Kluwer Copyright Blog

Generative AI: admissibility and infringement in the two US class actions against Meta's LLaMA

Gianluca Campus (University of Milan) · Wednesday, January 17th, 2024

The two US class actions against Meta

We have previously analysed US class actions against Open AI (here) and Google (here) for unauthorized use of copyright works in the training of generative AI tools, respectively ChatGPT, Google Bard and Gemini. To further develop this excursus on the US case law, in this post we consider two recent class actions against Meta launched by copyright holders (mainly book authors), for alleged infringement of IP in their books and written works through use in training materials for LLaMA (Large Language Model Meta AI). Such case law is interesting for the reconstruction of the technology deployed by Meta and the training methodology (at least from the plaintiff's perspective) but also because the court has had the chance to preliminarily evaluate the robustness of the claims. Given the similarity of the legal arguments and the same technology being at stake (Meta's LLaMA), upon the request of the parties, the Court treated the two class actions jointly (here).

Image via Pixabay

The plaintiffs' factual allegations in *Kadrey v Meta* and *Chabon v Meta*

The first class action, *Kadrey v Meta* (here), was filed on 7 July 2023, in U.S. District Court for the Northern District of California – San Francisco Division. The second class action, *Chabon v Meta,* was filed on 12 September 2023 before the same court (here). Both complaints are essentially based on the same arguments and factual allegations.

The plaintiffs are authors of books and did not consent to their use as training material for Meta's AI product, LLaMA. LLaMA is a large language model in the form of an AI software program designed to emit convincingly naturalistic text outputs in response to user prompts. Rather than being programmed in the traditional way, a large language model is "trained" by copying massive amounts of text and extracting information from it. This body of text is called the training dataset. A large language model's output therefore entirely and uniquely relies on the material in its training dataset. Thus, the decisions about what textual information to include in the training dataset are deliberate choices.

According to the plaintiffs, much of the material in Meta's training dataset came from copyrighted works – including works written by the plaintiffs – that were reproduced by Meta without consent, credit, and compensation. This was despite Meta having declared that the training dataset was a large quantity of textual data that was publicly available and compatible with open sourcing.

Such declarations are included in Meta's Paper "*LLaMA: Open and Efficient Foundation Language Models*" (the "Paper" available here) and are considered by the plaintiffs to be inconsistent with the table describing the composition of the LLaMA training dataset. In the Paper, Meta notes that 85 gigabytes of the training data come from a category called "Books." Meta further elaborates that "Books" comprises the text of books from two internet sources: (1) Project Gutenberg, an online archive of approximately 70,000 books that are out of copyright, and (2) the Books3 section of ThePile, a publicly available dataset for training large language models.

Meta's Paper on LLaMA does not further describe the contents of Books3 or ThePile. ThePile is a dataset assembled by a research organization called EleutherAI. In December 2020, EleutherAI introduced this dataset in a paper called "*The Pile: An 800GB Dataset of Diverse Text for Language Modeling*" (here). The EleutherAI paper reveals that the Books3 dataset comprises 108 gigabytes of data, or approximately 12% of the dataset, making it the third largest component of ThePile by size.

Books3 is a dataset of books derived from a copy of the contents of the "*Bibliotik private tracker*". According to the plaintiffs, *Bibliotik* is one of a number of notorious "shadow library" websites that have long been of interest to the AI-training community because of the large quantity of copyrighted material they contain and many of the plaintiffs' written works appear in the Books3 dataset.

Since the launch of the LLaMA language models in February 2023, Meta has made these models selectively available to organizations that request access under a non-commercial license focused on research use cases (access to the models to be granted on a case-by-case basis to academic researchers, organizations in government, and industry research laboratories) but in March 2023 the LLaMA language models were leaked onto a public internet site and have continued to circulate. Moreover, in the Summer of 2023 Meta released in open source the next version of LLaMA ("LLaMA 2" here) as commercially available.

The Cause of Action

The cause of action in both cases is the same and can be summarized as follows:

1. **Direct Copyright Infringement** (17 U.S.C. § 106, et seq): the plaintiffs never authorized Meta to make copies of their works and derivative works, publicly display copies (or derivative works), or distribute copies (or derivative works) during the training process of the LLaMA language models. LLaMA language models cannot function without the expressive information extracted from the alleged infringed works and the LLaMA language models are themselves infringing derivative works.

2. **Vicarious Copyright Infringement** (17 U.S.C. § 106): because the output of the LLaMA language models is based on expressive information extracted from the plaintiffs' works, every output of the LLaMA language models is an infringing derivative work. Meta has the ability to control the output of the LLaMA language models. Meta has benefited financially from the infringing output of the LLaMA language models. Therefore, every output from the LLaMA language models constitutes an act of vicarious copyright infringement.

3. **Digital Millennium Copyright Act ('DMCA') – Removal of Copyright Management Information** (17 U.S.C. § 1202(B)): the plaintiffs included one or more forms of copyright-management information ('CMI') in each of their works, including: copyright notice, title and other identifying information, or the name or other identifying information about the owners of each book, terms, and conditions of use, and identifying numbers or symbols referring to CMI. Meta used them as training data for the LLaMA language models and, by design, the training process does not preserve any CMI.

