Anna Bober-Kotarbińska
WSB Merito University in Gdańsk, Poland
anna.bober-kotarbinska@gdansk.merito.pl
ORCID: 0000-0003-0499-8253

https://doi.org/10.26881/gsp.2024.4.02

The Justification for Establishing Exceptions and Limitations to Copyright for Programs based on Artificial Intelligence

Introduction

The issue of artificial intelligence (AI) in the context of intellectual property law, including copyright, enjoys constant interest. Continuous innovation brings new challenges, and the stunning progress we have seen in this field in recent years, and especially over the past year, in AI systems and their rapid expansion of capabilities are attracting significant media and public attention. Among many AI models, one type of AI known as generative AI (GenAI) technology is particularly capable of generating output such as text, images, video and audio (including human voice emulation) using output in the form of copyrighted works. The adoption and use of GenAI systems have sparked widespread public debate about what these systems mean, raising important questions for the copyright system. In the wake of these questions, copyright holders have filed infringement lawsuits against AI companies, alleging copyright infridgement AI training processes and results obtained from GenAI systems.[1] These lawsuits raise justified concerns about the unauthorised use of copyrighted materials to create new creative content. However, at the same time, they may significantly slow down work on the development of AI, which, in addition to threats, brings many valuable solutions. In light of these challenges, a comprehensive and holistic approach is needed to address AI-related copyright issues, considering both the inputs and outputs of AI systems. Based on discussions, normative work, consultations, and guidance from experienced commentators (including Daniel Gervais, who noted that there is a broad spectrum of current regulatory issues for GenAI[2]), this article identifies one of the broad spectrum of copyright policy issues arising from on the development and use of AI regarding the

[1] For example: Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-135, ECF No. 13 (D. Del. Mar. 29, 2023) (alleging infringement based on the use of copyrighted images to train a generative AI model and on the possibility of that model generating images "highly similar to and derivative of" copyrighted images).

[2] D. Gervais, *Generative AI & IP: Gervais' checklist of issues*, https://www.linkedin.com/pulse/generative-ai-ip-gervais-checklist-issues-daniel-gervais/ [accessed: 2023.10.31].

legality of using copyrighted works for training AI models. The question posed is about the advisability of establishing a new system of copyright exceptions and limitations dedicated to AI systems. At the same time, the impact of existing limitations resulting from copyright exceptions and limitations on the development of AI is analysed.

To answer this question, this article examines the difference between AI and GenAI and describes their essential operation. Then, selected national legislation and normative responses of selected countries and draft acts dedicated to copyright in the context of AI are examined. Because of the developing concepts of establishing appropriate provisions of national law, attention is directed to the draft normative act unveiled in the autumn of 2023 in France. The next part presents a concept that denies the validity of a licensing system in favour of establishing a uniform, international copyright exception for using AI systems.

1. Introduction to technology – what are AGI, AI and GenAI, and how do they work?

Some time ago, there were serious discussions about the definition of AI (with a complete lack of understanding of the role and function of definition in practical logic, patterns from formal logic were then followed). Although no clear legal definition has been established to date, the concept of artificial general intelligence (AGI) has been developed as a computer program that will reach the level of human intelligence in all aspects of human thinking (and in many cases exceed it due to its significant computing potential greater than that of humans). AGI should be distinguished from AI systems (so-called specialised AI systems) that perform specific, limited tasks that require intelligence when performed by humans. These are usually very specialised tasks, but the quality of their performance often exceeds human capabilities. Currently, GenAI plays an essential role in the world of copyright law. GenAI uses generative models (i.e. for creating new content) based primarily on artificial neural networks, enabling the creation of new digital content, such as text, images, music, and film. Currently, the most famous and widespread tool based on GenAI systems is ChatGTP.[3] GPT is the abbreviation of generative pre-trained transformer, which describes the essential features of the design of this computer program. This tool produces texts using a computational technique called a transformer neural network, the network parameters of which are established by previously training it on examples from a huge text database. Generally, as input, the program receives a fragment of text in natural language, for example, a query, and its task is to generate another sensible, grammatically correct text that is most suitable as a continuation of the given fragment. Fitness is determined by the text database on which the program was trained. On a very general level, the process usually consists of two stages: extracting

[3] ChatGPT is a language model created by OpenAI, an IT company based in San Francisco, California, USA.

