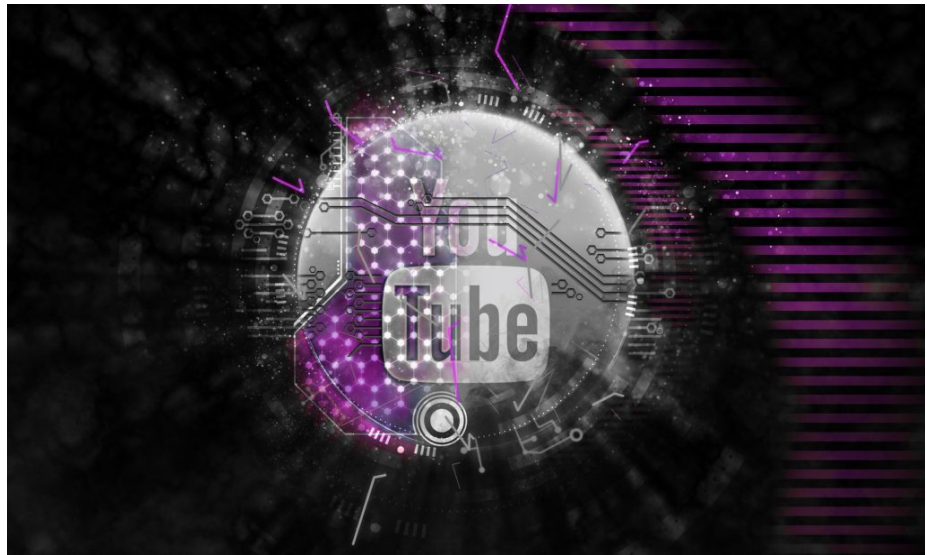


Kluwer Copyright Blog

Automated anti-piracy systems: A call for further evidence-based policies

Sabine Jacques (University of East Anglia) and Krzysztof Garstka (University of Cambridge) · Thursday, April 4th, 2019

The European Parliament has just approved the new text of the copyright directive, which will now go to the Council for a final vote on April 15th, 2019. This legislation not only modifies the copyright framework as set out in the Information Society Directive ([Directive 2001/29/EC](#))



but it will also modify the liability regime as enshrined in the E-Commerce Directive ([Directive 2000/31/EC](#)) for online content-sharing service providers. Against this backdrop, [a team of researchers](#) decided to investigate the impact of automated copyright enforcement mechanisms on cultural diversity.

The infamous article 13 of the Proposed new Copyright Directive ([COM\(2016\)0593 – C8-0383/2016 – 2016/0280\(COD\)](#)), now article 17 of the final text, renders the content-sharing platforms liable for the content uploaded by their users if the platform also gives public access to copyright materials. Whilst it is understandable why these platforms should obtain a licence for the copyright-protected content shared, the new obligation to adopt ‘best efforts to ensure the unavailability of [...] works and other subject matter [...]; and in any event, [act] expeditiously, upon [notification] from the right-holders, to disable access to, or to remove from, their websites the notified works or other subject matter, and made best efforts to prevent their future uploads’ has proven much more problematic to explain. It indirectly refers to so-called ‘upload filters’, which are implemented via complex algorithms identifying copyright infringements and deciding to block access to content, even prior to upload and without human oversight, in order to protect and promote cultural diversity. To counter this, the text attempts to offer some relief by adding that copyright exceptions (referring to the quotation and parody exceptions) must be respected and that

platforms should introduce an ‘effective and expeditious complaint and redress mechanism’ for users to challenge the decision of an algorithm.

The described technology carries a heavy bundle of promises, and there are significant concerns about its current state and ability to deliver upon the said promises. A major concern is respecting the careful balance struck by the legislator in the law between lawful and unlawful uses. Furthermore, there are doubts as to the efficiency of relying on private actors in order to enforce copyright, given how legal specialists already experience difficulties in deciding, for example, whether a copyright exception or limitation is applicable.

Without delving into the politically charged debate surrounding the adoption of the new copyright Directive, one element is striking. There is still a lot to be learned on the limits of technology for the future of creativity in the European Union.

Automation of copyright enforcement

Empirical research in the area of copyright algorithmic decision-making has mostly focused on notice and takedowns (see [Karaganis, Urban and Schofield](#)); and more recently, [Kretschmer and Erickson](#)). To contribute to this debate on the unavailability of content online, we decided to focus on automated anti-piracy systems such as [Content ID](#) or [Audible Magic](#). Where a notice and takedown requires online hosts to act expeditiously upon notification of a copyright infringement, the automated anti-piracy systems involve the interplay of a complex algorithm crosschecking all uploaded materials against an established database of (arguably) copyright-protected works. Crucially, these algorithms are capable of detecting wholesale as well as partial copying even if the work copied is transformed. Upon the finding of a match, the system automatically sends a notification to the supposed right-holder and he can select a course of action including the blocking of the content or its monetisation. The course of action can even be determined prior to the finding of any matches so that the system can simply apply the decision automatically to the finding of any matches as determined by the right-holder (which itself can be automated).

In a project focusing on YouTube’s [Content ID](#), we looked into the impact of this algorithm on the diversity of musical expressions over a limited period of 4 years in order to design an analytical framework for measuring cultural diversity online. We designed a theoretical framework to evaluate digital cultural diversity from a multi-dimensional perspective, taking into consideration variety (classifying videos per type), balance (the pattern of quantity across the types) and disparity (how dissimilar categories are within a type).

