



18 NOVEMBER 2024

Copyright and AI: For effective implementation of existing rules

Executive summary

The rapid growth of artificial intelligence (AI) has brought opportunities and challenges to copyright. AI technologies can generate creative content, streamline processes and offer new business models, pushing the boundaries of innovation. On the other hand, this advancement raises complex questions about the protection of creators' rights, the potential for infringement and the ownership of AI-generated content.

To ensure an effective balance, the European Commission should focus on:

- ▶ **Careful implementation of the Copyright Directive and AI Act:**¹ The Copyright Directive was agreed following heated debates between rightsholders and the tech industry, achieving a delicate balance between their interests. Reopening the legislation after the AI Act's finalisation would reduce legal certainty for all players – it would fail to resolve any of the current implementation issues, as negotiations would likely take years.
- ▶ **Operationalising the text-and-data mining (TDM) opt-out:** The Commission should continue to support the development of harmonised and internationally recognised standards to ensure a simple and effective opt-out mechanism for all rightsholders and AI model providers, regardless of their size, sector or type of work. Such work should be based on existing, widely used standards rather than the development of new standards.
- ▶ **Supporting transparency:** AI providers must comply with EU copyright laws, including respecting opt-outs, and provide a transparent summary of training data. The training data summary should be sufficiently detailed to be helpful to rightsholders, technically feasible, avoid excessive burden and safeguard trade secrets, confidential business information and model safety.

¹ Directive (EU) 2019/790 and Regulation (EU) 2024/1689, respectively.

- ▶▶ **Exploring copyright status of AI outputs:** AI-generated content without any human intervention should not receive copyright protection. However, AI-assisted content that demonstrates the required level of creative human input may qualify for copyright protection. No legislative changes are necessary as courts will clarify the situation on a case-by-case basis.

Table of contents

- **Executive summary**..... 1
- **Table of contents**..... 2
- **Implementing the Copyright Directive**..... 3
 - Operationalising the text-and-data mining opt-out..... 3
- **Implementing the AI Act provisions** 4
 - Respecting EU copyright law 4
 - Ensuring transparency of training data..... 5
- **Copyright status of AI outputs**..... 5



Implementing the Copyright Directive

The Copyright Directive should not be reopened. It represents a carefully negotiated compromise that balances the interests of various stakeholders. Reopening it now would delay implementation progress and create additional legal uncertainty, potentially hindering growth and innovation in the European digital economy.

Importantly, the Copyright Directive already addresses the use of TDM, which underpins the development of many AI models. The development of AI was widely discussed when the Copyright Directive was being negotiated in 2018, and TDM, as defined in the Directive, covers AI training.²³

The AI Act, which addresses copyright concerns related to generative AI, including by directly referring to the Copyright Directive, has recently been finalised, and its implementation has just begun. Therefore, efforts should be focused on properly implementing the AI Act and the Copyright Directive.

The Copyright Directive was agreed following heated debates between rightsholders and the tech industry, achieving a delicate balance between their interests. Given this, there is no need to reopen the legislation now. Revising the Directive would reduce legal certainty for all players and would fail to resolve any of the current implementation issues, as negotiations on a new proposal would likely take several years.

Operationalising the text-and-data mining opt-out

Developing general-purpose AI models requires substantial amounts of diverse data for training to function properly and to mitigate biases and risks in AI systems.⁴ This is why most AI model providers rely on large-scale collection of openly accessible data on the internet, known as ‘web crawling.’

According to the Copyright Directive, if rightsholders do not want their work collected and used for training generative AI but still want it available online, they must effectively express a machine-readable reservation.⁵ The robots

² Art. 2(2) Copyright Directive defines ‘text and data mining’ as ‘any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.’

³ See the 2018 open letter ‘Europe needs to adopt a broad and mandatory text and data mining (TDM) exception to guarantee a thriving European research and innovation sector,’ available at <https://cdn.digitaleurope.org/uploads/2018/12/FINAL-Open-Letter-Mandatory-TDM-Exception-Copyright-trilogues-13November.pdf>.

⁴ When it comes to high-risk AI systems specifically, Art. 10(3) AI Act provides that training, validation and testing data sets used whilst developing the high-risk AI system must be relevant, sufficiently representative and, to the best extent possible, free of errors and complete in view of the AI system’s intended purpose.

⁵ Art. 4(3) and Recital 18 Copyright Directive.

exclusion protocol ('robots.txt') provides a mechanism for a website owner to indicate which parts of its website should not be crawled.

Whilst it is a widely used web protocol for expressing opt-out preferences, it has room for improvement. Several stakeholders have called for the ability to indicate opt-out more granularly, and some rightsholders have expressed concerns that web crawlers may ignore expressed opt-outs.

The Commission should, therefore, support the protocol's refinement as well as the development of other internationally recognised standards that facilitate rightsholders' TDM opt-out in a machine-readable format and that can be implemented by model providers.

Model providers are engaging to explore effective machine-readable approaches to choice and control for web publishers. Initiatives like the TDM Reservation Protocol and 'Do Not Train' tools offer helpful contributions to the broader discussion.⁶ Additionally, a voluntary code of conduct on the responsible use of web crawlers could establish guidelines for responsible data collection practices.

