Can copyright bring AI to its knees?
Dr. Maja Bogataj Jančič, LL.M., LL.M., Open Data and Intellectual Property Institute ODIPI

Summary: "Can copyright bring artificial intelligence to its knees? Which other circumstances may cause that the "making" of generative AI can dramatically change in the (near) future. This short paper presents potential challenges that copyright poses to the training of the machines on large amount of data. Different jurisdictions address these issues differently. In the USA the legality of these activities is tested in several court cases. Do gentlemen's agreements and pragmatic symbiosis known from the "search engines business model" provide sufficient basis and/or incentive for the business model of "making" generative AI business model as well?

"Can copyright bring artificial intelligence to its knees?". Professor Pamela Samuelson posed this question already more than a year and a half ago in her lecture *Copyright meets AI*, at the CITRIS and Banatao Institute.[1] In the lecture, Professor Pamela Samuelson addressed the question of whether "making" of the generative artificial intelligence, which requires processing a huge amount of data, pose problems from the point of view of copyright law or not? Is it a permissible to use copyright protected works as a training data in this machine training process? Does this kind of use of data qualifies as a fair use or does it represent reproduction prohibited by copyright law?

At that time, many legal cases were already being filled in the USA courts, initiated by authors or rights holders against the companies that own generative AI technology: a group of artists filed a lawsuit against Stability AI for alleged vicarious copyright infringement and violation of the Digital Millennium Copyright Act[2], Getty Images also launched a lawsuit against Stability AI, because it allegedly copied more than 12 million photographs without permission or compensation from the Getty's collection to train its AI.[3]

Despite these lawsuits the world was marveled by the incredible capabilities of generic AI: many were fascinated by the results produced by the Chat GTP and similar generative AI models[4] initiated by human prompts. Many experts were reserved and critical, pointing out the potential harmful consequences that these powerful tools will bring for society and individuals and planet.[5]

1 Samuelson, P., Copyright meets AI, a lecture at CITRIS and Banatao Institute, April 26, 2023, https://www.youtube.com/watch?v=6sDGIrVO6mo;
2 Andersen vs. Stability AI, available at: https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.1.0.pdf;
3 Getty Images vs. Stability AI, available at: https://fingfx.thomsonreuters.com/gfx/legaldocs/byvrlkmwnve/GETTY%20IMAGES%20AI%20LAWSUIT%20complaint.pdf;
4 https://chatgpt.com/auth/login, https://www.midjourney.com/home, https://openai.com/index/dall-e-2/;
5 Bengio, Y., Major tech leaders call for six-month pause on advanced AI development in open letter, https://betakit.com/yoshua-bengio-major-tech-leaders-call-for-six-month-pause-on-advanced-ai-development-in-open-letter/;

Questions whether the machines can be authors, were discussed and debated by academics and experts and in different forums[6] [7] [8] [9] [10] [11] [12], and AI experts were explaining that at this stage of the art of the technology the AI is just a tool: extremely powerful and getting exponential more powerful, but without self-awareness.[13] Not even two years have passed since the Chat GPT tool was put on the market and the climate has changed: initial enthusiasm has disappeared: the general public and especially the affected stakeholders and, consequently, also the legislators, are significantly more critical and demand that the order that has been shaken by this new powerful technology needs to be restored (whatever this means).

Authors, or more correct - the copyright-holders, especially the copyright-holders in the entertainment industry that want "a piece of the cake" - are very upset and very loud. In the EU, their demands are most loudly voiced by collective management organizations that anticipate new lucrative opportunities for their old business model. Their potential role is strengthened, accidentally or maybe even purposely, by those who think that the complex copyright problems in the generative AI that have potential to heavily influence the making of the generative AI, could be easily solved if collective management organizations were entrusted with the management of copyright clearance issues. Proponents of such solutions have a wrong idea of how collective management organizations work today, and are completely neglecting their too often very non-transparent operation, high costs, unfair distribution of collected funds and their history of inefficiencies. But that is another issue that is not part of this short paper even if it is extremely important.

