WITNESS Input to the Second Draft of the Code of Practice on Transparent AI

WITNESS submits a high-level summary of our comments on the second draft and redlines for your consideration. These points are already reflected on our interventions during the calls and contribution to the latest survey, so we would appreciate it if this document is taken into account by the co-chairs and AI Office as concerns that deserve a stronger attention as we make our way to the last version of the Code of Practice.

The WITNESS submission, as detailed in the latest survey, focuses on core aspects in sections one and two. It aims to bring to your attention considerations around the need to strengthen privacy provisions, provenance chain integrity, improvements to compliance and public reporting, and the Neutral Icon, among other topics.

WG 1: Rules for marking and detection of AI generated and manipulated content applicable to providers of generative AI systems (Article 50(2) and (5) AI Act)

Our redlines document focuses on four areas where we believe the Code requires strengthening: privacy, provenance chain integrity, open-weight model requirements, and public reporting requirements.

Privacy

The Code acknowledges privacy concerns in Sub-measures 1.1.2 and 1.1.3 but not in Sub-measure 1.1.1, which is the primary and most broadly applicable compliance pathway. We do not consider GDPR coverage sufficient here: what constitutes personal data in signed metadata is genuinely contested and will produce inconsistent practice across signatories without explicit guidance at the measure level. We propose that 1.1.1 be brought into alignment: metadata markings should not include personally identifiable information by default, any inclusion of user-related data should be limited to what is strictly necessary or affirmatively chosen by the user, and control over such data should rest with the relevant data controllers and rights-holders. We also propose explicit language in the provision on metadata retention permitting signatories to remove personal information from provenance records in two circumstances: where the data subject has requested redaction, or where the signatory has determined that retention of that specific personal information is not necessary for the purposes of the provenance record. This exception is strictly limited to personal information and does not extend to other provenance data.

Provenance chain integrity

Measure 1.3 makes the entire provenance chain optional, conflating system-level transparency with privacy risk in a way that sacrifices the first to address the second. We propose a clear distinction between system provenance data, meaning information about tools, models, operations, and sequence, which should be mandatory, and personal or contextual provenance data, meaning data linking back to

individuals, original source files, or identifying circumstances, which should be protected by default. Signatories should be required to retain existing system provenance and add their own marking to the chain. This distinction will need to be defined in the relevant definitions section to be legally and technically operative.

Open-weight models

We urge the group not to remove the structural marking measure for open-weight models. Without it, the framework produces a structural gap no downstream obligation can close: individual and informal use of open weights carries no obligation at any point in the chain, and that stream includes motivated disinformation actors who face zero friction from this framework as currently designed. Feasibility concerns are real but addressable through a layered approach: base model releasers implement marking at training time, fine-tuners preserve existing marking where technically feasible, and hosting platforms ensure models distributed through their infrastructure retain or add appropriate marking. We consider this measure within legal scope and refer the group to existing analysis on that point.

Public reporting

Signatories operating under a voluntary code that confers a presumption of regulatory conformity bear a corresponding transparency obligation toward the public. We propose that summaries of compliance frameworks and testing and monitoring results be made publicly accessible, and that where full disclosure is not feasible, relevant documentation be made available upon request by any natural or legal person without requirement to demonstrate a specific interest, subject to protection of trade secrets and confidential commercial information. Equivalent redlines are proposed in both Measure 4.1 and Measure 4.2.

Working Group 2: Rules for labelling deepfakes and certain AI-generated and manipulated published text applicable to deployers of AI systems (Article 50(4) and (5) AI Act)

With regards to the second section of the Draft COP, we welcome and appreciate the changes made in order to bring the language closer to the AI Act, the addition of terms such as “Clear and user-friendly labelling” in terms of the COP direction and the stronger references in terms of accessibility and acknowledgements of vulnerable communities.

Neutral icon with interactive provenance disclosure vs. categorical labelling with a defined inference schema

Although the second draft has shown a lot of positive improvements, the ICON continues to be a contested point. In light of the phased approach to labeling introduced in this latest draft, we highlight a couple of points below. But above all, WITNESS urges the co-chairs to preserve the interactiveness of the icon, as well as the need for more context around content flagged as ai-generated or manipulated.

The current draft sits in an uncomfortable middle ground between two coherent approaches. Choosing one and committing to it fully is essential for consistency, enforceability, and accountability across the information ecosystem.

Option 1: Neutral icon with interactive provenance disclosure

A single, neutral EU icon functions as an entry point to underlying provenance data. On interaction, it surfaces the machine-readable marking information produced by providers under Section 1, letting users understand how content was made without the system rendering a categorical verdict about whether it is AI-generated or AI-manipulated.

- **Pros:** Can be applied immediately and consistently. Requires no prior agreement on a classification schema. Less technically and politically contested.
- **Cons:** Place the burden of interpretation on the user. From a UX perspective, a clear upfront signal is likely more impactful than expecting users to dig into provenance details. Lower immediate legibility may reduce real-world disclosure value.

Option 2: Categorical labelling with a defined inference schema

Deployers display a label –AI-generated, AI-manipulated, or out of scope– derived from the marking information produced under Section 1. This requires a transition layer: a shared schema that answers the questions a deployer must resolve in practice. **What qualifies as AI-generated? At what point does that become AI-manipulated? And at what point is content out of scope entirely?** Without answers to these questions baked into the Code, categorical labelling cannot function consistently.

- **Pros:** Clear and immediately legible for end users. Higher potential for awareness and impact.
- **Cons:** Requires developing and maintaining a shared inference schema -- a technically and normatively complex undertaking.

The second draft calls for categorical labelling while removing the taxonomy that would make such labelling coherent. **This produces the worst of both worlds.** Without a shared framework, each signatory will classify content according to its own interpretation of available signals. The same content could be labelled differently across platforms, generating confusion rather than transparency.

This inconsistency is compounded by the large volume of AI-generated and AI-assisted content that carries no markings at all. The result is a fragmented information environment: some content labelled as AI-generated in inconsistent ways, some labelled as AI-manipulated in inconsistent ways, and significant volumes of in-scope content left entirely unlabelled. **This does not just create confusion -- it creates a systematic bias in favor of labeled content**, with implications for how audiences perceive and trust unlabeled material.

WITNESS urges the co-chairs to commit to one of the options above. Absent that commitment, the Code must at minimum prescribe an industry-wide solution to prevent inconsistent implementation. Any approach adopted must preserve the following:

1. Need for more context around content flagged as ai-generated or manipulated content.
2. Possibilities for end users to interact with the icons adopted both at the EU level and the interim ones.
3. Interoperability of the labels and icons implemented in the interim of the development of the EU Common Icon.

