

Kluwer Copyright Blog

Open AI's vision for a social contract – of things to come...

Bernd Justin Jütte (University College Dublin) · Monday, June 3rd, 2024

On 7 May 2024, Open AI published its [approach to data and AI \(ADAI\)](#). This statement sets out OpenAI's vision for a 'social contract for content in AI'. In this vision, OpenAI shares its own perception of its contribution to creative ecosystems, but also its understanding of some of the legal implications of its business model.

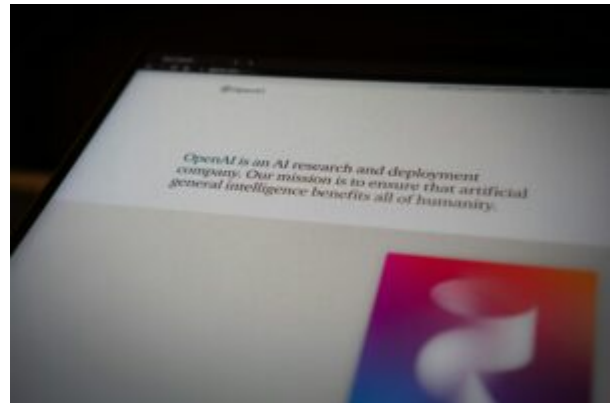


Photo by Jonathan Kemper on Unsplash

OpenAI's vision for a fair AI ecosystem

The ADAI mentions copyright only in passing, but it is fairly obvious that the desire to 'deal' with pending copyright infringement claims (largely in the US) lies at the core of this ambitious statement. In its second paragraph the statement makes strong allusions to copyright law – the term 'copyright' is used only once throughout the entire text. It speaks of benefits it strives to create for 'all of humanity' including 'creators and publishers'. OpenAI expresses that it believes that learning constitutes fair use (see [here](#)), referring to (unnamed) legal precedents, and implying that machine learning is equivalent to human learning processes (which it arguably is not, see [here](#)). However, OpenAI then states that it '[feels] that it's important we contribute to the development of a broadly beneficial social contract for content in the AI age.'

This part of OpenAI's ADAI reflects an almost subjective and emotional perception that something is not right in the current distribution of benefits between operators of GenAI systems and those whose 'creations' are used for training them. OpenAI therefore makes a commitment to respect the 'choices of creators and content owners' and to 'listen and to work closely' with stakeholders from these communities. For this purpose, OpenAI underlines that it is 'continually improving [their] industry-leading systems to reflect content owner preferences'.

In a nutshell, OpenAI recognizes that exploiting works of content creators without their consent – and without consideration – will antagonize these stakeholder groups. To avoid discontent amongst those who feed OpenAI's products with (indispensable) information, it implicitly commits to

complying with Article 4 of the [CDSM Directive](#), which requires that text and data mining (TDM) respects opt-outs from rightholders for commercial TDM activities; under US law, no such obligation exists. In other words, OpenAI makes compliance with mandatory EU law a mission statement that underpins its business model. For that purpose (i.e. to comply with that legal obligation) it further commits to liaise with the affected interest groups in what one might call a small stakeholder dialogue – the parallels to Article 17(10) CDSM Directive are uncanny.

Looking ahead, and having realized the practical limitation of a an opt-out in ‘an appropriate manner, such as machine-readable means’ (cf. Article 4(3) CDSM Directive), OpenAI announces the introduction of Media Manager ‘a tool that will enable creators and content owners to tell [OpenAI] what they own and specify how they want their works to be included or excluded from machine learning research and training.’ OpenAI adds that new features will be added in the future. En route to a planned roll-out in 2025 the company will collaborate with relevant stakeholders.

With Media Manager in development one can only speculate what opportunities it will bring for authors and rightholders. However, it becomes clear from ADAI that it is designed to make creators stakeholders in OpenAI with the aim of creating trust and enabling OpenAI to gain lawful access to vast amounts of information.

A rich metadata repository

One consequence of OpenAI’s plans will be the creation of a large pool of metadata. Having introduced its own [web crawler permissions system](#) in 2023, the company recognized the latter’s shortcomings. As opposed to traditional content recognition technology, opt-outs via [robots.txt](#) enable rightholders to opt out specific sources, i.e. websites under their control which contain their own works. The ability to opt out of TDM therefore lies in the hands of website operators and not necessarily rightholders. Once content is made available online its reuse, unlawfully or under a license agreement (or possibly as an exercise of a copyright exception), on TDM -permitting websites would enable OpenAI (in this case) to consume the content unless every website containing that content opts out of TDM.

