Google's response to the Second Draft of Art.50 Code of Practice under the AI Act

Introduction

The first draft (**TCoP 1**) of the Code of Practice on Transparency of AI-Generated Content (**TCoP**) suffered from numerous deficiencies. Among other things, it introduced new requirements with no basis in the AIA, contained numerous requirements that would violate the fundamental EU legal principle of proportionality, and took an inappropriately prescriptive and inflexible approach.

The second draft of the TCoP (**TCoP 2**) represents an improvement and we welcome many of the changes introduced. However, TCoP 2 still suffers from deficiencies: many provisions still lack a clear legal basis in the AIA, they are disproportionate or uncertain, or they are overly granular and rigid, and thus likely to be unworkable in the rapidly-evolving ecosystem of AI transparency.

We support the TCoP's core objective of limiting the harmful impact of deceptive or misleading AI-generated content, but are concerned that, without further changes to address these points, the TCoP risks becoming an unattractive option for signatories to demonstrate compliance with Art. 50 of the EU's Artificial Intelligence Act (**AIA**).

In this response, we explain the issues that remain in TCoP 2 and, to assist in arriving at a workable third draft, we set out proposals for proportionate and targeted improvements. These improvements focus on the following key themes:

1. **More Flexibility, Less Granularity.** The TCoP will be a voluntary instrument. To ensure broad adoption, it should focus on providing flexible guidance and examples of compliance, rather than setting inflexible, granular requirements which may not work in practice. Such an approach is particularly appropriate given the fast-moving and varied contexts in which the TCoP will apply. Generative AI technologies are developing rapidly, as are marking technologies and public awareness of the risks of AI-generated or AI-manipulated content. The requirements set out in the TCoP will apply in a wide range of different environments, from smart-phone apps to desktop browsers, and from televisions to art galleries and magazines. These circumstances call for flexibility. We suggest necessary improvements to address this issue, including in Points 2, 3, 8, 10, 11 and 12 below, particularly the introduction of single-level marking under Art. 50(2) (see Point 2) and provision for contextual disclosure of 'deep fake' content under Art. 50(4) (see Points 10-12).

2. **Focus on Core Requirements, Avoid Ancillary Measures.** The core goals of Arts. 50(2) and 50(4) are conceptually straightforward: to ensure that AI-generated content is marked in a machine-readable format and therefore detectable as such, and to ensure that certain types of AI-generated content are accompanied by further, visible disclosures. These points should be the key focus of the TCoP. However, TCoP 2 also sets out a wide range of additional procedural and other requirements that are, at best, ancillary to these core goals. For example, Measure 2.4 of

Section 1 sets out requirements on literacy on AI marking technologies, Measure 4.1 of Section 1 requires providers to create a "*compliance framework*", Measure 4.2 of Section 1 requires testing, verification and monitoring measures, and Measures 4.3 of Section 1 and 2.2 of Section 2 set out requirements for training. These requirements stray beyond the TCoP's legally-defined goal of "*facilitat*[ing] *the effective implementation*" of the marking and labelling requirements in Arts. 50(2) and 50(4). To ensure the TCoP facilitates the effective implementation of Art. 50, rather than introducing new requirements (which will be particularly dissuasive to organisations considering whether to follow the TCoP), we suggest targeted improvements and the removal of some requirements; in particular in Points 7, 8, 9, 13 and 15 below.

3. **Legal Certainty on Status of Commitments and Measures.** TCoP 2 improves on TCoP 1 by confirming that various measures are not essential for compliance (as further detailed in Point 1 below). While we welcome this increased flexibility, the current heterogenous language used to denote voluntary provisions means that the ultimate legal effect of many requirements is ambiguous. It is unclear, for example, whether there is intended to be a distinction between steps that a signatory is "*encouraged*" to or "*may*" take compared to explicitly "*optional supplementary*" measures. To avoid this uncertainty, we suggest that a small number of requirements be removed to make the TCoP workable in practice (see Points 4, 5, 14) and, in Point 1, we explain how we think this aspect of the TCoP could be improved, by reframing any remaining voluntary measures as explicitly "*optional*" (rather than "*encouragements*" or "*recommendations*") to make it clear that these steps are not required to demonstrate compliance with Art. 50 of the AIA. This will avoid confusion and allow providers and deployers to fully understand the importance of each Measure and prioritise those that are most necessary for compliance.

We invite the Chairs and the AI Office to introduce these much-needed improvements to ensure the TCoP's success as a lawful, proportionate and innovation-friendly tool for compliance with Art. 50 of the AIA.

Many of our previous comments in response to TCoP 1 also still stand, and we may share further comments when further drafts are provided. In addition, this Response proceeds on the basis that the scope of Arts. 50(2) and (4), e.g. the interpretation of exemptions and the definition of terms such as 'deep fake' (including the need for deceptiveness), 'human review and editorial control', and 'public interest', as well as the scale of text or "*edits*" deemed substantial enough to require labelling, will be addressed in the Commission's forthcoming guidance on Art. 50 rather than in the TCoP. This Response therefore does not address these topics.

Targeted improvements for TCoP 2

Issue No.	TCoP 2 Reference	Summary of Issue / Recommendation
1	**S1 M1.1.3, M1.3-1.4, M2.2, M2.4; S2 M1.3 – Optional / voluntary measures**	TCoP 2 uses inconsistent terminology for non-mandatory provisions. **Recommendation:** Standardise the language used for all non-mandatory provisions so it is clear they are optional supplementary measures, and state expressly at the outset that such optional measures are not required to demonstrate compliance with Art. 50(2) AIA.
2	**S1 M1.1, M1.1.3 – 'Multi-layered' marking / fingerprinting & logging**	Measure 1.1 still mandates 'multi-layered' marking by treating the Art. 50(2) quality criteria as absolute standards that can only be met through this approach; Measure 1.1.3 assumes that any "deficiency" should be addressed via fingerprinting/logging. This ignores Art. 50(2)'s qualifications of technical feasibility, state of the art and cost. It also fails to reflect the express exceptions for assistive/standard editing functions. **Recommendation:** Remove the mandatory multi-layer requirement in Measure 1.1 and require only a single layer of marking; state that it applies only where Art. 50(2) actually requires marking (explicitly excluding the Art. 50(2) exceptions); and amend Measure 1.1.3 to clarify that fingerprinting/logging is not appropriate or expected where limitations in marking techniques simply reflect the current state of the art.
3	**S1 M1.1.1-1.1.3 – Prescriptive marking techniques (e.g. metadata, digitally signed manifest, imperceptible watermark plus a publicly readable watermark, fingerprinting / logging)**	These techniques are prescribed in a one-size-fits-all way, contrary to Art. 50(2)'s requirement that providers choose appropriate marking methods considering content type, implementation costs and state of the art. Several also present significant technical and practical problems - e.g. requirements that markings be identifiable from a 'fragment'. The publicly readable watermark requirement manifestly exceeds Art. 50(2). **Recommendation:** Recast Measures 1.1.1–1.1.3 (other than the "publicly readable watermark") as an illustrative list of possible techniques; delete the publicly readable watermark requirement entirely and remove the "fragment"-based detection requirement.
4	**S1 M1.1.2 – Voluntary provisions on model-level marking**	Model-level approaches are currently inferior in robustness and output quality, and their promotion in TCoP 2 (even as voluntary measures) risks distorting incentives and encouraging undue downstream reliance on experimental upstream techniques. **Recommendation:** Remove this measure.