Disclosures of Artistic Work

On creative work, we are happy with the discussion about non-intrusive placement in artistic works - However, since the second draft provides broader examples for non-digital contexts, such as art galleries or physical media, where disclosures can be provided at the point of entry or on tickets; we would just like to remind deployers that the solutions continue to be non-intrusive and perceivable by the users and audiences.

Despite that, on the issue of the artistic work disclosure, we would like to reinforce the same comments submitted earlier:

> W*ITNESS previous work - such as the report launched in 2021, named "Just Joking: Deepfakes, Satire, and the politics of Synthetic media"; and an article published in 2023 - help highlight the importance of dealing with labels as an inherent part of the content. Added to that, we also recommend that the COP deals with this issue as more than an add-on functionality and set a minimum requirement for disclosures. Creative work should also not be exempted from the disclosures obligation, especially if the common EU Icon is more generic (i.e. the i for information), as this can enable compliance and enforceability that may not be achievable if exceptions for subjective understandings of satire and art are made.*
>
> *A label that feels clear and unmistakable in one setting may become far less noticeable or intuitive when the same content moves to short-form video apps, messaging channels, or platforms with different design norms. Added to that, the perceived obviousness of a content and/or disclosure can also be shaped by factors such as age, language, cultural expectations, and varying levels of media literacy of the target audience, as well as potential reshares and remixes of existing content. This emphasizes the need for multi-layered marking techniques to help protect satire/creative content on the visible or audible layer, and enable underlying disclosure. It is worth highlighting here again the need to ensure that these techniques, while required, do not infringe on privacy, but rather focus on non-personal provenance.*

Compliance and Accountability

On the commitments related to Compliance measures, Awareness and Training, WITNESS would like to see stronger recommendations in terms of proper transparency mechanisms for deployers.

On the point of awareness raising efforts, it seems the requirements were reduced in this version and we consider it an important point and part of the shared responsibility of both tiers of actors this guide directs itself to. With regards to accountability, we would like to request the reinstatement of stronger compliance mechanisms and improving public accountability. In this sense, we would like to recommend

the addition of measures such as public access to data and improved reporting mechanisms that could enable broader civil society and Academia with the necessary information for them to perform oversight on the implementation of the measures emerging from the AI Act article 50 and this COP.

Media Literacy

With regards to the Media Literacy language, we urge the Co-chairs to reintroduce some of the measures as part of an acknowledgement of the shared responsibility developers and deployers have towards end users. We believe that the COP could advance in mentioning some examples of training that should be taken into account by signatories, such as: Training on how to detect AI-generated and manipulated content, training on how to implement the icons' in content. Campaigns directed towards end users on how to access and understand the Icons and further labels in AI-generated or manipulated content.

Good capacity building measures is what will make the full implementation of the COP feasible and avoid having end-users being lost in the implementation moments, or not knowing how to identify the icon and disclosures provided.

Ensuring that the icons, watermarks and disclosure methods are clearly understandable for users is key, along with providing appropriate guidance on its use and clarifying its limitations. In this sense, capacity building plays an essential role in this process, much like with any new icon. Strengthening media literacy as a core component of the COP will therefore be crucial to support effective adoption and understanding.

***

We appreciate the time and efforts deployed by the Co-Chairs and AI Office, and remain at your full disposal in case this present document or our submission raises any questions.

WITNESS Input to the Second Draft of the First Code of Practice on Transparent AI

Second Draft		WITNESS Rationale for redlines
Section 1: Rules for marking and detection of AIgenerated and manipulated content applicable to providers of generative AI systems (Article 50(2) and (5) AI Act)		
Commitment 1: Multi-layered Marking of AI-Generated	**Sub-measure 1.1.1: Digitally signed metadata** If content is generated or exported in a data format that supports adding information as part of the metadata (e.g., an audio, image, video, or document file), Signatories will record and embed through the metadata information whether the content is AI-generated or whether it is AI manipulated, an interoperable identifier that can be referenced by other layers (e.g., watermark/fingerprint) and information regarding how to access the provider's marking detection tool specified in Measure 2.1. Signatories will ensure the metadata marking complies with the quality requirements specified in Commitment 3. All added information will be digitally signed and time-stamped in a secure and tamper-evident manner. Signatories shall ensure that metadata markings do not include personally identifiable information by default. Where the inclusion of user-related or contextual data is strictly necessary for the purposes of the marking, or where a user has affirmatively chosen to include such information, such data shall be limited to what is adequate and relevant for those purposes, and Signatories shall implement appropriate measures to protect it. Control over such data shall rest with the relevant data controllers and rights-holders. Signatories will adopt means for ensuring the secure usage of the signing certificates and the confidentiality of the associated private keys. In the case of free text and other output types that are not available in a format that hosts metadata, Signatories are encouraged to implement an option that allows the download of a digitally signed manifest containing a certified version of the output generated or manipulated by their AI system to certify the artificially generated or manipulated origin of the text content. Such a provenance certificate will enable deployers and other users to provide third parties with certificates that the content is AI-generated or manipulated, linking it back to the specific generative AI system.	**Adding privacy protections** Sub-measures 1.1.2 and 1.1.3 both include explicit privacy language, but Sub-measure 1.1.1, which is the primary and most broadly applicable compliance pathway, does not. The implicit assumption appears to be that GDPR coverage is sufficient. We consider this inadequate: what constitutes personal data in the context of signed metadata is genuinely contested, and silence at the measure level will produce inconsistent practice across signatories. We also flag a concern that goes beyond this Code's strict scope: this Code will function as a reference framework for other jurisdictions, many of which have no equivalent to GDPR. Gaps in the primary layer will travel. Bringing 1.1.1 into alignment with the privacy principles already present in 1.1.2 and 1.1.3 costs nothing here and could matter significantly elsewhere.

	Measure 1.2: Non-removal of machine-readable marking In addition to the requirement in Article 50(2) AI Act, as detailed in measure 3.3. that requires high robustness of the marking techniques against expected downstream processing and adversarial attacks, Signatories will make best efforts to preserve marks on content generated or manipulated by their AI system by applying the following cumulative measures: a) Signatories will retain and abstain from altering or removing existing metadata to the extent technically feasible, including where such content is used as input and subsequently transformed by their AI system into a new output, except where content transformation requires updating marks to maintain accurate provenance chain~~; and~~ or where provenance data contains personal information and the data subject has requested its redaction; or where the signatory has determined that retaining such personal information in provenance records is not necessary for the purposes for which those records are maintained, provided that any removal pursuant to this exception is strictly limited to personal information and does not extend to other provenance data. b) Signatories will include in the acceptable use policy, terms and conditions of or the documentation accompanying their generative AI system a prohibition for the intentional removal of or tampering with the marks by deployers or any other third party, unless removal is undertaken for the purpose of benchmarking the security of a marking solution or any content transformations and editorial control are recorded in the provenance chain, where available. For AI systems and models provided under free and open licenses, it is sufficient for Signatories to alert users to this requirement in the documentation accompanying the AI system or the AI model without prejudice to the free and open-source nature of the license. Measures specified in point (a) above do not imply responsibility of the Signatory for third-party markings. Enabling detection of those marks remains the responsibility of the original provider of the generative AI system. Signatories who operate an online platform or search engine or who otherwise disseminate content to the public are encouraged to ensure that the platform or the search engine preserves metadata and other marks for AI-generated or manipulated content.	**Adding non-removal exception to PII:** Sub-measure (a) establishes a strong default against altering provenance metadata, but does not accommodate cases where that metadata contains personal information about individuals who did not consent to its inclusion or who face harm from its disclosure. Absent an explicit exception, a signatory who removes personal information to protect an individual is technically in breach, even where removal is the only privacy-protective option available. The amendment addresses this gap by permitting removal strictly limited to personal information, in two conditions: where the data subject has requested redaction, or where the signatory determines retention is unnecessary for the purposes of the provenance record. All other provenance data must remain intact.

