

DOI 10.5817/MUJLT2023-1-5

# PLAYING THE SYSTEM: CONTENT RECOGNITION TECHNOLOGIES AND CREATIVE PROCESS OF SAMPLING MUSICIANS \*

*by*

IVAN DAVID † RUDOLF LEŠKA ‡

*In the first part of the study, we summarize the existing types of copyright bots. In the second part, we present the current state of legal research on the implementation of copyright bots and our own analysis which focuses on the Czech Supreme Court's decision concerning copyright bots. The core of this paper concerns the impact of copyright bots on the work of sampling musicians and how the creativity of musicians is shaped by their struggle to avoid detection of sampled music by bots in the online environment.*

## KEY WORDS

*Copyright, Bots, Technology, Sampling, Content Recognition*

## 1. INTRODUCTION

It is no secret that the improvement of copying technologies has always been a catalyst for conflict in creative industries as far as copyright law is concerned. It is also clear that legal concerns can shape an artist's creativity and work, mainly in copyright-intensive industries like the film and music industry, which depend on the online distribution of their products. It occurred to us that the introduction of automated content recognition technologies has

---

\* This work was supported by the Improving schematics of Doctoral student grant competition and their pilot implementation at the Palacký University Olomouc under grant no. CZ.02.2.69/0.0/0.0/19\_073/0016713. Specifically, the research was conducted within the "Law and Remix Culture: Aesthetics and Ethics of Musical Remix" project (project no. DSGC-2021-0015). Our thanks go to anonymous reviewers who substantially helped us to precise our paper and its findings. The data that support the findings of this study are available on request from the corresponding author, Ivan David. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

† ivan.david@kmvs.cz, PhD Candidate, Department of Media Studies, Palacký University Olomouc, Czechia, and Counsel, KMVS Law Office, Czechia

‡ leska@staidl-leska.com, Senior Assistant Professor, University of Finance and Administration, Czechia, and Counsel, Štáidl Leška advokáti, Czechia

had a palpable impact on the creativity of artists, mainly in the field of music production. As we will demonstrate, our research and this study have proven this hypothesis to be correct. Based on our 38 semi-structured interviews, 36 with professional musicians who routinely use sampled materials in their own work as well as one with a representative from an international record label and one from a global company that develops content recognition technologies, we present the attitudes of artists towards automated content recognition technologies (copyright bots)<sup>1</sup> and how they deal with them in their creative practice. Like the character Josef Švejk in the famous novel by Czech writer Hašek,<sup>2</sup> these artists tend to employ various strategies to play the system rather than submit themselves to it.

Our approach is interdisciplinary, as we believe it could be a meaningful contribution to the often abstractly technical legal research. We therefore conducted a qualitative sociological analysis and applied it to our findings based on desktop legal research. In the first part of our study, we summarize the existing types of copyright bots; in the second part, we present the current state of legal research on the implementation of copyright bots and our own analysis which focuses on the Czech Supreme Court's decision concerning copyright bots; the core of this paper then concerns the impact of copyright bots on the work of sampling musicians.

We do not address the issue of liability (including the well-researched sampling case law, such as the *Pelham*<sup>3</sup> case), management and mechanisms of automated takedown notices, nor the precision of copyright bots, as this has already been done by others – examples of anecdotal<sup>4</sup> as well as serious<sup>5</sup>

---

<sup>1</sup> Bot, "a computer program that works automatically, especially one that searches for and finds information on the internet" (*Cambridge Advanced Learner's Dictionary & Thesaurus*. Available from: <https://dictionary.cambridge.org/dictionary/english/bot> [Accessed 20 November 2022]).

<sup>2</sup> Hašek, J. (2005) *The Good Soldier Švejk and His Fortunes in the World War*. Translated from Czech by Cecil Parrott. London: Penguin Books.

<sup>3</sup> Judgment of 29 July 2019, *Pelham GmbH and others*, C-476/17, EU:C:2019:624.

<sup>4</sup> A 10-hour long video of white noise (available from <https://www.youtube.com/watch?v=VcQZAzDVT1A>) received five wrongful copyright claims. Cf. Tune, C. and Iverson, S. (2020) *The Rise of the Copyright Bots*. *Internet and Technology Law*. Available from <https://www.internetandtechnologylaw.com/copyright-bots/> [Accessed 16 November 2022].

<sup>5</sup> The Washington Post's story describes the difficulties classical musicians faced during the COVID-19 pandemic when automated bots misidentified their recorded content with other recordings. Cf. Brodeur, M. (2020) *Copyright bots and classical musicians are fighting online*. The Washington Post. Available from: [https://www.washingtonpost.com/entertainment/music/copyright-bots-and-classical-musicians-are-fighting-online-the-bots-are-winning/2020/05/20/all349c-98ae-11ea-89fd-28fb313d1886\\_story.html](https://www.washingtonpost.com/entertainment/music/copyright-bots-and-classical-musicians-are-fighting-online-the-bots-are-winning/2020/05/20/all349c-98ae-11ea-89fd-28fb313d1886_story.html) [Accessed 16 November 2022]. For an overview of similar cases cf. Berkowitz, A. (2022) *Classical Musicians v. Copyright Bots*. *Information Technology and Libraries*, 41(2). Available

flaws in the technology can be found in numerous sources; instead, we focus on how legal regulations, which are facing automated enforcement, are communicated, with the hope that this information may prove useful for policy makers, online content sharing platforms and copyright owners alike. Sociological findings serve for discussion of optimal or right legal regulation which should be based on the informed understanding of social expectations by the addressees of the regulation. Otherwise, the law becomes in tension with actual behavior of individuals which undermines the legal awareness and the rule of law as such.

Last but not least, data from our research have furthermore relevancy also for research in the field of musicology and aesthetics.

For the purpose of this study, we use the English term copyright as an umbrella term for author's rights and related rights (including phonograms) and the term copyright owner for any holder of copyright, whether it is the author and performing artist or entities so authorized, such as employers, assignees and licensees.