AI is changing the way we live, do business, learn, entertain, shop, work, exercise … and it will drastically continue to change economic and social development. Many talks about the great benefits it carries, and just as many warn of the risks. I think we should rethink everything in the future.[14] The situation that currently only corporations decide what kind of generative artificial intelligence is being created is worrisome and does not address the real needs humanity and planet need. Many lawmakers are trying to regulate artificial intelligence, the EU being among the first to tackle the problem

6 Gervais, D. (2020). Is Intellectual Property Law Ready for Artificial Intelligence? GRUR International, 69(2), 117–118. Available at https://doi.org/10.1093/grurint/ikz025;
Yanisky-Ravid, S. in Velez-Hernandez, L. A. (2018). Copyrightability of Artworks Produced by Creative Robots, Driven by Artificial Intelligence Systems and the Originality Requirement: The Formality-Objective Model. Minnesota Journal of Law, Science & Technology, 19(1). Available at https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1437&context=mjlst,
Vezina, B. in Moran, B. (2020). Artificial Intelligence and Creativity: Can Machines Write Like Jane Austen? Creative Commons. Available at https://creativecommons.org/2020/08/10/can-machines-write-like-jane-austen/;
White, C. in Matulionyte, R. (2020). Artificial Intelligence Painting The Bigger Picture For Copyright Ownership. Australian Intellectual Property Journal Update, 30, 4. Pridobljeno na: http://sites.thomsonreuters.com.au/journals/2020/08/04/australian-intellectual-property-journal-update-vol-30-pt-4/, Bogataj Jančiči, M., Ali je lahko umetna inteligenca avtor avtorskega dela?;

7 Guadamuz, A. (2017). Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works. Intellectual Property Quarterly, 2017 (2), 169–186. Available at http://sro.sussex.ac.uk/id/eprint/66693/;

8 European Commission. (2020). Trends and Developments in Artificial Intelligence, Challenges to the Intellectual Property Rights Framework. Available at https://ec.europa.eu/digital-single-market/en/news/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights-0;

9 Yanisky-Ravid, S. in Velez-Hernandez, L. A. (2018). Copyrightability of Artworks Produced by Creative Robots, Driven by Artificial Intelligence Systems and the Originality Requirement: The Formality-Objective Model. Minnesota Journal of Law, Science & Technology, 19(1). Available at https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1437&context=mjlst,

10 Vezina, B. in Moran, B. (2020). Artificial Intelligence and Creativity: Can Machines Write Like Jane Austen? Creative Commons. Available at https://creativecommons.org/2020/08/10/can-machines-write-like-jane-austen/;

11 White, C. in Matulionyte, R. (2020). Artificial Intelligence Painting The Bigger Picture For Copyright Ownership. Australian Intellectual Property Journal Update, 30, 4. Pridobljeno na: http://sites.thomsonreuters.com.au/journals/2020/08/04/australian-intellectual-property-journal-update-vol-30-pt-4/;

12 Bogataj Jančiči, M., Ali je lahko umetna inteligenca avtor avtorskega dela? Pravo in umetna inteligenca : vprašanja etike, človekovih pravic in družbene škode , Inštitut za kriminologijo pri Pravni fakulteti, 2021, available at https://www.ipi.si/wp-content/uploads/2023/03/Ali-je-lahko-umetna-inteligenca-avtor-avtorskega-dela-zadnja-verzija-pred-objavo.pdf;

13 European Commission. (2020). Trends and Developments in Artificial Intelligence, Challenges to the Intellectual Property Rights Framework. Pridobljeno na: https://ec.europa.eu/digital-single-market/en/news/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights-0, see also Gervais, D. (2020). Is Intellectual Property Law Ready for Artificial Intelligence? GRUR International, 69(2), 117–118. Pridobljeno na: https://doi.org/10.1093/grurint/ikz025;

14 Žerdin, A., Umetna inteligenca ne sme biti zgolj posel. Delovati mora v skupno dobro, interview dr. Maja Bogataj Jančič, Sobotna Priloga, December 2, 2023;

comprehensively with the EU AI Act which has also provisions on copyright.[15] Lawmakers are eager to act and it seems that of the old mechanisms already available - the copyright law protection (in additional to privacy and personal data protection rights) seems to be seen as being able to serve as a convenient tool to rebalance the existing privileges.

