Creative Commons Licenses and Generative Artificial Intelligence: A Synergy for Knowledge Sharing

Author

Deborah De Angelis

Introduction

Creative Commons (CC) licenses are designed to facilitate the sharing of knowledge and creativity, allowing copyright holders to specify the permissions granted to users, thereby enabling widespread use of their works. The main features[1] of CC licenses can be summarised as follows: 1. Irrevocability: once a work is licensed under Creative Commons, the license cannot be revoked, thus ensuring that the protected material remains perpetually available for the specified use; 2) Attribution: users must provide correct information regarding the rights regime of the work and include a link to the license, as well as indicate whether any modifications have been made. This transparency requirement also supports the ethical use of creative content during the training phase of generative artificial intelligence systems; 3) Prohibition of Technological Protection Measures: CC licenses prohibit the application of such measures, which would otherwise limit the freedoms granted by the license in promoting Open Access; 4) Respect for exceptions and limitations to copyright (as well as the doctrines of Fair use and Fair dealing in Common Law countries): CC licenses state that, in the presence of such cases, user rights are not affected by the licenses. Therefore, where exceptions, limitations, fair use, or fair dealing apply, CC licenses do not come into effect; 5) Modifications: as already mentioned, starting from version 4.0, users must indicate whether the licensed material has been modified, along with any prior modifications.

In recent years, particular attention has been given to the rapid evolution of artificial intelligence (AI), understood as a field of computer science that develops systems and algorithms capable of performing tasks that would typically require human intelligence.
Generative AI is a subclass of AI that focuses on creating new content, such as images, text, video, or sound, based on foundation models and large language

---

[1] D. De Angelis, V. De Vecchi Lajolo, Funzionamento delle licenze CC e, in particolare, della clausola NC. Una panoramica approfondita tra l'Italia e la Germania", su Diritto Industriale, n. 4/16.

models (LLMs). LLMs are large statistical models trained on enormous amounts of data using deep learning techniques[2].

Copyright-protected works are used to train these generative AI systems. In this context, reflections have emerged regarding the relationship between such uses and the protected content shared under CC licenses.

The role of open licenses in the development of generative Artificial Intelligence

In Artificial Intelligence (AI), particularly generative AI, the use of copyright-protected content released under CC licenses plays a fundamental role. This article explores the interaction between generative AI and Creative Commons licenses, highlighting the principles of knowledge sharing and addressing key legal considerations[3] in the field.

Generative AI models rely heavily on vast datasets to create text, images, and other media. These datasets often include copyright-protected works also released under CC licenses and non-protectable works in the public domain.

Creative Commons licenses are particularly significant because they allow the use of creative works for various purposes, including AI training, without the need for specific individual permissions, provided that the terms of the licenses themselves are respected.

---

[2] K. Tyagi, Copyright, text & data mining and the innovation dimension of generative AI, in *Journal of Intellectual Property Law & Practice*, Volume 19, Issue 7, July 2024, Pages 557–570, https://doi.org/10.1093/jiplp/jpae028

[3] Understanding CC licenses and generative AI, K. Walsh, https://creativecommons.org/2023/08/18/understanding-cc-licenses-and-generative-ai/

Legal considerations on Artificial Intelligence training

One of the main issues in using materials released under Creative Commons licenses for Artificial Intelligence training is determining whether such use falls within any of the exclusive rights reserved to rights holders under copyright legislation[4]. The interpretation of these exclusive rights can vary between civil law jurisdictions and common law systems.

It is also important to note that, in the context of generative AI systems, each system has its functionality, which, for the analysis mentioned above, requires a case-by-case evaluation to determine whether and which exclusive rights have been exercised during the training activity.

Exceptions for Text and Data Mining (TDM) in the EU

European copyright legislation, where the exception for text and data mining (TDM) is explicitly defined, provides a clear legal framework for using copyright-protected works during the AI training phase. The European Union Directive on Copyright in the Digital Single Market, 2019/790/EU (CDSM Directive) addresses TDM exceptions in Articles 3 and 4.

In particular, Article 3 permits TDM activities for scientific research by qualified beneficiaries (cultural heritage institutions and research organisations) with

---

[4] R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 European Intellectual Property Review, 2021/5, p. 322-337.

lawful access[5] to the works. Article 4, on the other hand, allows TDM activities by anyone with lawful access to such works for any purpose, provided that the rights holder has not expressly reserved such use[6].

The obligation to comply with the terms and conditions of the CC license for conducting TDM activities depends on whether the type of extraction activity involves the exercise of one of the economic use rights of copyright or sui generis right applicable to databases. If not, there is no need to rely on the contractual conditions of the license. However, since there are different methods of conducting text and data extraction, some types of such activity may involve the exercise of some of the exclusive rights granted by the license.