## 2. TYPES OF AUTOMATED TECHNOLOGIES USED TO DETECT ILLEGAL CONTENT

In today's online landscape, there is a large number of technologies used for the automated detection of illegal content. Such technologies are used for many purposes – from combatting child pornography to preventing the spread of terrorist content. Nonetheless, the primary focus of this chapter is on the technologies used by automated programs for recognizing copyright-infringing materials (“copyright bots”), which typically target films and music, photographs or pieces of literature<sup>6</sup> distributed without the appropriate licensing.

Many such technologies are of a hybrid nature. However, for the sake of simplicity, we can schematically divide the technologies into the following categories.<sup>7</sup>

---

from <https://ejournals.bc.edu/index.php/ital/article/view/14027> [Accessed 16 November 2022].

<sup>6</sup> Most often, this concerns e-books. See for instance: Rosenblatt, B. (2017) A Bounty Hunting Service for E-Book Piracy. Copyright and Technology. Available from <https://copyrightandtechnology.com/2017/01/30/a-bounty-hunting-service-for-e-book-piracy/> [Accessed 16 November 2022]

<sup>7</sup> In this chapter, we use the typology and terminology developed in the two recent EUIPO expert studies: European Union Intellectual Property Office (2020) *Automated Content Recognition: Discussion Paper – Phase 1, Existing Technologies and Their Impact on IP*. ISBN 978-92-9156-280-0. Available from: <https://data.europa.eu/doi/10.2814/52085> [Accessed 16 November 2022], and European Union Intellectual Property Office (2022) *Automated Content Recognition: Discussion Paper – Phase 2, IP Enforcement and Management*

## 2.1. HASHING

The method of hashing is based on the principle that simple identifiers consisting of a relatively short number of unique characters of a fixed-length value (“hashes”) are automatically generated for the identification of digital files. The new hash files – usually of a significantly smaller size than the original digital files – are then stored in reference databases. When determining whether a certain unknown digital file is identical to a file for which a hash has been generated, a hash is also generated for this unknown digital file using the same technology; the new hash and the original hash are then compared and a match is sought.

The clear advantage of this technology is that it is simple, easy to implement, fast and undemanding in terms of storage space and hardware performance. Furthermore, this technology can be used for a wide range of digital files regardless of their content (i.e., whether they contain pictures, music, texts or another type of content).

The obvious disadvantage is that the technology is primarily designed to match identical files. If the original file – in our case a file with copyrighted material – is altered in any way (for instance, by compression or change in format), a different hash will be generated (two distinct data will always deliver different hash) and no match will be found. What is more, hashes are only used to recognize specific digital files; they thus have no descriptive capacity and as such do not recognize the actual content of these files.

As far as copyright-infringing materials are concerned, hashing technology (as such) is most often successfully used by providers of platforms designed for sharing public and openly accessible content.<sup>8</sup> In such cases, it helps to ensure that if a digital file is flagged by any user or content owner as copyright infringing, other identical files on the platform are also disabled and the (identical) content is not reuploaded.

## 2.2. WATERMARKING

In the case of watermarking, the data used to identify digital files are not stored in separate reference databases, like in hashing, but such identifying data are incorporated directly into the digital files.

Therefore, it is not necessary to compare the hash of the original digital file with the automatically generated hash of the unknown file to find a potential match; instead, the information needed to identify the file can be found directly in the file itself.

---

*Use Cases*. ISBN 978-92-9156-326-5. Available from: <https://data.europa.eu/doi/10.2814/952694> [Accessed 16 November 2022]

<sup>8</sup> YouTube uses hashing technology in its Content ID tool to prevent identical files from being uploaded repeatedly.

Watermarks can either be generic, i.e., applied to large number of digital files and usually identifying their source,<sup>9</sup> or specific, thus identifying each unique digital file. We can also distinguish between watermarks that are clearly visible and thus generally readable by anyone using the file<sup>10</sup> and those that are invisible (typically found in the metadata of the marked file) and which generally require special software to be read. In terms of the “source” of the digital file that a watermark identifies, it can be either the original source in the sense of information on the rights holder, or the “illegal” source, i.e., the source from which the pirated (i.e., copyright-infringing) materials are spread.

While the advantage of watermarks is that there is no need for the creation of separate hashes and reference databases containing these hashes, the main disadvantage of this method is that only digital files that have been previously watermarked can be successfully recognized. Marking each digital file with a watermark can also be very demanding and resource-intensive, especially if watermarks are to be generated for every unique file. Finally, watermarking technologies are generally less standardized than hashing technologies, meaning that in the case of more sophisticated, i.e., invisible (embedded) watermarks, special software equipment is needed for these watermarks to be successfully read and recognized. On the other hand, the current state of art in watermarking technology – unlike hashing technology – is generally reliable even for marked files that are somehow altered.

### 2.3. FINGERPRINTING

Fingerprinting can simply be described as a more sophisticated form of hashing. Just like in hashing, identifying files (called “fingerprints”) are automatically generated resulting in digital files that are smaller in storage size than the originals. These newly created files are then stored in separate reference databases. When a match is sought, the pre-existing fingerprints are compared with the newly generated fingerprints for the unknown digital files.

The major difference between hashing and fingerprinting is that

---

<sup>9</sup> For instance, generic watermarks are used to mark audiovisual files for films intended for viewing by members of the Academy of Motion Pictures Arts and Sciences when voting for the “Oscar” winner in different categories. The same applies to audiovisual files used by members of Czech Film and Television Academy for choosing the winners of the “Czech Lion Awards”. In these and similar cases, the watermark informs the viewer that the digital file in question is intended solely for this purpose and cannot be used elsewhere – thus also identifying the source of the file.

<sup>10</sup> Most likely the best-known example of visible watermarks are the graphic logos of TV networks used to mark all programs broadcasted by them.

while hashing is only used for identifying identical digital files as such, fingerprinting can even identify identical or similar file content. Practically speaking, this means that fingerprinting technology is more resistant to the alternation of digital files, as it can still identify the major traits of the content contained in the files in question. In practice, fingerprinting software analyzes the content (music, films, texts etc.) and extracts statistical samples of the content (for instance, several samples for each second of a particular song); the more samples that are extracted, the more precise the fingerprinting recognition will be. On the other hand, the more precise the fingerprinting method is, the more demanding it is in terms of hardware (storage space, computational capacity) – mainly when compared to less demanding methods such as hashing and watermarking.

