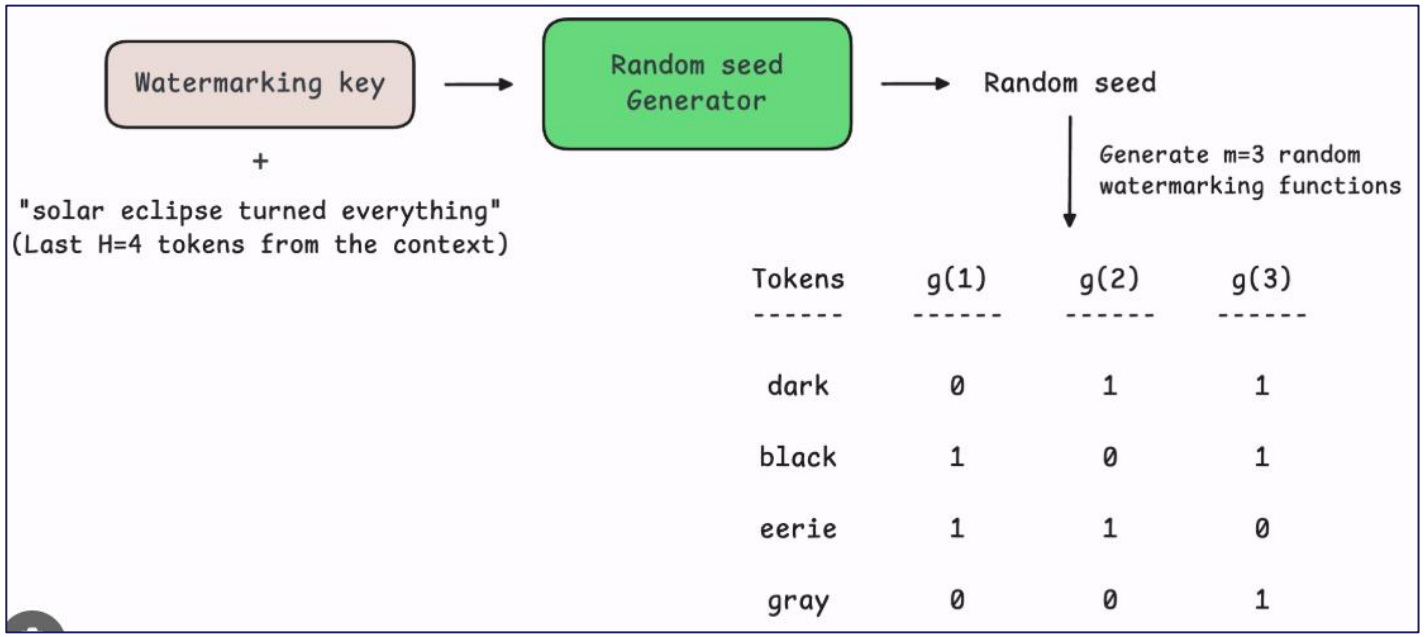
Text example

Claude example:

Claude adds an invisible statistical pattern to its word choices so that its text can later be identified as AI-generated without changing how the text looks.

Legal grey zone:

What happens if a deployer (user) uses „humanizer“ to redraft the text and hide watermarking?



SORAINEN

When does marking NOT apply?

Exeption examples (standard editing / no substantial alteration):

- Grammar and spellchecking
- Formatting and format conversion
- Minor image cropping, brightness/contrast adjustment
- Technical audio compression or noise reduction
- Video stabilization

Not exempt – marking required:

- Insertion or removal of objects/persons in images or video
- Face replacement in video
- Realistic voice synthesis in a named person's voice

SORAINEN

Deepfakes

SORAINEN

Article 50(4) Disclosure obligation for deep fakes

- **Applies to whom?**

Deployers of an AI system.

- **When does it apply?**

Where the AI system generates or manipulates image, audio or video content constituting a “deep fake”.

For deepfakes generated before 2 August 2026 no mandatory retroactive labelling but encouraged.

- **What is a “deep fake”?**

AI-generated or manipulated image, audio or video content that (i) resembles (ii) existing (iii) persons, objects, places, entities or events and would (iv) falsely appear to a person to be authentic or truthful.

- **What must be done?**

The deployer shall disclose that the content has been artificially generated or manipulated.

- **Exception - criminal offences**

The obligation does not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offences.

- **Limited obligation - artistic, creative, satirical or fictional works**

Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the obligation is limited to disclosing the existence of the AI-generated or manipulated content.