	Measure 1.3: Structural Marking for open-weight AI models and systems Signatories releasing open-weight AI models or systems will implement structural marking techniques encoded in the weights during model training. This will facilitate third parties who use these open-weight models or systems to build generative AI systems to comply with Measure 2.2.	**Reinsert 1.3 from Draft 1** We support the inclusion of this measure and urge the group not to remove it in response to open-source community pushback. The objections raised, primarily around technical feasibility and the limits of weight-based marking, are legitimate and should be addressed in implementation guidance. They are not grounds for removing the obligation entirely. Without model-level marking at base model release, the framework produces a structural gap no downstream obligation can close. Closed commercial systems will be marked and detectable. Individual and informal use of open weights will carry no obligation at any point in the chain. That third stream is not an edge case: it is precisely where motivated disinformation actors operate, and they face zero friction from this framework as currently designed. The feasibility concern is real but addressable through a layered approach: base model releasers implement marking at training time, fine-tuners preserve existing marking where technically feasible, and hosting platforms ensure models distributed through their infrastructure retain or add appropriate marking. On robustness: current techniques are not fully resistant to aggressive fine-tuning or model merging. This is a reason to invest in more robust techniques, not to exempt base model releasers from the obligation. Imperfect friction is still friction. On legal scope, we will not expand here as others have addressed it more fully elsewhere. We consider this measure to be within scope.
	Measure ~~1.3~~1.4: Transparency of the provenance chain ~~(optional)~~ ~~Signatories are encouraged to apply provenance standards providing further information about the provenance chain of AI-generated or manipulated content across workflows where technically feasible for the specific modality.~~ Signatories shall not remove or alter existing system-level provenance information when processing content, and shall add their own marking to that chain distinguishing the operations performed by their AI system from previous operations. This obligation applies to the extent technically feasible for the specific modality. Signatories are further encouraged to apply provenance standards providing additional provenance chain information across workflows where technically feasible. In addition to the information recorded in the metadata pursuant to	**System provenance data mandated in chain; personal provenance data optional** Measure 1.3 (1.4 in our redlines) as drafted makes the entire provenance chain optional, apparently in part to protect privacy. We believe this trade-off is unnecessary and that it risks losing the system-level transparency that Article 50 is specifically designed to deliver. The relevant distinction is between system provenance, meaning what tools were used, in what sequence, and what operations were performed, and personal or contextual provenance, meaning data that links back to individuals, original sensitive files, or identifying circumstances. The first serves the transparency interest and should be mandatory. The second creates privacy risk and must be protected by default. These are separable obligations, and treating them as a single optional package means that privacy protection comes

	measure 1.1.1, Signatories may add or record other relevant content provenance information within their AI systems in a way that distinguishes the additional operation(s) performed by their AI system from previous operations, by leveraging metadata to record and verify the provenance chain where technically feasible. The other provenance information that Signatories are encouraged to record includes the AI system and underlying model identifier, version number, company name of the AI provider, and a timestamp indicating when the content was generated or manipulated. For AI-manipulated content, it is recommended that the metadata contains information about the type of the operation performed by the AI system to modify the content (e.g., object removal). Multiple discrete processing steps carried out by the AI system are recommended to be encoded into a single marker to constrain complexity and to reduce burden. Where a human carries out an operation in AI-human workflows, the only information that is recommended to be recorded is that a human carried out a given operation (e.g. editing). For the purposes of this Measure, a distinction shall be drawn between system provenance data, meaning information about the tools, models, operations, and sequence involved in content generation or manipulation, and personal or contextual provenance data, meaning information that links content back to individuals, original source files, or identifying circumstances. System provenance data shall be retained and recorded as set out above. Personal or contextual provenance data shall not be included in the provenance chain by default and shall only be recorded where the appropriate data controller or rights-holder has exercised affirmative control over its inclusion. In such cases, the human author may also encode on a voluntary basis other descriptive information, such as the organisation name and copyright, as applicable.	at the cost of accountability. This distinction is not novel: emerging legislation in other jurisdictions is already drawing exactly this line. The Code should make it explicit rather than leaving it to inference. **Note:** The concept of system provenance data is introduced here for the first time. For this distinction to be legally and technically operative, it will need to be defined and reflected in the relevant definitions section of the Code.
	Measure ~~1.4~~1.5: Optional functionality for perceptible markings (for deep fakes and AI-generated and manipulated published text) In order to facilitate compliance of deployers of generative AI systems with their obligation to disclose deep fakes and certain AI-generated and manipulated published text pursuant to Article 50(4) AI Act, Signatories who are providers of generative AI systems that are capable of generating or manipulating such content are encouraged to provide an optional functionality in their system's interface and implement an integrated option that allows deployers and other	No comment

	users to directly – upon generation of the output – apply at their own discretion a perceptible machine-readable mark or label. Signatories are encouraged to implement such a functionality for perceptible marks and/or labels in consistency with the Commitments and Measures in Section 2 of the Code. It is also recommended that Signatories follow harmonised UX standards, to the extent possible, in an interoperable manner with existing standardised content management systems and workflows of media publishers and online platforms. Signatories are also encouraged to implement other supporting measures for displaying labels and provenance metadata that enable deployers and providers of online platforms and websites to implement display practices and policies that are appropriate for their use cases. This measure is without prejudice to the responsibility of deployers who remain responsible for the disclosure of deep fakes and AI-generated or manipulated published text in a clear and distinguishable manner in accordance with Article 50(4) and (5) AI Act.	
Commitment 2: Detection of the Marking of AI-Generated Content	**Commitment 2: Detection of the Marking of AI-Generated Content** In order to fulfil their obligations under Article 50(2) and (5) AI Act to ensure that the outputs of their AI system(s) are detectable as AI-generated or manipulated, Signatories commit to implement the following measures to enable the detection of audio, image, video or text content, or a combination thereof, as generated or manipulated by their AI system and to ensure this information is provided to natural persons concerned in a clear, distinguishable and accessible manner through tools or APIs, and ideally also as forensic detectors.	**We strongly support this commitment,** and reject formalistic objections as well as cybersecurity and user interpretability concerns (which may have validity, but are addressable without undermining the requirement). For more information, see WITNESS input for Draft 1.
	Measure 2.4: Support literacy on AI marking technologies and verification Signatories are expected ~~encouraged~~ to ensure that layperson-oriented documentation and other relevant information (excluding trade secrets) is provided to deployers and other users to support them in making informed decisions on what marking and detection mechanism(s) they may use, including helping them to understand how to access and apply detection mechanisms and to interpret the provenance data and the detection results. In addition to deployer-focused materials, Signatories are also encouraged to ensure that end-user literacy resources are provided, as appropriate, and calibrated to end-user needs where the AI systems serve populations with lower AI literacy or in sensitive	**Supporting literacy should be required, not 'encouraged'**. Without it, the intent of the law is undermined.