At the moment, fingerprinting appears to be the most popular and sophisticated method for the automated detection of illegal content as far as copyright infringement is concerned. Fingerprinting is, for instance, the main component of such software devices as YouTube Content ID<sup>11</sup> and Facebook Rights Manager.<sup>12</sup> The same principle is also the basis of the popular music recognition application Shazam.<sup>13</sup> When fingerprinting is used to find copyright-infringing materials (music, images and videos) on content sharing platforms like the above-mentioned YouTube or Facebook, content owners are typically given several options for dealing with copyright-infringing content, most often in the form of blocking,<sup>14</sup> tracking<sup>15</sup> or monetizing<sup>16</sup> the flagged content. In some cases, the automated technology also contains mechanisms for resolving conflicts concerning IP ownership (amongst IP owners or between an IP owner and a user) over specific pieces of content.<sup>17</sup>

In any case, it largely depends on the operator of the search technology as to where the accuracy threshold should lie. As is stands, the search tool can be set up in such a way that it can detect even very short extracts

---

<sup>11</sup> See <https://support.google.com/youtube/answer/2797370?hl=en> [Accessed 15 November 2022]. For Content ID analysis cf. Shinn, L. (2015) YouTube's Content ID as a Case Study of Private Copyright Enforcement Systems, *AIPLA Q. J.*, 43 (2/3). Perel, M and Elkin-Koren, N. (2016) Accountability in Algorithmic Copyright Enforcement. *STAN . TECH . L. REV.*, 19, 473, p. 510 et seq.

<sup>12</sup> See <https://rightsmanager.fb.com/> [Accessed 15 November 2022]

<sup>13</sup> See <https://www.shazam.com/> [Accessed 15 November 2022]

<sup>14</sup> Blocking means that the content is blocked from further access.

<sup>15</sup> Tracking means that content owner only receives statistics on how frequently the content is accessed.

<sup>16</sup> Monetizing usually means that advertisements are added to the content and the content owner receives (full or partial) revenue from the advertisers.

<sup>17</sup> This is the case of YouTube, for example. See "Resolve Asset Ownership Conflicts" at <https://support.google.com/youtube/answer/3013321?hl=en> [Accessed 15 November 2022]

of copyright-protected material (e.g., just a few seconds of sampled music) used in other works. Nevertheless, the operator of this technology can (and very often does) deliberately configure the search engine so that the targeted copyrighted extracts are not too short. Firstly, this reduces the demands placed on the hardware equipment needed to run the search engine; secondly, it undoubtedly helps in preventing too many false positives or too severe content restrictions, as they can be controversial when considering the existence of copyright exceptions.

#### 2.4. AI-BASED (ENHANCED) CONTENT RECOGNITION

The latest trend in the field of automatic content recognition is technological solutions based on artificial intelligence. Unlike all the aforementioned automatic recognition technologies, in this case the software solution actually “understands” what it sees or hears. In other words, AI-based technology can perform a much deeper analysis of potentially copyright-breaching content than hashing, watermarking or fingerprinting technologies ever could.

The development of these technologies is likely most pronounced in the automatic analysis of static and moving images – where artificial intelligence can (based on existing databases), for example, recognize the faces of specific people in an image and perform a predetermined action accordingly. Thanks to this technology, digital files can be sorted and categorized and keywords and other descriptors can be automatically generated.

Despite the obvious advantages posed by this advanced technology, as it can significantly improve accuracy in the detection of illegal content, its main disadvantage is the substantial demand on hardware performance and storage capacity. For these reasons, such advanced solutions are often combined or supplemented in practice with one or more of the above-mentioned “traditional” methods of automatic content recognition.

In summary, each of the above-specified technologies is suitable for a different purpose, mainly because they possess a different level of accuracy when searching for illegal content (even when the searched content has been in some way altered); furthermore, they require different demands in terms of the required storage space and performance of the hardware on which they run. These technologies are often combined to multiply their advantages and achieve higher precision.

### 3. COPYRIGHT AND AUTOMATED CONTENT

#### RECOGNITION TECHNOLOGY: THE DILIA CASE

Copyright, as any other right, deteriorates if unenforced. Rampant infringement in the online environment facilitated by the operation of non-labile platforms, which reap exorbitant profits on content they have neither produced nor paid for, is impossible to fight by human oversight. The legal rules that limit the liability of intermediaries – originally designed for traditional web hosting providers and by no means for social networks and similar services<sup>18</sup> – have allowed for such a broad interpretation that they provide a safe harbor (if not safe heaven) for platform operators, revolutionarily shifting the burden of copyright policing from those who monetize the content to those who create it. This has prompted these platforms to store hundreds of millions of protected authorial works.<sup>19</sup> Human policing of such vast amounts of content is clearly impossible and this situation has naturally given rise to automated solutions. While originally it was only copyright owners who were active in policing and enforcing copyright, newer case law in the US, EU and elsewhere has balanced the situation by forcing platform operators to adopt a more proactive approach to monitoring content on their websites.<sup>20</sup> YouTube – which is facing possibly ruinous effects of copyright infringement litigation – has become one of the most proactive services in assisting copyright owners with content policing through their proprietary Content ID system.

Although copyright bots are not appealing to the tech industry and some parts of academia,<sup>21</sup> they work substantially better than anything else

---

<sup>18</sup> For the origins of the rules that limit the liability of intermediaries and a critical assessment of how these rules have been misinterpreted to cover much broader types of services cf. the study by Leška, R. and Půr, M. (2022) Technologické společnosti a jejich odpovědnost za obsah na internetových platformách. *New Direction*. Available from: [https://newdirection.online/publication/technologicke\\_spolecnosti\\_a\\_jejich\\_odpovdnost\\_za\\_obsah\\_na\\_internetovch\\_pla](https://newdirection.online/publication/technologicke_spolecnosti_a_jejich_odpovdnost_za_obsah_na_internetovch_pla) [Accessed 20 November 2022]. For critical insight into US platform liability (or lack thereof), cf. Gabison, G. and Buiten, M. (2020) Platform Liability in Copyright Enforcement. *Columbia Science and Technology Law Review*, 21, pp. 237–280. For advocacy against amendments to the current US law, cf. Samuelson, P. (2021) Pushing Back on Stricter Copyright ISP Liability Rules. *Michigan Technology Law Review*, 27, pp. 299–344.