Focusing on parodies

Parodies constitute a [very particular type of artistic expression](#) that does not require prior authorisation from the original right-holders in the EU countries that have introduced the parody exception, provided that the [Deckmyn \(C-201/13\)](#) requirements are met. Their very nature involves the reworking of earlier works for entertainment, comment, homage or even criticism. Given that parodies (including pastiche, satire and caricature) rely on the use of other works (including copyright-protected works) to exist, these constitute an efficient way to gauge how content is reused online, what types of cultural expression flourish from these reuses and how algorithms respond to online parodies. In fact, parodies can be seen as one of the most difficult content characteristics to be discerned by a machine, the [prolonged struggle](#) over the definition of this term being evidence of this.

Measuring cultural diversity online

We decided to revisit and add to the data originally collected by our colleagues [Erickson, Kretschmer and Mendis](#) in relation to user-generated musical parodies on YouTube in 2012 based on UK hit charts. Although this dataset was created for different research purposes, it provides the advantage that we were then able to follow this cohort of works over a period of 4 years (up to December 2016). By relying on the content's location through the URL, we assessed the way content became unavailable on platforms to isolate the role of the algorithm as automated anti-piracy system, i.e. Content ID. The research team focused on evaluating the impact of content both supplied (that could be accessed on the platform) and consumed (what is actually being watched by viewers) on cultural diversity of user-generated musical work at three moments in time: in 2012 when the data was initially collected, a year later (2013) and at the end of the 4-year period (December 2016).

To do so, we decided to use a combination of the [Simpson's Index of Diversity](#) (which has been used in previous studies to measure diversity in relation to defined types) with what is traditionally referred to in competition analysis as the 'Numbers-Equivalent' to measure the number of types and their abundance.

Have algorithms used for preventing copyright infringements had an impact on cultural diversity?

One of our assumptions is that the unavailability of lawful content has the most harmful impact on cultural diversity. Therefore, if Content ID partners have elected other match policies such as monetisation or statistics tracking, such harm is decreased. Based on this, we found that algorithms significantly impact supplied and consumed diversity. For example, by analysing our sample at the end of the 4-year period, we see that only 59.2% of the content remained available and that Content ID accounts for 83.4% of content unavailable as opposed to 16.7% for takedown procedures. This should not be interpreted as a sign that Content ID is performing well. As we know from Erickson and [colleagues' study](#), only a very small percentage (5%) of the videos that formed part of the dataset constituted alleged copyright infringements.

Turning to consumed diversity, we determined that in 2013, only 27 videos were viewed by users out of a sample of 1471 videos. This number drops to 20 at the end of 2016. Delving even further, by looking at the number of individuals able to communicate effectively on the platforms, we saw that in our sample, only 13 YouTubers remained in 2016 as opposed to 940 in 2012. This is surprisingly low and could constitute an artefact of this particular sample, but it could be a more robust finding demonstrating the presence of strong network effects operating to determine popularity.

It is also particularly illuminating that audio-visual works were more likely to be blocked by the algorithms. This is not surprising given the functioning of the technology and how much easier it is to detect sound recordings and videos, than to detect other types of copyright-protected works such as literary (lyrics) or musical works (the composition).

A call for further evidence

At a moment in time when these automated anti-piracy systems are [about to become part of EU law](#), there is a crucial need to understand how these algorithms operate in order to preserve and respect the balance struck by the legislator in copyright law, as well as give due respect to cultural

diversity. The current drive for enforcement mechanisms relying on automation strategies in the hands of private actors creates concerns for cultural diversity. Whilst this research project aims to contribute to the debate by offering a theoretical framework which could be replicated with several different datasets, much still needs to be done to understand how diverse expressions are online. Also, in the absence of a robust legal framework promoting cultural diversity, freedom of expression and copyright are the main instruments to pursue this goal. As we know, in order to counter-balance the expansion of scope of copyright protection, legislators are introducing breathing space in the shape of exceptions and limitations to ensure their preservation. Therefore, not only is caution in this area necessary (perhaps limiting algorithms to operate on wholesale and not partial matches without human review) but strong and strict independent dispute resolution mechanisms are essential to respect copyright law and cultural diversity.

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

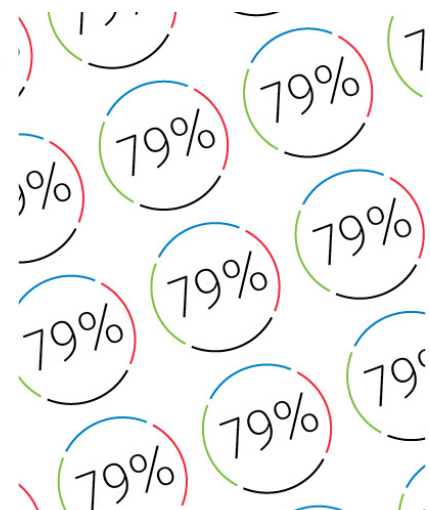
Kluwer IP Law

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 Thursday, April 4th, 2019 at 10:47 am and is filed under [Enforcement](#), [European Union](#), [Infringement](#), [Legislative process](#), [Liability](#), [Limitations](#)

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