

Copenhagen, November 2024

AI and copyrights

The Danish Rights Alliance's recommendations to safeguard copyright in relation to the development and use of artificial intelligence.

Since the launch of ChatGPT and subsequent AI services, The Rights Alliance has worked actively to control and protect our member's copyright protected content from being used illegally in connection with the training and use of artificial intelligence (AI).

As of today, it is neither possible to control nor protect the use of content in relation to the development and use of AI. In connection with our work on AI, we have learned that

- Providers of generative AI do not voluntarily disclose which copyright-protected material they have used to train their models,
- it is not possible to verify whether TDM reservations are being respected,
- legal action is required to get providers of generative AI models to acknowledge that they have violated copyright
- users upload copyright-protected material to AI services, and there are insufficient measures in place to prevent this, as well as
- the mechanisms of online content-sharing services to detect, report, and remove infringing AI outputs, such as deepfakes, are inadequate

It is based on this knowledge and experiences that The Rights Alliance puts forth a series of recommendations for measures that will help protect content from being used without permission from rightsholders in connection with the development and use of generative AI.

While the development calls for an update of legislation, it is also important to note that existing copyright regulations can already prevent the extensive, unauthorized use of content that is taking place. But this requires action to address infringements

and to put an effective stop to the illegal use of content found at all stages of AI development and use. This is not a task that rightsholders can undertake alone. Nor is it a task that can be resolved solely through the legal system, as this would result in lengthy cases and years of unclear legal standing.

The Rights Alliance's recommendations also incorporate the recommendations presented by the Danish Government's expert group on Tech Giants in a report from February 2024.¹

The Rights Alliance's recommendations are described below.

1. Comprehensive transparency about the training of AI models

Transparency is key to ensuring that content is not used illegally to train AI models and is therefore a prerequisite for being able to license or enforce copyrights.

The Rights Alliance has analyzed the transparency of AI developers by systematically reviewing what the largest AI models in text, image, video, and music generation, including GPT-4 and Llama, have been trained on. The review shows that none of these models meet the needs of rightsholders to verify whether their content has been used illegally.

Find the report here:

<https://rettighedsalliancen.dk/wp-content/uploads/2024/09/Report-on-AI-model-providers-training-data-transparency-and-enforcement-of-copyrights.pdf>

Rightsholders should be given access to review the datasets used for training AI models to ensure the best possible verification of whether their content is included in the dataset. This would require AI developers to grant rightsholders access to review datasets and provide search functionalities that enable them to search the datasets.

Alternatively, AI developers should provide information that allows rightsholders to obtain comprehensive details about the following:

- Which works and creations are included in the training data (titles, rightsholders, and other identifiers such as ISBN numbers).
- The source from which works and creations have been obtained, such as website address, URL, or similar,

¹<https://www.digmin.dk/Media/638447961946861250/Delafrapportering%202%20fra%20regerings%20ekspertgruppe%20om%20tech-giganter.pdf>

- Information about whether the works and creations used were sourced from restricted-access sources, and if so, what legal access the AI developers had to the works and creations, as well as
- The date of collection of works and creations from the sources.

The EU's AI Act imposes transparency obligations on providers of AI models in the European market. This is a positive development as it demonstrates good intentions regarding the regulation of transparency. However, there is a real risk that the practical effect of the regulation will not sufficiently ensure that rightsholders will actually be able to monitor the use of their content.

The AI Act stipulates that AI providers must provide a sufficiently detailed summary of the content used to train AI models and must develop policies on how they ensure compliance with the EU's copyright regulations.

It is unclear to what extent these obligations will apply to AI providers as the discussions are ongoing in working groups under the EU AI Office.

To monitor whether content is being used illegally to train AI models, including whether any reservations for text and data mining are being respected (see point 3), The Rights Alliance recommends the following.

Rightsholders should have access to review AI training datasets or, alternatively, have access to comprehensive information that allows them to obtain relevant details about works and creations in these datasets.

The Rights Alliance recommends that Danish authorities and policymakers actively work to ensure that the implementation of transparency requirements in the AI Act is as effective as possible.