information regarding the intention of the question (the intention to achieve some result) contained in the instruction indicated by a human user of the program and generating content consistent with the extracted intentions.[4] Thus, instead of linking to content previously posted on the Internet, GenAI draws on existing content to create new content. New content can appear in formats that include all the symbolic elements representing the premises inherent in human thinking: texts written in natural language, images, videos, music and even software code. The GenAI in GTP is trained with data collected from websites, social media conversations and other online media. The program generates the most likely continuation of the text based on the texts in the database by statistically analysing the distribution of words, pixels or other data elements it has absorbed and identifying and repeating typical patterns. As of Fall 2023, ChatGPT will soon support voice and image prompts (for paid users). Other entities, such as Spotify, launched GenAI that can translate and copy podcasts into other languages that match the speaker's voice/tone, and Getty Images launched a "commercially safe" AI image generator trained on licensed images.

As seen from the above, the development of AI-based programs implies the need to gain access to vast data pools, and the larger the pool, the more accurate the system becomes. From a copyright perspective, it is essential that while some data may include works that are in the public domain and can be freely used, others may be copyrighted works, especially those related to cutting-edge science and technology. Therefore, in order to develop, AI systems must perform activities on copyrighted works, e.g. reproduction, text and data mining (TDM) and the creation of derivative works,[5] and undertake activities related to copyrighted content that require at least reproduction, and sometimes even adaptation work that is within the exclusive control of the copyright owner. Thus, the conclusion that can be drawn is that when AI models are trained on data, there is a risk of copyright infringement in some, if not most, countries. Therefore, without the express permission from the copyright holder (such as a license), any reproduction, adaptation or other action, unless it falls within the list of permissible legal exceptions, may violate this right. For example, an AI system developed to discover new cancer treatments would need to process many copyrighted articles in the medical sciences by accessing, reading and copying (reproducing). Retrieving data from the Internet essentially involves creating a first copy that is needed for processing in order to access the data contained in its content and is a necessary step for further processing. There will undoubtedly be other steps to

[4] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. Yu, L. Sun, *A Comprehensive Survey of AI-Generated Content (AIGC): A history of Generative AI from GAN to ChatGPT*, "Journal of the ACM" 2018, vol. 37, no. 4, https://arxiv.org/pdf/2303.04226.pdf [accessed: 2023.10.31].

[5] In copyright law, a derivative work is a work based on one or more pre-existing works, such as a translation, adaptation, sequel, or a work based in some way on another work. A derivative work is considered to be a new work while retaining certain characteristics of the original work. A work may serve a transformative purpose even if it does not change the content of the original work. See R. Reese, *Transformativeness and the Derivative Work Right*, "Columbia Journal of Law & the Arts" 2008, vol. 31, iss. 4, p. 4.

prepare this data for use in training, but this first raw copy is a copy. In addition, these works would have to be adapted or compiled into new works or formats (derivative works). Therefore, to operate legally, the creators of an AI system would be required to obtain consent from authors.

2. Legislative concepts

2.1. Current legislative concepts

The universal copyright system does not contain established, uniform exceptions allowing data processing by electronic devices. Nevertheless, in international treaties and conventions, international copyright law allows and leaves room for states to establish their own exceptions and limitations, as long as they do not conflict with the so-called three-step test as described in art. 9 section 2 of the Berne Convention for the Protection of Literary and Artistic Works[6] (concerning only the right of reproduction), art. 13 of the TRIPS Agreement[7] (including any exclusive rights) or art. 10 of the WIPO Copyright Treaty[8] (referring to "the rights granted to authors of literary and artistic works under this Treaty").

Before the AI boom, national legal systems were moderately enthusiastic about establishing exceptions for using data resources by computer systems known as TDM processes.[9] Among the most well-known national exceptions allowing data mining for computer processing purposes was the world's first exception for computer-based TDM established in Japan in 2009. The Japanese Copyright Act[10] art. 30-4 allows access to copyrighted works for extraction, comparison, classification or other statistical analysis of language, sound or image data or other elements of which a large number of works or a large volume of data is composed and computer data processing. Article 47-4 exempts making incidental electronic copies of works from copyright infringement. Finally, art. 47-5 allows the use of copyrighted content for data verification purposes connected to research. Thus, Japan allows the flexible exception for purposes other than entertainment.