With Media Manager, it appears, Open AI intends to create a large database of protected works against which it can check its own products post-TDM. Control is thereby handed back to rightholders as opposed to website operators. Rightholders would of course be interested in ensuring that their preferences are respected. However, they would presumably require surrendering large amounts of metadata and content samples to OpenAI against a commitment not to use them for specific purposes. The database of content information that OpenAI could assemble would be vast.

The mechanism to construct this metadata-database foreseen resembles that effected by Article 17(4) CDSM Directive (see [here](#)), which incentivizes rightholders or rightholder organizations to provide information to online content-sharing service providers to ensure that unauthorized uploads are prevented. The result that OpenAI seeks to achieve is similar, to give rightholders the opportunity to decide, in a given situation, how their works will be used if they have been captured despite the employment of automated opt-outs on websites. Copyright enforcement is thereby moved to a private relationship; whereas this shift was legally mandated by the CDSM Directive,

OpenAI proposes the privatization of copyright enforcement on its own initiative.

Value beyond content

By gaining access to content information OpenAI would already generate significant value, disregarding whether access to content information comes along with permission to use this information for purposes other than implementing user preferences. Bypassing the opt-out model of the CDSM Directive, Media Manager would create a direct link to rightholders, enabling OpenAI to negotiate directly and likely more efficiently with rightholders. Again, the resemblances to Article 17 CDSM Directive are glaring. At least for large rightholders a permissions-platform offers an efficient tool to control the use of their works for machine learning. Inputs into different products can be managed and potentially competitive uses prevented. The opportunities (for OpenAI as well as rightholders) are immense, though it is too early to assess the concrete implications at this point.

The biggest value of the ADAI lies in an opportunity to forge [sustainable](#) relationships between rightholders and providers of GenAI systems. The tension that erupted in a multitude of lawsuits in the US can be better contained through a collaborative, consent-based model for the use of copyright protected subject matter. Creating acceptance through communication channels might indeed be the right way to go – at least for some stakeholders.

The long game

Other than promising to collaborate with creators (not authors) and publishers to create ‘mutually beneficial partnerships’ and support ‘healthy ecosystems’ with a view to explore ‘new economic models’ OpenAI’s approach to data and AI offers little in terms of concrete commitments. But already the collection of vast amounts of information would be a tremendous feat and enable OpenAI to develop business models based on a treasure trove of willingly surrendered content and metadata – how that content will be used is then another question.

Extending a hand to rightholders can also be seen as a strategic move. It demonstrates an ambition to use lawfully obtained training data as a workaround of the cumbersome opt-out solution (see [here](#)) and sets an example the obligations arising under Article 53 of the [AI Act](#) can be reasonably complied with.

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please [subscribe here](#).

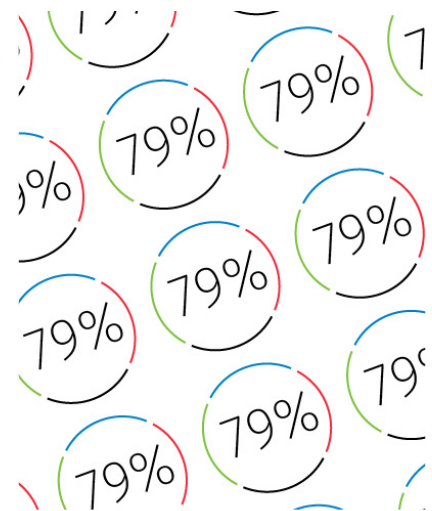
Kluwer IP Law

The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?

Learn how **Kluwer IP Law** can support you.

79% of the lawyers think that the importance of legal technology will increase for next year.

Drive change with Kluwer IP Law.
The master resource for Intellectual Property rights and registration.



2022 SURVEY REPORT
The Wolters Kluwer Future Ready Lawyer
Leading change

This entry was posted on Monday, June 3rd, 2024 at 8:09 am and is filed under [Artificial Intelligence \(AI\)](#), [CDSM Directive](#)

You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.