Intellectual Property Law – University of British Columbia

# Copyright Issues of AI Generative Text and Image Models in Switzerland: An Input and Output Analysis

## 1. Introduction

Recently, the legal debate surrounding artificial intelligence has intensified, leading to numerous lawsuits against developers of generative AI systems, alleging copyright infringements.[1] The following cases highlight the relevance and timeliness of this issue.

In *Tremblay v. OpenAI Inc.*, plaintiffs claim that OpenAI used their copyrighted books without permission to train ChatGPT. In the class-action lawsuit *Silverman et al. v. OpenAI Inc.*, the plaintiffs argue that OpenAI unlawfully used copyrighted works to train ChatGPT, which is now able to generate summaries of their novels. Similarly, the *Getty Images Inc. v. Stability AI* case has drawn considerable attention. In this case, the well-known photo agency Getty Images alleges that Stability AI, the developer of the Stable Diffusion image-generation tool, unlawfully scraped twelve million images from its database. Getty Images claims that the generated images even reproduce its watermark, without any permission or compensation.[2]

## 2. Foundations of Artificial Intelligence

**Large language models** (LLMs) like ChatGPT are trained on extensive datasets with the goal of learning word probabilities and language patterns in order to predict the next word in a sequence based on preceding words.[3] These models are characterized by their ability "agere sine intelligere". This means they can perform complex tasks and produce remarkably accurate results despite lacking a comprehensive understanding of the underlying processes.[4]

[1] LUCCHI, p. 13 f.
[2] LUCCHI, p. 13 f.; GOLLMER.
[3] GOZALO-BRIZUELA/GARRIDO-MERCHAN, p. 1 f.; BROWN, et al., p. 4 ff.
[4] LUCCHI, p. 2.

**Generative AI diffusion models** create images by initially adding noise to an initial image until it becomes pure noise. This process is then reversed, gradually transforming the noise back into a detailed image. This process is controlled by a neural network that learns how to generate realistic image content from noise. As a result, diffusion models can generate complex images from text descriptions.[5] One example of a modern image generator is DALL-E, which can also produce images in various styles and perspectives.[6]

The training data used as input for training AI models plays a critical role.[7] However, one of the concerning issues with generative AI is that its development is often shrouded in secrecy. Only the companies, like Meta and OpenAI (if anyone), know the full extent and content of the texts on which these programs are trained.[8] In general, it is assumed that these programs are primarily trained using publicly accessible data. AI providers would need to obtain copyright-protected data in the traditional way if they wanted to use it for training, which can be very costly. Books are also of great value for AI tools, but these too are limited to publicly accessible works.[9] To gather the necessary data, automated data extraction tools (bots, web crawlers) are used to comb through the web.[10]

## 3. Input Analysis - Qualification of Training Data

This section aims to address whether copyright issues arise at the input level. To determine this, it's necessary to first examine the sources of the training data for AI programs and whether AI operators have obtained licenses for these sources or can potentially rely on an exemption.

### 3.1. Training Databases

A paper published by OpenAI serves as the primary source of information about the contents of GPT-3's training datasets. According to the developers, the model was trained on 45 terabytes of text data from various sources, amounting to approximately 500 billion tokens. The model was trained on data from five major sources: Common Crawl, WebText2, Books1,

[5] ZHANG/WANG et al., p. 1 f.; CARLINI/HAYES, et al, p. 2.
[6] GOZALO-BRIZUELA/GARRIDO-MERCHAN, p. 5.
[7] KREUTZER/SIRRENBERG, p. 78.
[8] REISNER.
[9] ZHANG, p. 4.
[10] CHERPILLOD, p. 445 f.

Books2, and Wikipedia.[11] The Common Crawl data consists of billions of web pages collected by the Common Crawl Foundation. Books1 and Books2 are two internet-based book corpora. The WebText2 dataset includes text from web pages linked on Reddit with more than three upvotes, while Wikipedia constitutes 3% of the database.[12] Books1 is a complete edition from Project Gutenberg, an online publisher of around 70,000 books whose copyrights have expired or are licensed for non-commercial distribution. The contents of Books2 remain unknown, though some speculate it includes collections of pirated books, such as from Library Genesis.[13] Concerns have therefore been raised about the possibility that OpenAI may have used copyrighted material to train its model.[14] Other models, like Stable Diffusion and Midjourney, were trained on the LAION-5B dataset, which contains nearly six billion tagged images scraped indiscriminately from the internet and is now known to contain a substantial number of copyrighted works.[15] In August 2023, *The Atlantic*, a U.S. magazine, revealed that a dataset containing 191,000 books had been used to train generative AI systems by Meta and Bloomberg. This dataset, "Books3," includes a collection of pirated e-books, including novels by Stephen King.[16]