On the European continent, other countries followed Japan, including the United Kingdom in 2014, by establishing in CDPA:[11] 29A "Copies for text and data analysis for non-commercial research (1) The making of a copy of a work by a person who

---

6 Act of Paris of the Berne Convention on the protection of literary and artistic works, Paris, 24/7/1971.

7 Agreement on Trade-Related Aspects of Intellectual Property Rights, signed in Marrakesh, Morocco, 15/4/1994.

8 WIPO Copyright Treaty (WCT), Geneva, 20/12/1996.

9 Text and data mining (TDM) is a broad term used to describe any advanced techniques for computer analysis of large amounts of any type of data (numbers, text, images, etc.). It is a key tool in many areas of research, especially in the field of AI.

10 Japanese Copyright Act no. 48 of May 6th, 1970, as amended by Act No. 72 of July 13th, 2018.

11 Section 29A of the UK Copyright, Designs and Patent Act 1988.

has lawful access to the work does not infringe copyright in work provided that – (a) the copy is made so that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and (b) the copy is accompanied by a sufficient acknowledgment (unless this would be impossible for reasons of practicality or otherwise)." This exception was relatively broad in nature, allowing anyone (a person) to copy copyrighted works for the purpose of performing text and data analysis (computational analysis in the wording of the exception), although it requires that person to have lawful access to the materials, and at the same time that the analysis is exclusively for research and non-commercial.

Other countries, such as Singapore, adopted similar rules in 2021 guided by the desire to exist in the technology industry and to exploit the opportunities related to it. Section 8 and Art. Sections 243 and 244 of the Singapore Copyright Act 2021[12] provide that a permitted use is to make a copy of a work or recording of a protected work for computational analysis purposes, which includes "(a) using a computer program to identify, extract and analyse information or data from the work or recording; and (b) using the work or recording as an example of a type of information or data to improve the functioning of a computer program concerning that type of information or data."

In the American legal system, the most famous section is art. 107 of the Copyright Act. For some time, according to researchers,[13] it could be an answer similar to modern exceptions proposed in Asian countries or in the European legal system. It provides a legal framework for determining whether something is a permitted use and identifies certain types of uses such as criticism, commentary, news reporting, teaching, and scholarship and research as examples of activities that may qualify as a permitted use. The American doctrine of fair use has a very wide-ranging exception and for many years it has actually been the answer to many issues related to the use of works for education, transformation into available formats[14] or entertainment.[15]

In the regional system of European law frames like this exception are provided by art. 3 and art. 4 EU Directive 2019/790 of the European Parliament and the Council of 17 April 2019 on copyright and related rights in the Digital Single Market.[16] During the development stages of Directive 2019/790 (CDSM), the European Commission considered that a new mandatory exception or limitation on the use of TDM technologies in the field of scientific research would be beneficial in the development

---

12 Singapore Copyright Act 2021 (No. 22 of 2021), Art. 244.

13 S. Flynn, L. Schirru, M. Palmedo, A. Izquierdo, *Research Exceptions in Comparative Copyright*, PIJIP/TLS Research Paper Series no. 75, 2022; N. Scharf, *Digital Rights Management and Fair Use*, "European Journal of Law and Technology" 2010, vol. 1, iss. 2.

14 Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, June 10, Intervenor Defendants-Appellees. Yeah. 12-4547.

15 Sony Corporation of America v. Universal City Studios Inc, 464 US 417, 455, n 40 (1984).

16 Directive 2019/790 of the European Parliament and the Council of April 17th, 2019, on copyright and related rights in the Digital Single Market, OJ L 130, 17.5.2019, pp. 92–125.

of AI. Article 3 establishes a broad and specific TDM exception for nonprofit scientific research. In contrast, art. 4 provides a more general but narrower TDM exception, subject to certain limitations related primarily to the purpose of use. The fundamental difference between the above-mentioned norms applies to both the subjective and objective spheres – the key point is that scientific entities and non-profit activities (art. 3) have been given a wider opportunity to operate than commercial entities that meet the conditions for operating under art. 4. Tim same, the EU introduced significantly shallower versions of the exception for enterprises with a significant caveat that it may be overridden by the right to opt out, a concession to rights holders introduced during the very last stage of the Copyright Directive's adoption process. This is fraught with practical difficulties. It can also be argued that, unlike the US, the EU takes a protectionist stance and has established a degree of scientific responsibility for using training data. As pointed out by Nicola Lucchi,[17] under this provision, individuals such as commercial AI system developers and educators may make copies of works or databases to extract information from text and data. They may retain these copies for as long as they are needed for the AI training process.[18] However, Nicola Lucchi also asserts that "rights holders have the option to exclude TDM exemptions from their contracts in order to safeguard their commercial interests."[19] This particular provision has been criticised for "providing a copyright exception that is perceived as being too restrictive. In contrast to the traditional understanding of copyright, which generally focuses on protecting original expression, this provision appears to include factual information and data, and this aspect has drawn much criticism."[20] However, how this opt out option can be implemented and the extent to which AI developers will adhere to it are yet to be determined. It is important to underscore that the 2019 CDSM Directive allows training AI algorithms on other people's data sets. In the case of research and educational entities, it is virtually unlimited (art. 3 of the Directive). In relation to entities operating commercially, authorized entities have the opportunity to prohibit such use (the right to opt out), at least in the absence of determination of due compensation (art. 4 of the Directive). During the work on the directive until its entry into force in 2019, knowledge about the phenomenon and possibilities of GenAI did not extend beyond a narrow group of technology giants; thus, neither the text of the directive, nor the accompanying documents contain any phrases referring to this type of GenAI.