	contexts (e.g. educational contexts, youth or elderly users). These materials may either be developed by the Signatories themselves or created jointly through efforts involving other providers or by organisations or initiatives they belong to. This measure should be implemented in a proportionate manner, taking into account the level of awareness of the deployers and other users of the generative AI system and end-users of the content, the size and resources of the provider, in particular with regard to SMEs and SMCs. Signatories are encouraged to collaborate with academia, civil society, media and other relevant organisations to promote literacy and awareness regarding AI content provenance and verification, and to support EU-level initiatives to foster consistent understanding of provenance and detection across Member States.	
Commitment 4: Testing, verification and compliance	**Commitment 4: Testing, verification and compliance** In order to effectively fulfil and demonstrate compliance with their obligations under Article 50(2) and (5) AI Act, as well as with the Commitments and Measures specified in this Section of the Code, Signatories commit to set up, keep up to date and implement testing, verification and compliance processes, as specified in the following measures.	**We strongly support this commitment.** **However, public accountability requires public reporting.** Compliance documentation, including testing, verification, and monitoring records, serves no accountability function if it remains internal. Signatories operating under a voluntary code that confers a presumption of regulatory conformity bear a corresponding obligation of transparency toward the public whose information ecosystem they are shaping. We call on the Chairs to ensure that such documentation is made publicly accessible, or at minimum subject to access upon request by any natural or legal person without requirement to demonstrate a specific interest.
	Measure 4.1: Compliance framework Signatories will draw up, implement, and update, in line with the state of the art, a compliance framework that outlines the marking and detection processes and the measures that the Signatories implement to ensure compliance with Article 50(2) and (5) AI Act and the Commitments and Measures in this Section. The framework will contain a high-level description of implemented and planned processes and of measures to adhere to this Section of the Code and to maintain and keep up to date relevant documentation to be shared with competent market surveillance authorities upon request. A summary of the compliance framework, including a high-level description of implemented marking and detection measures and their outcomes, shall be made publicly accessible. Where full public disclosure is not feasible, the relevant	As above, **we strongly support this** but call for **public reporting** or access-to-information procedures.

	documentation shall be made available upon request by any natural or legal person without requirement to demonstrate a specific interest, subject to the protection of trade secrets and confidential commercial information. This measure should be implemented in a proportionate manner, taking into account the size and resources of the provider, in particular with regard to SMEs and SMCs. Signatories can demonstrate compliance through existing processes and compliance frameworks to the extent that they fulfil the measures in this Section of the Code. Where Signatories rely on marking and detection solutions provided by third parties or implemented at the level of the generative AI model, Signatories will employ solutions for which those parties adhere to the Code and have demonstrated compliance with this Section of the Code and Article 50(2) and (5) AI Act. This possibility of reliance is without prejudice to the ultimate responsibility of the Signatory as a provider of the generative AI system to ensure compliance with Article 50(2) and (5) AI Act.	
	Measure 4.2: Testing, verification and monitoring Prior to the placement on the market and regularly thereafter, Signatories will test the marking and detection solutions for their compliance with the requirements and the measures specified in this Section of the Code in real-world conditions. Signatories who are downstream providers of generative AI systems may rely on results of testing performed by an upstream model or a third party provider of marking and detection techniques, as long as they comply with the requirements specified in this measure. In the context of testing and evaluation, Signatories will take into account available state-of-the-art benchmarks and other measurement and testing methodologies, including benchmarks and frameworks developed or recognised by the AI Office in collaboration with the AI Board. Such benchmarks should be updated in accordance with the state of the art and reflect realistic transformations and adversarial scenarios. Signatories may involve independent experts in the testing or conduct such testing and evaluation under regulatory supervision in the context of AI regulatory sandboxes as provided for in Article 57 AI Act. To ensure that marking and detection solutions are future-proof, Signatories will implement an adaptive threat modelling approach, moving beyond generic robustness benchmarks by defining realistic and	As above, **we strongly support this** but call for **public reporting** or access-to-information procedures.

	use-case specific threat scenarios (e.g., recompression, transcoding, speech-to-speech revoicing) to support the development of adaptive defence mechanisms. They will also track real-world degradations, periodically re-evaluate detection thresholds and update detection mechanism(s) to keep false positive rates low, while preserving detectability. Signatories will implement and document appropriate follow-up corrective actions on compliance shortcomings reported by deployers, independent researchers, civil society and other third parties and observed or reported adversarial attacks. A summary of testing, verification, and monitoring results, including identified compliance shortcomings and corrective actions taken, shall be made publicly accessible on a regular basis. Where full public disclosure is not feasible, the relevant documentation shall be made available upon request by any natural or legal person without requirement to demonstrate a specific interest, subject to the protection of trade secrets and confidential commercial information.	
Section 2: Rules for labelling deepfakes and certain AI-generated and manipulated published text applicable to deployers of AI systems (Article 50(4) and (5) AI Act)		
Recitals	Whereas: **a) Detection and disclosure:** Signatories acknowledge that technological advances in generative AI systems can enhance the realism and persuasiveness of AI-generated or manipulated content, increasing the importance of transparency mechanisms to safeguard public trust and democratic discourse. AI systems capable of generating or manipulating image, audio or video content that appreciably resembles existing persons, objects, places, entities or events may produce content which falsely appears authentic or truthful, raising specific risks for individuals, the integrity of the information ecosystem and democracy. Moreover, AI systems capable of generating or manipulating text that is published with the purpose of informing the public on matters of public interest should also be disclosed to natural persons. Clear and distinguishable disclosure of the artificial origin or manipulation of such content is a necessary safeguard to mitigate the risk of deception and reputational harm and to uphold trust as a public interest. **b) Clear and user-friendly labelling:** Signatories acknowledge that as deployers of AI systems generating or manipulating deep fakes and AI generated or manipulated text falling within the scope of Article 50(4) AI Act, they are responsible for labelling the output accordingly and for disclosing its artificial	We welcome the addition of terms such as "**Clear and user-friendly labelling**", as well as stronger references in terms of accessibility and acknowledgements of vulnerable communities. In the general references to content under the scope of this COP, we would recommend the adoption of the same language present in art 50 (4): *Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake.*