What is a “deep fake”?

Content that (i) resembles (ii) existing (iii) persons, objects, places, entities or events and would (iv) falsely appear to a person to be authentic or truthful.

SORAINEN

Resembles

- Resemblance should be “appreciable”.
- The content does not need to be identical to the subject.
- Content appreciably resembles a subject if there is a high level of similarity between the deep fake content and the subject (including any of its recognisable elements) being simulated by the deep fake.
- Case-by-case assessment.

Existing

- AI-generated or manipulated image, audio or video content should resemble realistic subjects (e.g., persons, objects, places).
- Something that exists, can plausibly exist or could have plausibly existed in reality.
- By contrast, obviously unrealistic or impossible content (e.g. flying humans, dragons or elephants driving cars) that cannot mislead falls outside the transparency obligation.

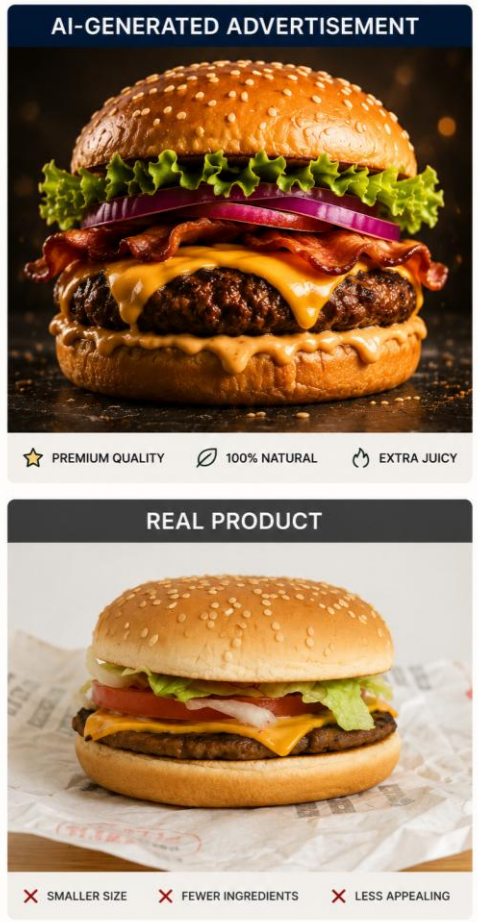
Persons, objects, places, entities or events

- **Persons**: realistic, human beings (including digital replicas of real persons, realistic AI-generated human avatars or personas, and personal characteristics or expressions, such as image, voice, behaviour, performances etc.).
- **Objects**: realistic, inanimate material items, including buildings, artworks, machinery, consumer goods etc.
- **Places**: realistic locations.
- **Entities**: realistic, non-human but animate beings including animals or other biological lifeforms.
- **Events**: realistic scenes or situations that can involve persons, objects, places and entities.

False appearance to a person to be authentic or truthful:

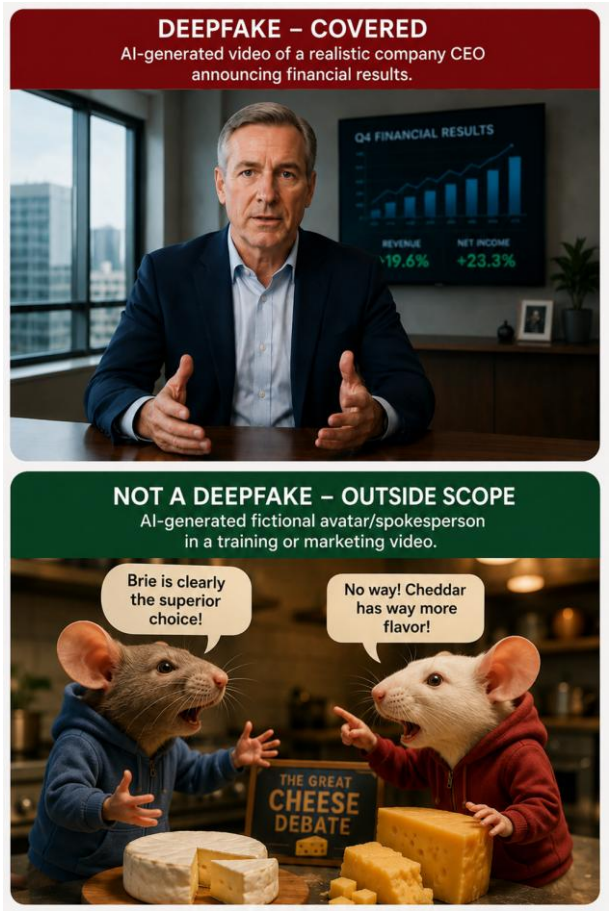
- Would the content mislead a person into believing it is authentic or truthful?
- Consider: level of resemblance, message/content of the material; context and environment of use; reasonably foreseeable audience and their expectations.
- No intent to deceive is required, the assessment is objective.
- Content used where audiences expect fiction (e.g. movies) is less likely to meet this criterion, while ssignificant AI manipulation is more likely to meet the criterion.
- Minor technical edits (e.g. colour correction, noise reduction, lighting, compression) generally do not meet the criterion.