It is therefore highly likely that many well-known AI programs have been trained, at least in part, with copyrighted material. The next step is to assess whether this constitutes copyright infringement. Only two scenarios in Swiss law might justify the use of copyrighted material in training: either the training process itself does not constitute a copyright-relevant action, or an exemption applies.[17] These two possibilities are examined below. Additionally, no copyright violation occurs once copyright protection has expired, which would need to be assessed on a case-by-case basis.[18]

### 3.2. Copyright Qualification of the Training Process

Whether the training of AI is copyright-relevant is disputed, as it doesn't directly result in the use or enjoyment of a work. However, Swiss copyright law generally subjects almost any use

[11] BROWN et al., S. 8 f.; ZHANG, p. 3 f.
[12] ZHANG, p. 1 ff.
[13] REISNER.
[14] ZHANG, p. 4.
[15] APPEL et al.
[16] REISNER.
[17] OEHRI.
[18] Art. 29 URG (Swiss Copyright Act).

of copyrighted works to the control of the copyright holder. For this reason, OEHRI argues that the extraction and storage of copyrighted texts and images in compressed and encrypted form constitutes a copyright-relevant reproduction under Swiss copyright law (URG).[19] CHERPILLOD also argues that the "tokenization" of text constitutes a temporary reproduction. However, once the text has been tokenized and no longer retains its original form, it is no longer considered a reproduction in the copyright sense, as the computer ultimately retains only tokens and the probabilities of connections between them.[20] Copyright-relevant reproduction includes data processing activities like loading into memory, displaying on a screen, or uploading to the cloud.[21] The copyright infringement therefore lies in the act of copying the data before it exists in tokenized form.[22] If texts or images are used to train AI programs, it is assumed that this constitutes a copyright-relevant reproduction under Art. 10, paragraph 2(a) URG under Swiss law.[23]

### 3.3. Exemptions

Swiss copyright law contains certain exemptions. Protected works may be used without the author's permission for private use under Art. 19 URG, but this does not include the training of AI tools. Nor does it cover temporary storage or reproduction under Art. 24a URG, as this provision only allows the reproduction of data for technically necessary, temporary storage during internet use (e.g., data in memory or on a computer hard drive). Another exemption is found in Art. 24d URG, which allows the use of copyrighted material for text and data mining for scientific research purposes. However, determining whether the Swiss exemption under Art. 24d URG applies is challenging, as the boundaries between research and science are difficult to define. Both CHERPILLOD and OEHRI reject the application of the exemption under Art. 24d URG, even though companies like OpenAI often present themselves as non-profit organizations. In many cases, organizations initially present themselves as non-profit research institutions but later derive commercial benefits from some applications.[24]

[19] OEHRI.
[20] CHERPILLOd, p. 446.
[21] Art. 10 URG; HILTY, Rz. 291 ff.
[22] CHERPILLOD, p. 446.
[23] CHERPILLOd, p. 446.
[24] CHERPILLOD, p. 447; OEHRI.

Since an exemption cannot be assumed, the conclusion is that using copyrighted material for training of generative AI constitutes copyright infringement.

### 4. Output Analysis

Another aspect to consider is whether copyright issues may arise at the output level. According to OpenAI's terms, users are responsible for ensuring that the content generated by AI does not infringe on third-party rights. This places users in a challenging position, as they cannot be certain whether the output generated for them has been directly copied from copyrighted material.[25] This raises the possibility of inadvertent copyright infringement, prompting questions about how AI systems function in the background. Specifically, it is worth investigating whether these systems merely replicate elements that may contain copyrighted content or if they generate something entirely original.

According to their developers, once these tools are trained, they theoretically no longer retain specific memories of individual documents or data sources used in the training set. Therefore, they are not supposed to access or retrieve specific documents. Instead, they only retain learned concepts, meaning the output should not match the original texts or images used for training. While this is technically true, WONG argues that generative AI programs can retain a surprising amount of information from the images in their training data—sometimes enough to nearly reproduce them perfectly.[26]

#### 4.1. Findings of Recent Studies

Several recently published studies challenge the assumption that text and image-generating AI tools cannot produce outputs identical to their inputs. The key findings from these studies, all examining the extraction of training data from well-known AI tools, are summarized below:

- One research team found that OpenAI models contain both public domain and copyrighted books in their data. The model's ability to reproduce content from these books is closely linked to the frequency with which passages appear online.[27]

[25] OEHRI.
[26] WONG; CHERPILLOD, S. 448.
[27] CHANG et al., p 1 ff.

