

ARTIFICIAL INTELLIGENCE and the DUBBING SECTOR

Analysis and perspectives

1. General Overview

The increasing importance of the Internet, and the advent of always more disruptive technologies such as **Artificial Intelligence (AI), Big Data, Blockchain, 5G and Internet of Things (IoT)**, are continuously transforming the world at an unprecedented speed, radically changing societies and economies and running a concrete transformation in the way we produce, distribute and access creative content. National and local authorities, policy makers and businesses are also becoming more aware of the important role and potential fulfilled by the creative industries in driving innovation and growth in the economy¹. Furthermore, an increasing number of policy makers are stressing that innovation, creativity and independent thinking are becoming increasingly crucial for the development of the global economy² placing creative industries at the heart of their national innovation and economic development agenda³, and including innovation and creativity in many national strategies that yearn for growth. At the same time, countries are adopting policies to support and develop their creative industries and local creative production and consumption⁴.

As technologies and regulation have a close relation – technology shapes legal developments and vice versa⁵ – **the latest digital technologies usually entail radical impacts on the existing legislative landscape**, especially in the case where the directives were predating the digital explosion. Even if digitalisation has enabled a large-scale reproduction and distribution of creative content, it has also facilitated the illicit reproductions and exploitation of works, increasing the difficulties of right holders to control the online uses of their works and the financial flows of their remunerations.

In the creative industries, digitalisation had a major impact on copyright and related rights as well as on the functioning of online licenses, through, *inter alia*, the aggregation of repositories, the creation of hubs, data interoperability, and databases. In particular, Copyright related challenges are raised by the increasing use of **Artificial Intelligence (AI)** in the creative industry.

¹ European Union Open Method of Coordination Expert Group on Cultural and Creative Industries, 2012, https://ec.europa.eu/assets/eac/culture/library/publications/cci-policy-handbook_en.pdf

² Moore, 2014, “Cultural and Creative Industries concept – a historical perspective”; Sean A. Pager, “The Role of Copyright in Creative Industry Development”, 10(2) L. Dev. Rev. 521 (2017)

³ Chapain, C., and R. Comunian. 2010. “Enabling and Inhibiting the Creative Economy: The Role of the Local and Regional Dimensions in England.” *Regional Studies* 44: 717–734; Chapain, C., N. Clifton, and R. Comunian. 2013. “Understanding Creative Regions: Bridging the Gap Between Global Discourses and Regional and National Contexts.” *Regional Studies* 47 (2): 1–4; Chapain, C., N. Clifton, and R. Comunian. 2014. “Creative Industries & Creative Policies: A European Perspective?” *City, Culture and Society* 5 (2014): 51–53. Elsevier.

⁴ Musterd, S., and Z. Kovacs. 2013. *Place-Making and Policies for Competitive Cities*. Chichester: Wiley- Blackwell; Musterd, S., and A. Murie. 2010. *Making Competitive Cities*. Chichester: Wiley-Blackwell; Pratt, A. C. 2009. “Policy Transfer and the Field of the Cultural and Creative Industries: What Can Be Learned from Europe? In *Creative Economies, Creative Cities: Asian-European Perspectives*, edited by L. Kong and J. O’Connor, 9–24. Dordrecht: Springer Press.

⁵ Finck, M., & Moscon, V. (2019). Copyright Law on Blockchains: Between New Forms of Rights Administration and Digital Rights Management 2.0. *IIC - international review of intellectual property and competition law*, 50(1), 77-108. doi:10.1007/s40319-018-00776-8.

In 2019 the President of the (EC) Ursula von der Leyen in her Political Guidelines⁶ stated the prioritisation of the investments in Artificial Intelligence and to put forward the legislation for a coordinated European approach. The usage of Emergent Technologies, AI and Robots among creative sectors has caused several legislative debates concerning, *inter alia*, protection of IP rights of the data taken as input by the technology, as well as the ownership of AI-generated outputs and AI-assisted outputs. The application of AI technologies within creative sectors is deeply rooted: RACTER, an AI computer program that generates English language prose at random dates to 1983, even though it still needed a human input⁷. Currently, there are algorithms capable of writing poems, novels, composing music, editing photographs, and even creating music videos in a completely autonomous way. We are hence faced with a **technological revolution that force us to review the interaction between computers, creative processes, and the rights thereto related**.