As seen from the above, all the exceptions in question have in common the purpose of essentially non-commercial use and the intention to process works for TDM purposes. Despite this, they cannot be considered sufficient for use in AI processing. The only permitted acts are to make a copy, storing, retaining and to communicate

---

17 N. Lucchi, *ChatGPT: A case study on Copyright Challenges for Generative Artificial Intelligence Systems*, "European Journal of Risk Regulation" August 2023.

18 *Ibid.*; art. 4(1) CDSM.

19 Article 4(3) CDSM.

20 T. Margoni, M. Kretschmer, *A Deeper Look into the EU Text and Data Mining Exceptions: Harmonization, Data Ownership, and the Future of Technology*, "GRUR International" 2022, vol. 71, iss. 8.

the work, and all references to computation data analysis itself are descriptive of the purposes for which the acts mentioned above are undertaken. While it is possible to over-generalise what is encompassed in the practical development of AI models, the operations generally undertook to include processing data to execute training, and my cautious observation is that there may be nuances and exceptions and that the steps to train AI include processing data into a form that can be used to train an AI model. The straightforward question in copyright is whether such processing to extract the data points is re-rendering the raw data into a form that can be used for analysis (e.g., comparable, even, to translation), or whether it is merely extracting ideas from expression. The answer may even be both, neither or something in between, but it is worth noting that on the only-translating side of the analysis, if the conversion of the copyright works into a format for training amounts to a form of adaptation, then those steps are a separately copyright-protected act, and adaptation it is not an act (or a verb) that appears to be permitted under the computation data analysis provisions. On the only-extracting-ideas-from-expression side of the analysis, there is an open question as to whether copyright is even engaged at all. If we land somewhere in between, would we have a transformation of the work such as to engage the fair use defence, which then is a segue to comparative reviews on transformative use under US copyright law, and even discussions on fair learning.[21] Despite the regulations established above, the copyright industry does not share the idea that the current activities of AI providers should fall within the established exceptions and limitations of national law. Regardless of the critical voices of the doctrine and literature, this is particularly illustrated by the legal situation developing in the USA, where defendants based on copyright claims include, among others, the legal case of Tremblay v. OpenAI Inc.[22] The plaintiffs assert that OpenAI employed their copyrighted books without obtaining proper authorisation in order to train ChatGPT. It is claimed that ChatGPT could effectively condense the content of literary works, meaning that the chatbot comprehensively interacted with and assimilated the information contained in these literary works. In the case of Silverman et al. v. OpenAI Inc.,[23] the claims assert that OpenAI engaged in unauthorised utilisation of copyrighted work for the purpose of training ChatGPT. Similar claims were made by Getty Images Inc. vs. AI Stability,[24] which disputes that their AI tool generated a file bearing the plaintiff's watermark for "the purpose of promoting, facilitating or concealing infringement of Getty's copyright Images." The resolution of all of these suits are pending and how they will be resolved remains uncertain at this time, but these cases clearly indicate that the long-popular Section 107 exception will no longer address AI issues.

Interestingly, the problem of the lack of exceptions is also being noted in China. Yudong Chen points out that although Chinese intellectual property laws exist very

---

21 M. Lemley, B. Casey, *Fair Learning*, "Texas Law Review" 2021, vol. 99, iss. 4, https://texaslawreview.org/fair-learning/ [accessed: 2023.10.31].

