Intellectual property, text and data mining and artificial intelligence training

The problem consisting of the use of works and other protected subject-matter for the purposes of mining and training of artificial intelligence is analysed in the light of the judgment of the Regional Court of Hamburg of 27 September 2024 (310 O 227/23)

ÁNGEL GARCÍA VIDAL
Professor of Corporate & Commercial Law, University of Santiago de Compostela
Academic counsel, Gómez-Acebo & Pombo

1. Introduction

The surge in artificial intelligence is sparking countless problems and uncertainties in several areas of the legal system, particularly in that of intellectual property law. One of the issues debated in this area concerns the use of works or other renditions protected by copyright or related rights when training artificial intelligence systems. Evidently, conflicting interests converge here: on the one hand, those of the rightholders, who consider that the use of their protected works or renditions for this purpose, if carried out without their consent, constitutes an act prejudicial to their rights; and, on the other hand, the interests of those responsible for the artificial intelligence systems that use the said works and renditions, who consider that such use is covered by the limitations provided for in intellectual property law. And it is precisely in this context that we find one of the current controversies, which concerns the possible application of the limitation in respect of text and data mining.

2. The European regulation on text and data mining as a copyright and related rights limitation

It is important to recall in this regard that Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market (the 'DSM Directive') introduced new limitations to copyright and related rights in EU law, referring to 'text and data mining', defined in Article 2 of the DSM Directive, 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". In particular, the DSM Directive incorporated a double regulation of the phenomenon of text and data mining from the perspective of intellectual property. Thus, in addition to an article devoted to "text and data mining for the purposes of scientific research" (Art. 3), another (Art. 4) addresses cases that do not fall under the previous one. Both of these articles are completed with the application of some of the provisions contained in Article 7, which are common to the different limitations regulated in the DSM Directive.

According to the first of these limitations, Member States shall provide for an exception to the right to reproduce any kind of work or other subject-matter, to the sui generis right in a database, as well as to the related right in respect of press publications. Under this exception, reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject-matter to which they have lawful access, are permitted. It is also provided that copies of works or other subject-matter made in compliance with the above provision shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results. In turn, rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. However, such measures shall not go beyond what is necessary to achieve that objective.

In addition to this limitation concerning mining for research purposes, Article 4 of the DSM Directive provides for a further limitation in cases where mining is carried out for other purposes, including commercial purposes. The copyright or related rights covered by this limitation are the same as in the case of the exception in Article 3, with the express addition of the right of reproduction and the right of translation, adaptation, arrangement and any other transformation of a computer programme. Again, Member States are obliged to provide for an exception or limitation to these rights in respect of lawfully accessible reproductions and extractions of works and other subject-matter for the purposes of text and data mining. It is also provided that reproductions and extractions made in this way may be retained, but it is clarified that this shall be "for as long as is necessary for the purposes of text and data mining". Moreover, the possibility for rightholders to object before the application of this limitation is provided for, as this exception or limitation shall apply on condition that the use of works and other subject-matter "has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online". Moreover, this exception can be altered by

contractual agreements, unlike in the case of mining for research purposes (Art. 7 of the DSM Directive).

The purpose of this European regulation, which has been incorporated with varying degrees of success by the Member States, is to allow the analysis of big data (including those protected by intellectual property rights) by means of complex algorithms, which generates knowledge that is difficult to obtain in other ways. The traditional study of data, based on the formulation of a cause-effect hypothesis which is then verified, is abandoned in favour of obtaining knowledge of the relationships between facts and data, often without knowing the cause-effect relationship, given the extraordinary number of existing variables. It is therefore a question of finding repetitive patterns in big data that can later be used to create predictive models and for decision making of all kinds, or for a better understanding of relationships that remained hidden in a traditional analysis.

The problem, however, lies in determining whether it is possible to apply the mining constraint when the aim is not so much to obtain concrete information from the analysis of big data, but to train a generative artificial intelligence system.

Well, in this context, a relevant judgment has just been handed down by a German court, to which the whole of Europe has paid attention, as it is one of the first pronouncements to analyse the issue. This is the judgement of the *Landgericht* (Regional Court) of Hamburg of 27 September 2024 (310 O 227/23[1]), the most relevant facts of which are set out below, together with a summary of the doctrine settled therein.

3. Text and data mining and the training of artificial intelligence in the judgement of the Hamburg *Landgericht* of 27 September 2024

3.1. The dispute in this judgment pits a photographer against a German non-profit organisation - LAION e.V. (Large-Scale Artificial Open Network) – that provides datasets, tools and models for machine learning research. In particular, one of LAION's projects is the so-called LAION 5B, a database of almost six billion image and text pairs describing the corresponding image, which has been used by third parties for the development of artificial intelligence systems.

Within the framework of this project, LAION downloaded a photograph from the internet available on a platform - Bigstockphoto - on which the photographer who took the photograph was marketing it, which led to the latter suing for infringement of his intellectual property rights. It is also relevant to note that that platform included a warning prohibiting users of the platform from using "automated programs, applets, bots or the like to access the [...]com website or any content thereon for any purpose, including, by way of example only, downloading content, indexing, scraping or caching any content on the website".

In this case, what is at issue is the po-ssible application of the text and data

[1] The judgement can be read in German at the following ***link***.

mining limitations set out in the German Intellectual Property Act (*Gesetz über Urheberrecht und verwandte Schutzrechte*), in particular § 60d (mining for research purposes) and § 44b (mining for other purposes), which transpose Articles 3 and 4 of the DSM Directive.

Similarly, the possible application of the limitation allowing temporary acts of reproduction of a work or other subject-matter which are transient or incidental, form an integral and essential part of a technological process, whose sole purpose is to enable either a transmission in a network between third parties by an intermediary or a lawful use of a work or other subject-matter to be made, and which have no independent economic significance (limitation laid down in Article 5 of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society and in §44a of the German Copyright Act), is also disputed before the German court.

