WilhelmGilliéron
AVOCATS

PHILIPPE GILLIÉRON
Avocat

Auteur: Wilhelm Avocats | Le : 30 January 2024

Artificial intelligence (AI) and copyright: an overview of recent cases

Since 2023, AI has been monopolizing discussions and monopolizing the spoken word, sweeping everything in its path like a tidal wave.

While the questions are multiplying, the answers are few and far between. The development and use of these tools has become a priority for companies, who rightly see it as a turning point that cannot be ignored. However, a governance framework is needed to provide answers to these many questions.

A. European regulations on artificial intelligence

In the coming weeks, I propose to take a tour of what has gone from being a "Proposal" to becoming the European AI Regulation, now known since it "leaked" on January 22, 2024.

Following the same structure, and at the rate of one publication per week, we will cover in turn: (1) general provisions; (2) general-purpose AI systems; (3) prohibited practices; (4) high-risk practices; (5) transparency obligations; (6) measures to encourage innovation; (7) governance; (8) the register; (9) monitoring and surveillance obligations; (10) codes of conduct; (11) sanctions; and finally (12) issues relating to the delegation of powers.

While we wait for the review of these regulations to begin next week, let's take a look this week at the current state of copyright issues raised by these systems, in the light of current legal cases.

B. NY Times v. OpenAI and Microsoft

The latest, [the action brought on December 27, 2023 before the *United States District Court Southern District of New York* by the NY Times](#) not only against OpenAI, but also against Microsoft, is undoubtedly the most emblematic since, for the first time, a major player is taking the case to court.

As with many other current actions, the NY Times is claiming massive infringement of its rights; with a 69-page application, the NY Times demonstrates through numerous examples that the results generated by ChatGPT lead in many cases, on the basis of the simplest prompts, to an almost complete reproduction of certain published articles.

In this case, therefore, it's not just the *input* data that's at stake – cases in which the issue of *fair use* is undeniably acute in the USA – but the result itself (*output*).

When it comes to *output*, however, the *fair use* exception appears far less convincing, at least in this case. The NY Times demonstrates with numerous examples that, in many cases, these results largely reproduce its articles.

For the NY Times, the risk is high that Internet users will give up their subscriptions to the daily newspaper – subscriptions which, as the NY Times points out, are invaluable in maintaining journalistic quality at a time when misinformation has become legion. While the NY Times is primarily concerned, the future of the press as a whole is at stake in this case.

On the face of it, it is hard to see how the conditions for the exercise of *fair use* could be met, and how, in the light of the claim, and at this stage at least, the defendants could escape recognition of their infringement of NY Times copyright.

On [January 8, 2024, OpenAI issued a lengthy letter](#) rejecting OpenAI's spurious allegations. Unsurprisingly, it points out that, notwithstanding authors' right to *opt out*, the entrainment of protected data is a case of *fair use*. She adds that the examples cited in the application are in fact only rare cases reflecting prompts with a deliberate intention to manipulate the system. To be continued.

C. Business in progress

Since last year, there has been a flurry of cases dealing with copyright issues, both in terms of *input* data and, to a lesser extent, *output*.

(I) Infringement of copyright on training data (input)?

In the United States, cases involving training data have included, but are not limited to, the following:

- January 13, 2023: *[Sarah Andersen et al. vs Stability AI, Midjourney Inc, Deviantart, Inc](#)* (*United States District Court Northern District of California*). In this case, which took the form of a *class action*, the plaintiffs, artists, argued not only that the defendants' systems violated their rights during training by reproducing their works as training data and removing the digital tattoo associated with their works, but also that the result of using these tools necessarily constitutes a derivative work, since this result derives from the compilation and aggregation of a multitude of images which are all protected by copyright, making these tools "*a 21st-century collage tool*" (§§ 90 et seq. of the application).

On October 30, 2023, following a "*motion to dismiss*" filed by the defendants, Judge William Orrick expressed the utmost doubt as to the merits of the plaintiffs' claim. In the absence of registration of the works with the *United States Copyright Office*, a prerequisite in the USA for bringing an action for copyright infringement, he dismissed all the claims except that of Sarah Andersen to the benefit of such registrations. Expressing grave doubts as to whether the results generated could constitute a derivative work, given the model's 5-billion-plus track record, the Judge gave the plaintiff time to amend her claim and provide further clarification as to the alleged infringement of her rights.

On November 30, 2023, Sarah Andersen filed an amended application that appears to convincingly establish the reproduction of protected images to which she (and her co-applicants) hold the rights, as well as the fact that in certain assumptions, the generated results could indeed be considered as derivative works otherwise no longer mentioning the attribution of rights (original work on the left, generated result on the right) :

- February 3, 2023: *Getty Images vs Stability AI* (*United States District Court for the District of Delaware).* Action similar to that brought on January 17, 2023 before the High Court of Justice of London (see below).