22 Tremblay *et al.* v. OpenAI, Inc. *et al.*, no. 4:2023-cv-03223 (ND Cal. Jul. 7, 2023).

23 Silverman *et al.* v. OpenAI, Inc. *et al.*, no. 4:23-cv-03416 (ND Cal. Jul. 7, 2023).

24 Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135-GBW (D. Del. Mar. 29, 2023).

close to the four-component premise exception from § 107 of the Copyright Act of the USA, based on opinions of the Supreme People's Court of China they are broadly consistent with the four factors of the US fair use statute, and the 13 additional exceptions (in Chinese Copyright Lat Act) do not cover AI issues. In Chen's opinion, the laws of both countries need to be further clarified and amended on this issue.[25]

2.2. New Legislative Concepts

Aware of the risks but also the benefits related to the use of AI in the areas of using works protected by copyright, some countries are trying to establish new provisions in their national law systems by modifying existing standards to include new technological trends. These include, among others, France, where on 12 September 2023 a proposal for a law was presented to reform certain norms in existing copyright law. To summarise the content, it should be noted that the preamble specifies the purpose of the act, which is "to protect authors and artists of works and interpretations based on the humanistic principle, in legal compliance with the Intellectual Property Code." From the perspective of the issue under analysis here, proposed art. 1 is important since it adds a paragraph at the end of art. L131-3 of the French Intellectual Property Act concerns the transfer of copyright: "the integration by artificial intelligence software of intellectual works protected by copyright into its system and a fortiori their exploitation is subject to the general provisions of this code and therefore to authorization from the authors or rights holders"[26] and thus their use is subject to the general provisions of this Code and therefore requires the authorisation of the authors or rights holders. As Andres Guadamuz rightly points out, adding such content is superfluous concerning the regulations, but adding this content emphasises the importance of authors' rights. Article 2 modifies existing art. L321-2, which concerns collective management organisations. Article 3 amends art. L121-2, obliging to mark the work generated in the program as a work generated by AI and to enter the authors' names of the works that led to the creation of such a work.[27] Article 4 of the proposed act deals with taxation issues.[28] However, the proposed content does not refer to establishing exceptions or limitations under copyright law.

---

25 Y. Chen, *The Legality of Artificial Intelligence's Unauthorized Use of Copyrighted Materials Under China and US Law*, "The Intellectual Property Law Review" 2023, vol. 63, iss. 2, pp. 241–279.

26 Proposition of art. L131-3 French Intellectual Property Act in original version: "l'intégration par un logiciel d'intelligence artificielle d'œuvres de l'esprit protégées par le droit d'auteur dans son système et a fortiori leur exploitation est soumise aux dispositions générales du présent code et donc à autorisation des auteurs ou ayants droit," the version in quote is author's own translation, https://www.lexing.law/avocats/proposition-de-loi-intelligence-artificielle-et-droit-d-auteur/2023/11/23/ [accessed: 2024.09.19].

27 According to A. Guadamuz (*idem*, *French lawmakers propose new copyright about generative AI*, 24/09/2023, https://www.technollama.co.uk/french-lawmakers-propose-new-copyright-law-about-generative-ai [accessed: 2023.10.31]) the main controversy on this subject concerns art. 2, namely the fact that the above assumes that AI works are subject to copyright.

28 More: *ibid.*

It is worth emphasizing that, at the regional level, in December 2023 the European Commission, the Council and Parliament reached a provisional agreement on the AIA,[29] which is a comprehensive legal act on the creation and operation of intelligent algorithms defined in the Act as AI. Although discussions and more or less advanced work on similar regulations are underway in countries around the world, the EU AIA is currently the most comprehensive regulation of this type. It is assumed that this will be an important landmark for similar legal acts in other places around the world and especially in the US. The AIA introduces limited exceptions for TDM, recognizing the importance of balancing copyright protection with promoting innovation and research. Recital 109 recognizes the need for proportionality of compliance requirements, particularly for small and medium-sized enterprises and start-ups. This provision aims to facilitate non-commercial research activities while ensuring adequate protection of the interests of rights holders. Recital 105 highlights the importance of obtaining permission from rights holders for any use of copyrighted content in AI training models unless appropriate copyright exceptions and limitations apply. Attention is drawn to the provisions of Directive (EU) 2019/790, which introduced exceptions and limitations allowing, under certain conditions, the reproduction and downloading of works for the purposes of TDM. However, it clarifies that rights holders may reserve their rights to prevent TDM (the right to opt out), unless it is for the purpose of scientific research. Furthermore, Recital 105 explicitly links the use of copyrighted works for training AI models with the TDM exception in art. 4 of the CDSM Directive. This link aims to put an end to disputes over the application of this exception to AI model training and confirms that even if the legislature did not expressly provide for such uses when discussing exceptions under TDM, the AIA recognizes that art. 4 of the CDSM Directive applies to such applications. GenAI was noted at the final stage of work on the AAI. Based on this regulation, providers of models capable of generating content will be obliged to provide information on how they train their models, and the data sets used as training substrate. They will also be obliged to create policies to respect copyright provisions (AI policy).