Is "this mechanism" really correct to help to "put things in order"? Does training machines on data, or more precisely on content protected by copyright, constitute an act relevant to copyright law or not? Or is training machines on content just the same as reading? Different jurisdictions deal with these questions differently. In the United States, the principle of fair use applies, meaning that if the use fulfills the four fair use conditions (the purpose and character of your use, the nature of the copyrighted work, the amount and substantiality of the portion taken, andthe effect of the use upon the potential market.) users do not need to notify or request permission from the copyright holder.16 Whether training the generative AI models on copyrighted works falls under fair use has been tested in several court cases. [17] The EU regulates the issue differently: European copyright legal systems do not recognize the concept of fair use but regulate the balance in copyright with exceptions and limitations to exclusive copyrights. The European wide TDM exceptions were introduced by Articles 3 and 4 of the Directive on Copyright in the Digital Single Market (DSM Directive)[18] that the EU members implemented in their copyright laws. Currently the EU is still developing an opt out process in connection to the general exception that will drastically influence conditions of generative AI making in EU.[19] Even before the introduction of the DSM exceptions several EU countries already recognized TDM exceptions. Germany allowed TDM enabling under certain conditions, primarily for scientific research; France had specific provisions DM for public research organizations; and Ireland provided a TDM exception geared towards non-commercial research.[20] The United Kingdom was the first in the EU that provided for the TDM exception for non-commercial research purposes.[21] Japan introduced a broad TDM exception in 2009 and amended it in 2018. The exception allows the use of copyrighted data for testing and analysis but does not explicitly differentiate between legally and illegally accessed content. Furthermore, the exception does not apply if the use of data allows 'enjoyment' or benefits from the copyrighted content, aligning with the Berne Convention's 'three-step test'. The importance of lawful data access is underscored, and a warning against using pirated content is given, indicating AI developers could be liable for infringement if they neglect this duty of care.[22, 23]

It seems that the regulation of the question of whether machine training is or will be allowed for all purposes depends primarily on the attitude towards AI in society and the power of the old entertainment industry: societies that perceive AI as a solution to their future problems strive to build a more friendly environment.

---

[15] Keller, P., A first look ar the copyright relevant parts in the final AI Act compromise, available here: https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/;

[16] Mark Lemley, How Generative AI Turns Copyright Law Upside Down, *Science and Technology Law Review*, volume 25, no 2 (2024). Published 5 June 2024., Henderson, Peter and Li, Xuechen and Jurafsky, Dan and Hashimoto, Tatsunori and Lemley, Mark A. and Liang, Percy, Foundation Models and Fair Use (March 27, 2023). Stanford Law and Economics Olin Working Paper No. 584, Available at SSRN: https://ssrn.com/abstract=4404340 or http://dx.doi.org/10.2139/ssrn.4404340, Matthew Sag, Fairness and Fair Urse in Generative AI, 92 Fordham L. Rev. 1887 available here https://ir.lawnet.fordham.edu/flr/vol92/iss5/7/, Matthew Sag, Copyright Safety for Generative AI, 61 Hous. L. Rev, 295 (2023), available here https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4438593;

[17] Klosek, Katherine: Training Generative AI Models on Copyrighted Works Is Fair Use, URL: https://www.arl.org/blog/training-generative-ai-models-on-copyrighted-works-is-fair-use/;

[18] 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, https://eur-lex.europa.eu/eli/dir/2019/790/oj;

[19] Keller, P., Considerations for opt-out compliance policies by AI model developers, Open Future Policy Brief, 2024, available at https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf;

[20] Margoni, Thomas; Kretschmer, Martin: A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3886695

[21] https://www.legislation.gov.uk/ukpga/1988/48/section/29A;

[22] Stephens, Hugh: Japan's Text and Data Mining (TDM) Copyright Exception for AI Training: A Needed and Welcome Clarification from the Responsible Agency, URL: https://hughstephensblog.net/2024/03/10/japans-text-and-data-mining-tdm-copyright-exception-for-ai-training-a-needed-and-welcome-clarification-from-the-responsible-agency/;

[23] Senftleben, Martin: AI Act and Author Remuneration – A Model for Other Regions?;

The answer to the question whether copyright can bring AI its knees is far from clear. The copyright may turn out to be a much more fatal for the generative AI making after the *New York Times* filed a copyright infringement lawsuit against Microsoft and OpenAI in which along with other allegations, the *New York Times* claims that Microsoft and OpenAI are infringing copyright when they train their large language models (LLMs) on material copyrighted by the *Times*.[24]

In addition to the copyright hurdles there seems that at least two additional circumstances can turn out to be even more fatal for the development of the generic AI: one is an enormous arrogance of the owners of the AI technology that they do not even try to hide anymore and the other is a lack of pragmatic benefit for their opponents in this process of making generative AI that they had previously in the business of search engine business model.