5	**S1 M1.4 – Voluntary provisions regarding functionality for deployers to add labels to systems' output**	Encouraging providers to provide functionality allowing deployers to add visible labels to outputs goes beyond Art. 50(2), may not be the most effective way to implement labels (deployers are better placed to determine how and when to label), and risks distorting roles along the value chain while diverting resources into early-stage, experimental features. **Recommendation:** Remove these provisions.
6	**S1 M1.1, M2.1, M4.1-4.2 – Reliance on third-party marking, detection and testing**	Allowing reliance on third-party marking/detection/testing only where that third party adheres to the TCoP effectively makes TCoP-adherence a precondition for supplying marking tools, distorts competition, and may deter investment in marking technologies, including by actors who are not "providers" under the AI Act. **Recommendation:** Delete conditions in Measures 1.1, 2.1, 4.1 and 4.2 that restrict reliance on third-party solutions to TCoP-compliant providers.
7	**S1 M1.2 – Non-removal of markings & contractual prohibitions**	Requirements to preserve existing metadata/marks and to prohibit removal of marks in terms/AUPs are over-broad, may conflict with legitimate functionality or privacy needs (e.g. removal of unreliable or personal-data-bearing metadata), may be technically infeasible, and improperly shift enforcement of Art. 50(2) onto providers (including via expectations placed on platforms and websites). **Recommendation:** Remove these provisions.
8	**S1 M3.1-3.4 – Quality criteria for marking and detection**	The interpretations of effectiveness, reliability, robustness and interoperability are over-prescriptive and sometimes circular, conflate quality criteria with particular technical solutions (e.g. detection mechanisms), understate operational constraints and implementation costs despite the reference to these in Art. 50(2), and prioritise interoperability (including shared repositories and a provider-agnostic detection interface) over robustness, security and protection of trade secrets. **Recommendation:** Refocus Measures 3.1–3.4 on clear, non-circular (and non-prescriptive) quality criteria; recognise that operational constraints may be taken into account when implementing marking / detection solutions; decouple "effectiveness" from specific measures such as detection tools; treat listed metrics as examples only; and delete obligations to cooperate on (i) an interoperable provider-agnostic detection interface and (ii) shared repositories of watermark/detector information.

9	**S1 M2.1-2.3 – Requirements to provide free detection mechanisms & disclose results in way that is comprehensible to lay persons**	The provision of detection tools and detailed disclosure of detection results, as well as EU-only hosting of solutions and a de facto ban on deprecating tools, all go beyond Arts. 50(2) and (5), and create cybersecurity risk by exposing marking performance and attack surfaces. **Recommendation:** Remove the requirement altogether or downgrade to voluntary. If retained, (i) allow providers to withdraw detection tools, (ii) require detectors only where they do not compromise the security or robustness of markings, (iii) remove requirements for signed downloadable reports and alignment with Measure 3.4, (iv) limit outputs to a simple Detected / Not Detected status, (v) drop the obligation to provide confidence scores and human-understandable explanations, (vi) remove the EU-hosting/"implementable locally" constraint, and (vii) delete the optional forensic detection measure in Measure 2.2.
10	**S2 M1.1-1.2 – Design and placement of deepfake disclosures**	Prescriptive design requirements (e.g. minimum contrast ratio) and rigid placement rules (disclosures must be "affixed on or directly embedded in" the content and "travel with" it) go beyond Arts. 50(4) and (5), are technically challenging and will obstruct users from identifying AI-generated information. **Recommendation:** Replace restrictive and unhelpful design and placement prescriptions in Measures 1.1 and 1.2 with a higher-level requirement for deployers to make disclosures in an appropriate clear and distinguishable manner based on guiding factors (suggested at Issue 11 below).
11	**S2 M1.2 – Modality-specific disclosure rules**	Highly granular modality-specific disclosure requirements often exceed Art. 50(5)'s requirement to disclose "*at the latest at the time of the first interaction or exposure*", are practically unworkable given the nature of certain content types, contain points of unclarity and are overly intrusive which may push users to less compliant content. **Recommendation:** Replace prescriptive modality-specific placement requirements in Measure 1.2 with a higher level requirement for deployers to implement appropriate clear and distinguishable disclosure based on guiding factors including modality as well as others, e.g. content duration, content size, livestreaming and disclosure prominence, which does not go beyond the AIA.

12	**S2 C3 – Deepfake disclosures for creative / artistic works**	Disclosure requirements for creative, satirical, fictional or analogous works are more prescriptive than Arts. 50(4) and (5) require, and leave little practical difference from non-creative use cases, are overly intrusive and hamper enjoyment, which may push users to less compliant content. **Recommendation:** Remove modality-specific disclosure requirements for creative works in Commitment 3 and apply the current more flexible "contextual disclosure" approach to all creative works disclosures, avoiding prescriptive requirements.
13	**S2 C4 – Human review / editorial control exemption**	Commitment 4 imposes onerous, disproportionate procedural requirements for deployers to justify reliance on the Art. 50(4) exemption which have no basis in the AIA. **Recommendation:** Remove Commitment 4 entirely. If the Code addresses the Art. 50(4) exemption at all, it should go no further than encouraging deployers to have a policy on human review and editorial control.
14	**S1 M3.5; S2 M1.3 – Use and development of transparency solutions and related taskforces**	Measures encouraging use and development of transparency solutions (including supporting related taskforces) are inappropriate for the Code's purpose of effective AIA implementation, dissuading organisations from signing the Code. They create confusion, particularly given their extensiveness, creating the risk of being inappropriately picked up by other regulatory bodies and undermining the state of the art. **Recommendation:** Remove Measures 1.3 (Section 2) and 3.5 (Section 1). If they are addressed at all, this should be in a separate document which is not part of AIA guidance / Code documentation.
15	**S1 M2.4, M4.1-4.4; S2 M2.1-2.3 – Compliance frameworks, training, monitoring & cooperation**	Measures on literacy, internal compliance, testing, verification and monitoring, training, incorrect disclosure feedback and review, and cooperation with authorities - which introduce governance and process obligations - have no basis in, or go beyond, the AIA. In any event, they are disproportionate and often unworkable, and dissuade organisations from signing the Code. **Recommendation:** Remove Measures 2.4 and 4.1 - 4.4 (Section 1) and 2.1 - 2.3 (Section 2), or alternatively, at the very least, make them optional.

1. Ensure clear delineation between mandatory and optional measures

TCoP 2 helpfully clarifies that certain measures are optional, supplementary measures that are not mandatory. However, TCoP 2 uses inconsistent language to indicate that measures are not mandatory

requirements, leaving it unclear what the expectations of regulators and consumers from AI providers and deployers should be. To avoid such confusion, TCoP 2 should use consistent, clear language to indicate where measures are optional rather than mandatory.

- **Issue:** TCoP 2 makes many of the provisions from TCoP 1 voluntary instead of mandatory, but does not do so consistently. Requirements are variously recast as "*optional supplementary*" measures (see Measure 1.1.3 and Measure 2.2 of Section 1), as simply "*optional*" (see Measure 1.3 of Section 1), as steps that providers are "*encouraged*" to take (see e.g. Measures 1.4 and 2.4 of Section 1 and Measure 1.3 of Section 2) or as steps that it is "*recommended*" that signatories should take (see e.g. Measure 1.4 of Section 1).
    - While we suggest removing several of these voluntary points entirely (see Issues 4 and 5 below), any optional requirements that do remain will, without clarification, give rise to unnecessary uncertainty as to what steps signatories *must* take to comply with the TCoP and demonstrate compliance with the AIA.
    - The current heterogenous language leaves the ultimate legal effect of many requirements ambiguous. It is unclear, for example, whether there is intended to be a distinction between steps that a signatory is "*encouraged*" to take compared to explicitly "*optional supplementary*" measures.
    - This lack of clarity creates unnecessary ambiguity as to the standard to which signatories may be held by regulators, deployers or consumers, and may therefore deter potential signatories and set unclear standards for compliance for non signatories. It is also likely to lead to divergent approaches by different signatories, who will inevitably interpret the various terminology used in diverging ways.
    - This risk is compounded in circumstances where the content of these measures may be picked up inappropriately by other bodies, across other jurisdictions, and potentially used as a benchmark or expectation about marking and detection requirements. This undermines the purpose of these provisions, which is to facilitate the advancement of state-of-the-art transparency solutions; not to set expectations.
- **How to fix:** To remedy these issues, the language used to denote requirements that are not mandatory should be standardised to make it clear that these measures are optional and supplementary. To ensure that the effect of these measures is clear, the TCoP should also clarify at the outset that such optional supplementary measures are not required in order to demonstrate compliance with Art. 50(2) of the AIA.

2. Align marking requirements with the text of the AIA: require only a single level of marking for content and reflect the exemptions in Art. 50(2) (Measure 1.1 of Section 1)

We welcome the acknowledgment in TCoP 2 that a single marking approach may be sufficient to comply in the future with Art. 50(2). Art. 50(2) and Rec. (133) require that marking techniques be assessed against the four quality criteria in Art. 50(2) by reference to technical feasibility and the *current* state of the art, rather than what might be available in the future. This flexibility must be reflected in the TCoP. In addition, the TCoP should include an express confirmation that marking requirements do not extend to AI systems falling within the exceptions for assistive functions and standard editing provided for in Art. 50(2).

- **Issue:** Art. 50(2) requires that providers ensure that their systems' synthetic outputs are marked in "*a machine-readable format*". It does not specify *how* providers should do this, providing only that technical solutions must be "*effective, interoperable, robust and reliable as far as this is technically feasible*", taking into account specificities and limitations of types of content, costs of implementation, and the generally acknowledged state of the art. Particularly in this light, the requirement in Measure 1.1 that signatories must adopt a multi-layered approach to marking poses a number of serious issues.