<sup>19</sup> Reportedly, YouTube alone stores approx. 800 mil. videos as of May 2022 (containing many authorial works and other protected content, often without a license). Cf. Hayes, A. (2022) YouTube Stats: Everything You Need to Know In 2022! Wyzowl. Available from: <https://www.wyzowl.com/youtube-stats/> [Accessed 16 November 2022].

<sup>20</sup> The evolution of this case law and industry agreements exceeds the scope of our study. Developments in EU law are summarized in the YouTube/Cyando judgment. *Frank Peterson v Google LLC and Others and Elsevier Inc. v Cyando AG*, Court of Justice of the European Union, joined cases C-682/18 and C-683/18.

<sup>21</sup> Pamela Samuelson, for instance, believes that the use of copyright bots by platform operators is inconsistent with copyright exceptions (user freedoms, as she calls it), incompatible with



currently available to us – unless we opt for some sort of blanket licensing, perhaps managed by collection societies. Of course, the content-matching technology used by these bots might often not recognize the context (although research in this direction is fast) resulting in mistakes (not unlike human error),<sup>22</sup> particularly when it comes to recognizing exceptions and limitations to copyright; nevertheless the system should be designed in such a way that allows the person asserting the exception to prove their claim (after the content has been disabled) and let the content be human-verified by the copyright owner with escalation to the platform operator. This is the only way this can work. Admittedly, the system makes mistakes when it comes to classical music, though technology is improving rapidly in this field, as confirmed by our interviewee (see below). To require human oversight for every single copyright claim for the tens of millions filed by copyright owners is unfeasible,<sup>23</sup> though this seems to be precisely the agreement that Lawrence Lessig came to with Liberation Music in a widely publicized lawsuit settlement.<sup>24</sup>

We do not share in condemnations of copyright bots as evil. In a situation where copyright infringement is rampant and personal policing impossible, it is hard to imagine that copyright owners alone would report infringing content while platforms would be free to erroneously keep millions of files with infringing content online – the approach should be fair and should not place stricter burdens on copyright owners than platform operators.

We are unaware of any case law which would specifically address the tolerated limit of erroneous takedown notices when assessing the rights of copyright owners. One exception is a landmark dispute in Czechia concerning the local collective management organization DILIA<sup>25</sup> and the

---

the no general monitoring obligation, in violation of rights to personal data and perhaps even the (EU) Charter of Fundamental Rights. Cf. Samuelson, P. (2021) Pushing Back on Stricter Copyright ISP Liability Rules. *Michigan Technology Law Review*, 27, pp. 299–344. We disagree with such a categorical stance.

<sup>22</sup> Even courts struggle to interpret exceptions and limitations and are often in disagreement with one another on the very same issues. It is therefore naive to believe that an assessment performed by a distant and under-educated employee would be a substantial improvement.

<sup>23</sup> DiCarlo, I. (2022) Copyright Bots Need a Tune-up. JTIP Blog. Available from: <https://jtip.law.northwestern.edu/2022/03/05/copyright-bots-need-a-tune-up/> [Accessed 10 November 2022]. Various studies examine the practice of prohibiting or at least making it very complicated for copyright owners to report infringing conduct, cf. Depoorter, B. and Walker, R. (2014) Copyright False Positives. *Notre Dame Law Review*, 89(1), p. 319. We believe this to be ill-advised for legislators as it would immensely benefit the technology companies without taking into account content creators.

<sup>24</sup> For details and an analysis of the case cf. Tan, C. (2014) Lawrence Lessig v Liberation Music Pty Ltd: YouTube's Hand (or Bots) in the Over-Zealous Enforcement of Copyright. *European Intellectual Property Review*, 36(6), pp. 347–351.

<sup>25</sup> Full legal name: DILIA, divadelní, literární, audiovizuální agentura, z. s. (Theatre, literary, audiovisual agency)

operator of the most popular Czech cyber-locker service “Ulož.to”.<sup>26</sup> DILIA filed a complaint in which it asked the court to order the cyber-locker operator to ensure the permanent stay-down of certain Czech motion pictures (including Jiří Menzel’s *Closely Watched Trains*, which received the Academy Award for Best Foreign Language Film in 1968).<sup>27</sup> What is important, DILIA also asked the court to order the defendant to remove their limit on the total number of requests made by DILIA from its IP address to the defendant’s servers. The case went to trial at the Municipal Court in Prague<sup>28</sup> and a judgment was issued by the High Court in Prague<sup>29</sup> and finally confirmed by the Supreme Court.<sup>30</sup> What is interesting about DILIA’s second claim is that the court investigated the correctness of DILIA’s takedown notices. DILIA originally made these notices manually and Ulož.to even concluded a special agreement with DILIA, under which DILIA was entitled to directly remove infringing content uploaded by users to the website. Once DILIA switched to automated data processing, the bots started to find – and remove – large amounts of (illegal) user data and Ulož.to withdrew from the agreement with immediate effect due to contract breach. Under the contract, DILIA agreed to only remove content to which they held the rights.

The number of false positives was under 1 %.<sup>31</sup> According to private information from DILIA and the letter of withdrawal from Ulož.to, which was read in a public hearing at court, in the period from June to October 2012, DILIA removed 2264 files of which Ulož.to identified 11 false positives.<sup>32</sup> This drives the number of false positives down further below 0.5 %. Formally,

<sup>26</sup> The operator’s legal name at the time the Supreme Court’s judgment was made was Ulož.to cloud a. s., later it was petacloud a. s. At the time of publication, the operator of the website has changed. Today, it is another entity named Cloud Platforms a. s. while the domain name itself was sold to yet different entity Meta Web Services a. s.

<sup>27</sup> Procedurally, the claim was combined. The plaintiff requested a a) permanent injunction (in case it was proven that the defendant was actively involved in illicit conduct), b) permanent stay-down based on certain word filters (in case the request for a permanent injunction was denied). This part of the proceeding is beyond the scope of our study but it can be noted that the plaintiff was only successful with claim b) since the court did not find (or was unwilling to find) enough evidence of structural infringement on the website.