The arrogant attitude is best demonstrated by the conflict that erupted when OpenAI used the Scarlett Johansson's voice without permission.[25] The actress said she was "shocked, angered and in disbelief" that the updated version of ChatGPT, which can listen to spoken prompts and respond verbally, had a voice "eerily similar" to hers. Open AI negotiated with her but she declined for personal reasons. This did not stop the OpenAI to proceed and the chief executive, Sam Altman, even braged about the connection when he posted on X on the day of the launch: "her" which is a reference to Johansson's signature role in the 2013 film Her. 2013 film Her.

The lack of pragmatic benefit is best explained by analyzing the current business model of search engines. It seems that search engines, at least in the EU, do not have a solid copyright legal basis for their operation. The exception for temporaryy reproduction cannot provide legal basis, because at least one condition, the condition that reproductions need to "have no independent economic significance« cannot be fulfilled.[26]

However, search engines continue to operate in the EU the same as elsewhere for over more than two decades. It seems that because of their »fundamental relevance to our 21st century life« and because »such a process is critical to economic and social prosperity« society and economy found a way to overcome a copyright hurdle. It is still diffcult, at least in the EU, to argue that their right to function is explicitly stated and protected in the EU legislation: there is no copyright exception in the EU copyright laws that would cover activities of search engines as fair use covers their activties under the US law. It seems that gentlements agreements substitute the absence of clear legal basis: the owners of content on the websites (the copyright - holders) allow that the search engines technology reproduce their every website (except those marked by robot.txt) and in return the content owned by copyright holders is found easier and quicker. In the economy of the attention when everybody is competing to be found this arrangement works efficiently and almost perfectly, the incentive for successful symbiosis is based on pragmatism: both sides benefit from this symbiosis. However, in the world of generic AI the content owners do not get any benefit, quite the contrary generative AI can produce content that competes with the original content of copyright holders. This gives copyright holders no incentive to tolerate copying their content to train the machines. Additionally, it seems to be more and more evident that the owners of the AI do not even respect the gentlemen's agreements based on the robot thx anymore.[27]

[24] https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf;
[25] Milmo, D., Scarlett Johansson's OpenAI clash is just the start of legal wrangles over artificial intelligence https://www.theguardian.com/technology/article/2024/may/27/scarlett-johansson-openai-legal-artificial-intelligence-chatgpt
[26] Searle, Nicola, Business Models and Copyright Reform: The Legal Business Model (May 13, 2020). Available at SSRN: https://ssrn.com/abstract=3599891 or http://dx.doi.org/10.2139/ssrn.3599891;
[27] https://www.businessinsider.com/openai-anthropic-ai-ignore-rule-scraping-web-contect-robotstxt;

Copyright may not bring AI to its knees. But the arrogant attitude of the AI technology owners and the lack of pragmatic benefit by the rightsholders may push the decision making into the direction that copyright will become a very powerful tool to tame generative AI. I am convinced that courts will not be influenced by these factors and will diligently evaluate all four conditions of fair use in the cases that they are currently deciding. Courts are bind by the rule of law that they will respect. Contrary to the judges, legislators follow different rules: they are inclined to listen voices of their demos: and demos in this case, especially the copyright industries, are very laud demanding their share at the table. They demand this regardless of the social cost that their demands can bring to others – namely the prohibitive increase of costs for making the generative AI.

How high the cost will be, will be determined by the winner in the clash between the old copyright industry, which has for years convinced legislators that the survival of the culture of nations depends on it, and the new AI industry, which's importance is inflated by the promises that it may bring to the progress of humanity.

Humanity and planet need a very reasonable and very knowledgeable legislators who will be able to balance interest of different stakeholders and not just these two industries and would have enough power to establish a new regime that would fulfill the promises of the AI and avoid risks. In this process copyright law would need to play a much more limited role and should definitely not be the one to bring AI to its knees.