    - Art 50(2) requires only that content be "*marked in a machine-readable format*". This wording clearly leaves open the possibility that content will be marked using a single machine-readable format, rather than multiple layers of marking. It follows that the starting point of Measure 1.1 - the requirement that outputs of generative AI systems must be marked "*with at least two layers of machine-readable active marking"* - is inconsistent with Art. 50(2).

    - While TCoP 2 acknowledges that signatories may be able to demonstrate compliance using a single layer of marking "*in the future*", the basis on which they may be able to do so is highly uncertain. TCoP 2 specifies that a single solution may be acceptable if a signatory can prove that it achieves "*at least the same, if not superior, level of robustness, reliability, effectiveness, and interoperability*" based on "*independently verified benchmarks*". It is entirely unclear what comparator should be used for this assessment. Put simply: At least the same as what? It is unclear, for example, if this wording contemplates a comparison between a single, novel solution and any two existing solutions, or if it contemplates a novel solution that is superior to any existing combination of solutions. This is entirely unworkable and, in either case, there is no basis in the AIA for such a test.

    - This confusion appears to stem from the fact that TCoP 2 seeks to reframe the requirements that marking solutions be robust, reliable, effective and interoperable as absolute standards set by reference to the use of multi-layered marking and without regard to the current nascency of marking technologies. Art. 50(2) explicitly qualifies the marking obligation by reference to what is technically feasible and the "*generally acknowledged state of the art*", i.e. the technical limitations applicable to marking *now* (rather than the technology that may be available in the future). Technical feasibility and the generally acknowledged state of the art are the only qualifications that should apply to the assessment of whether any marking technique or combination of techniques meets the quality criteria in Art. 50(2). If a single layer of marking meets the quality criteria by reference to these qualifiers, then that should be sufficient for the purposes of Art. 50(2).

    - Measure 1.1.3 of Section 1 (regarding fingerprinting or logging facilities) is affected by the same problem: while it is framed as an "*optional*" measure to address deficiencies where "*appropriate*" in respect of a technique's "*effectiveness, reliability and robustness*" or other limitations, it fails to acknowledge that Art. 50(2) does not require deficiencies to be addressed where the marking solution reflects the state of the art.

    - This is symptomatic of the maximalist, multi-layered approach to machine-readable markings adopted in TCoP 2. The introduction of such new requirements not found in Art. 50(2) is likely to substantially increase the compliance burden associated with signing the TCoP. Quite aside from going beyond the AIA, this is likely to deter providers from signing

and incentivise them to explore alternative means of complying with the Art. 50(2) obligation.

- In addition, TCoP 2 does not contain any express confirmation that AI systems performing assistive functions for standard editing or which do not substantially alter the input data provided by the deployer or the semantics thereof (i.e. functionality which is expressly carved out of Art. 50(2)) are exempted from Section 1's requirements. Measure 1 as currently framed would therefore appear to mandate that *all* synthetic content (i.e. including from such exempted AI systems) will need to be marked in the specified manner. This would not be consistent with Art. 50(2).

- **How to fix:** Remove the requirement in Measure 1.1 to apply a multi-layered marking approach, and instead require only a single level of marking of content. In addition, incorporate an express disclaimer that this Measure applies only to the extent that markings are required under Art. 50(2) (i.e. clarify that it does not apply to the excepted use cases in Art. 50(2)), and revise Rec. (b) to mirror this. In addition, revise Measure 1.1.3 to make clear that the implementation of fingerprinting or logging facilities will not be appropriate or otherwise expected where deficiencies in marking techniques reflect the current state of the art.

3. Confine marking requirements to one mandatory marking, providing a non-exhaustive list of suggestions, and remove "publically readable watermark" requirement (**Measures 1.1.1-1.1.3 of Section 1)**

Art. 50(2) gives providers the flexibility to determine the marking techniques appropriate to their AI system. We therefore welcome the revisions in TCoP 2 recasting as optional the requirements to apply a digitally signed manifest for modalities without embedded metadata (Measure 1.1.1 of Section 1) and to apply fingerprinting and logging (Measure 1.1.3 of Section 1). However, this same principle should be extended to other measures in TCoP 2; prescriptive requirements applicable to specific modalities and scenarios should be removed.

- **Issue:** Art. 50(2) states that providers should take into account "*the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art"* when determining what marking methods to apply. This wording requires that providers are granted flexibility to determine appropriate marking methods by reference to these factors. It would be inappropriate for the TCoP to prescribe a one-size-fits all approach that ignores this reality. However, this is precisely the approach adopted in TCoP 2.

  - TCoP 2 mandates that various forms of marking be applied to certain types of content, including digitally signed metadata or a downloadable manifest (Measure 1.1.1), both an "*imperceptible watermark*" and an "*additional publicly readable watermark*" (Measure 1.1.2), and optional additional fingerprinting or logging (Measure 1.1.3).

  - As well as being incompatible with the text of the AIA, there are also numerous practical difficulties in implementing the proposed techniques:

Technique	**Issue**
Metadata	• The carveout from this measure for text and other content which does not support metadata hosting is welcome, but does not give

Measure 1.1.1 requires signatories to record and embed prescribed information in content's metadata (where data format supports this), with a digital signature and timestamp in a secure, tamper-evident manner.	providers sufficient flexibility. For example, there may be instances where content is theoretically capable of hosting metadata, but this would materially impact the output, e.g. video playback optimisation. In circumstances where providers are required under Art. 50(2) to apply marking which complies with the quality criteria, it is unnecessary to mandate the use of a particular type of marking in this way. ● It is unclear why the metadata is also required to include directions to the provider's detection tool: the information required to be included in the metadata itself is sufficient to verify, as required by Art. 50(2), that the content is "*detectable as artificially generated or manipulated*". Additional verification via the tool itself is likely to serve little practical purpose - particularly in circumstances where any conformant C2PA tool would be able to detect C2PA metadata, regardless of which conforming generator product placed the metadata.
Digitally signed manifest / provenance certificate Measure 1.1.1 encourages signatories to facilitate the download of a digital document or 'provenance certificate' containing a certified version of the output to certify its origin as AI-generated or manipulated, including the specific AI system, where the output is not available in a format that hosts metadata.	● The industry standard methodology of applying digital manifests - C2PA - does not currently facilitate the addition of cryptographic manifests to raw unstructured text. The C2PA Technical Specification version 2.3 recently added methodology to achieve this capability but received feedback (including from the Unicode standards organisation[1]) that it is unworkable. ● To the extent that the manifest / certificate is intended to operate as a human-legible document, this would exceed the scope of Art. 50(2), which requires only machine-readable markings. ● The practical utility of this measure is also unclear: given that deployers are only required to label text when it is intended to inform the public on matters of public interest, it is doubtful that these additional certificates would ever be used outside of this context (it is unlikely, for example, that an end user would upload the signed certificate alongside a social media post for which they had used AI to generate a caption). This would also not be passed on via subsequent dissemination of the content. Given the lack of practical utility, the efforts required to implement this measure are disproportionate. ● As noted in our Response to TCoP 1, the digitally signed manifest will also be onerous for systems operating at scale, resulting in high latency and storage overheads.
Imperceptible	● The requirement that detectors should be capable of identifying

[1] https://www.unicode.org/L2/L2026/26042-embedded-metadata-in-plain-text.pdf

watermarks Measure 1.1.2 requires signatories to apply imperceptible watermarks to AI-generated content, which should be designed so that a fragment of the content suffices to detect the watermark. Signatories are required to ensure multimodal output is synchronised to allow marking to be recognisable when only one or a subset of modalities have been altered.	markings using only a "*fragment*" of the content in order to protect users' privacy is technically unworkable: ○ It could negatively impact the content - e.g. if it requires watermarking the whole of a lengthy video in order to ensure a very small AI-generated portion was detectable even if a user submitted a non-AI-generated fragment. ○ It could lead to significant false negative rates, if, for example, User 1 takes an AI-generated and compliantly watermarked image and manually incorporates it into a larger image, which User 2 then wishes to verify is AI-generated, so submits only a small portion of the collage which happens to not bear the watermark. Arguably, a watermarking and detection system which is vulnerable to false negatives in this way is more detrimental to the information ecosystem than no detection regime at all, if users are wrongly assured that content is not synthetic in origin. ○ Users may also have differing interpretations of how large a "*fragment*" should be, leading to uncertainty or inconsistency between providers. ● The technical feasibility of "*synchronising*" markings across modalities is also open to debate such that it should not be a mandatory requirement. To the extent this represents a component of reliability and robustness, this would already be covered by the existing quality criteria in Commitment 3, so an additional express obligation in this respect is unnecessary.
Publicly readable watermark Measure 1.1.2 requires signatories to add an additional publicly readable watermark to audio, image or video content indicating the provider for the marking detection tool and an identifier for the metadata.	● It is unclear whether this element is intended to require an additional layer of marking in addition to the two marking layers already required under Measure 1.1. In any event, to the extent it requires the addition of a human-readable watermark, it goes beyond Art. 50(2), which only requires machine-readable watermarking. ● This is not proven technology, particularly at scale, or for all modalities (e.g. audio). ● The layer of imperceptible marking required under the first paragraph of Measure 1.1.2 should be sufficient to act as a backup in the event of the metadata being stripped from the content. A further additional watermark is therefore unwarranted.

