

UNDERSTANDING ARTIFICIAL INTELLIGENCE

Copyright & AI – Explained

This explainer on copyright and artificial intelligence (AI) is part of CCIA Europe's 'Understanding AI' series, which aims to inform EU policymakers and the wider public about important concepts related to AI and the EU regulatory framework.



What is copyright in the AI context?

The relationship between copyright and AI has attracted a lot of political attention recently. Vast amounts of data, sometimes including copyrighted content, are used to train AI systems. And the **existing EU Copyright Directive already allows such use**, but it also **empowers rightsholders to object** to their content being used for AI training.

However, now that certain **AI systems are also able to generate content** – such as text, audio, video and visuals – from scratch, this also raises the question of **whether AI-generated output can be protected by copyright** and how this would work in practice.

These two issues come to the fore at different stages of the AI process: the first instance relates to input data, and the second to output data generated by an AI system. Currently, it is commonly agreed at international level that AI-generated outputs cannot be subject to copyright protection, as they lack the necessary element of human creation. This **explainer therefore focuses on the issue of using copyrighted data for training**.

What does this mean in practice?

Data is crucial for training AI models. It is used to finetune their performance, but also to allow them to adapt to new environments and make more accurate predictions. Training AI systems using **diverse and representative data also helps to reduce the risk of bias**.

In some cases, the datasets used to train AI include copyrighted content that is publicly available on the internet. This has policy implications, as described in the next section.

What are the implications for AI policy?

After four years of negotiations, the [EU Copyright Directive](#) was officially adopted in 2019. It should have been implemented by all EU Member States in 2021, although some [have been slow](#) to transpose the Directive into national law and others [have yet to do so](#). Under these European copyright rules, **protected works used to train AI systems require the developer to first seek the rightsholder's consent**, unless certain exceptions apply.

One **important exception to this, is in the case of so-called text and data mining (TDM)**. This essentially already **allows developers to copy large quantities of data or text** and automatically analyse it to identify patterns. That is, provided this information is publicly available on the internet and that respective rightsholders have not objected to such use.

Following calls from parts of the creative industry and questions from Members of the European Parliament, Thierry Breton (European Commissioner for the Internal Market) [clarified](#) that the 2019 directive also applies to AI developers, so “the **creation of art works by AI does not deserve specific legislative intervention**”.

Commissioner Breton also confirmed that the directive **strikes a good “balance between the protection of rightsholders, including artists, and the facilitation of TDM, including by AI developers”**. In parallel, leading AI developers such as [OpenAI](#) and [Google](#), are in the process of signing licensing agreements with press publishers to use their content in exchange for reasonable remuneration.

Nevertheless, the European Parliament recently [put forward](#) amendments to the [AI Act proposal](#), currently under negotiation, which **would oblige generative AI developers to disclose a summary of the copyrighted content used to train their systems**.

In reality, this requirement would be impossible to meet, as it **would essentially amount to disclosing a summary of all the content available on the internet**.

What are sensible AI copyright rules?

Introducing specific **copyright rules in the EU’s AI Act would go far beyond its scope** and objectives, and risks significantly delaying the progress made on this important file.

Furthermore, as Commissioner Breton made clear, **AI is already covered by the existing EU Copyright Directive**. Fostering industry collaboration to aid the development of state-of-the-art standards for efficient rights management could be explored separately.