GenAI Contracting Revisited – What's New?

February 27, 2024

Announcer
Welcome to Mayer Brown's Tech Talks Podcast. Each podcast is designed to provide insights on legal issues relating to Technology & IP Transactions, and keep you up to date on the latest trends in growth & innovation, digital transformation, IP & data monetization and operational improvement by drawing on the perspectives of practitioners who have executed technology and IP transactions around the world. You can subscribe to the show on all major podcasting platforms. We hope you enjoy the program.

Julian Dibbell
Hello and welcome to Tech Talks. Today, we're revisiting a topic we took on a year ago, back when it was first rising to the top of everyone's minds in and around the tech world – generative artificial intelligence, or GenAI, with a particular focus on the contracting and other legal issues it can create for users, data providers and others engaging with this still fast-evolving new technology.

I'm your host, Julian Dibbell. I'm a senior associate in Mayer Brown's Technology & IP Transactions practice. I'm joined today by Marina Aronchik and Rich Assmus. Marina is a partner in the Technology & IP Transactions practice. Rich is a member of the Intellectual Property Brand Management & Litigation practice and co-leads the firm's Technology & IP Transactions practice. Rich and Marina both have been advising clients for several years on emerging issues in AI, and both were on the podcast with us in March 2023 when we first took a look at this topic.

So Rich, let me ask you, when we were last discussing this topic, we were still in the early days of some of the copyright lawsuits around generative AI; in particular, the lawsuits related to copyright in the data used for training the models. Where are we now? I'd like to get a sense of how things have evolved since those early days.

Richard Assmus
Thanks, Julian. We're getting much closer to some interesting rulings, but generally we still don't have a lot of clarity. To take a step back, generative AI raises a host of IP issues, but I like to think of them mainly

as issues on the input side and the output side. Mainly, what's the consequence of using copyrighted materials as input and who, if anyone, owns the rights into the output? The training data cases are primarily about the input, although I would point out that it may well be that the fair use issues will be determined in part both by what is inputted and what is outputted. I think the first thing to notice that the pace of these new training data lawsuits has continued, if not accelerated. Just as an example, at the very end of 2023, the *New York Times* filed suit against OpenAI and Microsoft. The complaint alleged some near-exact copying that I don't think many people expected out of a large language model like ChatGPT. For example, if the complaint is to be believed, ChatGPT "memorizes *Time's* content and can be prompted to output it nearly verbatim." That's something that I think the model makers are trying to stop, but it's always a cat-and-mouse game between users and the developers of those models.

What we've seen so far in these rulings on preliminary motions is that unlike the *New York Times*, rightsholders are having a hard time making claims against output stick, but they appear likely to be able to get past a motion to dismiss, maybe after amending their complaint on training data use. As an example, a federal court in California recently dismissed vicarious copyright infringement claims with respect to the output of ChatGPT, but the copyright claims on training data use were not at issue in that motion; and what that means for users of AI tools is that these cases are likely to enter discovery before we get some important rulings on fair use. And that's an example of just one of the issues that could be dispositive in these cases.

Julian Dibbell

Okay, Marina, turning now to how these training data cases might impact contracts. Last year when we talked about this, your view was that it makes sense for the provider of the AI tool to take on the risk of infringement with respect to training data cases, but that providers were at the time finding it difficult to quantify the risk in the existing legal regime, and therefore were hesitant to offer protection against this risk to customers. How has that changed, if it has? Has it?

Marina Aronchik

Thanks, Julian. Yes and no, and this is where business folks who are listening to our podcast will say – but there have been numerous articles publicizing the willingness of leading GenAI providers to stand behind infringement claims. So customers and providers are aligned on the intent here. Now, from the legal standpoint, while there has been an unusual amount of advertising of these commitments, they do involve a lot of nuance. There is an emerging trend of distinguishing infringement claims with respect to use of data to train the model and infringement claims related to the output of the model. The former, so infringing use of data to train the model, is increasingly treated in a manner that's similar to potential other infringement risk of a technology product, meaning it's subject to an ordinary course infringement indemnity. The latter, the infringement of output, is generally subject to, and again, this depends on your deal, subject to very unusual and nuanced indemnity that may be quite difficult for a customer to enforce and for business purposes, may need to be treated as if there were no such indemnity.

Julian Dibbell

Can you give us an example of that?

Marina Aronchik

Sure. For one, ascertaining which models and tools are actually covered by the indemnity is often complex. Basic, as it sounds, the names of the models that are covered by the indemnity may not

correspond to the products that the business believes they're procuring. There might be subtle differences in nomenclature or an indemnity might refer to yet another product that incorporates the model versus the model itself.

Next, you may see carve-outs for fine-tuned models, and again, these are only examples; or if you're dealing with a so-called platform garden, where a user can access a variety of models, third-party models may be entirely carved out and subject to different terms than the so-called first-party models, meaning the models that the provider with whom you're contracting is making available as their proprietary products.

You'll also often see carve-outs or limitations based on customers' use of certain filters. These filters at their core may have been designed to filter out potentially problematic context, but a customer may not know how a particular filter may be related to the risk of infringement. So it's possible that infringement would have occurred regardless of whether a particular filter is enabled. So in summary, these indemnities do not work like our other ordinary course, infringement indemnities in technology contracts.

Julian Dibbell
Okay. So it sounds like there's been some encouraging movement on this front, but it's still a little bit of a caveat emptor situation. Staying on this topic of inputs and the risks related to inputs for just a few more minutes. Last year, Marina, you were also concerned with the risk of continued training by the tool on customer data and new risks that this might create based on limitations in the right to this data that the customer may or may not have; and in particular, how that might impact adversely any trade secret protection on the data that's processed in the AI tool. Can you talk a bit about that?