Fingerprinting / logging Measure 1.1.3 adds an option for signatories to implement fingerprinting or logging where "appropriate" to address deficiencies in other marking techniques.	● This Measure should be amended to clarify that any commitment by providers to implement it in a "*secure and privacy-preserving manner*" is confined to their own design and implementation of the system. The current framing of this in the passive voice, requiring providers to "*ensure*" that the fingerprinting / logging is implemented in this manner, appears to require providers to guarantee that deployers and other users also adopt secure and privacy-preserving approaches. This is excessive and disproportionate, and likely to deter providers from utilising these techniques. ● Reference to 'where appropriate to address deficiencies' should be removed.

The points outlined above illustrate that the appropriateness of any given marking measure is highly nuanced and context-specific. It is also reflective of a constantly changing state of the art - as recognised in Rec. (133), this field is defined by "*the fast technological pace*" of developments necessitating "*new methods and techniques to trace* [the] *origin of information*". In this fast-moving field, it is essential that the TCoP remains flexible and does not bind signatories to methods which may rapidly become outdated.

- **How to fix:**
    - Measures 1.1.1-1.1.3 should be amended to remove the requirement to use a multi-layered marking approach (per Issue 2) and to recast the prescriptive requirements (with the exception of the 'publicly readable watermark' requirement, see below) as an illustrative list of possible options available to satisfy the Art. 50(2) marking requirement. As providers are already required to take into account the current state of the art, there is no justification for the TCoP to be more prescriptive. Whilst the TCoP can (and should) highlight the breadth of potential options for marking techniques (and note the relative advantages and disadvantages of each option), it should not impose a one-size-fits-all approach by making these mandatory measures for all modalities and use cases.
    - The "*publicly readable watermark*" requirement in Measure 1.1.1 should be entirely removed in light of its incompatibility with Art. 50(2).
    - Given that detectors are already required by Commitment 2 of Section 1 to be "*privacy-protecting*", the requirement to detect content via a "*fragment*" in order to further safeguard user privacy should be removed. If the mandatory obligation to provide detection mechanisms is retained (contrary to our recommendation at Issue 9 of this Response), it should be sufficient to inform users that a verbatim copy of content will not be retained, and that the full piece of content can therefore be submitted, to address any privacy-related concerns.

4. Remove the voluntary provisions regarding model-level marking (Measure 1.1.2 of Section 1)

TCoP 1 contained requirements to implement model-level marking (Measures 1.1.2, 1.2.1, 1.3, 1.4, 1.5, 2.1, 2.2 and 2.3 of Section 1, TCoP 1) These measures went far beyond the scope of Art. 50(2), which is

focussed solely on providers of AI systems and does not impose requirements at the model level. While it is helpful that these measures are no longer mandatory, it is unclear why - given the recognition that they were beyond the scope of the AIA - they have not been removed in their entirety.

- **Issue:** TCoP 2 contains measures which encourage providers of generative AI systems to implement model-level marking. Their presence in the TCoP poses several issues:
    - As noted in Google's Response to TCoP 1, Art. 50(2) applies only to AI systems, not models. Although Rec. (133) envisages that marking techniques "*can be*" implemented at the level of the AI model, there is no legal basis to include measures to this effect in the TCoP, which (per Art. 50(7)) is to "*facilitate the effective implementation of the obligations"* in Art. 50.
    - Model-level markings have been demonstrated to produce inferior outcomes in terms of their robustness and impact on the quality of the ultimate output in comparison to system-level techniques. The TCoP should not recommend reliance on less effective methods which have little clear advantage over the system-level markings which are envisaged in the AIA.
    - This measure is liable to disrupt the AI supply chain. Although it should be open to AI model providers to experiment with techniques to facilitate downstream compliance with transparency obligations (e.g. as a commercial feature of their products), the TCoP should not encourage this while these techniques remain experimental and evolving to avoid fostering complacency or over-reliance by downstream entities. Avoiding prospective commitments of this kind is essential to ensure that the TCoP does not become rapidly (and needlessly) outdated. To the extent that these upstream compliance tools may emerge as the most appropriate options, the market will naturally support this.
    - There is also a real risk, particularly given the level of detail included in these provisions, that their content will be picked up inappropriately by other bodies, across other jurisdictions, and potentially used as a benchmark or expectation about marking requirements. This undermines the purpose of these provisions, which is to facilitate the advancement of state-of-the-art transparency solutions; not to set expectations. Again, this demonstrates just how far these provisions stray from the purpose of the TCoP.
- **How to fix:** These provisions should be removed.

5. Remove the voluntary provisions regarding providing functionality for deployers to add labels to systems' output (Measure 1.4 of Section 1)

TCoP 1 contained requirements to provide functionality for deployers to add visible labels to systems' output (Measure 1.7 of Section 1, TCoP 1). This measure went far beyond the scope of Art. 50(2), which is focussed solely on providers of AI systems and does not require providers to assist downstream deployers. While it is helpful that these measures are no longer mandatory, they should be removed in their entirety.

- **Issue:** TCoP 2 contains measures which encourage providers of generative AI systems to provide functionality for deployers to add visible labels to systems' outputs. Their presence in the TCoP poses several issues:

- Whilst providers may wish to explore this functionality as a feature to make their products more competitive and appealing to deployers, it is unnecessary for the TCoP to positively incorporate provisions to this effect. This may not be the most effective method of implementing labels, particularly given that deployers will naturally have significantly greater visibility of the content generated by the systems they use and the most appropriate manner in which it should be labelled. Encouraging providers to expend additional resources on developing labelling techniques which may ultimately prove ineffective is disproportionate and unjustified at this early stage in the technology's development.

- This measure is liable to disrupt the AI supply chain. Although it should be open to AI system providers to experiment with techniques to facilitate downstream compliance with transparency obligations (e.g. as a commercial feature of their products), the TCoP should not encourage this while these techniques remain experimental and evolving to avoid fostering complacency or over-reliance by downstream entities. Avoiding prospective commitments of this kind is essential to ensure that the TCoP does not become rapidly (and needlessly) outdated. To the extent that these upstream compliance tools may emerge as the most appropriate options, the market will naturally support this.

- There is also a real risk, particularly given the level of detail included in these provisions, that their content will be picked up inappropriately by other bodies, across other jurisdictions, and potentially used as a benchmark or expectation about marking requirements. This undermines the purpose of these provisions, which is to facilitate the advancement of state-of-the-art transparency solutions; not to set expectations. Again, this demonstrates just how far these provisions stray from the purpose of the TCoP.

- **How to fix:** These provisions should be removed.

6. Refine criteria for when providers can rely on third-party watermarking techniques, and ensure effective allocation of responsibility along the AI value chain (Measures 1.1, 2.1, 4.1 and 4.2 of Section 1)

TCoP 2 seeks to streamline providers' compliance with the requirements introduced in TCoP 1 by allowing them to rely on third parties' marking techniques, as well as those third parties' testing and detection mechanisms. However, it does so in a manner which is likely to drastically disadvantage these third parties, lacks any legal basis, and will significantly disincentivise investment in this essential field.

- **Issue:** Measure 1.1 of Section 1 allows providers to rely on third-party marking techniques only where those techniques are supplied by organisations that have adhered to this Section of the TCoP and have demonstrated compliance to the AI Office and competent market surveillance authorities. Similarly, Measures 2.1 and 4.1 allow signatories to rely on third party marking and detection solutions only insofar as they comply with Section 1 of the TCoP, whilst Measure 4.2 permits signatories to rely on testing results provided by a third party only insofar as they comply with that measure.