<sup>28</sup> Unpublished judgment of the Municipal Court in Prague of 22 February 2019, 34 C 5/2017. Available from: [https://www.dreport.cz/wp-content/uploads/TechLaw\\_judik\%C3\%Alt\\_Ulozto\\_Dillia\\_34-C-5-2017.pdf](https://www.dreport.cz/wp-content/uploads/TechLaw_judik\%C3\%Alt_Ulozto_Dillia_34-C-5-2017.pdf) [Accessed 1 November 2022].

<sup>29</sup> Unpublished judgment of the High Court in Prague of 20 January 2021, 3 Co 58/2019.

<sup>30</sup> Judgment of the Supreme Court of 8 June 2022, 23 Cdo 1840/2021. Available from the website of the court and numerous online sources, including <http://kraken.slv.cz/23Cdo1840/2021> [Accessed 1 November 2022].

<sup>31</sup> Srstka, J. (2014) Žaloba DILIA proti úložišti www.ulozto.cz. *Věstník DILIA*, , p. 41. Available from <https://www.dilia.cz/ke-stazeni?cat=ostatni> [Accessed 1 November 2022]. The court mentions “a minimal percentage” (paragraph 31 of the judgment of the municipal court).

<sup>32</sup> It should be noted that some of those files were manifestly illegal too, they just did not contain content by DILIA authors; some were suspicious Wikipedia articles which appeared

Ulož.to's withdrawal from the agreement under which DILIA was entitled to directly remove files on the Ulož.to service was legally valid and this fact was undisputed. On the other hand, even if DILIA had to later rely on standardized procedures under Ulož.to's policy and the framework of the E-Commerce Directive<sup>33</sup> and its Czech transposition,<sup>34</sup> the argument explaining DILIA's false positives did not convince the court when deciding about the "free pass" DILIA was given to be able to crawl, search and report infringing content. Unfortunately, the trial court was not specific in the legal justification stipulating that all of DILIA's requests are not to be blocked, including a defense against "denial-of-service" (DOS) attacks (flooding the target with information/traffic); nevertheless, judging by the context, it appears that the trial court found this obligation in the general tort law and in the general provision against causing unjustifiable harm to others (Art 2900 of the Czech Civil Code).<sup>35</sup> The appellate court also did not make reference to any specific statute, though it suggested that such an obligation is a sort of special claim made by the copyright owner in the spirit of recital 17 of the Enforcement Directive.<sup>36</sup> Finally, the Supreme Court upheld this position and further specified it in reference to Art 98c(1) of the Czech Copyright Act,<sup>37</sup> which provides collective management organizations with the right to ask users for information concerning uploaded content while per analogiam extending this claim to the platform operator (cyber-locker service), even if the platform operator is not a user in itself (the case was decided under pre-DSM Directive law).

The lesson learned from this interesting Czech case regarding the operation of copyright bots is that the use of copyright bots is fully legal, the intermediary is legally obliged to allow copyright bots unobstructed access to search through its service,<sup>38</sup> and that even if errors occur, they are not grounds enough to forbid copyright owners from performing searches in the service (provided that it does not constitute abusive conduct). That being said, copyright owners remain liable for actual damages caused by any such

---

as protected content, although it is unclear why someone would upload a Wikipedia article to a cyber-locker.

<sup>33</sup> Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.

<sup>34</sup> Act No. 480/2004 Coll., on certain information society services.

<sup>35</sup> Act No. 89/2012 Coll., the Civil Code.

<sup>36</sup> Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights.

<sup>37</sup> Act No. 121/2000 Coll., on copyright, on rights related to copyright and on the amendment of certain laws (Copyright Act).

<sup>38</sup> It must be noted that the operator of the service is also protected against possible abuses of such rights by the general provisions of the civil code prohibiting abusive conduct.

errors under applicable tort law. We believe this basic principle to remain valid even with the transposition of the new DSM Directive<sup>39</sup> in the laws of EU member states.

#### 4. IMPACT OF COPYRIGHT BOTS ON THE WORK OF SAMPLING MUSICIANS

In our research completed in 2022,<sup>40</sup> we conducted 36 semi-structured interviews with professional musicians who use music samples in their work. We spoke with 21 electronic dance musicians, 10 hip-hop and rap musicians and 5 alternative pop musicians. We also conducted an interview with a representative of a major Czech music label that uses YouTube Content ID technology, and one interview with a representative of a major international software company that develops globally successful automated software recognition technology.

In our interviews with the musicians, we focused on several topics concerning the impact of copyright on their creative practice, both from an aesthetic and ethical point of view. Importantly, all of the interviewed musicians – at least occasionally-release their records themselves on various platforms without interference from a record label, which would otherwise take care of the standard release protocol and rights clearance (if samples are concerned).

One of the main conclusions here was that even though all of the interviewed musicians were at least aware of the basic principles of copyright (though many were unsure of the specifics, for instance the conditions for copyright exceptions), the total majority admitted that they generally make no attempt to clear the rights to samples used in their projects; if they do seek permission, one of the main reported motivations was precisely the existence of copyright bots.

It appears from the interviews that copyright bots were often the only confrontation these artists had with actual copyright enforcement, and as such they represented the whole copyright system “in a nutshell”.

According to our findings, the existence of copyright bots has a profound impact on the work of sampling artists from a creative point of view, as these

---

<sup>39</sup> 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 and amending Directives 96/9/EC and 2001/29/EC.

