

WG 1 - Transparency and copyright-related rules: Provide your feedback to the third draft of the General-Purpose AI Code of Practice

Fields marked with * are mandatory.

Provide your feedback to the third draft of the general-purpose AI Code of Practice!

Thank you for your continuous support and participation in the drawing-up of the **first general-purpose AI Code of Practice**.

As participants of Working Group 1, upon receiving the **third draft**, you are encouraged to express your comments on the content via this survey for those **Commitments and Measures related to Working Group 1, deadline Sunday 30 March 2025 EOD**.

As for the first round, your feedback is essential in tailoring the Code of Practice to meet diverse stakeholder needs to achieve a final Code of Practice that best reflects the different submissions. This will lead to a proper application of the rules of the AI Act for general-purpose AI models. Please be aware that the survey does not cover Art. 53(1)(d) issues.

The survey is structured per commitments, measures, and it presents two types of standard questions for the **Transparency section**, namely:

- **Opinion ratings** (close-ended feedback): express the level of agreement with the content of the commitments, measures choosing among different options.
- **Open-ended questions**: to provide suggestions on how to improve the commitments and measures.

On the other hand, for the **Copyright section**, only open-ended questions are available.

For the ease of engaging with the survey, please have the Code of Practice open for reference.

You have the option to respond to all questions if you wish, however, you do not have to. Each stakeholder can only **submit one answer** to this survey. **If you forward the link, it will be possible for others to edit or download your response.**

Your responses will be kept confidential in accordance with the Vademecum.

The AI Office is looking forward to this inclusive and transparent process!

Stakeholder Information

Please provide your name, surname, email address, and the name of your organisation (if applicable). Please note that if your contact information does not correspond to an eligible participant or to the organisation's Point of Contact, your response will be discarded.

	Name	Surname	Email address	Organisation (if applicable)
Stakeholder	Teresa	Nobre	teresa@communia-association.org	COMMUNIA Association for the Public Domain

* Which stakeholder category would you (or your organisation) consider yourself in?

- Academia (in a personal capacity)
- Civil society organisation
- Downstream provider of an AI system based on general-purpose AI models, or acting on behalf of such providers
- EU Member State representative
- European or international observer
- Other independent expert (in a personal capacity)
- Other industry organisation, or acting on behalf of such organisations
- Other organisation with relevant expertise
- Other stakeholder organisation
- Provider of a general-purpose AI model, or acting on behalf of such providers
- Rightsholder organisation

Transparency Section Preamble

Please provide below your thoughts on the preamble of the Transparency Section of the Code of Practice.

5000 character(s) maximum

Commitments by providers of general-purpose AI models: Transparency

Transparency, Commitment I.1: Documentation

To what extent do you agree with this commitment? (Commitment I.1)

- The commitment should be removed in its entirety
- The commitment should be substantially edited and/or further clarified
- The commitment should be lightly edited and/or further clarified
- The commitment is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the commitment and drafting improvements (Commitment I.1)

3000 character(s) maximum

Transparency, Measure I.1.1: Drawing up and keeping up-to-date model documentation

To what extent do you agree with this measure? (Measure I.1.1)

- The measure should be removed in its entirety
- The measure should be substantially edited and/or further clarified
- The measure should be lightly edited and/or further clarified
- The measure is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure I.1.1)

3000 character(s) maximum

Transparency, Measure I.1.2: Providing relevant information

To what extent do you agree with this measure? (Measure I.1.2)

- The measure should be removed in its entirety
- The measure should be substantially edited and/or further clarified
- The measure should be lightly edited and/or further clarified
- The measure is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure I.1.2)

3000 character(s) maximum

Transparency, Measure I.1.3: Ensuring quality, integrity, and security of information

To what extent do you agree with this measure? (Measure I.1.3)

- The measure should be removed in its entirety
- The measure should be substantially edited and/or further clarified
- The measure should be lightly edited and/or further clarified
- The measure is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure I.1.3)

3000 character(s) maximum

Model Documentation Template: Transparency

Transparency, Item 1: General information

To what extent do you agree with this item? (Item 1)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 1)

3000 character(s) maximum

Transparency, Item 2: Model properties

To what extent do you agree with this item? (Item 2)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 2)