1) Deepfake



2) Not a deepfake

3)

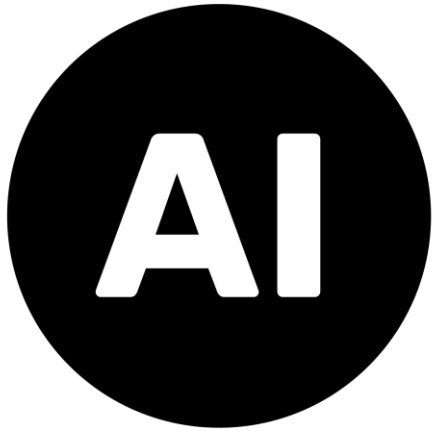


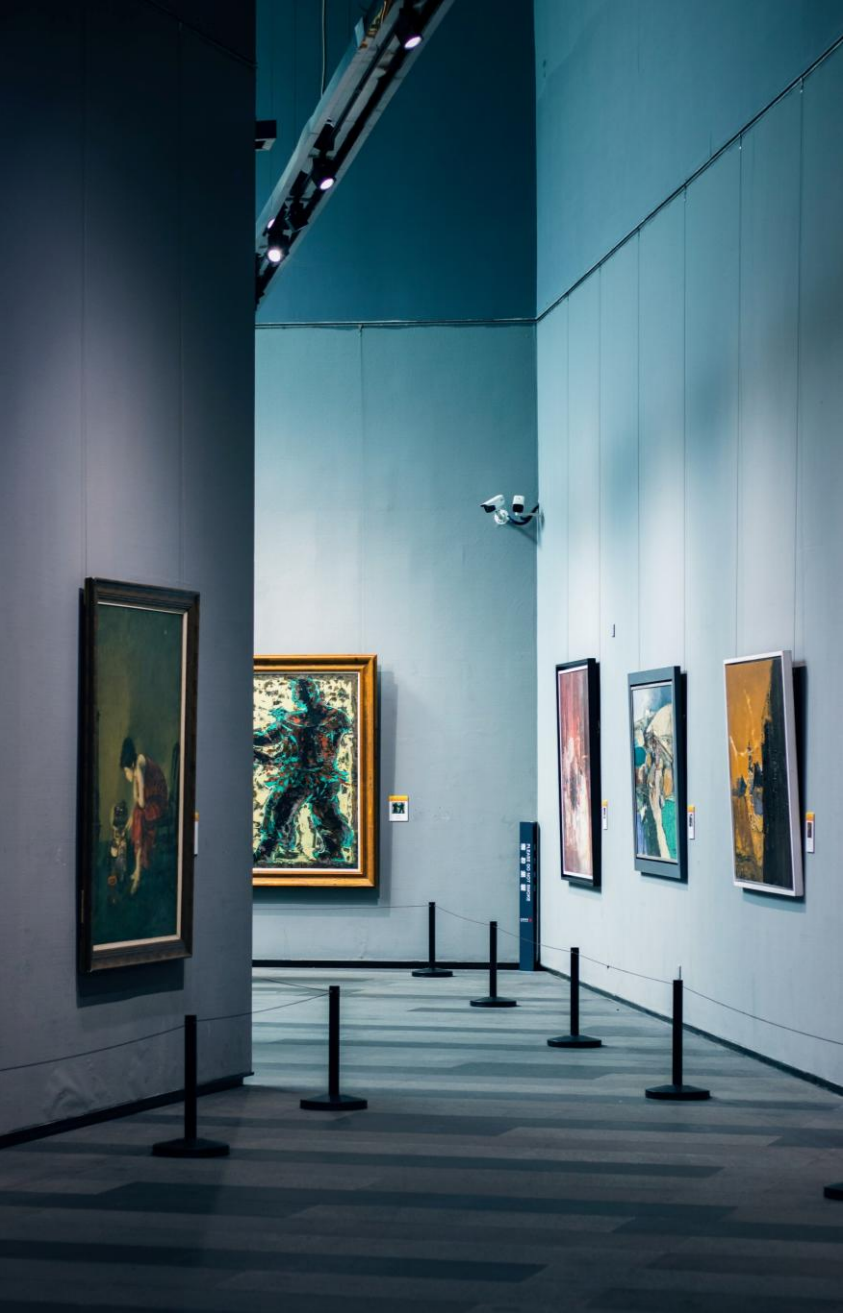
A deepfake is AI-generated or manipulated image, audio or video content that realistically resembles a person, object, place, entity or event that exists or could realistically exist, and could make people believe that it is real or truthful.

