

Kluwer Copyright Blog

Generative AI: Remunerating the human author & the limits of a narrow TDM Exception

Kalpana Tyagi (Maastricht University) · Wednesday, December 13th, 2023

In his classic work, ‘Capitalism, Socialism and Democracy’, Josef Schumpeter referred to the ‘waves of creative destruction’ to describe how monopoly rents incentivise entrepreneurs to take risk and innovate. The monopoly rent that the entrepreneur derives from his innovation is short-lived, as another wave of creative destruction soon replaces this wave, and gives way to another wave of innovation. In the era of Artificial Intelligence (AI), Schumpeter’s ‘waves of creative destruction’ have become shorter and taken the shape of, what I call, ‘waves of creative ‘digital’ destruction’. And like every ‘wave of creative ‘digital’ destruction’, the latest wave of Generative AI, too first knocked on the gates of copyright.



Photo by Markus Spiske on Unsplash

Generative AI offers an unprecedented potential to create any kind of work – be it written, audiovisual, or even programming code – upon a mere prompt. As the entire value chain of Generative AI seems centered around copyright, it raises questions on both the input side and the output side (*see* [here](#), [here](#), [here](#) and [here](#)). From the output side, the relevant issues are whether the output is copyright-protected, and whether it infringes the copyright of ‘works “ingested” during the training stage of the AI system’ (*see* Quintais, [here](#); *see* also [here](#) and [here](#)). It is also relevant to consider whether ‘the output [is] a “derivative work” of the “ingested” copyrighted work’ (*see* [here](#)).

The vertebra of this Generative AI value chain is text and data mining (TDM) (*see* [here](#), [here](#) and [here](#)). To facilitate the process of TDM, data is the key. Data is the food that algorithms need to digest and regurgitate in order to radiate patterns and insights. While data *per se* is ‘not’ copyright-protected, it is the ‘creative form’ and expression of the author, namely the work that is protected

by copyright (*see* [Geiger, Frosio and Bulayenko \(2018\)](#)). In the Generative AI debate, thus, one of the most crucial ‘knots to untangle is the relationship between “TDM and access to content”’ (*see* [Ducato and Strowel \(2021\)](#)). The ongoing lawsuits against generative AI tools in the US and the UK allege that tools such as ChatGPT directly infringe copyright-protected works (*see* [here](#), [here](#) and [here](#)). They include allegations that generative AI models, such as that of OpenAI, are ‘trained [on] caches of pilfered copyrighted works’, which constitutes a ‘systematic theft on a mass scale’ (*see* [Authors Guild v. Open AI](#)). In the EU, lawful access to content is a pre-requisite to benefit from the available TDM exceptions (Cf Arts 3 and 4, and supporting Recitals 17 and 18, 2019 CDSM Directive). As long as the AI-generated output reflects the author’s intellectual creation, their personal touch, the current framework may duly account for the ‘romanticised human author’ that sits at the core of copyright. However, what happens when the Generative AI goes a step further, such that there is no longer any ‘direct resemblance to a specific pre-existing work?’ (*see* [Senftleben \(2023\)](#)). This may, for example, be the case with outputs that come from artificially generated synthetic data, such as is the case with deep fake videos (*see* [Tyagi \(2023\)](#)). Synthetic data may be defined as ‘artificial data that mimics real-world observations’ (*see* [DataScience \(2022\)](#)). The challenge that emerges with the rise of synthetic data is that it may become increasingly difficult to establish a correlation between the pre-existing work and the outputs generated by advanced TDM techniques. It is for this reason that timely and adequate remuneration of the human author becomes crucial. Introduction of statutory licenses, as suggested by scholars such as [Senftleben](#) and [Geiger and Iaia](#), to balance the interests and right to culture and science on the one hand and the freedom of artistic expression on the other, are a highly attractive policy recommendation to account for different rightholders and users in the generative AI debate.