- In 2023, a number of similar actions were brought, often in the form of *class actions*, in which authors claimed copyright infringement resulting from the reproduction of their works for training purposes, and sometimes infringement of their

rights in the results generated, such as: *Paul Tremblay and Mona Awad v. OpenAI* (*United States District Court Northern District of California*, June 28, 2023); *Kadrey v. Meta and Silverman v. OpenAI* (*United States District Court Northern District of California*, July 7, 2023, where it should be noted that the Judge dismissed the claim against Meta on November 20, 2023, on the grounds that the plaintiff could not conclude that any result necessarily constituted a derivative work of his simply because his work had been reproduced as training data); *J.L. v. Alphabet* (*United States District Court Northern District of California*, July 11, 2023); *Chabon v. OpenAI & Chabon v. Meta* (*United States District Court Northern District of California,* September 12, 2023); *Authors Guild v. OpenAI Inc*. (*United States District Court Northern District of California,* September 19, 2023, where plaintiffs claim, among other things, that prompts could be used to generate sequels to their works or produce detailed summaries of them); *Huckabee v. Meta* (*United States District Court Southern District of New York,* October 17, 2023); *Concord Music Group, Inc. v. Anthropic PBC* (*United States District Court for the Middle District of Tennesse*, October 18, 2023, where the plaintiffs argue that the system would be trained on the basis of the lyrics of music that is, of course, protected, suggesting as generated results music with lyrics infringing their copyrights); *Sancton v. OpenAI* (*United States District Court Southern District of New York,* November 21, 2023).

- The most recent case to our knowledge was *Nicholas Basbanes and Nicholas Gage v. Microsoft, OpenAI et al* (*United States District Court Southern District of New York*), dated January 5, 2024. The plaintiffs, journalists by profession, accused the defendants of reproducing their works for training purposes without their consent, without the case presenting any particularity whatsoever.

Outside the United States, we can also mention :

- January 17, 2023: *Getty Images v. Stability AI (High Court of Justice of London)*, in which Getty Images criticized Stability AI not only for exploiting its image database to drive its system, without having sought a license as other players have apparently done, but also for infringing its rights with regard to the images generated by the tool, which substantially reproduced images in its database.

On December 1[er] 2023, Justice Joanna Smith dismissed two applications from Stability AI, one of which sought a preliminary declaration that the Court had no jurisdiction to hear the action for infringement of copyright in the training data, on the grounds that Stability AI had allegedly been developed in the USA. Considering that there was some doubt as to whether Stability AI had seen its model trained on servers located in the UK, Justice Joanna Smith decided to take up the action, while suggesting that if the training had indeed only taken place in the USA, then Getty Images' action would be dismissed.

(II)   Copyright on output?

- On March 16, 2023*,* the *United States Copyright Office* launched an initiative to examine whether the results generated by these tools merited copyright protection. In its guidelines published the same day, the USCO notes that protection can only be granted if there is a *human contribution*, i.e. "*[...] an author's own original mental conception, to which the author gave visible form. The answer will depend on the circumstances, particularly how the AI tool operates and how it was used to create the final work*."

For the USCO, the result generated on the basis of a simple prompt (*one shot prompting*) does not satisfy the requirement of sufficient human contribution to the generated result; in general, the USCO assimilates prompts to instructions communicated to the artist in charge of creating a work. When all the expressive elements of the result are determined by the tool, the necessary human contribution is lacking, and the result cannot be registered with the USCO. In any case, the use of a generative tool must be transparently disclosed.

Continuing its momentum, on August 30, 2023 the USCO published a call for public comment aimed at obtaining as much feedback as possible on a host of issues relating to the challenges posed by artificial intelligence systems in copyright law, ranging from training data to the results generated, via transparency obligations and possible solutions, whether legal such as the introduction of a right to remuneration, extended collective licensing, or technical such as digital watermarking. The
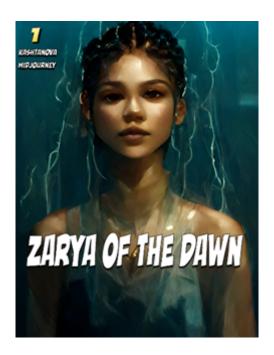
public consultation was due to close at the end of 2023, and it will be interesting to see the results.

- Over the course of 2023, the USCO has repeatedly rejected applications to register results generated by generative tools, on the grounds that human intervention was insufficient to consider that it had played a role in making the work appear to be that of the user.

This was the case for the following results:

A Recent Entrance to Paradise (February 2022)

Zarya of the Dawn (February 2023)

Space Opera Theater (rejected notwithstanding the fact that the painting, which had won first prize at the *Colorado State Fair* in September 2022, had required over 80 hours of work and thousands of prompts. The author had argued that MidJourney had been used as a brush-like tool, an argument that did not convince the USCO).