<sup>40</sup> The research was realized within the OP VVV project "Improving the schemes of the Doctoral student grant competition and their pilot implementation", reg. No. CZ.02.2.69/0.0/0.0/19\_073/0016713. Specifically, the research was conducted within the “Law and Remix Culture: Aesthetics and Ethics of Musical Remix” research project (project no. DSGC-2021-0015). The results of these interviews are reported in this submission for the first time.

bots force them to either (less often) clear all third party rights, or (more often) radically shorten, transform or modify music samples so that they cannot be recognized by the ubiquitous copyright bots.<sup>41</sup> Some of these artists even admitted to using apps for automated content recognition (like Shazam) to ensure that copyright bots do not recognize the shortened and altered samples after they are uploaded to YouTube or a similar content sharing platform. Other artists, for instance, “test” the copyright bots on YouTube by uploading a music video containing uncleared samples, though without making the video public – they then simply wait to see if they will be notified of copyright infringement by the YouTube administrator.<sup>42</sup>

Only a small minority of artists who do not use popular content sharing platforms or online distribution tools, but instead only release their records on vinyl or MC, seemed to be “immune” to the influence of copyright bots on their work; such artists thus tend not to limit themselves in their use of uncleared samples, exhibiting no need to significantly shorten or modify these samples just for the sake of going unrecognized by copyright bots.<sup>43</sup>

Similarly, artists who do not want their music to be overly screened by copyright bots but still want to publish it online, choose online platforms which (to their knowledge) have lenient policies regarding the automatic detection of illegal content.<sup>44</sup>

Another strategy that some sampling artists have developed in response to copyright bots is to only use samples of music they consider to be “old”, “underground”, “forgotten”, “unknown” etc. – and which they (usually correctly) anticipate will not be screened by automated technologies.<sup>45</sup>

<sup>41</sup> One of the artists – a member of a successful alternative pop band – told us: *“We try to make the result unrecognizable. We release it on every possible media platform there is. Either we release some things ourselves, some through independent labels. [...] There was never a problem with the algorithm finding something that was taken from somewhere else [...]”* Another artist – working mostly as a DJ – told us: *“[...] some years ago I was looking for other music myself and I never had a problem with that because there was always some creative input [from my side] into it and it’s not like a copier. It’s [...] at least usually edited in such a way that the algorithm can’t recognize it. The listener recognizes it there, the excerpt, and says to himself: ‘Yeah, I know this’, only it’s slower and transposed a few degrees elsewhere, but the algorithm doesn’t bother with it, because it’s only interested in copy-paste.”*

<sup>42</sup> One of the hip-hop artists told us: *“It is possible, I advised the same to [anonymized name of another artist] when he asked me how I do it. And I know [anonymized name of another artist] will also upload it to YouTube and find out that way. So it’s probably a good idea to lock it so it’s not public and upload it to YouTube. And they’ll let you know if there’s a problem.”*

<sup>43</sup> Another hip-hop artist told us: *“I can put absolutely anything I want on vinyl and I’m good, because we’re in the Czech Republic and we’re completely off the radar. No digital robots check it, and [...] no one will notice, you can put whatever you want there.”*

<sup>44</sup> One of the interviewed creators of electro dance music told us: *“I put it, for example, on Bandcamp, because on Bandcamp I feel that they don’t care much about copyright unless someone reports it to that person, but I was afraid to put it on Spotify, for example, because I know that their algorithms are probably much more trained to capture those copyright breaches.”*

<sup>45</sup> For example, one of the interviewed hip-hop artists expressed it thusly: *“I think that about 80*

Almost all of the interviewed artists that publish their music online had some real-life experience with copyright bots flagging their work for containing samples with uncleared rights; what usually happened was that the IP owner requested monetization, i.e., the partial or complete redirection of advertising revenue generated by the content to the IP owner's account.<sup>46</sup> Most of the artists we talked to found this practice to be correct and fair.<sup>47</sup> At the same time, most of the artists seemed to be "at peace" with the current state of affairs and some even found copyright bots to serve as an instigator of creativity, a sort of a game.<sup>48</sup>

Our interview with a representative of a major Czech record label essentially confirmed our findings from our interviews with the musicians as well as our general knowledge of how copyright bots operate (as described above). YouTube's 2022 policy for Content ID allows copyright owners to choose the length of the sample with potentially uncleared rights that YouTube's automated recognition software should search for – ranging from 1 second to the entire song. Our respondent confirmed that the record label he represents chooses to monetize the content in 90 % of infringing cases and very rarely decides to block the content in question. We can assume that similar economic behavior is also characteristic of other major labels.<sup>49</sup> The interviewee also explained that ContentID allows copyright owners to choose whether the content will be automatically blocked or whether it

---

*% or 90 % of the samples [that I use in my work] are absolutely unrecognizable because I distort them and use them in such a way that it's impossible [for copyright bots to recognize them]. And the samples that are, let's say, more sloppy, or that are more readable, are basically from underground metal bands. I talked about it recently with a couple of people, and basically the feedback was that they [the copyright bots] are not searching for this." The same artist also told us: "When I use a sample that is, let's say, a bit more readable, I think about where the line is and how to use it, so that I can use it and not consider it some form of theft. But of course, I realize that what I'm doing is stealing in a way." Another hip-hop artist we interviewed told us: "[...] all I can think of is that the only thing that cannot be sampled is something that is paid out, monetized on YouTube. So if we use a sound from an old blues thing on YouTube that nobody profits from, then it's cool."*

<sup>46</sup> One of the rappers told us: "Well, it happened to us with an album [that copyright bots detected that we used an uncleared sample], and I think it's easily solvable [...] and that it was resolved in such a way that we didn't pay any money, of course, but that they [i.e., the IP owners] revoked our monetization and it all went to the other party."

<sup>47</sup> The same interviewed rapper commented on the same situation: "Meaning the sample was just stupidly overused and so on. So it was just a tax on stupidity."

<sup>48</sup> One of the hip-hop artists told us: "So, in a way, it annoys me that it's a little bit more complicated, how the robots watch it more closely, how everything is online and digital, but on the other hand, it's still the same and it's actually an interesting game, like it was before. [...] So it just depends on your creativity, how you can hide it so that no robot finds it, so that no one knows where you got it from, so you have a good unique source that no one knows and therefore no one can track it, that's the best approach and the magic of it."

<sup>49</sup> 90 % of matched content being monetized is also mentioned by Perel, M. and Elkin-Koren, N. (2016) Accountability in Algorithmic Copyright Enforcement. *Stanford Technology Law Review*, 19, p. 512 (quoting Lev-Aretz, Y. (2012) Second Level Agreements. *Akron Law Review*, 45, p. 152.

will first be reported to the copyright owner, who then decides how to proceed with such content based on the circumstances. This record label representative corroborated our finding that YouTube's automated content detection technology cannot detect samples that have been materially altered. Finally, he also confirmed that the vast majority of musicians who use uncleared samples in their work, which are then detected by copyright bots, usually do not protest against advertising revenue being redirected to the owner of the sampled recording, or even against the blocking of such content.