As regards the legal framework for TDM, different jurisdictions have diverging approaches (*see* [Flynn et al \(2022\)](#)). The US’ fair use provision seems most permissive (at least until current litigation is resolved), followed by the Japanese provision on TDM under Article 30-4 that allows use of works for non-enjoyment purposes (*see* [here](#)). In the UK, which currently offers TDM for non-commercial purposes, the discussion on a permissive TDM exception for any purpose was quickly silenced, as the Publishers Association voiced concerns over a broadly-worded exception (*see* [here](#), [here](#) and [here](#)). In the EU, the TDM exception for research organisations and cultural heritage institutions does not permit an opt-out under Article 3 CDSM Directive (*see* [here](#) and [here](#)). However, TDM for commercial purposes under Article 4 may be opted-out (*see* [here](#)). This is a significant limitation in the current EU approach because, as is increasingly visible, research is not undertaken only by universities and research institutions; notable disruptive digital innovations in recent times have come mostly from the private players. The restrictive provisions of Articles 3 and 4 CDSM, thus, may emerge as a ‘significant competitive disadvantage for the EU economy’ (*see* [here](#)).

For copyright to achieve its fundamental objective, which is to enhance creativity, each new generation of authors and creators must enjoy similar freedom to their predecessors to use ‘pre-existing works as building blocks for new creations’ (*see* [Senftleben \(2012\)](#)). Here, exceptions and limitations (E&Ls) have an important role to play in balancing the interests of users as well as rightholders. If there were to be a broader E&L, then what should be the design of such a framework? Should the scope of the broader E&L be confined only to TDM?

As regards TDM, there has been a recurrent call for a broadly framed TDM exception in the EU, designed along the lines of the Japanese concept of the [non] enjoyment of a work under Article 30-4 of the Japanese Copyright Act (*see* [Ueno \(2021\)](#)). In my view, however, we should have an even more broadly worded (but well-defined) general exception in the EU. While the connection

between a targeted TDM exception and a general and more broadly worded E&L may be slender and thin, it is nonetheless highly tensile. While a closed-ended framework undoubtedly offers certainty, the digital reality calls for flexibility in the interpretation of E&Ls. This can be explained on the grounds that technology is disruptive, and digital developments can follow an incalculable number of paths. An open and flexible E&L framework offers the space and scope to accommodate these unpredictable paths followed by innovation and creativity. In its letter to the Commissioner for Internal Market, the European Copyright Society too calls for ‘a reassessment of the existing exceptions and limitations in particular for research including text and data mining’ (see [European Copyright Society \(2023\)](#)).

The narrower TDM exceptions under Articles 3 and 4 CDSM Directive are generally indicative of the limited efficacy of the closed-ended E&Ls. A broadly framed E&L may not only take account of uses such as TDM, but also suitably accommodate any further unanticipated demands of the digital economy. In practice, the restricted and closed-ended nature of E&Ls has on several occasions led to unexpected outcomes. Sampling of music is a case in point. In the landmark *Pelham* case, Sabrina Seltur’s use of a two-second sample from Kraftwerk’s classic electric music work, ‘Metall auf Metall’, led to a 20+ year-long dispute, and a preliminary reference to the CJEU (see [Senftleben\(2020\)](#) and [Jütte and Quintais \(2021\)](#)).

Considering that we have had a recent revision to the copyright *acquis Communautaire*, following a long, drawn out legislative process, and many Member States have only recently implemented the provisions of the CDSM Directive (see [here](#)); how practicable and doable is one such policy recommendation? As it merits a reflection not only on the design of a broadly-worded E&L, but also its coherence with the EU’s core common values, the Charter of Fundamental Rights, that is a thought for a follow-on post...

This blog post summarizes the main research and findings of the article “The Copyright, Text & Data Mining and the Innovation dimension of Generative AI” (forthcoming, 2023, pre-print, available [here](#)).

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

Kluwer IP Law

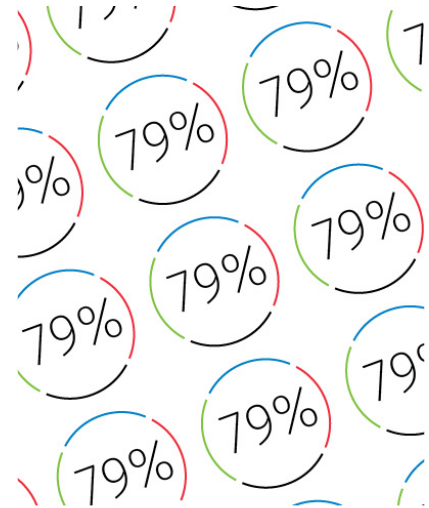
The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred 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 Wednesday, December 13th, 2023 at 8:58 am and is filed under [Artificial Intelligence \(AI\)](#), [European Union](#), [Text and Data Mining \(TDM\)](#), [USA](#)

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