	origin or manipulation in a manner that is appropriate to the type of modality and context of dissemination. These responsibilities are complementary to the technical solutions implemented by providers under Article 50 (2) AI Act, contributing to increased transparency and trust along the AI value chain. Transparency measures should be user-friendly across the Union to strengthen the ability of the public to distinguish AI-generated or manipulated content and to support the resilience of the information ecosystem. **c) Artistic creation:** Signatories emphasise that, where the AI-generated or manipulated deep fake content forms part of an evidently artistic, creative, satirical, fictional or analogous work, transparency requirements apply in a proportionate manner. The disclosure of the existence of such AI-generated or manipulated deep fake content should therefore be implemented in a way that does not hamper the display, enjoyment, normal exploitation or creative quality of the work, while preserving appropriate safeguards for the rights and freedoms of third parties as enshrined in the Charter. **d) Accessibility:** Signatories emphasise the relevance of ensuring accessible disclosure to end-users exposed to the content, particularly in relation to different needs and vulnerabilities. Icons, labels and disclaimers should be designed in a way that ensures they are easily perceivable and understandable by persons with disabilities. This includes, for instance, providing alternative text for screen readers, audio disclosures for visually impaired users, sign language or captioned disclosures for hearing-impaired users, and ensuring sufficient colour contrast and readability. **e) AI literacy:** Signatories recognise that clear disclosure of AI-generated or manipulated deep fake content and AI generated text publications of public interest within the scope of Article 50(4) AI Act is essential for individual awareness and for supporting AI literacy. Public awareness and clear labelling of such AI-generated or manipulated content can further strengthen individuals' ability to distinguish synthetic content, thereby enhancing the practical impact of the transparency measures set out by this Code. **f) Additional safeguards under other Union and national law:** Signatories acknowledge that transparency obligations apply alongside, and do not replace, other legal responsibilities that may apply to the creation, distribution or use of AI-generated or manipulated content under applicable Union legislation on data protection, consumer protection, digital services	

	(Digital Services Act 1), intellectual property, media law (Audiovisual Media Services Directive2 and European Media Freedom Act3), political advertising, criminal law and other relevant regulatory frameworks	
Commitments	This Section of the Code applies only to Signatories in so far as they are deployers of AI systems that generate or manipulate deep fakes or published text with the purpose of informing the public on matters of public interest falling within the scope of Article 50(4) AI Act. Each reference below to "content" implies content that qualifies as a deep fake under Article 3(60) AI Act (referred to as 'deepfake') or, respectively, text published with the purpose of informing the public on matters of public interest, without human review or editorial control and where no natural or legal person holds editorial responsibility for the publication of the content (referred to as 'published text' or 'text published on matters of public interest').	We recommend the adoption of the full language of art. 50 (4) when it comes to the content generated by the deployers. I.e. *Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake.*
Commitment 1: Disclosure of AI-Generated and Manipulated Deep Fakes and Published Text	In order to fulfil their obligations under Article 50(4) and (5) AI Act, Signatories commit to ensure consistent disclosure of the artificial origin of AI-generated or manipulated deep fakes or published text on matters of public interest by using the uniform EU icon (once available) or choosing an alternative icon or labelling solution that follows the design and placement requirements specified in the following measures. The disclosure and labelling process may be integrated into existing disclosure and labelling practices of the Signatory to the extent that compliance with this Section of the Code and Article 50(4) and (5) AI Act is ensured. Signatories recognise that the use of labelling does not exempt them from other Union and Member States' laws, such as those related to the protection of third parties' rights and the fundamental freedoms, including applicable legal requirements with regard to obtaining consent of the depicted person or rights holders.	WITNESS would recommend that the COP language brings as much detail as possible in terms of the adoption of the Icon. In order to avoid discrepancies in the implementation and adoption of the Icon, or further disclosure options, it would be important for the COP to recommend an industry wide solution and format. Therefore, we trust that the language should be concerned on preserving the following: (a) need for more context around content flagged as ai-generated or manipulated content. (b) possibilities for end users to interact with the icons adopted both at the EU level and the interim ones. (c) Interoperability of the labels and icons implemented in the interim of the development of the EU Common Icon.
	Measure 1.1 Design requirements for icons, labels or disclaimers For content to which a visual icon or label can be applied, Signatories will implement the following design requirements: The icon or label will hold as the main visual element the capitalized acronym "AI" in the English language (e.g., an AI icon), possibly supplemented, where appropriate, with a short text label regarding the type of involvement of AI (e.g. "Generated with AI", "Made by AI" or "Manipulated with AI"). Where technically feasible and available this can	WITNESS appreciates the attempt present in this draft of providing minimum design requirements for the labels and icons developed in the interim of the EU Common Icon, as well as the addition of children, elderly people, and persons with disabilities as part of the target audiences for this measure. WITNESS would recommend that the COP language brings as much detail as possible in terms of the adoption of the Icon. In order to avoid discrepancies in the implementation and adoption of the Icon, or further disclosure options, it would be important for the COP to