Rightsowners' policies cannot be confirmed by YouTube Copyright Transparency Report,<sup>50</sup> since it only provides general information on the number of content matched and taken down, not the ration of monetized and taken down content.

In our interview with a representative of a company that develops globally successful automated software recognition technology, we learned that major global copyright owners prefer to intentionally configure YouTube copyright bots or their own bots so that very short samples go undetected. This prevents erroneous content recognition (false positives) – generally, the longer the detected segment of content, the lesser the chance of an error made by the copyright bot – and also avoids the recognition of content subject to copyright exemption, as users tend to exploit exceptions for short sampled fragments; furthermore, the reporting of potentially legitimate content is generally reputationally undesirable. Our respondent also confirmed that copyright bots still have difficulty identifying materially altered samples – especially if the song's tempo has been changed. This appears to be in line with reports of our interviewees experiencing issues with longer samples rather than very short or substantially modified samples and it is confirmed by YouTube Copyright Transparency Report.<sup>51</sup>

<sup>50</sup> In its last version, Google published this report in June 2022 for the previous year. Cf. Google Inc. (2022) YouTube Copyright Transparency Report H2 2021. Available from [https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22\\_2021-7-1\\_2021-12-31\\_en\\_v1.pdf](https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2021-7-1_2021-12-31_en_v1.pdf) [Accessed 6 June 2023]. Certain dynamic data (though with little relevance for this paper) can be found online at <https://transparencyreport.google.com/youtube-policy/removals> [Accessed 6 June 2023].

<sup>51</sup> *"Despite the power of technology, there are some cases where Content ID fails to identify a match with a user video. This can be due to uploaders' efforts to evade Content ID, or due to the fleeting use of the copyrighted work. For videos missed by automated identification, many Content ID partners have the ability to issue claims manually. While this tool covers an important gap, it accounted for fewer than 0.5% of Content ID claims made in the second half of 2021. For music rights holders in particular, the automation rate is even higher. Finally, all channels on YouTube also have access to our copyright removal webform to request removal of any content not captured by another tool to which they have access."* In: Google Inc. (2022) YouTube Copyright Transparency Report H2 2021. Available from <https://storage.googleapis.com/transparencyreport/report->

## 5. CONCLUSION

In this study, we did not advocate for specific statutory solutions and did not judge the existence of automated content recognition technologies as generally good or bad – our aim was to summarize their current legal framework using the example of the DILIA vs. Ulož.to case and to portray how these technologies have changed the creative attitudes of artists and what kind of playing field it creates, as these technologies will surely be with us for long.

In today's world, it is technology (and how it is configured by humans) which dictates what is legal and what is not on a daily basis. While in the past, the legitimacy of certain practices – for example when it comes to copyright exceptions – was derived from the conduct of individuals using the copyrighted content (publishers), it is now largely derived from the conduct of platform operators. "Automated processes designed to protect copyright-protected material"<sup>52</sup> thus clearly impose limits on the content available online,<sup>53</sup> though as demonstrated by our practical research, human creativity proves very resourceful and although one would think you cannot outsmart technology,<sup>54</sup> you actually can. In another words: "Do something inventive with it!"<sup>55</sup> One of the interviewed artists called this process "an interesting game" and it is indeed a "metagame"<sup>56</sup> when sampling artists try – not without success – to play the system in their creative work, whether it is copyright compliant or not.

## LIST OF REFERENCES

- [1] "Resolve Asset Ownership Conflicts" at <https://support.google.com/youtube/answer/3013321?hl=en> [Accessed 15 November 2022].
- [2] Act No. 121/2000 Coll., on copyright, on rights related to copyright and on the amendment of certain laws (Copyright Act).
- [3] Act No. 480/2004 Coll., on certain information society services.
- [4] Act No. 89/2012 Coll., the Civil Code.

---

downloads/pdf-report-22\_2021-7-1\_2021-12-31\_en\_v1.pdf [Accessed 6 June 2023], p. 12.

<sup>52</sup> Eakman, A. (2015) The Future of the Digital Millennium Copyright Act: How Automation and Crowdsourcing Can Protect Fair Use. *Indiana Law Review*, 48 (2) p. 631, 634.

<sup>53</sup> *Ibid.*

<sup>54</sup> Douglas, N. (2018) You Can't Fool YouTube's Copyright Bots. LifeHacker. Available from <https://lifelhacker.com/you-cant-fool-youtubes-copyright-bots-1822174263> [Accessed 20 November 2022].

<sup>55</sup> Starsky, C. How to Avoid YouTube Copyright Claims. Available from [https://www.youtube.com/watch?v=w-7TMj\\_7UYU](https://www.youtube.com/watch?v=w-7TMj_7UYU) [Accessed 20 November 2022]

<sup>56</sup> Bailey, J. (2021) Copyright in the Age of Bots. Available from <https://www.plagiarismtoday.com/2021/09/23/copyright-in-the-age-of-bots/> [Accessed 20 November 2022]