Suryast (December 11, 2023, the work on the left being a photograph of a sunset taken by Indian artist Sahni, the work in the middle representing a work by Van Gogh whose style was claimed by Sahni to transform his work and generate the result on the right. Human contribution deemed insufficient by the USCO)

- The USCO's severity was confirmed by the *United States District Court for the District of Columbia,* which, in a decision handed down on August 18, 2023 in the case of [*Thaler v. Perlmutter, upheld the USCO's decision of February 14, 2022. Perlmutter*](#), upheld the [USCO's February 14, 2022 ruling](#) that the human contribution was insufficient to consider the painting *A Recent Entrance to Paradise* copyrightable. Stephen Thaler nevertheless [appealed this decision on October 11, 2023](#), and it will be interesting to see how the *United States Court of Appeals for the District of Columbia Circuit* decides this case.

- Other jurisdictions, however, have been more open to recognizing that the sequence of prompts could lead to sufficient human input to give the result an individual character that the use of the tool alone would not have achieved, whether at the level of intellectual property offices or the courts.

This was the case with intellectual property offices in both India and Canada, where these offices agreed to register the work entitled *Suryast* by the aforementioned artist Sahni. The same was true in South Korea, where the company Nara Knowledge Information obtained registration for its film "AI Suro's Wife", for which it had used the Midjourney and Stable Diffusion tools. Here too, the Korean office considered that the selection and arrangement of images in the course of editing the film demonstrated sufficient individual human contribution.

And so it has been judicially this time in China, where on November 27 2023 the *Beijing Internet Court* admitted an action for copyright infringement brought on the basis of a work generated using Stability Diffusion, recognizing that the sequence of prompts and the resulting work testified to a work requiring rigorous selection and arrangement of both the prompts and the results generated over the course of these prompts :

D. Conclusion

What can we conclude from this overview of the situation at the start of 2024?

(I) Training data

It's astonishing that OpenAI should take offence at the criticisms levelled at it, when it is clearly seeking partnerships with major players capable of providing it with a massive flow of data.

Thus, in July 2023, it signed a licensing agreement with both Associated Press and Shutterstock, enabling it to exploit their content. On December 13, 2023, it did the same with publishing giant Axel Springer (and we wonder whether it's shooting itself in the foot, even if the terms of the transaction are obviously unknown to us).

In any case, it's hard to see why OpenAI would sign such agreements if it didn't recognize the illicit nature of the reproductions it makes of these data to drive its model.

In any case, the cases currently underway already testify to the procedural difficulties that plaintiffs in such actions are likely to encounter in winning their cases.

The debate is not limited solely to the question of whether these training data infringe the rights of the owners, and whether the developers of these generative tools can claim an exception. First and foremost, the court seized must have jurisdiction to do so. While most of the cases have been brought in the United States, where the developers are based and the question of jurisdiction does not arise, the issue is different when the owners wish to act abroad. The *Getty Images* case pending before the *High Court of Justice of London* underlines the fact that the jurisdiction of this Court is an open question, on which the *High Court* will have to rule. Little examined to date, aspects of private international law should not be underestimated.

Secondly, assuming that the Court has jurisdiction, the plaintiffs must have active legitimacy, i.e. they must demonstrate that works in which they hold rights have actually been reproduced for use as training data; a mere allegation that such exploitation appears more than likely in view of the number of images reproduced does not seem sufficient. The bar could therefore prove particularly high.

We can be sure that solutions will be found in the years to come, for example in the form of remuneration rights or extended collective licenses, without necessarily going so far as to offer a blank check for such practices, as Japan, Singapore or even possibly Israel have chosen to do.

(II)   Results generated

Although Ukraine opted in 2023 for the introduction of a *sui generis* right protecting the results generated by generative tools, this choice seems to remain an exception. More generally, the decisive question is whether the user has made a sufficient contribution to the sequence of prompts and their arrangement to be considered an "author".

While this criterion seems to be the rule, we have to admit that its interpretation varies widely from one country to another:

The USA is setting the bar particularly high at the moment; if 80 hours of work and thousands of prompts weren't enough to convince the USCO that Jason Michael Allen's contribution was sufficient to be considered the author of *Theatre of Space Opera*, one wonders what level the USCO expects.

China, on the other hand, seems generous in admitting copyright protection for an image of a young woman, although it is questionable whether it is truly sufficiently "original" in the sense required in principle by copyright.

So it all depends on where you want to place the cursor to recognize sufficient causality between the user and the generated result to see in the former the author of the latter.

The fundamental question of whether such results deserve protection in view of the raison d'être of copyright, i.e. to encourage creation, should also be asked. Should I be granted such rights to an "artistic" result, despite the fact that I have no real skills to do so, and that my only qualities lie in the ability to do effective prompting?

Could prompting be considered the equivalent of a paintbrush, as Jason Michael Allen argued, an argument the USCO refused to hear? But wouldn't this be taking the risk of levelling copyright downwards, by conferring such rights on the vast majority of users? What's more, wouldn't this be tantamount to levelling down the required level of creativity, by ultimately granting copyright over millions of works generated every day? How high should the bar be set? These are all questions that deserve to be asked, and to which no definitive answers have yet been found (for a discussion of these issues, see Damian Flisak's interesting contribution).