Unique Issues To Look Out For In Generative AI Transactions

By **Aaron Rubin and Heather Whitney** (May 16, 2023)

Aaron Rubin

Heather Whitney

Because transactions involving artificial intelligence technologies can resemble those involving traditional software — like software as a service agreements — parties often assume that their expectations from those standard agreements about what is reasonable and "market" should also apply to AI-related technology transactions.

And, in some cases, this approach is appropriate.

However, it is important not to overlook the unique issues raised by certain AI technology transactions, particularly those that involve generative AI models.

Training Data

Generative models typically undergo two stages of training — initial training and fine-tuning.

Initial training comes first and exposes the generative model algorithm to massive data sets to help it gain a broad understanding of the data and the relationships between the data.

In most cases, the provider of the generative model completes this initial training before the model is made available to customers and end users.

For their initial training, many generative models — particularly large language models — are trained on data scraped from the internet.

There have been years of litigation on the permissibility of web-scraping, which is independent of scraping for training purposes,[1] and there are ongoing cases addressing whether training on copyrighted materials without permission qualifies as a fair use.[2]

Fine-tuning typically involves a smaller and more specialized data set that is designed to refine the generative model, so it is better able to generate outputs for specific use cases.

Some model providers offer customers the ability to fine-tune an instance of a generative model for the customer's specific use case. In these cases, sometimes the model customer provides the fine-tuning data, sometimes the model provider gets the data from elsewhere, and sometimes the parties work together to create the data.

Agreements concerning generative models should clarify the ownership and permissible uses of training data and allocate liability arising from its use.

For instance, in cases where the model customer provides data for use in fine-tuning, the agreement should clarify whether the model provider is permitted to use the data only for the model customer, or also to train the generally available version of the generative model.

Generally speaking, it may be reasonable for each party to be responsible for training data that it provides or sources and to indemnify the other party from claims arising from such data's use.

That said, where a generative model is initially trained on large swaths of data scraped from the internet, the model provider may be reluctant to indemnify the model customer for claims arising from such data, particularly given that the law around such training is unsettled. The model provider may argue that, at least for large language models, this type of training is a necessary aspect of developing the technology and that there is no practical way for the model provider to affirmatively secure rights to the data.

Without any contractual risk allocation, the model provider will likely bear most of the risk arising from use of such data for initial training, given that the model provider will be the party sourcing the data — e.g., through scraping — and performing the activities that could give rise to claims, like making copies.

User Prompts

User inputs, or prompts, are the questions, queries and other inputs that users enter into generative models to generate outputs.

Which party owns prompts is a question that may arise in negotiations. Typically, as between the model provider and model customer, the model customer generally retains any ownership rights it may have in its prompts.

But a word of caution: Model providers should avoid language that suggests that they are assigning to the model customer any preexisting rights the model provider may have in prompts.

If a model customer copies patent claims from the model provider's patents and uses them as a prompt, surely no reasonable person would think the model provider intended to assign its rights in those patents. The model customer had no rights in the model provider's patents beforehand, and its inputting the patent claims into the generative model should not give it those rights either.

Another issue that often arises in transactions involving generative models is confidential treatment of prompts.

Model customers may want their prompts to be treated as their confidential information for a variety of reasons, including to maintain trade secret protection. While these concerns are understandable, model providers often do not know what kinds of data customers are inputting into the model and are not in a position to do a case-by-case analysis of prompts to determine which prompts may contain confidential information.

Given this, taking on onerous obligations — and potential liability — for trade secrets or other confidential information contained in prompts may mean, in practical terms, that the model provider has to comply with those obligations for every prompt by every model customer.

Moreover, most commercially available generative models are not trained on prompts from enterprise customers, and model providers often provide nonenterprise users an opt-out right if they do not wish to have their prompts used for training (frequent alarm about this possibility notwithstanding).

Another concern model customers may have is whether prompts containing information about new inventions could constitute a public disclosure — specifically, a printed publication under Title 35 of the U.S. Code, Section 102 — that would prevent patenting the inventions.[3]

Again, generative models are generally not trained on enterprise customer prompts, though the possibility that a model provider could access and disclose a prompt is a theoretical possibility. Even in that case, though, the prompt would not constitute a public disclosure of the invention unless it describes the invention in enough detail for someone skilled in the field to replicate or use it, i.e., enablement.

The good news for model customers is that, at least under U.S. patent law, even if their prompts were to somehow constitute a public disclosure, they have a one-year grace period from the date of disclosure to file a patent application.[4]

In an attempt to avoid the above issues, where feasible, model providers may tell model customers to not include in their prompts confidential information, like personal identifiable information, or information about inventions for which patent applications have not yet been filed.

Some use cases may require model customers to input highly sensitive confidential information, like proprietary source code. In those cases, the parties may need to negotiate specific terms to address the treatment of such information contained in prompts.

Generative Model Outputs

The defining feature of generative models is that they generate outputs in response to prompts.

In most cases, agreements between model providers and model customers should expressly address ownership of outputs. The threshold question, though, is whether outputs constitute intellectual property that can be owned at all, and the answer to that question remains unclear.

What is relatively clear is that a generative model is neither an inventor under the Patent Act[5] — as the U.S. Court of Appeals for the Federal Circuit held in a case the U.S. Supreme Court recently refused to review, Thaler v. Vidal[6] — nor an author under the Copyright Act.[7] What is less clear is when a human's use of a generative model can strip the human of inventor or author status.