  In practice, this would require signatories to use only techniques developed by TCoP signatories, turning the TCoP into a soft gatekeeping tool or *de facto* certification scheme. This would distort competition and constrain commercial freedom by making TCoP adherence a *de facto* precondition for supplying marking techniques to signatories.

- By substantially limiting the marketability of marking techniques, the TCoP risks deterring innovation and investment in watermarking, an emerging and strategically important field for the wider information ecosystem. The EU should be encouraging, not restricting, such investment. If leading developers are deterred from signing the TCoP (for reasons set out elsewhere in this Response), signatories will face a very narrow set of options or be forced to develop their own techniques. Reliance on a small number of techniques increases systemic vulnerability, as defects in widely used methods could have outsized impacts. The cost and time of in-house development is likely to be prohibitive, particularly for smaller providers, leading to weaker solutions. The status of multi-member schemes such as C2PA is also left unclear.

- This requirement is unnecessary. Developers will in any event test their techniques before placing them on the market, and providers are already required by Art. 50(2) to use state-of-the-art, robust, reliable and effective techniques. There is no clear rationale for the TCoP to overlay additional testing obligations. Further, developers of marking techniques will not necessarily be providers of the relevant AI systems under the AIA, as they occupy different points in the value chain. TCoP 2 does not address whether such non-provider actors would nonetheless be expected to comply with extensive provider-level obligations on documentation, user literacy, or governance. This extends the TCoP beyond its legitimate scope, analogous to imposing GPAI Code of Practice requirements directly on microchip manufacturers, with limited practical benefit.

- **How to fix:** References in Measures 1.1, 2.1, 4.1 and 4.2 that restrict reliance on third-party techniques or testing to parties or solutions that are themselves compliant with the TCoP should be removed. Providers should be free to use any techniques they consider appropriate, provided these meet the four quality criteria. The TCoP should safeguard commercial freedom and avoid unintentionally discouraging investment in marking techniques, which is essential for European competitiveness and the resilience of the information ecosystem.

7. Remove provisions on non-removal of markings (Measure 1.2 of Section 1)

We welcome the clarification in Measure 1.2 of TCoP 2 that providers are not responsible for third parties' markings. However, in maintaining requirements in relation to (i) the inclusion of a prohibition on removal of or tampering with marks in contract terms and acceptable use policies (AUPs), and (ii) the preservation of existing metadata, this provision still manifestly exceeds what is necessary or useful to address the core risks outlined in Rec. (133) of "*misinformation and manipulation at scale, fraud, impersonation and consumer deception*".

- **Issue:**

  - Measure 1.2(a)'s requirement for providers to preserve existing metadata and other markings is problematic:

    - Even if technically feasible or within the scope of providers' nominal "*best efforts*", providers should not be required to preserve metadata where this would obstruct functionality (especially if, contrary to the points raised at Issues 2 and 3, multiple layers of other marking types are also required).

- The retention of metadata in the C2PA manifest will be onerous for systems operating at scale, resulting in high latency and storage overheads.
- Providers may remove metadata for numerous legitimate reasons, including because existing marking is considered to be unreliable, or to avoid displaying personally identifiable information (PII) (e.g. contained in the C2PA manifest). The TCoP should acknowledge that, in these circumstances (among others), removing the metadatais a better alternative.
- Relatedly, an inability to remove metadata could lead to abuse by malicious actors - e.g. the tracking of PII, device IDs or certificates embedded in a manifest (which could then be used for doxxing and harassment), or social engineering (whereby an attacker layers a deepfake over an authentic, signed image, which will still appear 'verified' because of the metadata).
- The requirement to "*update*" third party marks to reflect content transformation to maintain the provenance chain may not be technically feasible if, for example, those markings rely on closed proprietary technologies.

- Measure 1.2(b)'s requirements regarding the use of AUPs or contract terms to prevent counterparties altering or removing marks fails to acknowledge that third parties (including substantial modifiers / "*deemed providers*" of the AI system) will have a variety of entirely legitimate reasons for removing the implemented watermarking techniques in order to deploy their own instead - e.g. to better suit the specific modalities or use cases in which they operate.
  - Furthermore, parties looking to disable marking techniques to use their own will likely already be subject to the provider obligations under Art. 50(2) in their own right. Where these parties go on to disseminate unmarked content, they will be in breach of the AIA and appropriate enforcement action may be brought against them by the relevant regulator(s). There is no basis in the AIA to make signatories responsible for enforcing this requirement on regulators' behalf, or bearing the burden of the cost and time required to do so.
  - The "*encouragement*" for providers of online platforms and websites to preserve markings is also ambiguously drafted: this demonstrably exceeds the scope of Art. 50(2) and the purpose or rationale for this inclusion is unclear.

- **How to fix:** Remove the provisions relating to the use of contract terms and policies, the requirement to refrain from removing metadata, and the provisions relating to online platforms and websites.

8. Refine the definitions of the core quality criteria for marking techniques, to focus on the techniques themselves rather than ancillary aspects such as detection mechanisms (Commitment 3 of Section 1)

As with TCoP 1, the interpretations of the quality criteria (effectiveness, reliability, robustness and interoperability) in Commitment 3 and its Measures still extend beyond their ordinary meaning and the scope of the AIA. While some TCoP 2 amendments usefully reflect Rec. (133) (e.g. taking account of

content-specific limitations and technological and market developments), several elements remain prescriptive, disproportionate and unnecessary to Art. 50(2)'s purpose of achieving the quality criteria.

- **Issue:**
    - TCoP 2's interpretation of effectiveness, reliability, robustness and interoperability must be clear and proportionate, so that providers can assess with reasonable certainty whether a proposed marking or labelling technique is an adequate means of compliance.
    - The recognition of "*operational constraints*" as a factor moderating Commitment 3 is welcome. However, the statement that such constraints cannot be used to circumvent compliance with Commitments 1 and 2 is over-simplified in that it fails to reflect (and is therefore incompatible with) Art. 50(2), which explicitly allows providers to take into account "*the costs of implementation*" (an operational constraint) when deploying marking solutions. This narrow view of operational constraints is therefore inconsistent with the AIA.
    - The requirement that providers "*strive to achieve the highest possible*" effectiveness, interoperability, robustness and reliability should be removed. It does not reflect the wording of the AIA and ignores the complex trade-offs between these factors. Providers will often be unable to achieve the "*highest possible*" level across all dimensions simultaneously, and it would be unclear when this standard is met, undermining legal certainty.
    - Further adjustments to the specific criteria are needed:
        - **Effectiveness:** It is positive that references to energy efficiency and compute cost have been removed. However, explicitly tying effectiveness to the presence of detection mechanisms is problematic: (i) detection is not itself mandated by the AIA, and (ii) embedding specific measures (such as detection) into the definition of effectiveness is circular and unhelpful to providers' compliance assessments.
        - **Reliability:** TCoP 2 retains extensive testing obligations and adds granular requirements specific to GPAI systems, which are disproportionate and lack any basis in the AIA.
        - **Robustness:** Signatories are required to assess robustness of their marking and detection solutions in terms of resilience to adversarial attacks, which is impractical to use as a metric. Given all machine learning systems are inherently prone to such attacks, it is not possible to provide any guarantee of resilience against them. Robustness should instead be framed by reference to the state of the art, focusing on performance under common transformations (e.g. compression, resizing, format conversion), as proposed in Google's Response to TCoP 1.
        - **Interoperability:** TCoP 2 continues to prioritise interoperability, as in TCoP 1. While interoperability and common standards are valuable long-term objectives, this prioritisation:

- overlooks significant cybersecurity benefits of closed, proprietary techniques, which are harder for malicious actors to analyse, reverse-engineer and circumvent; and

- downplays trade secret and confidentiality concerns arising from extensive obligations to cooperate with other providers (e.g. on a "provider-agnostic detection interface") and with the Commission on shared information repositories. These requirements are likely to deter stakeholders from signing the final TCoP.

- **How to fix:** The TCoP should prioritise robust and reliable markings so that Art. 50 obligations can fulfil their primary practical purpose of reinforcing trust in the information ecosystem in the short to medium term. This goal should not be subordinated to interoperability where interoperable techniques are materially less secure, more vulnerable to removal, or otherwise weaker. Over-reliance on interoperable but less reliable and robust techniques risks eroding confidence in content markings.

