SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org

EANA: Comments and suggestions on the second draft of the Code of Practice for the Artificial Intelligence Act (AI-Act)

January 22nd, 2025

by Verena Krawarik, Jiri Majestr, Mark Westerhoff, Lilian Hübner, on behalf of the EANA Copyright Committee

Because the Code of Practice (CoP) is intended to specify the provisions and obligations of the AI Act, to build on them and to help make them implementable - and above all enforceable in practice - it must be considered in conjunction with the measures, which must have the necessary level of detail to make the CoP fit for purpose.

From our point of view, we appreciate that after the first draft of the Code of Practice and the resulting feedback, several clarifications, details and improvements have already been made. As a representative of a community of rights holders, for the members of the EANA, the protection and appropriate handling of copyright is of outstanding importance. This includes clear rules, measures and regulations on transparency in connection with the guidelines for all types of AI model providers.

While the Code of Practice will certainly have a major impact on big AI model providers, it is important to emphasize that it applies to all companies that develop or apply AI systems. From our perspective, all providers must therefore guarantee that copyright is respected and that no exceptions can be made. This requires a detailed regulation in particular for the use of copyright protected material for the purposes of AI training or RAG solutions.

SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org

Detailed feedback on the transparency and copyright related provisions is provided below:

Comments and suggestions:

TRANSPARENCY

Steakholder Group “Rights Holders”

Measure 1.1:
According to the reference table for measure 1.1 (drawing up, keeping up-to-date and providing the relevant information), currently shows two reference groups. The signatories undertake to provide the specific information listed there, but only to make it available to the AI Office and the national competent authorities upon request and/or to the downstream providers. More transparency is needed here, according to us rights holders and, as far as legally permitted, the public should also have access to this information.
We therefore suggest that the reference table lists the missing stakeholder groups, in particular the rights holders, accordingly, to make it clear that information is also provided for them. It is important for rights holders to know which body or organization will provide the information required for copyright management in the future. Especially the following information from the reference table for measure 1.1 is always of relevance to rights holders:
- Annex XI §1 2.(c) and Annex XII 2.c) Information on the data used for training, testing and validation.

SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org

COPYRIGHT

Rights enforcement across the content supply chain

The Code of Practice has to take into account that the current methods (robots.txt etc.) proposed to express opt-outs, are not sufficient and do not reflect the needs of all copyright owners that are part of the content supply chain.

This is based on the following considerations:

a) The initial situation of the news agencies:

On traditional media sites, copyrighted content is frequently published. The rights to this content are owned by with the directly publishing medium, but also very often by third-party providers, e.g. news agencies, which generally grant their media customers publication and distribution rights, but no further rights. News agencies, which are directly addressed in the Copyright Directive, cannot currently enforce their rights via a robots.txt on a third-party website. This is already in the essence of the matter. It can be assumed that the copyrights of news agencies are always affected when crawling media websites, regardless of whether the media website has attached a rights reservation or not. We therefore suggest developing a passage within the Code of Practice that specifically addresses the content supply chain and how to deal with it. At this stage, we assume that all model providers will have to enter into agreements with each individual EANA partner respectively news agency in the future if they want to remain compliant.

EUROPEAN ALLIANCE OF NEWS AGENCIES
EANA

SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org
–

b) Technical methods and standards:

i. There are currently no standardized procedures or processes in place that document and express the copyrights of all rights holders in given media content (texts, images, videos, graphics...) along the supply chain in a machine-readable way. We are therefore very interested in working on machine-readable means of appropriately expressing and, above all, enforcing a reservation of rights for news agency material. Attempts from the standardization organisation IPTC through the metadata scheme RightsML are already there but not widley used by now: (http://dev.iptc.org/RightsML-Newspaper-Licensing-Agency-NLA-case-study)

ii. The core problem with robots.txt is, that it was designed in an era where the scope was for search and related index systems, that would solely facilitate linking to the original content and thus bringing traffic/human eyeballs. AI (LLM) “answering” systems are more focused towards training and massively copying knowledge and represent that in a completely different way to automatically create new content forms and variants without bringing back any direct or indirect benefit (traffic) to the content owner. Robots.txt does not facilitate making specific reservations on specific use and as a tool is too limited. We see other systems, emerge like TDM reservation policy (W3C) were this could be possible in the future https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240510/. As there are no systems in place right now, where news agencies can express their rights towards crawlers, we are very interested to work on machine-readable means to appropriately express this rights reservation for news agency material.