3. Establishing a harmonized exception

The impact of AI and GenAI is significant for global economies, especially in the areas of customer service, marketing, software engineering and research and development. In developed countries, AI and GenAI are also used in the banking, advanced information technology and insurance sectors, where employee work automation processes have been significantly improved. At the same time, AI is used in developing countries and countries with a significant degree of digital exclusion to improve education and

[29] Proposal for a Regulation of the European Parliament and of the council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM/2021/206 final, Document 52021PC0206.

teaching processes where the primary problem is the lack of books. As access to data becomes more and more essential for the economic development of countries and for the digital opportunities of individuals, those countries and people who do not have access or cannot afford data remain in a situation of data poverty.[30]

Since, like it or not, AI technologies exist and the direction of future development for science and technology depend on them,[31] it is necessary to include AI's unauthorised use of copyrighted materials in the scope of fair use as much as possible. Continuing to remain in a legal situation in which the legal system does not respond to real problems related to developing the technology of the future not only constitutes a weakness of law as an instrument responding to actual needs, but it is also a real blow to the development of technology.

With this goal in mind, the literature considers several strategies for establishing a legal security framework, the most popular of which includes entering into explicit data-sharing agreements with data providers.[32] I am, however, sceptical about this concept as this solution has several fundamental drawbacks. The first is the fact that copyrights are sometimes separated from authors, and training AI on transformed works would require making decisions that fall within the sphere of the creator's personal rights. Significantly, the copyright holder usually does not have the legal rights to make decisions that infringe the sphere of personal copyright rights. Second, it should be pointed out that there are potential, although significant, difficulties in concluding the contract in question. Assuming that the license agreement covers all works that are not in the public domain, it would be necessary to conclude appropriate agreements with copyright holders. This would imply the need to identify, locate and contact the authorised entity, which, in view of the standards governing the protection of personal data, could constitute a significant obstacle in the process of concluding a contract. Regardless of the standards governing the protection of personal data, it sometimes happens that even if the author's personal data is known, contacting them may be difficult or impossible. The issues in question constitute the basis for the existence of the doctrine dealing with orphan works, the authors of which cannot be identified or located. Even if such a position were accepted as justified, it could lead to the biased treatment of orphan works and to a distortion of

---

30 T. Marawala, *Artificial Intelligence, Game Theory and Mechanism Design in Politics*, London 2023, pp. 41–58.

31 M. Kop, *AI & Intellectual Property: Towards an Articulated Public Domain*, "Texas Intellectual Property Law Journal" 2020, vol. 28, no. 1.

32 In this way, among others M. Kop, *The right to process data for machine learning purposes in the EU*, "Harvard Journal of Law & Technology" 2021, vol. 34; M. Senftleben, *Generative AI and Author Remuneration*, "International Review of Intellectual Property and Competition Law" 2023, vol. 54; proposing to introduce remuneration mechanisms that ensure the payment of compensation for the use of generative AI systems in the literary and artistic field); *idem*, *A Tax on Machines for the Purpose of Giving a Bounty to the Dethroned Human Author – Towards an AI Levy for the Substitution of Human Literary and Artistic Works*, SSRN, January 2022; G. Frosio, *Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?* [in:] *A Research Agenda for EU Copyright Law*, eds. E. Bonadio, C. Sganga, Cheltenham 2023.

the principles set out in the Directive on specific permitted uses of orphan works.[33] It is difficult, if not impossible, for the creator of AI to determine in advance in what fields of use the AI system will use the work. As mentioned earlier, AI systems require input (text, data, etc.) to create output, i.e. to learn. P. Bernt Hugenholtz goes so far as to suggest that "much of current and future AI development depends on TDM."[34] Given the current state and development of technology, it is impossible to clearly determine whether, within a few months, the sphere of use of the work will not include completely new, hitherto unknown fields of exploitation. Third, there is no doubt that the current trend and development of AI systems applies not only to the so-called "big players" and international IT corporations but also to small developing companies. The investment system for entities just starting out on the market is completely different from that of huge corporations with significant financial resources at their disposal. It seems that establishing a system of highly licensed exceptions could limit access to the technology market for entities with fewer financial resources. Fourth, viewing the use of AI in the context of society's access to a broad information framework is both an argument against the licensing system and an encouragement for a broad exception. The use of narrow data resources, limited to sets covered by license agreements, may result in the generation of limited and biased results by AI. The antidote to such algorithmic limitations would be a system of broad exceptions and restrictions on AI, allowing access to a broader pool of data unrestricted by the barriers to licensing identified above.