Marina Aronchik
These topics remain top of mind. So generally, to the extent that a customer is able to negotiate terms related to AI products, which is a complex question on its own, and express prohibition against training and retraining models based on customer data is a productive conversation for both parties. And I will use this question, if I may Julian, as an opportunity to highlight a topic that's nuanced, but it's critically important here, which is that customers' agreements for AI products may be bolted on to existing cloud agreements; and cloud agreements typically include commitments by customer with respect to customers' own data, which the customer is then processing in the cloud services. So if you're designating output of an AI tool or product as customers' data, which you may want to do for the purpose of allocating IP rights, then a customer will want to make sure that the indemnification obligations that might exist in the underlying cloud agreement did not become circular through the definitional issues of tying up these concepts of input and output when we now come to the question of providers' indemnification obligations for infringement claims relating to the output.

Julian Dibbell
Okay. I think I followed that, but my head is spinning a little bit, and I'm sure customers' will. So it's important to sit and think these issues through very carefully. All right, but so speaking of that input and output dichotomy and turning back to the outputs, you talked about the input cases, training data cases that have unfolded in the past year. What about the output cases? The copyrights in what is actually produced by the AI? Have the courts or Congress perhaps provided any guidance on the underlying IP issues there?

Richard Assmus

So we actually do have greater clarity on the *output* side. I think that's reflective of the fact that that question can be answered as a matter of law, whereas the issues in the training data cases are much more either factual issues or mixed issues of fact and law. So now we have decisions from the Copyright Office and the courts on the output of generative AI. And what those decisions have said is that without more, that output is not subject to copyright protection. The same holds true on the patent side, namely an invention that's solely the result of an AI process is not patentable. In each case, it's for lack of a human being. On the copyright side, it's for lack of a human author. On the patent side, it's for lack of a human inventor. So as with copyright, it's basically been decided by the USPTO and the Federal Circuit that you can't have a sole AI inventor of a patentable invention.

Julian Dibbell

Ah, so on the output side, does this all mean then that innovators trying to leverage AI to invent, to create, are kind of out of luck with respect to sort of retaining any rights in what's produced?

Richard Assmus

The short answer there is really no—there are workarounds. So those inventors, those innovators are not totally out of luck. On the copyright side, while the unedited output of generative AI may not be predictable, if that output is sufficiently modified by human authors or combined with other human-authored content, the resulting work as a whole is very likely to be subject to copyright protection. And as a practical matter, I think that may be enough for many users.

Just to give you an example, if a company is using an assistive coding tool like CoPilot to create computer code, the fact that snippets of that code may have been AI-generated probably doesn't have a big commercial impact on that company's ability to assert rights in the code as a whole. And to be clear here, I'm setting aside the issue of whether the coding tool has introduced any viral open-source code into the mix, which does have the potential to infect the entire work.

Similarly on the patent side, just a couple weeks ago on February 12th, the USPTO issued guidance on AI-assisted inventions. And it's really a similar story there. The Patent and Trademark Office has indicated that AI-assisted inventions can be patentable so long as a human being was responsible for a significant contribution to the invention. That's actually the same test that's applied to any invention when there's some question as to who qualifies as a joint inventor. It's come up quite a bit in litigation over joint inventorship. Although the PTO didn't quite put it this way, you and an AI tool can be co-inventors, but the law only allows the human to be named. In that guidance, however, the PTO made it clear that a person simply overseeing an AI system that is used in the creation of invention that does not provide a significant contribution to the conception of that invention, that does not make a person an inventor.

Julian Dibbell

Okay. Interesting. So Marina, on this output side of things, any final remarks here from the perspective of how does this impact contracting?

Marina Aronchik

Well, on output, technology lawyers have long understood the challenges with protectability of data under IP laws. When we talk about output, much of the output is data. Therefore, contractual allocation of rights and responsibilities for data remains important. It probably becomes even more important now

when we talk about output given, the developments that Rich just described. Now keep in mind that the issue of allocating rights to data and to the output remains subject to rights to the underlying data based on which the model may have been trained and that might be part of the output. So again, the nuance here of allocating rights as between the parties is important relative to the commitments and limitations that might exist with respect to the underlying third-party data.

But Julian, this was really a nuance to the contracting point and output. I do want to end with a different remark as my final remark for this podcast. So, the market seems to have shifted from thinking from, dare I say, generic issues in AI and generative AI to industry and use case-specific risks. So the same GenAI product will raise materially different issues, risks, and concerns if implemented by a bank versus a manufacturing company versus a life sciences company, or yes, even a law firm. There are also different risks that are involved in deploying a code-generating AI tool like the one that Rich was describing earlier versus a GenAI model that's used in the sales functions or in drug development, going back to the point about the industry. So these so-called use case reviews and the broader analysis of strategic use development and implementation of AI and generative AI requires a sophisticated and interrelated set of interdisciplinary, meaning legal, technical, business compliance, and other knowledge, to identify, assess, and evaluate risk, ideally in a consistent manner that takes into account AI and non-AI regulations and frameworks, meaning the existing laws, which continue to apply when you use AI and I hope that we have an opportunity to discuss this additional layer of complexity in one of our upcoming Tech Talks.

Julian Dibbell
Well, I hope so too. I thank you and Rich for updating us on the latest developments and complications in the ongoing saga of generative AI and contracting.

Listeners, if you have any questions about today's episode or if you have an idea for an episode, you'd like to hear about anything related to technology and IP transactions and the law, please email us at techtransactions@mayerbrown.com. Thanks for listening.

Announcer
We hope you enjoyed this program. You can subscribe on all major podcasting platforms. To learn about other Mayer Brown audio programming, visit mayerbrown.com/podcasts. Thanks for listening.