2. Reversal of the burden proof

In addition to rules on comprehensive transparency, further measures are needed to ensure that rightsholders of copyright-protected content genuinely have the ability to enforce the illegal use of their works and creations for the training of AI models. This is due to the aforementioned concerns about the implementation of transparency obligations for AI providers under the AI Act, as well as the unique circumstances surrounding AI models, where only the developers are aware of which works and creations are included in their AI training datasets.

To address this unequal legal standing, new legislation on the reversal of the burden of proof should be introduced into Danish law.

In cases where rightsholders do not have access to comprehensive information, it should therefore be presumed that their content has been used for training.

Thus, a rule on the reversal of the burden of proof should be introduced in cases involving the training of generative AI. The rule can be incorporated into chapter 57a of the Danish Administration of Justice Act, which already contains special rules on the securing of evidence in cases of infringement of intellectual property rights, or it can be added to the Danish Copyright Act, as it specifically concerns copyright.

3. Text and data mining – and reservations

It is The Rights Alliance's view that text and data mining exemptions have limited application in relation to generative AI. For example, text and data mining solely involves the analysis of existing data to identify patterns. In the development of generative AI, however, data and content are additionally stored in the generative AI's vectors, which are used to create new content.

The use of content for the development and training of AI is therefore far from limited to text and data mining but goes far beyond it. Consequently, it is misleading to equate text and data mining with AI development.

At the same time, a number of protective measures in international regulations safeguard rightsholders from exceptions in the law being used to undermine the value of copyrights. This is highlighted, among other things, by the Berne Convention's three-step test.

This implies that, to the extent that works are used and reproduced in connection with AI services, it is unlikely to solely constitute text and data mining. Rules regarding exceptions and reservations for text and data mining cannot therefore be used as an exemption from complying with copyright regulations.

To the extent that text and data mining is conducted in connection with the development of AI models, it may be legal to use content for this purpose if the person using the content has lawful access to it and respects any reservations made by rightsholders. In principle, rightsholders thus have the opportunity to make reservations ensuring that their content can only be lawfully used for text and data mining.²

The uncertainties surrounding what practically constitutes text and data mining creates legal uncertainty for both rightsholders and those who use content to develop AI. This can hinder innovation and may lead Danish entrepreneurs and established businesses to refrain from developing AI solutions. This situation is unsustainable, and it would be insufficient to only await legal precedents in this area, as this can at best take several years. To ensure progress and innovation, Danish authorities should actively address these issues.

There is a need for authorities to guide Danish institutions and businesses on when text and data mining may be conducted, when reservations must be respected, and in which cases agreements must be made with rightsholders for the use of content to train AI.

To the extent that text and data mining is involved, rightsholders, as mentioned, have the opportunity to make reservations in connection with agreements with distribution channels, as well as utilize the technical options provided by AI developers.³

However, it is not practically possible for rightsholders to make reservations across all necessary distribution channels, nor can rightsholders practically monitor

² See the Danish Copyright Act §§ 11b and 11c, which implement the rules on text and data mining from the DSM Directive.

³ For example, many website owners use a solution called robots.txt to indicate to so-called crawlers that they do not want their content copied by these crawlers. Many AI developers use such crawlers to collect content for training their AI models. Robots.txt is not a technical solution that prevents copying, but rather a kind of sign or message to crawlers visiting the website. It has been shown that some AI companies do not respect these signs or messages in robots.txt where website owners request that their content not be copied.

whether a reservation they have made is being respected. Content is distributed through many different services and channels, and it is not feasible for rightsholders to know which distribution channels their content is spread from, nor can they ensure that any conditions they set regarding reservations in licensing agreements are actually upheld.

The practical reality is that rightsholders' ability to make reservations and verify whether AI developers adhere to these reservations does not exist.

Regardless of the practical ambiguity about when text and data mining apply and when it does not, there is a need to find practical solutions to the fact that the system around reservations does not work in practice.

A dialogue should be established between authorities, rightsholders, distributors, and users of text and data mining on how to best implement reservations in practice and how to ensure that they are respected. It is essential that rightsholders can make reservations without incurring burdens in terms of time and costs.