	further be elaborated in a second layer detailing (e.g., what has been modified). ● The letters in both the acronym and the text will have the same vertical dimension; ● The icon or label can appear in different styles (e.g., colour and typography), as long as it remains clear, accessible, and distinguishable, i.e. readable and recognisable to all, including vulnerable categories of end-users that may be exposed to the content, such as children, elderly people, and persons with disabilities; ● The icon or label may appear in a different size depending on the context, while ensuring that it is clear and distinguishable. If resized, the proportions of the letters must be preserved; ● The contrast ratio will be maintained of at least 4:5:1 against the background. For audio-only content, Signatories will include, within the audio content itself, a short audible disclaimer, sound or signal in plain and simple natural language, in the same language as the content (where applicable), disclosing the artificial origin of the audio. Where appropriate, information regarding the type of involvement of the AI system may be provided (e.g. “Generated with AI”, “Made by AI” or “Manipulated with AI”). When applying the design requirements of this measure, Signatories will consider the potentially diverse composition of the audience exposed to the content (including diverging levels of AI and digital literacy, language proficiency or general knowledge) and the potential sensitive nature of the context in which the content is used (e.g. in the financial, medical, education or other high-risk sectors). Signatories will ensure accessible disclosure in different modalities in accordance with applicable Union law, including but not limited to: ● audio descriptions or alternative cues for visual disclosure elements; ● tactile cues for audio-only content, when the device used allows for such cues (e.g. a vibration alert before play), taking into account the needs of end-users with hearing impediments ● high contrast icons or labels and screen-reader compatibility, including for end-users with colour vision deficiencies;	recommend an industry wide solution and format. Therefore, we trust that the language should be concerned on preserving the following: (a) need for more context around content flagged as ai-generated or manipulated content. (b) possibilities for end users to interact with the icons adopted both at the EU level and the interim ones. (c) Interoperability of the labels and icons implemented in the interim of the development of the EU Common Icon. The suggested approach in this measure has been recently adopted by Meta and phased out after the label was proved to be misleading when added to content minimally edited (reference: https://www.crikey.com.au/2024/06/03/photographers-meta-instagram-made-with-ai-label/). In this sense, we would insist on a neutral approach that could eventually avoid acronyms and opt for a symbol instead. WITNESS will provide an additional set of comments regarding this point to the Co-Chairs. Although, we are concerned that the possibility opened by Commitment 1 and Measure 1.1, might introduce the possibility of non-standardized disclosure methods. In this sense, we would appreciate it if the design requirements could be expanded in order to include aspects such as interoperability, interactiveness and need for deployers to enable the AI-recipe to travel between platforms. Here, we would like to strongly emphasize that the approach suggested in Commitment one needs to work in tandem with the multilayered approach mentioned in Section one of this COP. This would allow us to help ensure that the disclosures persist even when content is reshared, edited, or moved across platforms, without relying solely on visible labels. Such an approach gives users the right level of information at the right time, while enabling deployers and developers to move beyond simplistic distinctions between “benign” and “malicious” content. It acknowledges the diversity of AI-generated and AI-modified expressions. The goal is not to police intent, but to provide consistent, trustworthy signalling at scale so people can understand how content was created and assess it appropriately.

	● detectability of the icon or the label by assistive technologies. Signatories are encouraged to implement any available relevant accessibility standard or guideline, including but not limited to the harmonised standard ETSI EN 301 549 "Accessibility requirements for ICT products and services" or the W3C Web Content Accessibility Guidelines 2.1, to the extent the Signatories' services or products fall within the scope of such standards or guidelines. The appendix provides sample icons and labels to be further discussed with the Code of Practice participants in order to support the development of a uniform EU icon that may be used by deployers in the execution of this measure. Once finalised, the EU icon will be made available under the Europe Union Public Licence and will be made available for download on an EU website.	
	Measure 1.2 Placement requirements for icons, labels or disclaimers To meet the legal requirements of labelling in a clear and distinguishable manner at the latest at the time of first exposure under Article 50(5) AI Act, Signatories will display the icon, label or disclaimer in an appropriate and perceivable position, in accordance with the content format and dissemination context, taking into account the following overarching principles applicable to all content modalities: ● The icon, label or disclaimer should be affixed on or directly embedded into the AI-generated or manipulated content. ● Where technically feasible, signatories will ensure that the icon, label or disclaimer always travels with the content to which it was applied. Deployers will collaborate on a best effort basis with actors whose services or products are used to further distribute or disseminate the content (e.g. publishers, online platforms or retail) to preserve applied icons, label or disclaimers consistently (including those applied in accordance with Commitment 3). ● Icons, labels or disclaimers should be clearly perceivable at the latest at the time of first exposure of a natural person to the content. ● For deep fake content that is part of artistic, creative, satirical, fictional or analogous works, the requirements detailed in Commitment 3 apply.	WITNESS is concerned with the possibility opened by Commitment 1 of non-standardized disclosure methods. In this sense, we would appreciate it if the design requirements could be expanded in order to include aspects such as interoperability, interactiveness and need for deployers to enable the AI-recipe to travel between platforms. Still on this point, we would like to highlight the relevance of the multilayered approach as it can help ensure that the disclosures persist even when content is reshared, edited, or moved across platforms, without relying solely on visible labels. This means that visible labels can provide simple, lightweight cues, while cryptographically signed provenance metadata—such as C2PA—anchors authenticity in a tamper-evident, machine-readable layer. Even when surface labels are altered or removed, embedded signatures allow platforms and tools to verify provenance. By combining the steps above, we believe that the disclosures model adopted by deployers and developers can make this process more meaningful, accessible, and resilient—while also addressing the higher levels of content available due to the AI Slop. Such an approach gives users the right level of information at the right time, while enabling deployers and developers to move beyond simplistic distinctions between "benign" and "malicious" content. It acknowledges the diversity of AI-generated and AI-modified expressions. The goal is not to police intent, but to provide consistent, trustworthy signalling

	Further specifications per type of modality are provided below. **Real-Time Video (multiple modalities)** For real-time deep fake video (including live television broadcasts or livestreaming), Signatories will display an icon or an alternative label consistently throughout the exposure where feasible. In the case where disclosure is done through audio, it should be presented simultaneously with the icon or label. Alternatively, Signatories can use a visual or audio disclaimer at the latest at the beginning and at regular intervals during the exposure that discloses that the content includes deep fakes. Such a disclaimer should be displayed or broadcasted for an appropriate duration to ensure Perceivability. **Non-Real-Time Video (multiple modalities)** For non-real-time deep fake video, Signatories will disclose that the video contains deep fakes with an icon or label. The Signatories may choose among the following disclosure options, individually or combined, as appropriate to the context: ● Long videos: icon or label at the latest at the beginning and repeated at regular intervals (e.g., when the specific deep fake content is shown and/or after commercial breaks). ● Short videos: icon or label consistently throughout the exposure from the beginning of the exposure. If the content of the video is entirely AI-generated and manipulated, this must be indicated throughout. The icon or label should clearly stand out and not be hidden (e.g. it should stand out from the background of the video and not be too close to other overlaying text and icons). ● In both cases: where disclosure is done through audio, it should be presented simultaneously with the visual icon or label. ● A disclaimer in the credits at the end of the video can be inserted. This measure always needs to be accompanied by one or more of the previous options. **Image (single modality)** Recognizing the cross-platform and cross-media transferability of deep fake images, Signatories will place an icon or label consistently at the latest at the first exposure and at any subsequent exposure to the image. The icon or label should be clearly distinguishable, particularly from the image itself, prominently visible, and not obscured or hidden (e.g. embedded within image layers, placed too close to	at scale so people can understand how content was created and assess it appropriately.