SORAINEN

What must be done if a deepfake is used?

- Disclose that the content has been artificially generated or manipulated and indicate its artificial origin.
- Disclosure must be clear and noticeable to users (e.g. through a visible or audible label), and thus cannot rely solely on machine-readable markings.
- May follow an adequate code of practice under Article 50(7) AI Act to demonstrate compliance with the disclosure obligation.





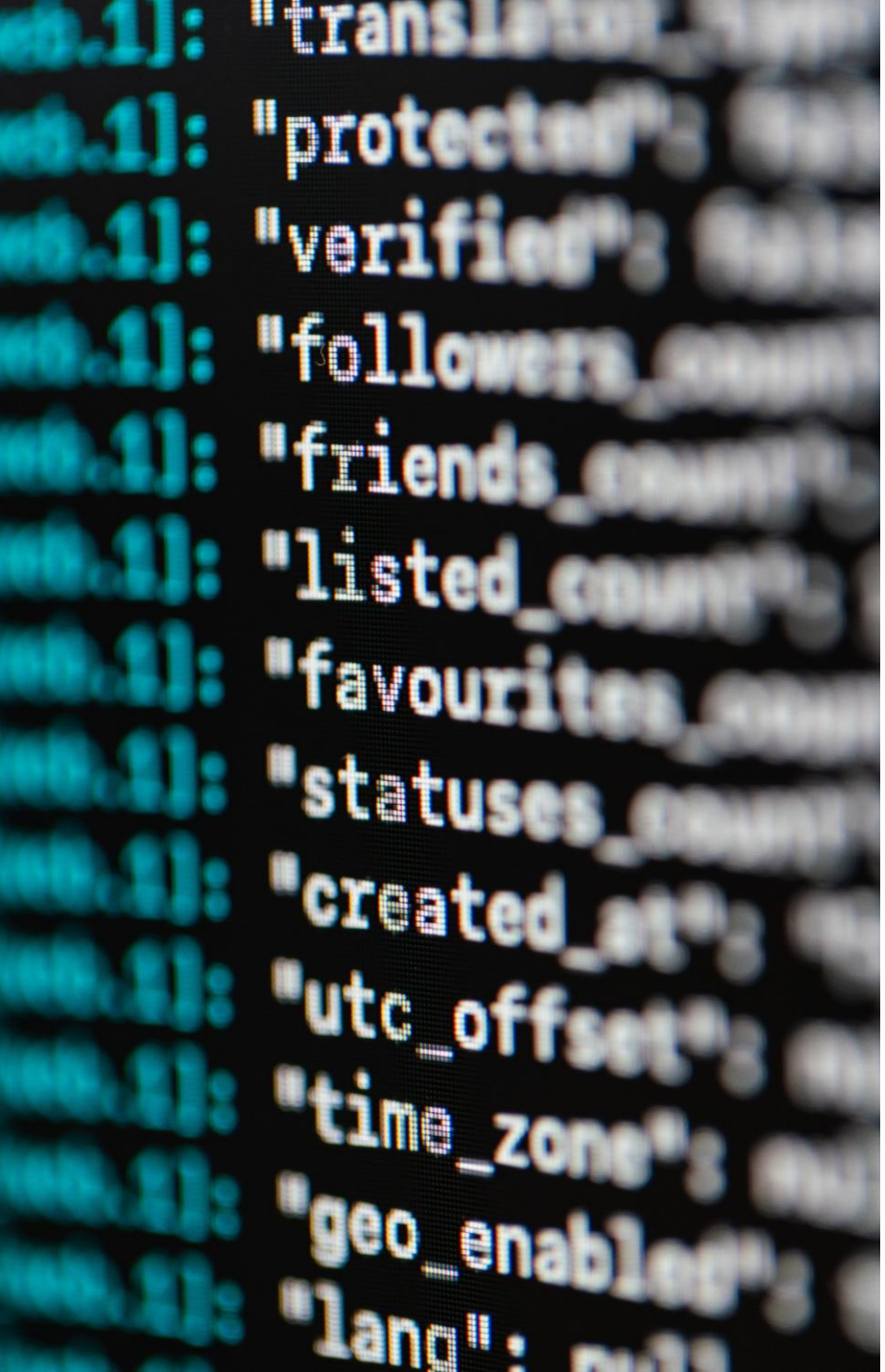
Artistic, creative, satirical, fictional or analogous deep fake content