- Another team of researchers presented an attack on GPT-2 to extract training data, showing that modern language models (LMs) can memorize their training data and that attackers can retrieve this data using simple techniques. Hundreds of verbatim text sequences, including personal contact information (names, phone numbers, and email addresses), IRC conversations, program codes, and 128-bit UUIDs (Universally Unique Identifiers), were extracted. These sequences were extractable even when they appeared in only a single document in the training data, indicating that LMs can memorize both frequently repeated and rare content, potentially exposing it later. The study also found that larger models are more susceptible to memorization than smaller ones. A critical insight is that the extent of content memorized by GPT-2 has been significantly underestimated, and it is likely that prompts could be used to uncover even more memorized content, including longer text passages, which could present copyright issues.[28]
- Another publication examining various open-source language models demonstrated that it is possible to extract very long strings of over 4,000 characters. Verbatim paragraphs from novels, exact copies of poems, excerpts from research papers, full summaries of publications, and bibliographic data from hundreds of papers were extracted. The study suggests that attackers with more resources could potentially access even more training data.[29]
- The main finding of yet another study was that diffusion models like DALL-E 2 can retain and reproduce images from their training data. It was shown that over a thousand training examples, including copyrighted logos, could be extracted. Images duplicated at least 100 times were the most likely to be reproduced. Additionally, the study found that modern diffusion models retain more information than earlier models, suggesting that generative image models may become more vulnerable over time. The study highlights the need for advancements in privacy-preserving training.[30]

[28] CARLINI/TRAMER et al., p. 1 ff.
[29] NASR/CARLINI et al., p. 1 ff.
[30] CARLINI/HAYES et al, p. 1 ff.

These publications are highly relevant to copyright discussions, as they demonstrate that original data (including copyrighted texts and images) contained in training data can be reproduced verbatim. This suggests the potential for long-term retention or storage, or at least the capacity for nearly perfect reconstruction. The true extent of memorization remains unclear and is left open in all studies, though it is estimated to be far greater than previously assumed.

### 4.2. Legal Qualification

Since the studies mentioned above are quite recent, there has been little discussion in the literature regarding their potential impact on copyright law. However, if an AI tool generates output that closely resembles the original training data, this may lead to copyright infringement. In such cases, the copyright aspects to consider would be reproduction (Art. 10 para. 2 lit. a URG), distribution (Art. 10 para. 2 lit. b URG), and possible infringement of the right of attribution (Art. 9 para. 1 URG).

The author has a moral right to attribution under Art. 9 para. 1 URG, which includes the right to choose the form of attribution (name, initials, pseudonym, or anonymity) and to be credited whenever the work is used.[31] If the AI-generated output reproduces a copyrighted work, or parts of it, without acknowledging the author, this could violate Art. 9 para. 1 URG, as it is similar to plagiarism, where citations or sources must be indicated under Art. 25 para. 2 URG.

The author's economic rights allow them to achieve commercial or non-commercial success with their work and to decide if, when, and how the work is used. Usage occurs when works are reproduced, distributed, or otherwise made perceptible, as stated in Art. 10 URG.[32] If it is possible to reconstruct the original image or text after processing, this could be considered a permanent, copyright-relevant reproduction, as it essentially creates a copy of the work.[33] This also constitutes distribution if the copies are made available to new audiences. The author's consent would be required, as these rights exclusively belong to the copyright holder.

[31] HILTY, Rz. 381 ff.
[32] HILTY, Rz. 291 ff.
[33] CHERPILLOD, p. 446.

However, if only the stylistic technique is replicated, it does not constitute copyright infringement, as style itself is not protected by copyright.[34]

### 5. Summary and Conclusion

The literature analysis reveals that, regarding input, the training process of AI tools may indeed infringe copyright, as it involves the legally significant act of reproduction and distribution. Technically, these tools should not be capable of reproducing content identically. However, recent examples such as the Getty Images case and research studies indicate that AI programs can replicate parts or even entire works. Therefore, the second question—whether copyright infringement may also occur at the output level—can be answered affirmatively. The reproduction of original data constitutes a copyright-relevant act of duplication and distribution. Additionally, there could be a violation of the copyright holder's right to recognition.