- [5] Bailey, J. (2021) Copyright in the Age of Bots. Available from <https://www.plagiarismtoday.com/2021/09/23/copyright-in-the-age-of-bots/> [Accessed 20 November 2022].
- [6] Berkowitz, A. (2022) Classical Musicians v. Copyright Bots. *Information Technology and Libraries*, 41(2). Available from <https://ejournals.bc.edu/index.php/ital/article/view/14027> [Accessed 16 November 2022].
- [7] Brodeur, M. (2020) Copyright bots and classical musicians are fighting online. The bots are winning. *The Washington Post*. Available from: [https://www.washingtonpost.com/entertainment/music/copyright-bots-and-classical-musicians-are-fighting-online-the-bots-are-winning/2020/05/20/all349c-98ae-11ea-89fd-28fb313d1886\\_story.html](https://www.washingtonpost.com/entertainment/music/copyright-bots-and-classical-musicians-are-fighting-online-the-bots-are-winning/2020/05/20/all349c-98ae-11ea-89fd-28fb313d1886_story.html) [Accessed 16 November 2022].
- [8] Cambridge Advanced Learner's Dictionary & Thesaurus. Available from: <https://dictionary.cambridge.org/dictionary/english/bot> [Accessed 20 November 2022]
- [9] Depoorter, B. and Walker, R. (2014) Copyright False Positives. *Notre Dame Law Review*, 89(1).
- [10] DiCarlo, I. (2022) Copyright Bots Need a Tune-up. *JTIP Blog*. Available from: <https://jtip.law.northwestern.edu/2022/03/05/copyright-bots-need-a-tune-up/> [Accessed 10 November 2022].
- [11] 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 and amending Directives 96/9/EC and 2001/29/EC.
- [12] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.
- [13] Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights.
- [14] Douglas, N. (2018) You Can't Fool YouTube's Copyright Bots. *LifeHacker*. Available from <https://lifelifehacker.com/you-cant-fool-youtubes-copyright-bots-1822174263> [Accessed 20 November 2022].
- [15] Eakman, A. (2015) The Future of the Digital Millennium Copyright Act: How Automation and Crowdsourcing Can Protect Fair Use. *Indiana Law Review*, 48 (2).
- [16] EUIPO expert studies: European Union Intellectual Property Office (2020) *Automated Content Recognition: Discussion Paper – Phase 1, Existing Technologies and Their Impact on IP*. ISBN 978-92-9156-280-0. Available from: <https://data.europa.eu/doi/10.2814/52085> [Accessed 16 November 2022].

- [17] European Union Intellectual Property Office (2022) *Automated Content Recognition: Discussion Paper – Phase 2, IP Enforcement and Management Use Cases*. ISBN 978-92-9156-326-5. Available from: <https://data.europa.eu/doi/10.2814/952694> [Accessed 16 November 2022].
- [18] Frank Peterson v Google LLC and Others and Elsevier Inc. v Cyando AG, Court of Justice of the European Union, joined cases C-682/18 and C-683/18.
- [19] Gabison, G. and Buiten, M. (2020) Platform Liability in Copyright Enforcement. *Columbia Science and Technology Law Review*, 21.
- [20] Google Inc. (2022) YouTube Copyright Transparency Report H2 2021. Available from [https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22\\_2021-7-1\\_2021-12-31\\_en\\_v1.pdf](https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2021-7-1_2021-12-31_en_v1.pdf) [Accessed 6 June 2023].
- [21] Hašek, J. (2005) *The Good Soldier Švejk and His Fortunes in the World War*. Translated from Czech by Cecil Parrott. London: Penguin Books.
- [22] Hayes, A. (2022) YouTube Stats: Everything You Need to Know In 2022! Wyzowl. Available from: <https://www.wyzowl.com/youtube-stats/> [Accessed 16 November 2022].
- [23] <https://rightsmanager.fb.com/> [Accessed 15 November 2022].
- [24] <https://support.google.com/youtube/answer/2797370?hl=en> [Accessed 15 November 2022].
- [25] <https://transparencyreport.google.com/youtube-policy/removals> [Accessed 6 June 2023].
- [26] <https://www.shazam.com/> [Accessed 15 November 2022].
- [27] Judgment of 29 July 2019, Pelham GmbH and others, C-476/17, EU:C:2019:624.
- [28] Judgment of the Supreme Court of 8 June 2022, 23 Cdo 1840/2021. Available from the website of the court and numerous online sources, including <http://kraken.slv.cz/23Cdo1840/2021> [Accessed 1 November 2022].
- [29] Leška, R. and Půr, M. (2022) Technologické společnosti a jejich odpovědnost za obsah na internetových platformách. *New Direction*. Available from: [https://newdirection.online/publication/technologicke\\_spolecnosti\\_a\\_jejich\\_odpovdnost\\_za\\_obsah\\_na\\_internetovch\\_pla](https://newdirection.online/publication/technologicke_spolecnosti_a_jejich_odpovdnost_za_obsah_na_internetovch_pla) [Accessed 20 November 2022].
- [30] Perel, M and Elkin-Koren, N. (2016) Accountability in Algorithmic Copyright Enforcement. *STAN. TECH. L. REV.*, 19, 473.
- [31] Rosenblatt, B. (2017) A Bounty Hunting Service for E-Book Piracy. Copyright and Technology. Available from <https://copyrightandtechnology.com/2017/01/30/a-bounty-hunting-service-for-e-book-piracy/> [Accessed 16 November 2022].

- [32] Samuelson, P. (2021) Pushing Back on Stricter Copyright ISP Liability Rules. *Michigan Technology Law Review*, 27.
- [33] Shinn, L. (2015) YouTube's Content ID as a Case Study of Private Copyright Enforcement Systems, *AIPLA Q. J.*, 43 (2/3).
- [34] Srstka, J. (2014) Žaloba DILIA proti úložišti www.ulozto.cz. *Věstník DILIA*, , p. 41. Available from <https://www.dilia.cz/ke-stazeni?cat=ostatni> [Accessed 1 November 2022].
- [35] Starsky, C. How to Avoid YouTube Copyright Claims. Available from [https://www.youtube.com/watch?v=w-7TMj\\_7UYU](https://www.youtube.com/watch?v=w-7TMj_7UYU) [Accessed 20 November 2022].
- [36] Tan, C. (2014) Lawrence Lessig v Liberation Music Pty Ltd: YouTube's Hand (or Bots) in the Over-Zealous Enforcement of Copyright. *European Intellectual Property Review*, 36(6).
- [37] Tune, C. and Iverson, S. (2020) The Rise of the Copyright Bots. *Internet and Technology Law*. Available from <https://www.internetandtechnologylaw.com/copyright-bots/> [Accessed 16 November 2022].
- [38] Unpublished judgment of the High Court in Prague of 20 January 2021, 3 Co 58/2019.
- [39] Unpublished judgment of the Municipal Court in Prague of 22 February 22 2019, 34 C 5/2017. Available from: [https://www.dreport.cz/wp-content/uploads/TechLaw\\_judik\C3\%A1t\\_Ulozto\\_Dillia\\_34-C-5-2017.pdf](https://www.dreport.cz/wp-content/uploads/TechLaw_judik\C3\%A1t_Ulozto_Dillia_34-C-5-2017.pdf) [Accessed 1 November 2022].