  The TCoP should therefore:

  - set out clearly defined, non-circular quality criteria against which signatories can assess their marking and detection solutions, removing prescriptive elements such as the reference to detection mechanisms in the definition of effectiveness, and making clear that examples of metrics that may be used to measure reliability are just that - examples, rather than requirements;

  - revise Commitment 3 and its Measures to recognise explicitly that providers are entitled under Art. 50(2) to take operational constraints, including costs of implementation, into account when implementing marking and detection solutions; and

  - delete the requirement for signatories to cooperate: (i) with other providers to develop an "*interoperable provider-agnostic detection interface*" and (ii) the Commission on a *"shared repository*" of watermark and detection information, given practical problems these are likely to present.

9. Adopt a flexible and resilient approach to detection mechanisms and disclosure of results (Measures 2.1-2.3 of Section 1)

Whilst it is positive that Measure 2.1 is now framed as requiring mechanisms to detect active marking applied under Commitment 1 (rather than generally to verify AI-generated content), our view remains as in our Response to TCoP 1: (i) such tools are not required by the AIA, and (ii) their availability poses serious cybersecurity risks. These requirements should be removed or downgraded to voluntary. If retained as mandatory measures, their scope must be narrowed so that compliance is only required where technically feasible and does not undermine the reliability or robustness of the marking regime.

- **Issue:** As noted in our Response to TCoP 1, the proposed detection mechanisms are both unnecessary under the AIA and liable to create practical and security problems:

  - Art. 50(2) requires only that outputs are "detectable" as AI-generated or manipulated; it does not oblige providers to offer detection tools. Rec.(135) merely notes that the TCoP

may support practical arrangements to make detection mechanisms accessible, not that signatories must supply them.

- The obligation in Measure 2.1 for tools to be "*implementable locally or hosted within the European Union*" is unnecessary for privacy or security reasons. Any tool processing personal data of individuals in the EU will be subject to EU data protection law and robustness requirements (e.g. Measure 3.3) regardless of hosting location. The restriction unjustifiably limits EU access to leading detector technologies.

- The inclusion of Measure 2.2, which encourages collaboration to support the development of forensic detection mechanisms remains problematic (even downgraded from mandatory to voluntary), given its significantly weaker reliability compared to watermarking. The TCoP should not actively endorse flawed techniques in this way, particularly while also mandating other, more effective techniques in parallel.

- Publicly available detection tools, especially when accompanied (as envisaged in Measure 2.3) by information on methodology and confidence levels, give malicious actors a way to probe and circumvent markings. This risk is compounded by restrictions on withdrawing or deprecating tools, which could oblige providers to maintain mechanisms that are insecure or obsolete.

- Measure 2.3 suggests that signatories must ensure detection and verification results are presented in a way that is clear and easily comprehensible to laypersons, so as to comply with Art. 50(5). However, this does not find any basis in the AIA:

  - Art. 50(5) applies to "*the information referred to in [Arts. 50(1)-(4)]*"; Arts. 50(1)-(4) do not refer to the results of a detection mechanism.

  - Art. 50(5) also cannot have a bearing on detection results, since it requires information to be provided "*at the latest at the time of first exposure*", which is not possible for detection results which require the user to have encountered the content in order to input it into the interface.

- **How to fix:** The requirement to make available detection mechanisms should be removed or, at least, amended and reframed as an optional commitment. If it is to be retained, then the following amendments should be made:

  - Removing the element of Measure 2.1 preventing providers from deprecating or withdrawing detection tools, to ensure adequate flexibility and resilience in the context of this evolving field, and including an express statement that these detectors are only required in so far as they do not impair the security of markings.

  - Removing the requirement in Measure 2.1 that signatories must ensure that detection and verification results are provided in accordance with Measure 3.4 and can be downloadable on request in a digitally signed document, on the basis that this is not supported by the AIA.

  - Limiting the required output of detectors to Detected / Not Detected, which mirrors the requirements of the AIA, which simply require content to be identifiable "*as*" AI-generated or otherwise. More complex mechanisms are likely to be excessive and unnecessary for

the average consumer (which is all that should be necessary for a free tool, to be developed at providers' expense).

- The requirement to provide confidence scores and human-understandable explanations should be removed from Measure 2.3.
- The requirement for the detection tool to be *"implementable locally or hosted within the European Union*" should be removed.
- The optional forensic detection measure in Measure 2.2. should be removed entirely.

10. Replace overly restrictive and unhelpful design and placement requirements for deep fake disclosures (which go beyond the AIA) with a requirement for deployers to disclose in an appropriate manner based on modality (Measures 1.1-1.2 of Section 2)

We welcome the move away from a common EU-wide icon for deep fakes. However, TCoP 2 still goes beyond the AIA by imposing overly restrictive requirements on the design and placement of deep fake disclosures, which are ultimately likely to be unhelpful for users.

- **Issue:**
  - TCoP 2's requirements on the design and placement of deep fake disclosures in Measures 1.1 and 1.2 of Section 2 go beyond the AIA's requirements, and will serve only to push deployers into finding alternative means of compliance rather than signing the TCoP.
    - Neither Art. 50(4) nor Art. 50(5) contain specific requirements beyond disclosing, in a clear and distinguishable manner, that deep fake content has been AI-generated or AI-manipulated.
    - The purpose of the TCoP is to facilitate effective implementation of the AIA's requirements; not to introduce new, rigid design requirements such as a contrast ratio "*maintained of at least 4:5:1 against the background*". There are a plethora of other ways in which a disclosure can be "clear and distinguishable" without meeting this unnecessarily restrictive ratio requirement.
    - The requirement in Measure 1.2 for disclosures to be "*affixed on or directly embedded in*" and to "*travel with"* the content also go beyond the AIA. These requirements are clearly not necessary for a disclosure to be "clear and distinguishable" and, had the drafters of the AIA intended to include these sorts of requirements, then they would have expressly done so in Art. 50(4). In some cases, alternatives to an affixed icon (e.g. an overlay) would allow deployers to better align their solutions, especially where different deployers have different technical means. There are also technical challenges with a requirement to affix or embed disclosures in the content itself - this may not be possible for deployers who do not have the ability to access the underlying generative AI technology to be able to do this. It may prove particularly challenging (and confusing) where deployers are faced with different disclosure requirements across jurisdictions for the same content.

        - Furthermore, requiring the disclosure to travel with the content creates a potentially endless chain of onward actors with whom a deployer would need to collaborate. Requirements like this will serve only to dissuade organisations from signing the TCoP.

    - Aside from their lack of legal basis, these requirements will actively obstruct a disclosure from being clear and distinguishable, ultimately making it harder for users to identify AI-generated information and thereby undermining the aim of this disclosure as explained in Rec. (133):

        - Requiring disclosures to be affixed on / embedded in and to travel with the content will lead to overcrowding of disclosures e.g. where other disclosures are required under different regulations or where there is clustering of other content. This will make it more difficult for users to understand disclosures and could lead to label fatigue. This may, in turn, push them towards non-compliant content with less intrusive labelling.

        - Due to limited content space and the need to avoid content obstruction, the requirement for disclosures to be affixed on / embedded in content is likely to mean that users receive less helpful information in circumstances where, for example, an adjacent label with additional information would have been appropriate. For smaller content (e.g. thumbnails or partial views), in-content disclosures may be difficult to make clearly visible and affixed / embedded disclosures are also more at risk of being accidentally cropped out when content is resized for different formats.

    - As a final point, which we hope is capable of straightforward resolution, revised Measure 1.1 in TCoP 2 is drafted ambiguously. It refers to the main visual icon or label as being "*possibly supplemented, where appropriate, with a short text label*" and says that "*[w]here technically feasible and available this can further be elaborated in a second layer detailing (e.g., what has been modified)*". We assume that both of these requirements are optional (which makes sense) but, to the extent these points remain in the next TCoP draft, we would be grateful if this could be clarified.

- **How to fix:**

    - More generally, we reiterate that the overly prescriptive design and placement requirements for deep fake disclosures in Measures 1.1 and 1.2 - i.e. the fixed contrast ratio, the "affixed on / embedded in content" requirement, and the "travel with the content where feasible" requirement - should be removed.

    - These should be replaced with a broader requirement, reflecting the AIA, for deployers to adopt an appropriate design and placement of their deep fake disclosure, based on guiding factors (e.g. consideration of content duration, rather than prescriptive requirements), which could include content modality, as we propose in more detail in Issue 11 below.