4. Content recognition tools in the same style as "DSM Article 17 tools"

Users of AI services are widely uploading copyright-protected content to these services without permission from the rightsholders, in order to use AI technology on the works and creations.

Today some AI services function like online content-sharing platforms similar to YouTube, where users can upload protected works and creations to the service, after which these are made available to the public.

A common issue with these AI services is the lack of measures to prevent users from uploading illegal copies of copyright-protected content to the services.

Another issue is that AI services are used to generate output that infringes upon copyright-protected content, such as output "in the style of," which includes protected elements. Output can also infringe on the personality rights of Danish individuals, such as using a person's voice, image, or other personal characteristics in the output.

It is problematic that there are no measures in place to counteract such illegal use of content. Therefore, measures should be integrated into AI services to ensure that

users cannot upload protected content that is not permitted by the rightsholders. Similarly, it should be ensured that AI services cannot generate output that is copied or imitated from protected works. It must also be ensured that content cannot be created "in the style of" if it would constitute a copyright infringement.

Danish authorities should work to ensure that AI services incorporate content recognition and enforcement tools that can filter illegal content.

5. Regulation of personality rights

The right to one's own voice, image, and personal characteristics has become more relevant than ever before. Because the development of AI makes it easy and quick to create content using other people's voices, images, etc., there is a need to codify personality rights.

This would make it possible to enforce the misuse of voices that are cloned and distributed on the internet, including on online content-sharing platforms such as TikTok, YouTube, and Instagram.

The regulation of personality rights can also serve as a foundation for clear information and communication about what is allowed and not allowed with other people's voices and similar characteristics.

Legislation should be introduced to ensure that personality rights are regulated when it comes to digitally generated content. The legislation should apply broadly and cover all individuals' voices, images, and personal characteristics.

Campaigns and educational materials should be developed to inform about the rules in this area and to encourage the lawful use of AI.

6. Effective tools to block and stop the distribution of deepfakes including voice clones on online platforms

Illegal AI-generated content is widely found on social media and other online platforms. Unfortunately, there are currently no effective tools to ensure that such content is not distributed on these online platforms.

Therefore, it must be ensured that online platforms provide effective tools to stop and remove illegal AI-generated content that is distributed on these platforms.

Obligations to ensure that illegal content is not spread on online platforms are already regulated in the legislation.

Clear requirements should be set for online platforms regarding the authorities' expectations for blocking and stopping this type of illegal content.

7. Information and guidance on AI and copyright

The Rights Alliance has identified a significant need for legal guidance for businesses and institutions that wish to use copyright protected content to train their own AI models or finetune and interact with existing large generative AI models.

Additionally, there is an urgent need to raise awareness of the legal limitations for institutions, educators, and students using customized chatbots at Danish educational institutions.

Deepfakes including voice clones are also areas where The Rights Alliance sees a large need for information.

As AI spreads and continues to develop, the need for information and education on AI and copyright will only increase.

There is a high likelihood that AI-generated content infringes on rights, as the large and popular generative AI models currently on the market have already been trained on large amounts of content without the rightsholders' permission.

Therefore, awareness initiatives should be launched as soon as possible to help increase users' (businesses, institutions, and individuals) understanding of what is right and wrong regarding the use of copyright protected material in the context of generative AI.

Funds should be allocated for this purpose.

8. Resources for police and prosecutors

Illegal training of AI models and the spread of deepfakes, voice clones, and similar content are expected to reach a speed and scale that rightsholders will not be able to manage on their own.

It is assumed that a large portion of the infringements that will occur will take place on online platforms, where rightsholders will face significant challenges in prosecuting the responsible parties. It should be more of a public responsibility than it is today to address the illegal use and distribution of content on the internet.

In cases where illegal content is systematically disseminated, such as deepfakes, voice clones, music, films, and books, the police should be provided with resources to handle these cases.

Dedicated resources should be allocated for the police and prosecution office to deal with these issues and it should be considered if there should be a higher degree of public prosecution in cases involving the spread of illegal material on online platforms.

Authorities in Denmark should, to a much greater extent than is currently the case, monitor and ensure that Danish laws are respected on online platforms.

/The Danish Rights Alliance