3000 character(s) maximum

Transparency, Item 3: Methods of distribution and licenses

To what extent do you agree with this item? (Item 3)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 3)

3000 character(s) maximum

Transparency, Item 4: Use

To what extent do you agree with this item? (Item 4)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 4)

3000 character(s) maximum

Transparency, Item 5: Training process

To what extent do you agree with this item? (Item 5)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 5)

3000 character(s) maximum

Transparency, Item 6: Information on the data used for training, testing, and validation

To what extent do you agree with this item? (Item 6)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 6)

3000 character(s) maximum

Transparency, Item 7: Computational resources

To what extent do you agree with this item? (Item 7)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 7)

3000 character(s) maximum

Transparency, Item 8: Energy consumption

To what extent do you agree with this item? (Item 8)

- The item should be removed in its entirety

- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 8)

3000 character(s) maximum

Transparency, Item 9: Additional information to be provided by providers of general-purpose AI models with systemic risk

To what extent do you agree with this item? (Item 9)

- The item should be removed in its entirety
- The item should be substantially edited and/or further clarified
- The item should be lightly edited and/or further clarified
- The item is close to where it needs to be

Please explain your rating and suggest improvements, clearly distinguishing between concerns regarding scope/level of the item and drafting improvements (Item 9)

3000 character(s) maximum

Copyright Section Preamble

Please provide below your thoughts on the preamble of the Copyright Section of the Code of Practice.

5000 character(s) maximum

Commitments by providers of general-purpose AI models: Copyright

Copyright, Commitment I.2: Copyright policy

Please suggest improvements, clearly distinguishing between concerns regarding scope/level of the commitment and drafting improvements (Commitment I.2)

3000 character(s) maximum

Copyright, Measure I.2.1: Draw up, keep up-to-date and implement a copyright policy

Please suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure 2.1)

3000 character(s) maximum

We are concerned with the dilution of the public disclosure commitments foreseen in the previous versions of the draft code. Paragraph 2 is one of the transparency measures that has been weakened in this new draft: whereas in the 2nd draft Signatories committed to publicly disclose a summary of their internal copyright policy, in the new draft they are only encouraged to publicize such information.

The public disclosure of a summary of the copyright compliance policy would go a long way in demonstrating compliance with the obligations to put in place such a policy. While the AI Act only foresees the disclosure of a summary of the training data, having additional public disclosure measures regarding the obligations included in Article 53(1)(c) would contribute to the proper application of the AI Act.

The disclosure of a summary constitutes a proportional approach to the different interests at stake, without creating an excessive burden on the model providers. Due consideration is given to the need to protect trade secrets and confidential business information (model providers only need to disclose a summary), while still protecting the legitimate information needs of rights holders and other parties interested in exercising and enforcing their rights under Union law.

Therefore, we recommend reverting back to the text of measure 2.2 included in the 2nd draft.

Copyright, Measure I.2.2: Reproduce and extract only lawfully accessible copyright-protected content when crawling the World Wide Web

Please suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure 2.2)

3000 character(s) maximum

We welcome the deletion of the previous measure 2.4 (2nd draft) and the approach taken in paragraph (a) of this new draft to introduce a commitment to not circumvent effective technological measures as defined in Article 6(3) of Directive (EC) 2001/29. However, we recommend the deletion of the reference to paywalls. There is no unified definition of what constitutes a “paywall” and adding such an example without further qualification could once again risk introducing interpretative elements to autonomous concepts of EU law, which is not recommended. In some cases a paywall may constitute an effective technological measure, but in others it does not (e.g. when a “paywall” makes certain information difficult to read for a human, but it is still accessible by a web crawler).

With regards to paragraph (b), we welcome the qualification of “piracy domains” introduced in the first sentence, which restricts these Internet domains to those that have been “recognized as such by courts or public authorities” in the EU and the EEA. We are, however, concerned with the language introduced in the second sentence of this paragraph. While we welcome the attribution of an active role to the European Commission (which, in our understanding, would be responsible for listing such piracy domains “on an EU website”), it is not clear if that list will only contain qualified privacy domains, as mentioned in the first sentence, or if it will extend to any Internet domains listed in “relevant piracy domain lists issued by the relevant bodies in the European Union and the European Economic Area”. If the latter, this commitment would represent a setback and it would be even more problematic than the already problematic language in the 1st draft, since there is no guarantee that those “relevant lists” issued by “relevant bodies” would have followed the due process of law. If the approach based on a list of hyperlinks is maintained, then it must be made clear that such lists can only be made available on a EU website when they meet the criterion set forth

in the first sentence.