Conclusions

The rapid democratisation and emergence of GenAI has shed greater light on the reality that developing and refining AI is data-intensive. The more data, the better its quality, and therefore, the more robust and accurate AI. However, the development of AI encounters legal limitations and legal uncertainty, namely, the use of literary and artistic works and other materials protected by copyright and related rights, including computer programs and databases, by AI systems is still subject to objections from copyright holders.

This article examines copyright issues related to GenAI in general. Current legal frameworks, such as those in Japan, China, Singapore and fair use in the US and the TDM exemption in the EU, provide some guidance on using copyrighted material to train AI models. However, this framework may need to fully address the complexities inherent in GenAI. This is evidenced by new legislative actions taking place in France,

[33] Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, OJ L 299, 27.10.2012, pp. 5–12.

[34] P.B. Hugenholtz, *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, 2019, http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/ [accessed: 2023.10.31].

among other countries, but also by budding court disputes regarding the use of copyright content by OpenAI USA programs.

The analysis conducted supports the idea that establishing a new, appropriate and technologically evolving exception dedicated to AI in the field of copyright may provide many benefits. This concept, based on previous experience, is expected to have a positive impact on the industry. Clear rules and legal certainty can send a positive signal to the market and help avoid protests related to the development of AI. Legislative chaos does not inspire confidence in either program users or investors, and the sense of legal certainty and security is one of the fundamental principles of economic development. Certainly, regulating this issue would be consistent with the concept of promoting a culture of the rule of law that follows and adapts to the new information society. A situation in which the law does not offer a broad exception to the widespread development of AI-enabled technology is similar to one in which we would not have a fair use exception in the copyright system in the form of quotation rights. As evidenced by observations of the everyday world, the lack of proper calibration of a system of exceptions and limitations in copyright law does not hinder the development of technology, and what is more, it creates an image of law as an inflexible tool that is not adapted to a changing world.

Literature

Cao Y., Li S., Liu Y., Yan Z., Dai Y., Yu P., Sun L., *A Comprehensive Survey of AI-Generated Content (AIGC): A history of Generative AI from GAN to ChatGPT*, "Journal of the ACM" 2018, vol. 37, no. 4, https://arxiv.org/pdf/2303.04226.pdf.

Chen Y., *The Legality of Artificial Intelligence's Unauthorized Use of Copyrighted Materials Under China and US Law*, "The Intellectual Property Law Review" 2023, vol. 63, iss. 2.

Flynn S., Schirru L., Palmedo M., Izquierdo A., *Research Exceptions in Comparative Copyright*, PIJIP/TLS Research Paper Series no. 75, 2022.

Frosio G., *Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?* [in:] *A Research Agenda for EU Copyright Law*, eds. E. Bonadio, C. Sganga, Cheltenham 2023.

Gervais D., *Generative AI & IP: Gervais' checklist of issues*, https://www.linkedin.com/pulse/generative-ai-ip-gervais-checklist-issues-daniel-gervais/.

Guadamuz A., *French lawmakers propose new copyright about generative AI*, 24/09/2023, https://www.technollama.co.uk/french-lawmakers-propose-new-copyright-law-about-generative-ai.

Hugenholtz P.B., *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, 2019, http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/.

Kop M., *AI & Intellectual Property: Towards an Articulated Public Domain*, "Texas Intellectual Property Law Journal" 2020, vol. 28, no. 1.

Kop M., *The right to process data for machine learning purposes in the EU*, "Harvard Journal of Law & Technology" 2021, vol. 34.

Lemley M., Casey B., *Fair Learning*, "Texas Law Review" 2021, vol. 99, iss. 4, https://texaslawreview.org/fair-learning/.

Lucchi N., *ChatGPT: A case study on Copyright Challenges for Generative Artificial Intelligence Systems*, "European Journal of Risk Regulation" August 2023.

Marawala T., *Artificial Intelligence, Game Theory and Mechanism Design in Politics*, London 2023.