## References

All internet sources cited in this paper were last accessed on November 5th, 2024.

APPEL GIL / NEELBAUER JULIANA / SCHWIDEL DAVID, Generative AI Has an Intellectual Property Problem, Harvard Business Review 2023, available at: <https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem> (cited as: APPEL et al.).

BROWN TOM / MANN BENJAMIN / RYDER NICK / SUBBIAH MELANIE / KAPLAN JARED / DHARIWAL PRAFULLA, et al., Language Models Are Few-Shot Learners, Advances in Neural Information Processing Systems 33 (NeurIPS) 2020, pp. 1877–1901 (cited as: BROWN et al.).

CARLINI NICHOLAS / HAYES JAMIE / NASR MILAD / JAGIELSKI MATTHEW / SEHWAG VIKASH / TRAMER FLORIAN / BALLE BORJA / IPPOLITO DAPHNE / WALLACE ERIC, Extracting Training Data from Diffusion Models, in: Joseph A. Calandrino, Proceedings of the 32nd USENIX Security Symposium, Anaheim 2023, p. 5253–5270 (cited as: CARLINI/HAYES et al.).

CARLINI NICHOLAS / TRAMER FLORIAN / WALLACE ERIC / JAGIELSKI MATTHEW / HERBERT-VOSS ARIEL / LEE KATHERINE / ROBERTS ADAM / BROWN TOM / SONG DAWN / ERLINGSSON ULFAR / OPREA ALINA / RAFFEL COLIN (2020) Extracting Training Data from Large Language Models, arXiv, available at: <https://arxiv.org/abs/2012.07805> (cited as: CARLINI/TRAMER et al.).

[34] CHERPILLOD, p. 448.

CHANG KENT / CRAMER MACKENZIE / SONI SANDEEP / BAMMAN DAVID (2023) Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4, arXiv, available at: <https://arxiv.org/abs/2305.00118> (cited as: CHANG et al.).

CHERPILLOD IVAN, Intelligence Artificielle et Droit d'Auteur, sic! 2023, p. 445–452.

GOLLMER PHILIPP, Wem gehört das Chat-GPT-Gedicht? NZZ, April 29, 2023.

GOZALO-BRIZUELA ROBERTO / GARRIDO-MERCHAN EDUARDO, ChatGPT Is Not All You Need: A State of the Art Review of Large Generative AI Models, Global Review of AI Community Ethics 1/2023.

HILTY RETO, Urheberrecht, 2nd ed., Bern 2020.

KREUTZER RALF / SIRRENBERG MARIE, Künstliche Intelligenz verstehen, Wiesbaden 2019.

LUCCHI NICOLA, ChatGPT, A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems, European Journal of Risk Regulation 2023, p. 21–23.

NASR MILAD / CARLINI NICHOLAS / HAYASE JONATHAN / JAGIELSKI MATTHEW / COOPER FEDER / IPPOLITO DAPHNE / CHOQUETTE-CHO CHRISTOPHER / WALLACE ERIC / TRAMER FLORIAN / LEE KATHERINE (2023) Scalable Extraction of Training Data from (Production) Language Models, arXiv, available at: <https://arxiv.org/abs/2311.17035> (cited as: NASR/CARLINI et al.).

OEHRI ISABELLE, ChatGPT & Co. – Was sagt das Urheberrecht? Management and Law Blog HSLU 2023, available at: <https://hub.hslu.ch/management-and-law/2023/04/05/chat-gpt-co-was-sagt-das-urheberrecht/>.

REISNER ALEX, Revealed: The Authors Whose Pirated Books Are Powering Generative AI, The Atlantic, August 19, 2023.

WONG MATTEO, Artists Are Losing the War Against AI, The Atlantic, October 2, 2023.

ZHANG DONG (2023), Should ChatGPT and Bard Share Revenue with Their Data Providers? A New Business Model for the AI Era, arXiv, available at: <https://arxiv.org/abs/2305.02555>.

ZHANG TIANYI / ZHENG WANG / HUANG JING / MOHIUDDIN TASNIM / SHI WEI (2023), A Survey of Diffusion Based Image Generation Models: Issues and Their Solutions, arXiv, available at: <https://arxiv.org/abs/2308.13142> (cited as: ZHANG/WANG et al.).