EUROPEAN ALLIANCE OF NEWS AGENCIES
EANA

SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org

c) Publicly available datasets and lawful access to copyright-protected content:

We would like to emphasize that, for the reasons described above, publicly available data sets such as Laion5B or Common Crawl as well as data sets that were derived from it like C4 include a large amount of copyright protected content, that was originally provided by news agencies. While the mere collection may fall under the TDM exception for research, this does not cover the exploitation of these data collections. Access to these publicly available large data sets does therefore not imply the right to exploit or commercially use them. To execute rights, it could be a solution to establish an obligation to compile indexes (e.g. by crawlers or operators of data set holders) to websites which are crawled, so rightsholders are able to check up, if their content is crawled. A human readable database with all crawlers or large-scale data sets used by AI model providers with their indexed websites could be provided by the AI Office to the rightsholders or the information is given upon formal request of the rightsholders by the AI Office. An example for such a searchable database was introduced by Washington Post on C4 dataset in 2023: (https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/)

Our thoughts and explicit suggestions on the different Measures can be found below:

Measures 2.1, 2.2:

We see the Copyright Directive being limited to internal use without any justification or legal basis in the AI Act itself. Instead, the policy needs to be publicly available, as an internal policy is not

EUROPEAN ALLIANCE OF NEWS AGENCIES
EANA

SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org
–

subject to external audit and therefore does not effectively address the problem the law is intended to address at all.

The idea to establish an obligation to publish a summary of the policy in Measure 2.2 is for sure well intended but misses the point. There is no reason why a legal obligation to adopt a policy on respect for copyright and related rights should not be fully available.

As the provisions on copyright are often mistakenly associated only with the training of GPAI models, it is important to clarify that the introduced policy must ensure compliance with copyright law and related rights for all possible purposes, including training. The current wording limits its scope to the training and development stages. Also, it is important that changes are documented, and the policy is always updated, so any changes and newer versions must also be made public. Incidentally Art. 53(1)(c) does not refer to an internal policy. There is therefore no legal basis for withholding important information from rightsholders.

Measure 2.3:

Although it is to be welcomed that the obligations are set out in more detail, an even more precise definition of the incentives would be desirable. If signatories are merely obliged to make “reasonable efforts to assess copyright compliance”, it is entirely left to the providers to decide how and to what extent they carry out this due diligence. This leaves rights holders at the discretion of GPAI providers, who could avoid potential copyright infringement charges by relying on their “reasonable efforts”. It should be clear that signatories are required to verify that there is no copyright infringement when they enter contracts for content and datasets. If a third party cannot provide (contractual) assurances regarding compliance with copyright and related rights, a

SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org

signatory should not use content from unsafe sources. We would suggest deleting the wording of "reasonable".
As already stated, the request to the signatories to make reasonable and proportionate efforts is quite inadequate. This problem could be avoided by demanding that signatories obtain contractual assurances from the third party providing the data set. In addition, it is not comprehensible why SMEs should be excluded from measures to ensure copyright compliance.

Measure 2.5:

We consider that the "reasonable and proportionate" measures regarding piracy websites from crawling activities to train their general-purpose AI models are still too vague and leave too much scope for discretion to AI companies. Therefore, we suggest tightening this measure.

Measure 2.6, 2.7:

We are critical about the preference for robots.txt (also see above) in the Code of Practice, although we welcome the clarification in Measure 2.7. It must therefore be made clear that no appropriate opt-out can in any way lead to exclusion. In addition to the limitation to robots.txt, it is also insufficient to merely encourage signatories to take such appropriate measures.

Measures 2.6 and 2.7 should be grounded upon the right of the rights holders affected to expressly reserve the use of lawfully accessible works and other subject-matter for the purposes of TDM in an appropriate way, as provided for in Art. 4(3) of the DSM Directive. The wording on findability should also be extended to the full range of opt-out options, not just robots.txt.

SECRETARIAT
Alexandru Ion Giboi
Switzerland
Mobile: +41764.756.663
E-mail: secretarygeneral@newsalliance.org
Website: www.newsalliance.org

Measure 2.9:

Notwithstanding the affirmative obligation in Measure 2.10, it is not entirely clear what is meant by the term “overfitting”. We therefore suggest a clarification.

Measure 2.11:

It is not comprehensible to us why the obligation to give rights holders the opportunity to submit complaints should not apply to SMEs. In this respect, we would like to suggest deleting this restrictive sentence to strengthen rights holders.

ENDS