We further recommend not to limit the commitments contained in this measure I.2.2 to training data that has been obtained by “crawling the World Wide Web.” Web crawling is one of many data acquisition strategies adopted by AI model developers. Such limitation would mean that training data that is being obtained via any other way is not covered. For example, paragraph (b) would not cover a scenario where a model provider would use bit-torrent to download data sets that include large numbers of pirated works. In our view, there is no reason to limit all the rights reservation-compliance commitments to a narrow web-crawling context. This measure should definitely apply to any acts of collection of data from public sources.

Copyright, Measure I.2.3: Identify and comply with rights reservations when crawling the World Wide Web

Please suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure 2.3)

3000 character(s) maximum

We welcome the attempts to future-proof the commitment in paragraph (b), but we consider that the new draft text still falls short of expectations. Any standards other than the Robot Protocol Exclusion Protocol (REP) that may emerge over time still continue to be subject to a lower level of commitment than that required for the REP. As argued in previous submissions, the code should apply the same level of commitment to all standardized machine-readable means to express rights reservations. In our view this should be done by simply requiring Signatories to commit to comply with all standardized machine-readable means to express rights reservations.

The argument provided by the chairs during the WG meetings that the code can only mandate compliance with rights reservations that are based on an international standard is unconvincing. Such a requirement can neither be derived from the relevant provisions contained in the CDSM Directive (which only requires machine readability) nor from the provisions in the AI Act (which refer back to the CDSM Directive). In addition, it is worth pointing out that while the REP is indeed an international standard, the only way to express a horizontal (i.e., not applying to specific crawlers that need to be individually addressed) rights reservation is to disallow all crawling, which cannot be reconciled with the commitment contained in paragraph (5) of this measure. REP may be an international standard, but it was never designed to express rights reservations related to TDM, including the use of protected works to train GPAI models.

We are also concerned with the dilution of the previous public disclosure commitments regarding rights reservation compliance information. While the AI Act only foresees the disclosure of a summary of the training data, having additional public disclosure measures regarding the obligations included in Article 53(1) (c) would contribute to the proper application of the AI Act. The disclosure of such information constitutes a proportional approach to the different interests at stake, without creating an excessive burden on the model providers. Therefore, we recommend reverting back to the text of measure 2.8 included in the 2nd draft.

Copyright, Measure I.2.4: Obtain adequate information about protected content not web-crawled by the Signatory

Please suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure 2.4)

3000 character(s) maximum

The title of the measure is misleading. The measure does not apply to “protected content not web-crawled by the Signatory” but rather to “protected content that has been independently web-crawled by parties other than the Signatory.”

As we noted in our previous answer, we consider the limitation of copyright compliance obligations to content that has been obtained by web crawling to be a structural fault of the approach underpinning the 3rd draft. The copyright compliance obligations should apply to all publicly available protected content that is used to train models, regardless of the data acquisition method.

Copyright, Measure I.2.5: Mitigate the risk of production of copyright-infringing output

Please suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure 2.5)

3000 character(s) maximum

We welcome that this new draft no longer targets output similarity, in line with our previous comments. We also welcome that open source AI model providers are now explicitly excluded from the commitment to prohibit copyright-infringing uses in their acceptable use policy, terms and conditions, or other equivalent documents.

Copyright, Measure I.2.6: Designate a point of contact and enable the lodging of complaints

Please suggest improvements, clearly distinguishing between concerns regarding scope/level of the measure and drafting improvements (Measure 2.6)

3000 character(s) maximum

The clarification that repetitive complaints should be considered excessive does not serve a justifiable purpose. The mere fact that complaints are repetitive should not automatically qualify them as excessive, since repetition of complaints that are not acted upon is a perfectly legitimate course of action for parties exercising their rights. The wording “in particular because of their repetitive character” does more harm than good and should be removed.

Contact

[Contact Form](#)