	other icons or text elements, or displayed against multiple backgrounds that reduce visibility). **Audio (single modality)** For deep fake audio-only content shorter than 30 seconds (e.g. commercials or advertisements), Signatories will include a short audible disclaimer at the latest at the beginning of the content. For longer deep fake audio formats, real-time as well as non-real-time (e.g. audio-only social media content, AI-generated phone calls, AI-generated podcasts or radio broadcasts). Signatories will provide repeated audible disclaimers at the beginning, at appropriate intermediate phases, and at the end of the content. Where deep fake audio content is delivered through a user interface and/or screen (e.g. car or smartphone display), Signatories will also display a visual icon or label via the elements of the user interface under their control, at the moment of the first exposure of the natural person, or upon initial access to the audio content (e.g. an icon or label embedded into a cover image or the title of the audio content). **Other Multimodal Content** For other multimodal deep fake content, Signatories will ensure that the multimodal content containing a deep fake is consistently disclosed using an icon or label, ensuring that the disclosure is clearly perceivable to the natural person without any further interaction on their part. Other multimodal content includes, but is not limited to, the following combinations of static or dynamic content: ● Image-text-audio; ● Text-audio; ● image-audio; ● Image-text. **Text** For AI-generated or manipulated text publications within the scope of Article 50(4) AI Act, Signatories will place the icon or label in a consistent position. This may be, for example, above or at the top of the text, near the headline of the text, or in the colophon at the beginning of the text, as long as placement is clear and distinguishable for the end-user within the type of text content published by the Signatory. If only part of the text publication is AI-generated or manipulated, it is sufficient to label only the part that is AI-generated or manipulated.	

	For short-form texts (single words or brief phrases), where labelling the text outputs would degrade readability, Signatories may ensure disclosure through contextual notice in the user interface or session (e.g., an indicator adjacent to the output, or session-level or page-level disclosure that AI was used).	
	Measure 1.3 Optional use of an EU icon and participation in its development Recognising the fast pace at which technology develops, Signatories are encouraged to use the EU-wide icon (to be specified in the Annex) and to support the further development of an optional uniform EU label designed to provide more advanced and usable information on the AI-generated or manipulated elements of content. For this purpose, Signatories are further encouraged to support the work and activities of a dedicated task force to be facilitated by the AI Office aimed at advancing the development, usability and development of such a label, in accordance with the following requirements: ● The taskforce will be established after the publication of this Code of Practice with the participation of Signatories and relevant stakeholders from various sectors and fields of expertise, national competent authorities and existing Chairs/Vice-chairs of the relevant working group. ● The taskforce will assist and advise the AI Office in the further development and possible testing of the icon, including aspects related to accessibility and modality-specific solutions, such as an audio-only content. ● The taskforce's Signatories who make use of the icon will support usability and user feedback to ensure the icon remains clear and distinguishable for all natural persons, while following the design requirement set out in Measure 1.1 and allowing state-of-the-art technological advancements for future iterations. ● The icon will be freely available to all deployers and other users, AI system providers and online intermediary service providers. Technical elements of the EU icon will be provided under free and open-source licenses allowing distribution and use (such as the European Union Public Licence). ● The taskforce will support the refinement of the EU icon in a manner that avoids information overload and ensures that the	WITNESS appreciates the expansion of the development process, as well as the possibility of link testing. While we still have doubts regarding the need of an acronym-guided icon, and would prefer to have a more neutral signal - we welcome the specific design requirements introduced in the text and would require the preservation of the interactiveness from the beginning of the icon implementation. We would like to reinforce again, the need for industry-wide and standardized methods of disclosure in order for users to fully understand the types of content they are being exposed to and how to address them. Having said that, we trust that in order for such an initiative such as an EU-Common icon to work, the approach needs to be made mandatory or have stronger incentives for the adoption and implementation by deployers. Added to that, we trust that the idea of convening a Taskforce could be an interesting space for exchanging ideas and learnings in the broader process of the icons, watermarks and other forms of disclosure approaches. In this sense, we would like to reinforce the need for these efforts to also take into account the expertise of Civil society and academic experts, in true Multistakeholder fashion. Lastly, as highlighted in a previous comment, it is of utter importance for the Icon to be able to carry additional interactive elements that can help avoid icon fatigue, and for users to be able to have access to relevant information at the first exposure moment, In this sense, we would like to recommend that the EU-Icon, once adopted, makes it mandatory for deployers to provide the additional interactive element.

	disclosure remains meaningful and usable for all end-users. ● The task force will discuss the ~~explore the possibility of~~ development an additional interactive element linked to the EU label (a "second layer") that will provide more detailed information about what has been generated or manipulated by AI. This second layer may be accessible by hovering over or clicking on the icon and may appear as a clear information field. Such interactive icon should follow the same design requirements as Measure 1.1, making it very easy for end-users at all levels of literacy and digital skills to discern the information provided. Possible characteristics of such interactive icon are: ○ Disclose at the very top or beginning of the information field (second layer) that content has been AI-generated or manipulated and specify the type of manipulation (e.g., fully AI-generated content or specific modifications such as the removal of an object), through machine-readable marking techniques implemented in accordance with the Section 1 of this Code, where available, and/or through deployer-provided information; ○ The information in the second layer can be displayed in English and be available in all the languages of the Member States through a translation plugin so it can be displayed in the native language of the natural persons exposed to the content. ● To the extent technologically and practically feasible, the refined EU icon will be designed to work in tandem with and to further integrate the machine-readable marking and detection solutions as described in Section 1 of the Code. Its implementation will remain practical and proportionate across deployers of all sizes and operational contexts, while avoiding over-labelling. ● For audio-only disclosures, the taskforce will aim to test the accessibility and usability of disclosures (i.e. audio disclaimers, sound logos or signals), considering various vulnerable categories and the proportionality of disclosure time relative to total content duration.	

	• Further the task force could function as a forum for exchanging good practices of AI literacy for promoting the labels in all Member States with the support of the Signatories.	
Commitment 2: Proportionate compliance, awareness and review	To ensure effective compliance with their obligations under Article 50(4) and (5) AI Act and the commitments and measures as specified in this Section of the Code, Signatories commit to implement proportionate internal processes, awareness measures and review mechanisms for the proper implementation of the labelling of deep fakes and text publications within the scope of Article 50(4) AI Act, taking into account their size and resources.	WITNESS would like to see stronger recommendations in terms of proper transparency mechanisms for deployers. On the point of awareness raising efforts, it seems the requirements were reduced in this version and we consider it an important point and part of the shared responsibility of both tiers of actors this guide directs itself to.
	Measure 2.1: Internal compliance Signatories will establish, adapt or maintain proportionate (existing) internal documentation or equivalent internal processes that specify how they implement the disclosure obligations under Article 50(4) and (5) AI Act. Such documentation or processes may include: • A general description of the disclosures applied across services or products, in accordance with Commitment 1; • A general description and representative, concrete and real examples of how disclosures are implemented in practice in accordance with Commitments 3 and 4, including deep fake content forming part of artistic, creative, satirical, fictional or analogous works. This description can clarify how the disclosure obligation under Article 50 (4) AI Act is applied to artistic, creative, satirical, fictional or analogous works, in accordance with Commitment 3. Or it can clarify when human review and editorial responsibility is involved in AI-generated or manipulated text publications on matters of public interest, in accordance with Commitment 4 and the related decision-making processes.	WITNESS would appreciate stronger support for the reinstatement of stronger compliance mechanisms. This could be directed towards public access to data on compliance levels and improved levels of public reporting.
	Measure 2.2: Awareness and Training Signatories will make reasonable and proportionate efforts to ensure awareness of the disclosure obligations under Article 50(4) and (5) AI Act among personnel, including employees and external contractors, directly involved in the implementation of labelling measures or overseeing compliance with the	WITNESS considers the point on capacity building and awareness one of the core ones in this document. Good capacity building measures is what will make the full implementation of the COP feasible and avoid having end-users being lost in the implementation moments, or not knowing how to identify the icon and disclosures provided.