In this context, the question has been raised as to whether the No Derivative Works (ND) and No Commercial Use (NC) clauses in CC licenses can be interpreted as the exercise of the reservation, governed by Article 4(3) and Recital 18 of the CDSM Directive (and any national implementing legislation).

The answer can only be no, considering, first and foremost, that the rationale behind CC licenses is to encourage the circulation of content rather than to impose restrictions. In the European Union Member States context, the use of a work licensed under a CC license for TDM purposes for commercial purposes falls under the exception established by Article 4 of the CDSM Directive. Therefore, as previously noted, the CC license has no effect when an exception applies.

CC licenses do not reduce, limit, or restrict the rights provided by exceptions and limitations to copyright[7]. Therefore, the contractual terms of a CC license cannot constitute a reservation of rights within the framework of a copyright exception. Consequently, no internal element of CC licenses can be logically or systematically interpreted as a reservation of rights under Article 4.

---

[5] In the Italian implementation of Article 3 of the CDSM Directive, within Article 70-ter of the Copyright Law, this legal requirement was transposed using the term "lawful access," which differs from the meaning of "legitimate access.

[6] D. De Angelis, *Le eccezioni per scopo di ricerca: il mondo guarda all'UE,* in Crisi e resilienza del diritto d'autore. Il recepimento italiano della direttiva 2019/790, Giappichelli, pagg. 21- 38.

[7] Art. 2, a, 2 CC *legal code*.

This argument leads to the corollary that any "explicit reservation" of the use of the work under Article 4(3) of the CDSM Directive must be formally exercised outside the framework of the adopted CC license. Such a reservation would render the exception ineffective, and, as a result, the terms of the CC license would once again apply.

In this sense, a work licensed under CC BY-NC (Attribution, Non-Commercial), where an opt-out reservation under Article 4(3) of the CDSM Directive is explicitly and separately expressed from the content of the license itself, can still be subject to extraction for non-commercial purposes (but not for commercial purposes). The reason is that the sole function of the reservation is confined to the scope of the said article of the Directive and cannot contradict the broader obligations of a CC license, in this case, the "NC" clause.

A similar argument can be made for the ND (No Derivatives) and SA (Share-Alike) clauses when the outcome of the extraction results in adapted material[8].

Common Law Countries, Fair Use for TDM, and CC Preference Signals

In Common Law jurisdictions, instead of the system of exceptions and limitations to copyright, the doctrine of fair use applies—a more flexible legal framework but one that is inherently unpredictable. The fair use doctrine assesses the legality of use based on the analysis of four factors: 1) the purpose and character of the use, including whether it is commercial or non-commercial; 2) the nature of the copyrighted work; 3) the amount and substantiality of the portion used; and 4) the effect of the use on the potential

---

[8] K. Szkalej, M. Senftleben, *Generative AI and Creative Commons Licenses: The Application of Share Alike Obligations to Trained Models, Curated Datasets and AI Output,* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4872366

market for the original work. Based on a case-by-case evaluation, this approach could create legal uncertainty for AI developers, potentially slowing down the development of the target market.

Creative Commons is exploring a new method to enable rights holders to express their preferences regarding using protected works for training generative AI systems. As mentioned, such preference signals cannot be enforced through the terms of CC licenses. Instead, they aim to allow creators to express a broader range of intentions and encourage better sharing of their content while respecting the values and principles of the Open Access movement.

The so-called "Preference Signals" for AI represent a way for creators to indicate their choices regarding using their works for training AI models[9]. Introducing preference signals is a means to support and promote sharing materials that might otherwise not be shared, identifying new ways to reconcile current tensions[10].

Unlike the opt-out reservation provided in the TDM exception for commercial uses, preference signals would not have a legal effect and, therefore, would not be legally enforceable. In the context of European copyright law, in the case of exercising the opt-out option and thus maintaining the full validity of CC licenses, the exception would no longer apply, and the contractual terms of the license would regain their effect, such as the NC or ND clauses. In this sense, preference signals will likely have more significance and a broader scope to the contractual terms of more open licenses (CC BY and CC BY-SA).

In conclusion, Creative Commons licenses play a crucial role in enabling the use of creative works for training generative AI models and promoting knowledge and innovation sharing. While the legal frameworks of civil law countries provide practical guidelines through TDM exceptions, common law

---

[9] A. Tumadóttir, *Question for consideration for AI & the Commons*, qui: https://creativecommons.org/2024/07/24/preferencesignals/

[10] P. Keller, A. Tarkowski, *The Paradox of Open*, in https://paradox.openfuture.eu/

countries rely on the flexible yet uncertain fair use doctrine. As AI technology progresses, ongoing dialogue and legal clarity will be essential to balance the rights of creators with the benefits of open knowledge sharing. CC believes it is possible to balance creators' rights with the development of AI systems by developing the aforementioned preference signals, on which it is focusing its efforts in research and interpretation and for which it is expected to provide further details shortly.