2. The use of creative works to train AI algorithms

The first aspect to consider when assessing the interaction among the dubbing sector and AI is the fact that already existing dubbing material (i.e. voices of dubbers present in an indefinite number of works) could be used to feed AI technologies and allow the creation of new dubbing content.

Within the machine learning context, machines are repeatedly trained by means of specifically designed learning algorithms that use a corpus of examples in the form of data sets as training material. Very often, in particular with respect to algorithmic creativity, the training material is mainly composed by copyrighted works, such as texts, images, paintings, audio-visual works and – in our case - voices. Machine Learning workflow involves the realization of (multiple) reproductions of any protected work used as training material.

a) *Does such kind of use infringe copyright?*

The training typically includes activities relevant for copyright protection. For example, the individual data must be saved on a server and stored in the working memory, which implies a reproduction of the work in terms of copyright⁸. If the original data is pre-processed, an adaptation relevant under copyright law will often take place. In this respect, one needs to consider the value of creative content and how any creative endeavors is monetized, especially when third parties wish to use copyright protected material for their own benefit and/or for offering new services on the market.

As a consequence, the training can only be carried out in conformity with intellectual property rights if either a license is obtained, or a specific exception is provided for the respective intellectual property right.

Such an exception has now been enacted in the fully harmonizing Art. 3 et seq. of Directive 2019/790 on copyright in the digital single market (DSM Directive). Hence, the question arises to what extent the European legislator has succeeded in achieving an adequate balance between the exploitation interests of the rightholders and innovation interests, i.e. to what extent the risk of blockage has been properly addressed.

⁶ U. von der Leyen, 2019, “Political Guidelines For The Next European Commission 2019-2024” https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

⁷ B. Chamberlain, 1984, “The Policeman's Beard Is Half Constructed”.

⁸ Geiger, Frosio and Bulayenko, The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects, Briefing for the JURI committee of the European Parliament, 2018, 6; Raue, ‘Rechtssicherheit für datengestützte Forschung’, ZUM (2019), 684 (685); Obergfell, ‘Big Data und Urheberrecht’, in: Ahrens et al. (eds.), Festschrift für Wolfgang Büscher, 2018, 223 (226); Spindler, ‘Text und Data Mining’, GRUR 2016, 1112 (1113); cf. also recital 8(6) and recital 9(2) CDSM Directive.

According to Art. 3(1) CDSM-Directive, Member States must establish an exception to the right of reproduction and extraction when the use for text and data mining is made by *‘research organizations 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’*. Thus, if such an actor has legally acquired access to the data, all acts of reproduction, but also pre-processing are allowed for the purpose of automated data analysis.

However, Art. 2(1) CDSM Directive defines the term ‘research organisation’ such that the organization must not operate for profit, with full reinvestment of profits in research or with a mission in the public interest recognized by the State.

Therefore, profit-oriented companies, cannot, even if they pursue research objectives and publish their results (as is not uncommon) in leading international journals, invoke the implementation of Art. 3(1) exception.

To sum up: while Article 3 of the DSM Directive⁹ provide the legal framework to carry out text and data mining within the AI context without the right holder’s authorization, the final text of the Directive limits the exception of Article 3 for text and data mining activities carried out by a specific category of beneficiaries i.e. research organizations and cultural heritage institutions.

Profit-oriented uses of AI training are therefore only covered by Article 4 of the CDSM Directive which provides for a general limitation or exception “for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining”. However, according to the third paragraph of the Article, the respective right holders can oppose pre-emptively the use of their works for text and data mining activities, communicating this decision in an ‘appropriate’ manner.

On this basis, it should be clarified what measures would qualify as ‘appropriate’ in order to assess how the opt-out mechanisms should be implemented in a practical manner for both rightsholders and right users.

The text of the DSM Copyright Directive does not clarify whether such exception allows the processing of large sets of data to produce creative outputs generated – autonomously or not – by an artificial intelligence. On this basis, the dubbers associations should suggest to the local legislator transposing the DSM Copyright Directive to clarify the scope and application of the text and data mining exception of Articles 3 and 4 of the DSM Copyright Directive, notably:

- In which situations and how the exception of text and data mining applies?
- Does the exception of text and data mining allow all use of lawfully accessible sets of data (among which human voice) to train an AI algorithm regardless of its purpose?
- Does the exception of text and data mining cover all use of copyright protected content to train an AI algorithm to produce creative content?
- What is an appropriate manner to oppose the use of a content for text and data mining activities?