	measures in this Section of the Code. Signatories are encouraged to provide training or equivalent guidance covering situations in which disclosure is legally required, how disclosures are implemented in the workflow, cases when editorial responsibility is involved or cases of artistic, creative, satirical, fictional or analogous work; accessibility considerations and procedures for correcting missing or incorrect labels when these have been identified. Training should be proportionate to the size and resources of the Signatory and applied only to the extent to which personnel (considering their technical knowledge, experience, and education) are involved in creating, modifying, and disseminating relevant content. Signatories remain free to determine the training formats and their frequency.	Ensuring that the icons, watermarks and disclosure methods are clearly understandable for users is key, along with providing appropriate guidance on its use and clarifying its limitations. In this sense, capacity building plays an essential role in this process, much like with any new icon. Strengthening media literacy as a core component of the COP will therefore be crucial to support effective adoption and understanding. Lastly, with regards to which types of training would suffice the scope of this COP, we would recommend things such as: Training on how to detect AI-generated and manipulated content, training on how to implement the icons' in content. Campaigns directed towards end users on how to access and understand the Icons and further labels in AI-generated or manipulated content.
	Measure 2.3: Review, feedback and cooperation with authorities Signatories will support effective implementation of the disclosure obligations through review and feedback mechanisms. Specifically, Signatories are expected ~~encouraged~~ to provide channels that allow individuals or third parties (trusted flaggers, independent fact-checkers etc.) to flag missing or incorrect disclosures, preferably through existing reporting mechanisms (e.g., trusted flagger mechanisms, interfaces for third-party fact-checking services or notice and action mechanism). Signatories will review cases that have been reported or observed as mislabelled or non-correctly labelled and remedy disclosures without undue delay. Signatories will cooperate with competent authorities in accordance with applicable European Union and national laws.	Our evaluation is that the current language can be strengthened in the direction of stating stronger compliance mechanisms and improving public accountability. WITNESS would like to recommend the addition of measures such as public access to data and improved reporting mechanisms that could enable broader civil society and Academia with the necessary information for them to perform oversight on the implementation of the measures emerging from the AI Act article 50 and this COP.
Commitment 3: Appropriate Disclosure for Artistic, Creative and similar Works	To fulfil their obligations in Article 50(4) and (5) AI Act, Signatories commit to implement measures to disclose deep fake content that forms part of evidently artistic, creative, satirical, fictional or analogous work or programmes. Pursuant to Article 50(4) AI Act, such disclosure is limited to 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, including its normal exploitation and use, while maintaining the utility and quality of the work. Where feasible, the disclosure should always travel with the content.	We welcome the references to **EU accessibility law and standards**. Despite that, on the issue of the artistic work disclosure, we would like to reinforce the same comments submitted earlier: *WITNESS previous work - such as the report launched in 2021, named "**Just Joking: Deepfakes, Satire, and the politics of Synthetic media**"; and an article published in 2023 - help highlight the importance of dealing with labels as an inherent part of the content. Added to that, we also recommend that the COP deals with this issue as more than an add-on functionality and set a minimum requirement for disclosures. Creative work should also not be exempted from the disclosures*

	Signatories will use an icon, label or disclaimer following the design requirements in Measure 1.1. and will place it in a manner appropriate to the type of artistic, creative, satirical, fictional or analogous content and to the context in which it is presented. Such placement needs to be clear and distinguishable to end-users and provided at the latest at the time of first exposure to the content. Where relevant, this can be complemented with end credits, contextual or creative disclosure methods, and post-viewing disclaimers. The disclosure will be placed in a non- intrusive yet effective (i.e. clear and distinguishable) position, which may include, but is not limited to, the following: ● Real-time or near real-time video: at the latest at the time of the first exposure in the top or bottom corners for at least five seconds without further warnings throughout exposure (e.g. during opening credits); ● Video: the disclosure will be placed for a timing sufficient to inform the viewer at the latest at the time of first exposure without interfering with the experience (e.g. during opening credits); ● Other multimodal content: the disclosure will be displayed at the latest at the time of first exposure, ensuring that the disclosure is clearly visible to the natural person without requiring any further interaction on their part; ● Image: at the latest at the time of the first exposure in an appropriate place with the possibility of integrating it into the image or the background of the image while preserving the ability for the end-user to discern the labelling; ● Audio: an audible disclaimer should be inserted at the latest at the time of the first exposure. Signatories can also conceive and implement 'contextual' disclosure solutions where the disclosure options mentioned above are not available or would affect the display or enjoyment of the work. When deep fake content is made available in a digital and/or interactive manner (e.g. on websites, apps or other user interfaces), the icon, label or disclaimer can be placed outside but adjacent to the video or image frame, or adjacent to the audio content and integrated into user interface elements or overlays under the control of the Signatories. Such a contextual disclosure solution should be perceivable by the end-user without the need for scrolling or additional engagement.	*obligation, especially if the common EU Icon is more generic (i.e. the i for information), as this can enable compliance and enforceability that may not be achievable if exceptions for subjective understandings of satire and art are made.* *A label that feels clear and unmistakable in one setting may become far less noticeable or intuitive when the same content moves to short-form video apps, messaging channels, or platforms with different design norms. Added to that, the perceived obviousness of a content and/or disclosure can also be shaped by factors such as age, language, cultural expectations, and varying levels of media literacy of the target audience, as well as potential reshares and remixes of existing content. This emphasizes the need for multi-layered marking techniques to help protect satire/creative content on the visible or audible layer, and enable underlying disclosure. It is worth highlighting here again the need to ensure that these techniques, while required, do not infringe on privacy, but rather focus on non-personal provenance.*

	Where content is made available in a non-digital or non-interactive manner (e.g., exhibitions, art galleries, festivals or comparable contexts, audio or video on a physical carrier), disclosures can be provided, e.g. at the point of entry, when tickets are sold, or as part of introductory information or information provided via a physical carrier. Disclosure should be clear, accessible and understandable to all audiences. Where feasible, disclosure methods applicable to deep fake content made available digitally should remain attached to or travel with the content when it is shared or distributed.	