- The transparency obligation is limited to the disclosure of the deep fake in an appropriate manner that does not hamper the display or enjoyment of the work.
- Categories should be interpreted strictly, justified by the fact that a lighter disclosure regime applies due to the potential impact on the freedom of expression.
- Still requires a clear AI-generated/AI-manipulated label, but it may be placed flexibly, e.g. in credits, accompanying notes, adjacent to the content, or on packaging.

SORAINEN

Public-interest text disclosure

SORAINEN



The duty & two exceptions

- **Article 50(4), second subparagraph:**

*"**Deployers** of an **AI system that generates or manipulates text** which is **published** with the **purpose of informing the public on matters of public interest** shall **disclose** that the* text has been artificially generated or manipulated.*"*

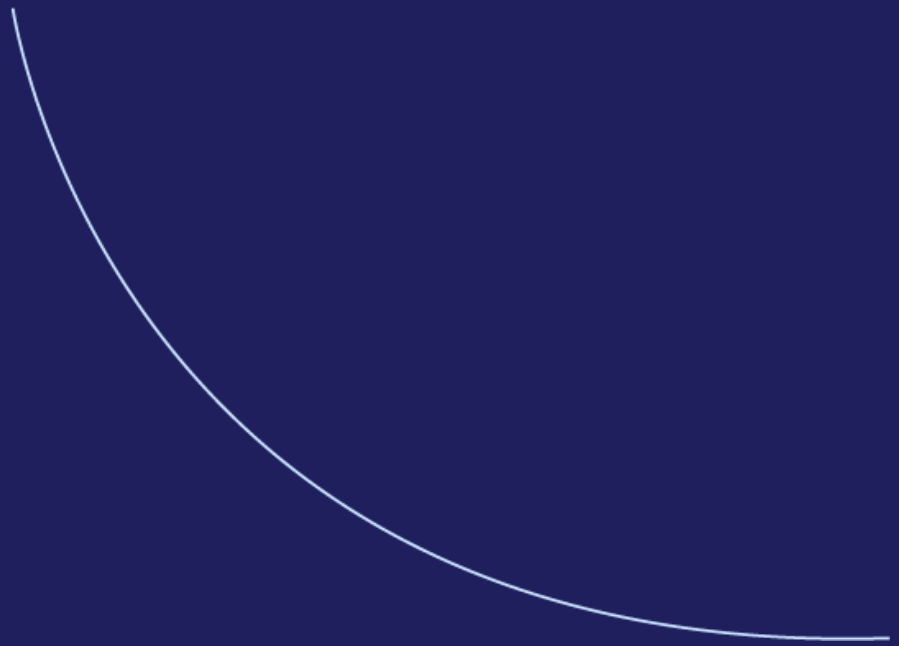
- **Exception 1 – Law Enforcement**

Use authorized by law to detect, prevent, investigate or prosecute criminal offences.

- **Exception 2 – Editorial**

The content has undergone human review or editorial control, and a natural or legal person holds editorial responsibility for the publication.

The trigger: three cumulative criteria



1. “Published”

Accessible to an indeterminate, fairly large number of unrelated readers — whether free or against payment.

2. “Informing the public”

The text is intended to communicate knowledge, opinions or facts to its audience

3. “On matters of public interest”

Matters relevant to society at large, meriting public debate or scrutiny.

SORAINEN

“Public interest” is broad and open-ended

Politics & democratic processes	**Public administration & services**	**Administration of justice & law enforcement**
Fundamental rights	**Public security**	**Public health**
Environmental protection	**Consumer safety**	**Economic & financial developments**
Scientific developments	**Cultural developments**	**… any development that may merit public debate (=open-ended list)**

The editorial exception

1. **Human review**

Deliberate, substantive examination of the content by one or more people with relevant subject-matter knowledge and professional judgement

2. **Editorial control**

Control actually exercised by a responsible editorial function — e.g. an editor-in-chief — with authority to approve, alter or reject the substance of the text.