11. Replace modality-specific disclosure requirements with factors to guide deployers to determine an appropriate, clear and distinguishable disclosure for deep fake content (Measure 1.2 of Section 2)

Measure 1.2 of Section 2 contains modality-specific placement requirements for deep fake disclosures. We welcome the efforts in TCoP 2 to make these measures more workable, including providing alternatives to consistent disclosures, removing the need for disclosures to be in a fixed place on content, and the increased clarity on disclosures in varying lengths of video. However, these requirements continue to suffer from a number of issues (as we explain below) and should be replaced by a broader requirement for deployers to adopt an appropriate design and placement of their deep fake disclosure, based on guiding factors that we address at the end of this section.

- **Issue:** The modality-specific placement requirements:
    - Go beyond the AIA; in particular around the required timings for disclosure (which, under Art. 50(5), are only required "*at the latest at the time of the first interaction or exposure*"). This overreach will disincentivise deployers from signing the TCoP and instead push them to opt for alternative, more workable ways of demonstrating compliance with the AIA.
    - Are practically unworkable, not taking into account the nature of content of that modality (e.g. the unpredictability of real-time content) or how multiple disclosures would impact each other.
    - Are too intrusive, which risks pushing users towards non-compliant content elsewhere, undermining the aim of disclosure under Rec. (133) and potentially penalising compliant deployers with reduced traffic.
    - Are often unclear, introducing vague terms or inconsistent requirements, creating legal uncertainty, increasing the risk of uneven application amongst deployers in the information ecosystem.
    - Specific examples of each of these issues are set out, by modality, in the following table:

Real-time video	The requirement for disclosure to be "*consistently throughout the exposure where feasible*", or the alternative requirement to use a visual or audio disclaimer "*at the latest at the beginning and at regular intervals during the exposure*", goes beyond Art.50(5) which only requires disclosure "*at the latest at the time of first interaction or exposure*". Such consistent or repeated disclosures would interfere with the content (especially where users access livestreams intermittently), possibly obstructing key live details and increasing latency. This may push users towards less compliant content elsewhere. While the flexibility for deployers to decide what disclaimer intervals are "*regular*" or of an "*appropriate*" duration is welcome, these undefined terms create legal uncertainty. In any event, these requirements are also practically unworkable given the very nature

	of real-time videos. They can be broadcasted for an unknown overall length of time, which makes it very difficult for deployers to know how regularly they need to include disclosures or what would be an appropriate duration.
Non-real-time video (long)	While we welcome the point that clear and distinguishable disclosure will depend on the length of the video, the terms "long" and "short" videos are not defined (in contrast to audio, where short audio content has been defined as under 30 seconds). This introduces subjectivity and legal uncertainty, but also highlights why a broader requirement is more appropriate here, instead of granular modality-specific requirements. The requirement for disclosure to be "*at the latest at the beginning and repeated at regular intervals*" goes beyond Art. 50(5) which only requires disclosure "*at the latest at the time of first interaction or exposure*". Such repeated disclosures could interfere with the content and may increase latency. This may push users towards less compliant content elsewhere. Where non-real-time video also contains audio content, the measure is also not clear on whether only visual disclosure would suffice. This further complexity underlines the need to avoid such granular, modality-specific requirements, but to the extent this point remains in the next TCoP draft, we would be grateful if it could be clarified that visual disclosure would suffice in such circumstances..
Non-real-time video (short)	The requirement for disclosure to be "*consistently throughout the exposure from the beginning of the exposure*" goes beyond Art. 50(5) which only requires disclosure "*at the latest at the time of first interaction or exposure*". The requirement for disclosures to "*not be too close to other overlaying text and icons*" may also be practically unworkable for content where multiple labels are needed under different regulations. This is particularly problematic for short-form, vertical videos where visible, unpopulated screen area is already very constrained. The inclusion of a disclosure requirement for short non-real-time videos which are "*entirely AI-generated and manipulated*" is inconsistent with the removal of a requirement to distinguish fully AI-generated and AI-assisted content.
Image	The requirement for disclosure to be included "*consistently at the latest at the first exposure and at any subsequent exposure*" to the

	image goes beyond Art. 50(5) which only requires disclosure "*at the latest at the time of first interaction or exposure*". The requirement for disclosure to not be *"placed too close to other icons or text elements"* will also be practically unworkable for content where multiple labels are needed under different regulations.
Audio	The requirement for longer audio to include audible disclaimers "*at the beginning, at appropriate intermediate phases, and at the end of the content*" goes beyond Art. 50(5) which only requires disclosure "*at the latest at the time of first interaction or exposure*". This is also practically unworkable for real-time audio given its nature, i.e. it may be broadcast for an indefinite duration, which makes it very difficult for deployers to know what "appropriate intermediate phases" would be. In any event, while the flexibility for deployers to decide what phases are "*appropriate"* is welcome, the lack of definition creates legal uncertainty and underlines the need to avoid such prescriptive requirements. Further, given this nature of real-time audio, intermediate disclosures are not feasible without obstructing the audio being broadcast. This is especially the case for users who are accessing a livestreamed audio on and off, where repeated disclosures could interfere with the content and may increase latency. As a result, users may be pushed towards non-compliant content elsewhere. For audio of real-time video (see the real-time video requirements in Measure 1.2), any audio disclosures have to be "*simultaneously with the icon or label*", which is required to be "*consistently throughout the exposure where feasible*". This, in turn, suggests that any audio disclosure also has to be included consistently throughout. This is both inconsistent with the intervals in the audio-specific measure, and, again, goes beyond Art.50(5) which only requires disclosure "*at the latest at the time of first interaction or exposure*".
Short-form text	The requirement for short-form text to have a "*contextual notice in the user interface or session*", given the small amounts of text involved ("*single words or brief phrases*"), is both disproportionate and confusing for users. It would be difficult for users to connect such small snippets to an overarching contextual disclosure (especially where the disclosure refers to several, separate bits of text), making it hard for users to distinguish sections (e.g. the rest of the page) which are not AI-generated.

- **How to fix:**

- The modality-specific placement requirements in Measure 1.2 should be removed and replaced with a broader requirement, reflecting the AIA (which does not require disclosure beyond the "*first interaction or exposure*"), for deployers to adopt an appropriate clear and distinguishable design and placement of their deep fake disclosure, based on guiding factors (rather than prescriptive requirements).

- These guiding factors could include modality (which we agree is an important contextual factor), together with the factors that are currently addressed in Measure 1.2, such as duration of content and disclosure, whether the content is livestreamed, and prominence of disclosure, as well as other factors such as (but not limited to) distance to content and size of content.

- This guidance would avoid the various issues with the modality-specific requirements noted above and be far more workable for deployers as it would give them flexibility to take into account the possible variations between content even within the same modality or within the same "factor" (e.g. videos of the same length), in turn providing users with more effective disclosures of deep fake content.

12. Apply flexible "contextual disclosure" approach to all disclosures for deep fake creative works (Commitment 3 of Section 2)

We welcome the increased flexibility in Commitment 3 of Section 2 to allow deployers to label creative works unobtrusively. However, Commitment 3 remains more prescriptive than Art. 50(4) and Art. 50(5) require, hindering deployers from placing disclosures in a non-intrusive way, hampering users from enjoying creative content and ultimately pushing them towards less compliant content.

- **Issue:**

  - The modality-specific requirements in Commitment 3 are far too prescriptive for the more limited disclosure for creative work permitted by Art. 50(4). In practical terms, TCoP 2 presents very little difference between the requirements for disclosing deep fakes in creative versus non-creative work. For example, Commitment 3 requires creative work disclosures to "*follow the design requirements in Measure 1.1*", to be "*clear and distinguishable*", to be "*complemented*" with other disclaimers where relevant and placed in a "*nonintrusive yet effective*" position.

  - Not only do these requirements go beyond Art. 50(4), but by ignoring that deployers are best placed to determine an appropriate disclosure that "*does not hamper the display or enjoyment of the work*" (Art. 50(4)), these requirements are ultimately likely to push users to less compliant content when enjoying creative work.

  - Positively, Commitment 3 does recognise the possibility of "contextual disclosure", which appears to acknowledge the concerns above and provides deployers with more flexibility in their approach to disclosure. In our view, a "contextual disclosure" approach should be the default for creative work. Indeed, Commitment 3 even acknowledges that a "contextual disclosure" would be appropriate where the more prescriptive requirements would "*affect the display or enjoyment of the work*". This being so, the overly prescriptive, modality-specific requirements for creative works should not be needed at all.

- That said, the "contextual disclosure" requirements in Commitment 3 should be even more flexible, to reflect the "appropriateness" benchmark envisaged by Art. 50(4). For example, the requirements e.g. to place disclosures "*adjacent*" to the video "*without the need for scrolling or additional engagement*" don't accommodate differences in the format, presentation and enjoyment of creative content: e.g. this requirement would not be feasible where there is limited space around the content and where an expanded description (or other click through) would be more appropriate.