The Commission should continue to support the development of harmonised and internationally recognised standards to ensure a simple and effective opt-out mechanism for all rightsholders and AI model providers, regardless of their size, sector, or type of work. Such work should be based on existing, widely used standards rather than the development of new standards.



Implementing the AI Act provisions

Respecting EU copyright law

Providers placing general-purpose AI models on the EU market must ensure compliance with the relevant copyright legislation. To this end, the AI Act requires providers to establish policies to ensure they comply with EU law on copyright and related rights, particularly to identify and comply with opt-outs expressed by rightsholders under Art. 4(3) Copyright Directive.⁷

This will help ensure that AI providers respect existing copyright law. We hope the forthcoming code of practice for providers of GPAI models clarifies and provides helpful guidance on the elements that should be included in the copyright policy that providers must implement. However, guidance on implementing this provision must not go beyond what is required by the AI Act, Union copyright legislation and existing CJEU case law.

⁶ TDM Reservation Protocol (TDMRep), Final Community Group Report, February 2024, available at <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/>, and the 'Do Not Train' tool developed by spawning.ai.

⁷ Art. 53(1)(c) AI Act.

Ensuring transparency of training data

A key feature of the AI Act is its transparency obligations for providers of general-purpose AI models. The transparency obligation regarding training data should help copyright holders identify whether their content has been used in the training of AI models so that they can exercise and enforce their rights. The AI Act mandates that providers of general-purpose AI models make publicly available a ‘sufficiently detailed’ summary of the content used for model training.⁸ The AI Office will create a template to assist providers in implementing this requirement.

The summary should be ‘generally comprehensive’ yet accessible to make the transparency obligation meaningful and actionable.⁹ It must provide sufficient detail to help rightsholders enforce their rights and operationalise the reservation of their content. The summary should include the main data collections or sets, including large public and private datasets, and provide narrative explanations of the other sources used in training.

Given the highly confidential and competitive nature of the training dataset composition, preparation and licensing agreements, the template for the summary must protect trade secrets and confidential business information.

For instance, it is crucial that information on data filtering, processing and enrichment, and statistical properties and training parameters (such as weights) are not disclosed. They constitute confidential and strategic information that competitors could exploit. Similarly, where a provider cannot disclose information related to confidential licensing agreements, it is reasonable that such information should remain outside the scope of disclosure obligations.

The training data summary template must strike a balance to ensure it is technically feasible, avoid placing excessive burden on providers, safeguard trade secrets, confidential business information, model safety and facilitate copyright holders to exercise their rights.

The AI Act’s copyright-related provisions will be further detailed through a code of practice for general-purpose AI model providers.¹⁰ DIGITALEUROPE will contribute to the code’s drafting process over the coming months to ensure the code brings legal certainty to all stakeholders.



Copyright status of AI outputs

⁸ Art. 53(1)(d), *ibid.*

⁹ Recital 107, *ibid.*

¹⁰ <https://digital-strategy.ec.europa.eu/en/news/ai-act-participate-drawing-first-general-purpose-ai-code-practice>.

It is important to distinguish between machine-generated content created autonomously without any human intervention ('AI-generated content') and content created by humans using AI tools ('AI-assisted content').

Purely AI-generated content without any creative human input does not benefit from copyright protection under current legislation, and EU law should not be modified to offer such protection.

AI-assisted content could be eligible for copyright protection, provided it meets the required level of creative human input. This can come in the form of editing, arranging, or combining with other human or AI-generated work. The presence or absence of creative human input is a nuance that will need to be addressed on a case-by-case basis. Increasingly, creators are integrating AI-based tools into their creative process. In these cases, the final product may qualify for copyright protection.

At this stage, there is no need for legislative changes to clarify the conditions under which AI-assisted content should be protected by copyright. Existing copyright law is sufficiently flexible to address the current challenges posed by AI. National courts and the Court of Justice of the EU will continue to provide additional clarity through their rulings in the coming years.

FOR MORE INFORMATION, PLEASE CONTACT:



Hugh Kirk

Senior Manager for Consumer, IP and Platforms Policy

hugh.kirk@digitaleurope.org / +32 490 11 69 46



Julien Chasserieau

Associate Director for AI & Data Policy

julien.chasserieau@digitaleurope.org / +32 492 27 13 32



Alberto Di Felice

Policy and Legal Counsel

alberto.difelice@digitaleurope.org / +32 471 99 34 25



About DIGITALEUROPE

DIGITALEUROPE is the leading trade association representing digitally transforming industries in Europe. We stand for a regulatory environment that enables European businesses and citizens to prosper from digital technologies. We wish Europe to grow, attract, and sustain the world's best digital talents and technology companies. Together with our members, we shape the industry policy positions on all relevant legislative matters and contribute to the development and implementation of relevant EU policies. Our membership represents over 45,000 businesses that operate and invest in Europe. It includes 108 corporations that are global leaders in their field of activity, as well as 41 national trade associations from across Europe.