3.2. The Hamburg Regional Court considers that the aforementioned limitation does not apply, because in the present case the images were downloaded from the internet for analysis by means of specific software, without meeting the conditions laid down by the Court of Justice (Case C-5/08 *Infopaq*, EU:C:2009:465), which requires, for an act to be classified as transient, that there be an automated process which automatically removes the temporary reproduction without human intervention.

3.3. As regards mining limitations, the Hamburg Regional Court accepts that the defendant organisation's action falls within the legal concept of text and data mining, stressing that downloading the image in order to compare it with a description of that image is an automated analytical technique intended to generate a correlation (which fits within the legal concept of 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" (Art. 2 of the DSM Directive). Thus, the Hamburg Regional Court denies the validity of the narrow interpretation that mining only takes place when information is obtained from the analysed data, but not when the analysed data can be used to create new creations (such as results generated with artificial intelligence), which may compete with the reproduced works or performances.

Similarly, the Hamburg Regional Court also does not accept the argument that the mining limitations provided for in the DSM Directive could not be applied to artificial intelligence because in 2019, when the DSM Directive was adopted, the technological leap experienced with generative artificial intelligence had not yet taken place, so that the limitations could not be applied to a factual situation not foreseen by the European legislature. In addition to stressing that this is not decisive and that the limitation applies to all conduct that fits within the text of the rule, the Hamburg Regional Court also insists that the recent regulation on artificial intelligence - Regulation (EU)

The Hamburg Regional Court applies the limitation of text and data mining for research purposes.

2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence - provides in Article 53(1)(c) that providers of general-purpose artificial intelligence models shall "put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790" ¸ which confirms that mining limitations apply to artificial intelligence.

3.4. Having acknowledged the existence of mining acts, the Hamburg Regional Court considers it doubtful whether the general limitation in Article 4 of the DSM Directive (and § 44b of the German Act), which also covers mining for commercial purposes, can be applied. This is because the German legislature, in transposing the DSM Directive and providing for the possibility for rightholders to object to the processing of their works or other subject-matter for mining purposes, provides that the reservation of use in the case of works available online is only effective if it is in a machine-readable form ("*Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt*"). And in the case in question it is doubtful whether the reservation of rights contained in the app in which the claimant's photograph appeared complied with this requirement, as the notice was in human language. And although the Hamburg Regional Court seems to consider that it might be possible to admit that it was machine-readable text (given that advances in artificial intelligence are enabling machines to understand natural language), in reality it leaves the question open, because the Hamburg Regional Court considers that all the conditions for the application of the limitation of mining for research purposes are met.

In this respect, the Hamburg Regional Court argues that the concept of research should not be limited only to those cases in which mining directly enables new knowledge to be obtained, but also covers the generation of correlations, patterns, etc., which can then be used for research purposes. And that is what would happen in the case in question, since the corpus generated by LAION was published free of charge and made available to researchers in the field of artificial intelligence, it being irrelevant that, in addition, commercial companies have access to the corpus and use it for the further development of their artificial intelligence systems.

3.5. Furthermore, the Hamburg Regional Court also analyses the application to the case of the so-called 'three-step rule' (contained in Article 5.4 of Directive 2001/29/EC and applicable to the limitations to text and data mining, by express provision of Article 7(2) of the DSM Directive), according to which such limitations "shall only be applied in certain special cases which do not conflict with a normal exploitation of

the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder".

In the Hamburg Regional Court's view, the defendant's use of the photograph is limited to analysing the image to check whether it corresponds to a pre-existing description of the image, which does not affect the use and exploitation of the photograph. The Hamburg Regional Court acknowledges that the data set thus created can be used to train artificial neural networks and that, as a result, artificial intelligence can create content that may compete with human works. But this possibility alone is not enough to understand that mining involves an impairment of the exploitation rights of the works, among other things, because, as highlighted in the judgment, this would mean that any mining act should be prohibited, which contrasts with the express acknowledgment's intention.

4. Assessment of the judgment and implications for the Spanish case

Without prejudice to the fact that the judgment addresses some issues of specific interest for German legislation, the judgment is relevant because much of its analysis is applicable to the rest of the EU Member States that have transposed the DSM Directive. In any case, it remains to be seen to what extent the European courts follow or not the position of the Hamburg Regional Court, since, as has been said, it is a debatable issue and there is no lack of reports and studies that reach contrary positions (as is the case, notably, of the report presented in the same month in which the judgment was handed down, September 2024, by Dornis and Stober on 'Copyright and the Training of Generative AI Models - Technological and Legal Foundations').

In any case, and as explained above, the judgment insists on the irrelevance of the prohibition of mining by the rightholder when it comes to mining for research purposes (which is what is established in the DSM Directive). However, the transposition of the DSM Directive into Spanish national law - carried out by Royal Decree-law 24/2021 of 2 November - is clearly unsatisfactory on this point. Indeed, Article 67 of this Royal Decree-law introduces the limitation in respect of text and data mining to allow its analysis in digital format in order to generate information that includes patterns, trends, correlations or similar elements, and merges into a single article the two exceptions on the matter contained in the DSM Directive, which distinguishes "text and data mining for the purposes of scientific research" (Article 3), from mining carried out for other purposes (Article 4). And this joint treatment of both exceptions means that the limitation "shall not apply where rightholders have expressly reserved the use of the works to machine-readable or other appropriate means". Therefore, the Royal Decree-law allows the opt-out or exclusion of mining also in relation to mining carried out for research purposes, which is contrary to the wording of the DSM Directive.