

Litigators of the Week: Cooley Secures an Important Early Win for Meta in Copyright Test Case for AI

By Ross Todd

December 4, 2023

When comedian Sarah Silverman and authors Richard Kadrey and Christopher Golden sued Meta Platforms this summer claiming the company infringed their copyrights by training its LLaMA set of large language models using data sets that included their works, it was no laughing matter. Whether there's a fair use right to use copyrighted texts to train LLMs such as LLaMA is one of the central legal questions facing companies developing generative artificial intelligence.

But when a lawyer for the authors tried to convince U.S. District Judge Vince Chhabria last month that LLaMA's outputs—convincingly natural text responding to user prompts—somehow infringed the author's copyrights, the judge said the argument was making his "head explode."

Chhabria knocked out a significant chunk of the plaintiffs' initial claims late last month calling the argument that the LLaMA language models are themselves infringing derivative works "nonsensical"—a win for Meta's team led by Bobby Ghajar, Mark Weinstein and Judd Lauter of Cooley.

Litigation Daily: What's at stake here for Meta and for the AI community more broadly here?

Mark Weinstein: The plaintiffs, in this lawsuit and various near-identical lawsuits filed



L-R: Bobby Ghajar, Mark Weinstein and Judd Lauter of Cooley.

in California and New York, seek a ruling that it's copyright infringement to use the text of copyright-protected books to train an AI Large Language Model (LLM). The plaintiffs have also, more generally, alleged that the LLMs themselves are infringing works because they were allegedly trained using the text of the plaintiffs' copyright-protected books, and that such LLMs violate the copyright management information (CMI) provisions of the Digital Millennium Copyright Act (DMCA), and California unfair competition law, and give rise to claims

for unjust enrichment and negligence under common law.

LLMs require large and diverse volumes of data in order to understand the relationships between words and concepts. As many have observed, at a broader level, if any piece of text out there that was potentially subject to copyright protection could no longer be used for AI training, that could handicap efforts to build LLMs that have a more complete and accurate understanding of human communication. That, in turn, could result in less useful (and more biased) models and a weaker domestic AI industry.

While the long-term implications and stakes of these claims are hard to predict at this point, some of the possible implications of the copyright claims in the AI community have been publicly discussed by the media.

Who is on your team and how have you divided the work thus far? (I saw that Professor Mark Lemley has signed on **to the team.**)

Bobby Ghajar: I handed the initial brief with Cooley litigation partner Mark Weinstein, former partner Angela Dunning, special counsel Judd Lauter, and senior associate Colette Ghazarian. Cooley partner Kathleen Hartnett joined the team to work on the reply briefing and I led the argument in front of Judge Chhabria. Professor Mark Lemley, the William H. Neukom Professor of Law at Stanford Law School and the director of the Stanford Program in Law, Science and Technology, also assisted with the briefing.

Judge Chhabria isn't usually one to hold back. At the hearing last month **he said the plaintiff's argument that the LLaMa language models themselves were infringing derivative works made his "head explode" and that he couldn't wrap his brain around them. What stood out from the hearing to you?**

Ghajar: Judge Chhabria was extremely engaged and, as usual, well prepared. He understood the

arguments we'd made and challenged plaintiffs repeatedly to explain their positions. At the hearing, the plaintiffs had no adequate response to the defects in the complaint, which Judge Chhabria acknowledged both at the hearing and in the order.

In your motion to dismiss, you write that "[u]se of texts to train LLaMA to statistically model language and generate original expression is transformative by nature and quintessential fair use" but you reserve that issue "for another day, on a more fulsome record." Can you give us a bit of a preview of that fair use argument about LLaMA's "training"?

Judd Lauter: Meta and many other developers of Generative AI models, including Open AI, Google and Microsoft, have submitted comments to the U.S. Copyright Office that preview some of the considerations around fair use.

What's important in this ruling for Meta and others in the AI community?

Weinstein: The core issue in this case is whether training on copyrighted works is an infringement of the copyright of those works, and if so, whether that is a fair use. This decision may set the parties on a path to being able to focus on that issue. In addition, the remaining claims in the complaint, which the court dismissed, have been repeated in most of the other lawsuits against Generative AI developers. Others may benefit from the court's analysis providing guidance on claims pertaining to the "derivative works right," CMI under the DMCA, and vicarious infringement, among other issues.

What will you remember about getting this particular ruling for Meta?

Ghajar: Artificial intelligence may be the technology that defines the 21st century. Cooley is often at the center of precedent setting rulings, but this is an especially thrilling experience to be part of a case involving such important, cutting edge technology and legal issues.