- Ultimately, consistent with the AIA, deployers should be given more flexibility, especially with regard to creative works, to determine an appropriate disclosure (which might be an icon in the content itself or an adjacent label or a click-through or menu button providing more information).

- **How to fix:** The modality-specific requirements for disclosing deep fake content in creative works in Commitment 3 should be removed, allowing for the more flexible "contextual disclosure" approach to be applied to all creative works. This approach should also avoid prescriptive requirements (some of which are currently included even for contextual disclosure in Commitment 3), giving deployers the flexibility to disclose in an appropriate manner that does not hamper the display or enjoyment of the creative work, consistent with the AIA.

13. Remove procedural requirements on deployers to be able to rely on Art. 50(4) exemption for public interest content that has undergone human review (Commitment 4 of Section 2)

Commitment 4 of Section 2 goes beyond the AIA, containing onerous procedural requirements for signatories to justify their reliance on the exemption from Art. 50(4) for content that has undergone human review or editorial control.

- **Issue:**

  - Commitment 4 has been softened in TCoP 2, but it still fundamentally requires signatories to produce documentation - whether on an output-by-output basis or more generally - to demonstrate their reliance on the exemption in Art. 50(4). There is no basis in the AIA, however, for any such requirement.

  - Moreover, as a general proposition, the AIA does not prescribe governance or compliance requirements. It is up to deployers to demonstrate that they have complied with AIA provisions if and when challenged on this by regulators. It is inappropriate for the CoP to introduce requirements of this nature.

  - Such an onerous requirement is also disproportionate (especially where the TCoP requires labelling for even "*single words or brief phrases*"), even if it does allow deployers to rely on existing processes and documents. Further, it wouldn't engender any practical benefit compared to, for example, an encouragement on deployers to adopt a policy of human review and editorial control.

  - The additional requirement to publish the contact details of the individual with editorial responsibility also has no basis in the AIA.

- **How to fix:** Commitment 4 should be removed. Deployers should be free to evidence compliance with this exemption however they wish. To the extent that the finalised TCoP addresses this

exemption at all, it should be limited to encouraging deployers to have a policy regarding human review and editorial control.

14. Remove inappropriate and unnecessarily detailed provisions relating to further development of transparency solutions (Measure 3.5 of Section 1 and Measure 1.3 of Section 2)

Measure 1.3 of Section 2 encourages signatories to use the EU-wide common icon and to support the further development of a uniform label through "*the work and activities of a dedicated task force… in accordance with the following requirements…*". Measure 1.3 goes on to include extensive requirements about this task force.

Similarly, Measure 3.5 of Section 1 encourages signatories to: invest in scientific R&D, to collaborate with various stakeholders to advance marking and detection mechanisms, to participate in another task force, and to cooperate on the development of watermarking, forensic models and fingerprinting techniques, as well as shared benchmarks and red-teaming exercises.

While we welcome the adjusting of these requirements to "*encouragements*" in TCoP 2, these provisions are still wholly inappropriate for inclusion in the TCoP and create various issues, as we explain below.

- **Issue:**
    - As explained in Issue 1 above, measures which are stated to be "encouragements" risk adding unnecessary confusion to the TCoP. This issue is exacerbated in this case because Measure 1.3 of Section 2 and Measure 3.5 of Section 1 include numerous and extensive provisions.
    - For the same reason, encouraging signatories to invest in R&D, to support various task forces and to collaborate with stakeholders is far removed from the effective implementation of the AIA, which is the sole purpose of the TCoP pursuant to Art. 50(7). We do not seek to suggest that these are not worthwhile endeavours, but using the TCoP as a vehicle to encourage organisations to engage in them is inappropriate. There is a real risk of dissuading organisations from signing the TCoP when it strays beyond its statutory purpose.
    - Furthermore, the level of detail that TCoP 2 includes for some of these measures underlines the concern here. The "requirements" for the task force contemplated in Measure 1.3 of Section 2, for example, is one of the most detailed measures in TCoP 2. As well as being unnecessary, the provisions in this measure are evidently also inappropriate. For example, they appear to preemptively dictate the work and goals of the task force and signatories; they address further prescriptive requirements for labelling not otherwise addressed in TCoP 2; and they appear to bind signatories to future actions ("*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*").
    - There is also a real risk, particularly given the level of detail included in these provisions, that their content will be picked up inappropriately by other bodies, across other jurisdictions, and potentially used as a benchmark or expectation about labelling requirements. This undermines the purpose of these provisions, which is to facilitate the advancement of state-of-the-art transparency solutions; not to set expectations. Again, this demonstrates just how far these provisions stray from the purpose of the TCoP.

    - Notwithstanding the above, measures on the use and development of a voluntary common icon undermine the aim of developing state of the art disclosure solutions, which, in order to keep up with technology, not only needs flexibility on technical requirements (e.g. how layered disclosures would work), but also needs room for deployers to have different solutions.

- **How to fix:** Measures 1.3 of Section 2 and 3.5 of Section 1 should be removed from the TCoP. If they are to be addressed at all, they should be addressed in a separate, more appropriate document, which is not part of AIA guidance / Code of Conduct documentation.

15. Remove measures which have no legal basis in the AIA i.e. measures relating to compliance frameworks, governance, training, testing, verification, monitoring and cooperation with regulators (Measures 2.4 and 4.1-4.4 of Section 1 and 2.1-2.3 of Section 2)

TCoP 2 contains various measures which have no legal basis in the AIA. These are:

Section 1:

- M.2.4 - Support literacy on AI marking technologies and verification (now voluntary)
- M.4.1 - Compliance framework
- M.4.2 - Testing, verification and monitoring
- M.4.3 - Training
- M.4.4 - Cooperation with market surveillance authorities

Section 2:

- M.2.1 - Internal compliance
- M.2.2 - Awareness and training
- M.2.3 - Review, feedback and cooperation with authorities

While some of these measures have been softened from TCoP 1 (for example, becoming voluntary or incorporating a “proportionate efforts” qualifier), ultimately these measures introduce new requirements which go beyond those set out in the AIA and/or they address points that are not addressed in the AIA.

- **Issue:**

    - Fundamentally, these measures cannot legitimately be included in the TCoP because they introduce new requirements or provisions that have no legal basis in the AIA. The purpose of the TCoP is to “*facilitate the effective implementation of the obligations regarding the detection and labelling of artificially generated or manipulated content*” (Art. 50(7), emphasis added). These measures are supplementary to the obligations in Art. 50; they do not facilitate their effective implementation.

    - The inclusion of these measures is likely to dissuade organisations from signing the TCoP. Organisations will not want to endorse and potentially subject themselves to additional requirements when there would be no question at all as to their applicability to non-signatories. These requirements would be irrelevant.

    - Unworkable new requirements such as Measure 2.3 of Section 2 (which encourages feedback channels for missing / incorrect disclosures and requires the review of such cases) will particularly dissuade organisations from signing the TCoP. Where signatories distribute other deployers’ content, it would be very difficult for them to distinguish

between reports of missing / incorrect disclosures which the signatory has a responsibility to review as the deployer, and reports of missing / incorrect disclosures by another deployer which the signatory has no responsibility to review (as individuals and third parties would unhelpfully use the feedback channels to report both).

    - Even though some of these measures have been softened in TCoP 2, they are still disproportionate, particularly those requirements which impose onerous, on-going obligations on signatories (e.g. Measure 4.2 of Section 1, which imposes extensive, burdensome testing requirements for their marking and detection solutions). They are also inappropriate: beyond contexts in which compliance documentation is explicitly required or envisaged, the AIA gives operators flexibility as to how they comply with their obligations. It is for operators to determine how best to comply with those obligations and, if challenged, to demonstrate compliance. The TCoP should not seek to prescribe how operators comply and, in doing so, create yet more requirements that go beyond the AIA.

- **How to fix:** We reiterate that the Measures listed above should be removed in their entirety, alternatively - at the very least - they should be made optional. They have no basis in the AIA and cannot legitimately be included in the TCoP, nor are they practically necessary to facilitate compliance with the Art. 50 obligations in any case. Indeed, they are, if anything, likely to hamper providers and deployers by imposing burdensome procedural requirements that occupy resources which could more effectively be employed in complying with the AIA obligations themselves. It is ultimately likely to make it difficult for many organisations to sign the TCoP if these measures are retained.