Meta's motion to dismiss

On 18 September 2023, Meta filed a motion to dismiss (here) (for the notion of this US procedural mean, see here). More specifically, they argued the following points:

1. **Direct copyright infringement**: the plaintiffs' claim for direct infringement is premised on a theory that LLaMA is itself an infringing "derivative" work. This is supported by a single allegation: that LLaMA "*cannot function without the expressive information extracted from Plaintiffs' Works and retained inside* [it]." Meta has argued that the plaintiffs do not explain what "information" this refers to – the mere use of "information" from a copyrighted text is not the standard for infringement. From Meta's point of view, the only pertinent question is whether the software comprising LLaMA is, itself (more specifically, in terms of outputs), substantially similar as protected expression to the plaintiffs' books.

2. **Vicarious copyright infringement**: the plaintiffs seek to hold Meta vicariously liable for purportedly infringing outputs generated by others using LLaMA. Yet, they do not identify a single output ever generated by anyone that supposedly infringes their books. Instead, the plaintiffs advance the fallacy that every output generated using LLaMA is "*based on expressive information extracted from*" the plaintiffs' books and, therefore an "*infringing derivative work*" of each of those books.

3. **DMCA**: the plaintiffs allege that Meta provided false CMI in violation of 17 U.S.C. § 1202(a)(1) by asserting copyright in the LLaMA models. However, such claims are only actionable where the allegedly false CMI is included in an exact copy of a work. The plaintiffs' CMI was never included, much less intentionally removed by Meta with wrongful intent. The plaintiffs' allegations fail to state a claim under the DMCA.

Court's order on the motion to dismiss

On 20 November 2023, Judge Vince Chhabria issued his order on Meta's motion to dismiss (here). In the Judge's view, Meta has moved to dismiss all claims except the one alleging that the unauthorized copying of the plaintiffs' books for purposes of training LLaMA constitutes copyright infringement. On this basis, the Judge granted the dismissal, recognizing that the remaining theories of liability, at least as articulated in the complaint, were not admissible. More specifically, Judge Chhabria challenges the argument according to which the LLaMA language models are themselves infringing derivative works because the models cannot function without the expressive information extracted from the plaintiffs' books. A derivative work is "*a work based upon one or more preexisting works*" in any "*form in which a work may be recast, transformed, or adapted.*" (17 U.S.C. § 101) but the plaintiffs offered no evidence that the LLaMA models themselves can be considered recasting or adaptation of any of the plaintiffs' books.

With regards to the vicarious liability argument, the complaint offered no allegation of the contents of any output to support the statement that "*every output from the LLaMA language models constitutes an act of vicarious copyright infringement.*" To prevail on a theory that LLaMA's outputs constitute derivative infringement, the plaintiffs would have had to allege and ultimately prove that the outputs incorporate in some form a portion of the plaintiffs' books.

Also, the plaintiffs' DMCA claims were dismissed, because there were no facts to support the allegation that LLaMA ever distributed the plaintiffs' books, much less did so without their CMI.

Conclusion

As compared to the other class actions against Open AI and Google, this class action has arrived to a more mature step. The Judge concentrated on the very core of the copyright issue in the generative AI tools – their alleged training via resources made public on the internet and/or protected under copyright laws.

These class actions remain a timely occasion for parties to clarify and judges to assess legitimacy of generative AI tools based on a deep analysis of the technical functioning and composition of the training datasets. The main point to be assessed would likely be whether (or to what extent) the training process can benefit from an exception or limitation to copyright rules, such as for text and data mining or, where applicable, the fair use doctrine (see more here and here).

It is worth noting that on 9 December 2023 representatives of the European Parliament, EU member states and the European Commission reached a provisional agreement on the proposed AI Act (here). In this context, a newly introduced article on "*Obligations for providers of general-purpose AI models*" was proposed, with two distinct requirements related to copyright: (i) Section 1(c)[1] requires providers of GPAI models to: "*put in place a policy to respect Union copyright law in particular to identify and respect, including through state of the art technologies where applicable, the reservations of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790*" and (ii) Section 1(d) requires them to: "*draw up and make publicly available a*

sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office".

All in all, these discussions demonstrate that the main controversies are the transparency of the training data (with the aim of clarifying the access to legitimate resources) and the respect of the reservation of rights by copyright holders where there is an alleged recourse to the text and data mining exception under Article 4 of Directive (EU) 2019/790.

_____

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe here.

Kluwer IP Law

The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?

Learn how **Kluwer IP Law** can support you.
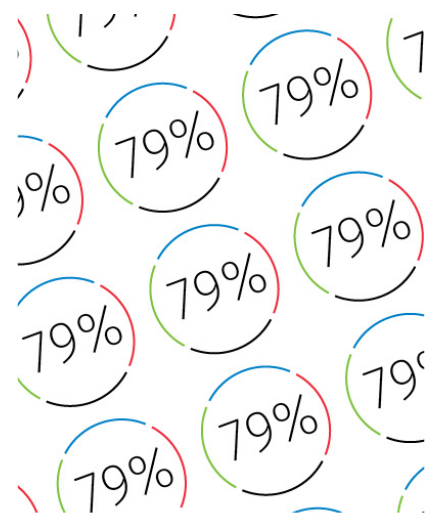
This entry was posted on Wednesday, January 17th, 2024 at 8:24 am and is filed under Artificial Intelligence (AI), Exceptions and Limitations, Infringement, Text and Data Mining (TDM), USA
You can follow any responses to this entry through the Comments (RSS) feed. You can leave a response, or trackback from your own site.