The use of copyright-protected content as input for AI technologies raises thus questions on the conditions to obtain the right holder’s authorization, whether to ensure remuneration for right holders in some circumstances, and how to grant transparency concerning such use.

⁹ 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: <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex:32019L0790>

AI technologies may render the traceability of IPRs and their application to AI-generated output difficult, thus preventing human creators whose original work is used to power such technologies from being fairly remunerated; the issue of the protection of IPRs in the context of the development of AI and related technologies has not been addressed by the Commission, despite the key importance of these rights.

b) Does such kind of use infringe privacy rules (GDPR)?

Data protection law does provide for certain requirements in the area of data quality. For example, Article 5(1)(d) GDPR specifies that personal data must be '*accurate and, where necessary, kept up to date*'. These prerequisites, however, are very vague and remain undefined in the GDPR. Even more importantly, the applicability of data protection law to training data is highly doubtful because it is often anonymized. Hence, training data sets in supervised learning may often fall between the cracks of the GDPR. In training environments for reinforcement learning, its applicability is even more clearly excluded if synthetic data is used.

3. Legal protection of AI generated contents

As concerns legal protection of AI-generated outputs, the latter might not be eligible for copyright protection, failing the authorship (as expression of a human being) and originality requirements.

Whether copyright can be attributed to an AI-generated or AI-assisted work will largely depend on the amount of input provided by the author. In the absence of copyright protection, a certain degree of protection shall be ensured according to the trade secrets directive and/or contractual provisions. The legal doctrine has evidenced that protection by means of a sui generis/related right, owned by the person who made the necessary investments for the AI invention, might be a valid legal instrument to stimulate and safeguard the investment in AI for creative purposes and allow for the establishment of a system reflecting the peculiarities of AI creativity.

If considered eligible for copyright protection, ownership of copyright on the AI-generated work would be the first challenge: the rights of economic exploitation could be vested in whoever commissioned the work to the AI, as it happens for commissioned works, where the results belong to the party who commissioned them, or in the creator/inventor of the AI. Potential challenges of joint ownership and the appropriate duration of copyright protection afforded to AI-generated works could also arise.

Some countries have already implemented legal provisions dealing with machine created content¹⁰, admitting the protection of intellectual works created by AI and thus allowing companies to continue investing in the process of creative technology and development. However, in many situations drawing the line between AI-generated or AI-assisted creative output is not straightforward. This entails an ethical problem of AI transparency on how humans are informed when they are accessing AI-produced and not human-created works. For instance, the website "BOT OR NOT"¹¹ implements a Turing test for poetry and the user is called to guess whether the poem is written by a human or by a computer: the test shows an average of 35% of failure in distinguishing AI generated content.

Finally, there is the issue of liability of robots/software agents, which could be legally treated as mere machines, instruments in the hands of their human creators who would remain responsible

¹⁰ For example, Section 9 of the UK Copyright, Designs and Patents Act 1988 (CDPA) states that: "In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken", while Section 178 of the CDPA further defines a computer-generated work as one that "is generated by computer in circumstances such that there is no human author of the work".

¹¹ <http://botpoet.com/>

for their activities. However, the growing autonomy of AI are generating unknown risks, hence, new rules are needed to provide sufficient safeguards for the efficient allocation of liability, also concerning IP infringement cases.

In the light of the above, there are multiple challenges to face in connection to AI in the cultural industries:

- **Protection for data used by AI:** AI systems have a close relation with data to train algorithms, and these data could be protected by third-party rights. Under certain circumstances data could be protected by copyright. Remuneration for the right holders of such data according to copyright laws might be due, but the exception related to text and data mining and other copyright limitations may apply;
- **Regulation of AI-generated outputs and AI-assisted outputs:** most copyright legislations across EU Member States are very much dependent on human-centred concepts, for establishing (i) the right owner (i.e. the author); (ii) the conditions for protection (e.g. creativity, originality); and (iii) the rights granted (economic, but also moral rights). Copyright might thus not be most efficient legal mean to protect AI-generated contents. A suitable legal approach for protecting AI generated creative content should be laid down at EU level;
- **Adaptation of traditional rules regarding IP infringement and liability:** some commentators have suggested to qualify AI (or robots) as a new form of legal person (so-called “ePerson”) as a first step towards the question of whether the actions of AI can be attributed to the human being using it, or if the machine should be liable for its actions¹². However, AI in its present forms cannot determine the preferences or goals to be achieved; it is still up to the human being implementing and using AI to define the areas and goals for it. In other words, AI may improve ways to achieve a goal but cannot change it. Ultimately, the real question regarding all legal areas concerns to a greater extent the issue of whether the actions of AI can be attributed to the human being using it¹³;
- **The application of AI technologies within creative sectors:** AI technologies are applied within the creative industry in different manners and with different purposes, such as IP enforcement and creation of new cultural contents. There is an entire start-up ecosystem (For instance Splice¹⁴; Amadeus Code¹⁵) around services that give artists automated song writing recommendations or enable the average internet user to generate customised instrumental tracks at the click of a button (for instance “Flow Machines Technology by Sony”¹⁶). These cases are characterized by a legal complexity which relies in determining whether human artists using AI instruments could be considered the real “authors” of the end works, or whether they are just “programming” the AI tools. In any case, as currently AI cannot set its own goals and preferences, but it is still the author who defines the framework within which the specific work will be created, there might be still good arguments to attribute the work created by AI to the individual making use of it depending on how the preferences and data training was set¹⁷;
- **Commercial implications of EU policy:** the digital society will be increasingly characterized by the interaction of human actors and non-human technological actors or virtual agents.

¹² J. Goodman, 2016, “Robots in Law: How Artificial Intelligence is Transforming Legal Services”

¹³ Gerald Spindler, “Copyright Law and Artificial Intelligence”, International Review of Intellectual Property and Competition Law volume 50, pages 1049–1051(2019)

¹⁴ <https://techcrunch.com/2018/04/16/splice-sounds/?guccounter=1&renderMode=ie11>

¹⁵ <https://amadeuscode.com/top>

¹⁶ <http://www.flow-machines.com/>

¹⁷ Gerald Spindler, “Copyright Law and Artificial Intelligence”, International Review of Intellectual Property and Competition Law volume 50, pages 1049–1051(2019)

In order to boost technological and industrial capacity & AI uptake for the EU, a crucial challenge will be to foster discussion on the evolution of EU policy in an increasingly AI-driven world, to ensure an appropriate ethical & legal framework.

Given the identified opportunities and challenges, this specific contract concerns the modelling of a solution for a correct data management in compliance with the Intellectual Property Right (IPR) legislative context.

4. The white Paper of the Commission

On 19 February 2020, the European Commission unveiled three core documents for its digital agenda: a data strategy¹⁸, a report on security and liability issues in emerging technologies¹⁹, and the White Paper on AI²⁰. The latter is characterized by a clear dichotomy of promotion and regulation of AI. Overall, the White Paper attempts a difficult, but generally worthwhile, balancing act of making Europe a centre of AI development and application, while at the same time adequately addressing the risks of this technology and ensuring that European fundamental rights and values are adequately enforced.

However, the two major parts of the White Paper remain almost unconnected. The Commission starts with a clear and factually accurate description of the importance of AI for research, the economy and the society at large, stresses the need for the intelligent use of diverse data sources, and then sets out, on four pages, dimensions of enabling and promoting AI development, especially for SMEs.

However, IP law is conspicuously absent from the entire discussion, and the Commission hence fails to mention that the new copyright exception for text and data mining in Art. 3 of the CDSM Directive.

Overall, a legal framework for training data affords the advantage of actively shaping AI applications ex ante, at the stage of their technical design, in such a way that elementary legal norms and social values are respected. In contrast to human decisions that can hardly be explicitly controlled, this possibility, of consciously determining the relevant parameters, also demonstrates the considerable promise of responsible AI for socially desirable decisions.

¹⁸ European Commission, A European strategy for data, Communication, COM(2020) 66 final

¹⁹ European Commission, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64 final, 8 et seq.

²⁰ European Commission, On Artificial Intelligence - A European approach to excellence and trust, White Paper, COM(2020) 65 final.