Margoni T., Kretschmer M., *A Deeper Look into the EU Text and Data Mining Exceptions: Harmonization, Data Ownership, and the Future of Technology*, "GRUR International" 2022, vol. 71, iss. 8.

Reese R., *Transformativeness and the Derivative Work Right*, "Columbia Journal of Law & the Arts" 2008, vol. 31, iss. 4.

Scharf N., *Digital Rights Management and Fair Use*, "European Journal of Law and Technology" 2010, vol. 1, iss. 2.

Senftleben M., *Generative AI and Author Remuneration*, "International Review of Intellectual Property and Competition Law" 2023, vol. 54.

Senftleben M., *A Tax on Machines for the Purpose of Giving a Bounty to the Dethroned Human Author – Towards an AI Levy for the Substitution of Human Literary and Artistic Works*, SSRN, January 2022.

Summary

Anna Bober-Kotarbińska

The Justification for Establishing Exceptions and Limitations to Copyright for Programs based on Artificial Intelligence

The issue of artificial intelligence ('AI') in the context of intellectual property law, including copyright law, has attracted continued interest. Progressive innovation brings new challenges, and the advances we have seen in recent years - particularly in the development of generative artificial intelligence ('GenAI') systems – are attracting media and public attention. The adoption and use of generative artificial intelligence systems has sparked widespread debate about their relevance to the copyright system. In the wake of emerging questions, copyright holders have begun to file copyright infringement lawsuits against artificial intelligence companies targeting the process of training artificial intelligence with the results obtained from generative artificial intelligence systems. As a result of these questions, copyright holders have begun filing copyright infringement lawsuits against owners of programs trained on the basis of data protected by copyright and data protection law. Drawing on analysed discussions, normative proposals, consultations and recommendations from experienced practitioners, this article identifies one of the broad questions of contemporary copyright policy towards artificial intelligence, concerning the legality of using copyrighted works to train artificial intelligence models. It also poses the question of the desirability of establishing a new system of copyright exceptions and limitations dedicated to artificial intelligence systems, while analysing the impact of existing limitations under copyright exceptions and limitations on the development of artificial intelligence.

Keywords: copyright, exceptions and limitations, artificial intelligence, generative artificial intelligence.

Streszczenie

Anna Bober-Kotarbińska

Zasadność ustanowienia wyjątków i ograniczeń prawa autorskiego dla programów bazujących na sztucznej inteligencji

Problematyka sztucznej inteligencji (AI) w kontekście prawa własności intelektualnej, w tym prawa autorskiego, cieszy się niesłabnącym zainteresowaniem. Postępujące innowacje niosą ze sobą nowe wyzwania, a postępy, które obserwowaliśmy w ostatnich latach – zwłaszcza w zakresie rozwoju systemów generatywnej sztucznej inteligencji (GenAI) – przyciągają uwagę mediów i opinii publicznej. Przyjęcie i wykorzystanie generatywnych systemów sztucznej inteligencji wywołało szeroko zakrojoną debatę na temat ich znaczenia dla systemu praw autorskich. W następstwie pojawiających się pytań posiadacze praw autorskich zaczęli kierować pozwy o naruszenie praw autorskich przeciwko właścicielom programów trenowanych na podstawie danych chronionych przepisami prawa autorskiego i prawa o ochronie danych osobowych. W świetle tych nakreślonych w nich wyzwań konieczne staje się kompleksowe i holistyczne podejście do kwestii praw autorskich związanych ze sztuczną inteligencją, uwzględniające zarówno legalność korzystania z danych wejściowych, jak i wyjściowych. Opierając się na przeanalizowanych dyskusjach, propozycjach normatywnych, konsultacjach i zaleceniach doświadczonych praktyków, niniejszy artykuł identyfikuje jedno z szerokiego spektrum pytań współczesnej polityki praw autorskich wobec sztucznej inteligencji, dotyczących legalności wykorzystywania utworów chronionych prawem autorskim do szkolenia modeli sztucznej inteligencji. Stawia się w nim również pytanie o celowość ustanowienia nowego systemu wyjątków i ograniczeń prawa autorskiego odnoszącego się do systemów sztucznej inteligencji, analizując jednocześnie wpływ istniejących ograniczeń wynikających z wyjątków i ograniczeń prawa autorskiego na rozwój sztucznej inteligencji.

Słowa kluczowe: prawo autorskie, wyjątki i ograniczenia, sztuczna inteligencja, generatywna sztuczna inteligencja.