3. **Editorial responsibility**

A natural or legal person holds ultimate legal responsibility for the publication — including for the review or control itself. The Guidelines expect the holder's identity and contact details to be easy to find

Formula: (1 OR 2) + 3 = no need to label

Not “editorial control”

- Spell-checking or grammar correction alone, superficial, purely formal or procedural checks
- An editorial policy that exists on paper but is not applied to the text at hand
- Automated or AI-performed review, including AI review followed only by a human grammar pass
- Sign-off without substantive engagement with the content (rubber-stamping)

Passes. Standard newsroom editing: a desk editor who fact-checks the piece and has the authority to change or spike it.

Fails. A well-documented process of superficial checks. The test is if substance actually engaged with.

Example scenarios

Scenario	Label?	Why?
Newsroom article generated with AI, fact checked and editorial responsibility clear	NO	Editorial exception
Same article machine translated into 15 languages	YES	Clearance does not travel
AI storm warning auto-posted by a meteorological institute	YES	Public interest, no review
AI summary of human-written article, auto published in a news site	YES	No substantive review
AI generated marketing text or product description without health, safety or ‘green’ claims	NO	Outside scope

Note labelling requirements in Art 50 (5)

What deployers should do now

- Map what you publish — which AI-touched text meets the three trigger criteria
- Decide per channel: an editorial workflow, or a visible label
- Make review substantive and documented — fact-checking is the floor
- Re-review after any AI change
- Review every language version, not just the source text (or add label)
- Publish who holds editorial responsibility — with contact details

SORAINEN

- Ensuring compliance when using AI in your company
- AI agreements
- Intellectual property protection in AI
- AI use policy development and employee trainings
- Engagement with regulators and public authorities regarding the use of AI tools
- Dispute resolution and litigation in AI-related matters

Turn to our team if you have any questions about the use of AI by your company

Mihkel Miidla
Partner at Sorainen Estonia,
Head of TMT sector group

+372 56 622 985
mihkel.miidla@sorainen.com

Ignas Sidaras
Counsel at Sorainen Lithuania

+370 67 049 625
ignas.sidaras@sorainen.com

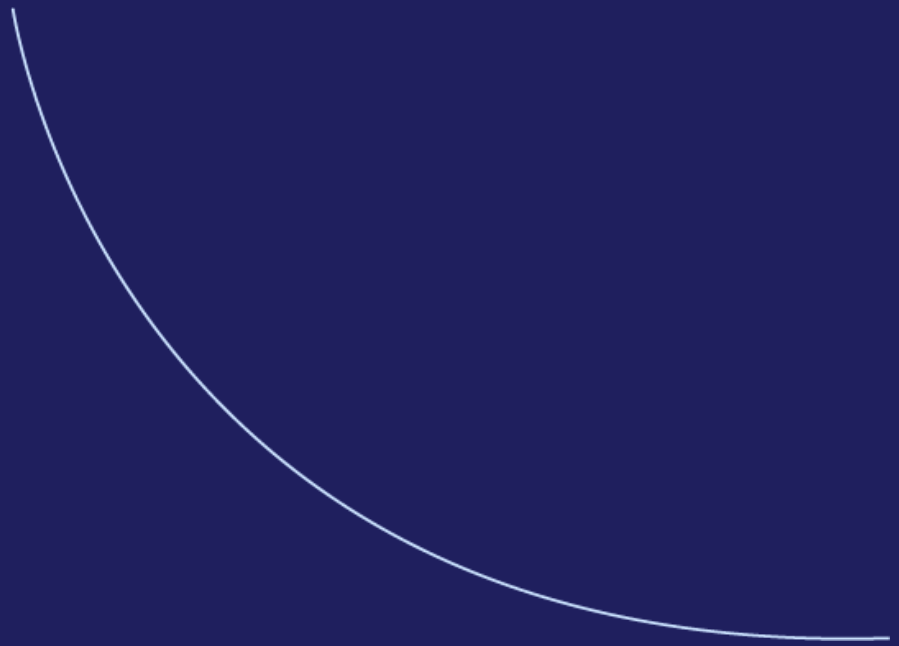
Linda Sarāne-Reneslāce
Partner at Sorainen Latvia

+371 27 877 241
linda.reneslace@sorainen.com

Krišjānis Knodze
Associate at Sorainen Latvia

+371 20 515 137
krisjanis.knodze@sorainen.com

Thank you!



SORAINEN