With respect to copyrightability, the guidance the U.S. Copyright Office published in March suggests several tests for authorship that, in certain cases, lead to conflicting outcomes. The tests in the guidance also differ from the test the U.S. Copyright Office applied when it revoked a copyright in images created by Kris Kashtanova using Midjourney.[8]

Parties to transactions involving generative models should keep in mind that there are no clear answers regarding the circumstances in which outputs can be owned, in the intellectual property sense.

Moreover, even if outputs are protectable intellectual property, allocating intellectual property rights appropriately between the parties can be more challenging than it first appears.

In traditional transactions where vendors create customized content for customers, vendors often assign all right, title and interest in that content to the customer, but that may not be appropriate in the generative AI context.

Outputs may, for instance, incorporate or be substantially similar to third-party material, including material contained in the data the generative model was trained on. Obviously, the agreement between the model provider and the model customer cannot allocate ownership of intellectual property that belongs to third parties. Therefore, the agreement should at least carve out such third-party material from any terms purporting to allocate ownership of intellectual property as between the model provider and the model customer.

Parties should also consider what happens if outputs are substantially similar to the model provider's preexisting intellectual property.

Imagine a case where the model customer is able to get a generative model to disclose parts of the generative model's own source code; the model provider presumably would not be willing to assign those intellectual property rights to the model customer. To deal with this problem, the parties may choose to omit any assignment language from their agreement and rely on background intellectual property law to allocate output ownership.

However, if the model customer insists on the model provider assigning whatever intellectual property rights it may have in outputs, the model provider may consider carving out its preexisting intellectual property rights.

While these carveouts may help address the model provider's concerns, such carveouts can leave the model customer uncertain whether it owns its outputs.

There is no easy solution to ownership issues, but on balance it may be reasonable for the model customer to be responsible for doing the necessary due diligence — e.g., code scans, clearance searches, etc. — to determine the provenance of outputs. After all, the model customer is the one generating the outputs through its prompts, and the model provider will typically not be aware of the specific outputs being generated.

The issues with outputs do not end there. Setting aside preexisting intellectual property, a generative model could create virtually identical outputs for two or more model customers. Model providers should make model customers aware of this possibility and clarify that, while outputs generated by the model customer may belong to that customer, that does not mean that the model provider is guaranteeing that other customers will not generate similar or identical outputs — and if they do, those other customers' outputs may belong to them.

When model customers own outputs, the question arises as to what rights, if any, the model provider retains in them — for example, the right to train a generative model using them.

As noted above, many generative models are provided under terms that do not permit the model provider to use prompts for training — at least for enterprise customers — and the same approach may make sense in many cases for outputs. But, of course, this can be negotiated on a case-by-case basis.

Another important issue in any transaction involving generative models is allocation of liability for claims arising from outputs.

The issue of "hallucination," where a generative model confidently provides incorrect information, has been much in the news lately. Such hallucinations could result in a variety of claims — e.g., defamation, false advertising or product liability. In addition, lawsuits have been filed alleging that outputs infringe preexisting copyrighted works. If such claims arise, who should be responsible, the model provider or the model customer?

A model customer who lacks a sophisticated understanding of how generative AI works may see the generative model as simply a way to obtain content. And the model customer may have a general expectation that, when it obtains content from a content provider, the content provider should take responsibility for that content through contractual mechanisms that allocate risk.

But viewing the model provider as merely a content provider is an oversimplification of how a generative model works.

In reality, it is quite difficult for a model provider to take on this kind of liability for outputs. Certainly, model providers can incorporate guardrails into their models; for example, a model provider could prevent its model from generating images of well-known copyrighted or trademarked characters. But designing a generative model that is incapable of producing any problematic outputs is very difficult, if not impossible.

Moreover, as noted above, model providers do not ultimately control which prompts model customers enter. Yet those prompts go a long way toward determining what outputs the generative model generates. Indeed, in most cases where generative models have generated outputs that include what appears to be third-party content, users went to great lengths to push the model into generating it.

At the same time, though, while model customers can control what prompts they enter, they do not control whether the generative model takes their innocuous prompt and provides a potentially problematic output.

Nonetheless, in practical terms, it may be more reasonable for the model customer to be the party that is ultimately responsible for deciding whether the outputs are suitable for use.

One topic we have seen raised is whether statutory safe harbors — Section 230 of the Communications Decency Act, specifically — apply to outputs.[9] Justice Neil Gorsuch raised this issue during oral arguments in Gonzalez v. Google LLC in the U.S. Supreme Court this past term.[10]

Section 230 provides broad immunity to online platforms and other providers and users of interactive computer services for liability arising from third-party content.[11]

While some have argued that Section 230 immunity should extend to outputs because outputs are derived from third-party content, — i.e., training data — Section 230 immunity does not apply when the platform operator itself is responsible in whole or in part for the illegality of the content at issue.

A plaintiff is likely to argue that a model provider is responsible at least in part for the outputs of its generative model.

Other Issues

Sometimes, a model provider may provide model customers with access to a third-party generative model — or an application programming interface for a third-party generative model — including as part of a larger services offering. In such cases, the model provider and model customer should review the terms that apply to the third-party generative model to confirm that the model customer's intended use case is consistent with those terms.

Some generative models are distributed under terms that prohibit commercial use or the creation of derivative works, e.g., AlexaTM 20B, while others impose specific use restrictions intended to enforce the responsible use of generative models, which must be flowed down to end users — e.g., models licensed under a Responsible AI License.

Finally, parties need to consider how to address compliance with the growing body of laws and regulations targeting the use of generative models and generative models themselves.

While it is straightforward to say that each party is responsible for its own compliance with law, in some cases the model customer's use of the generative model may trigger compliance obligations on the model provider that the model provider would not otherwise have had.

Concluding Thoughts

By staying up to date on the rapidly evolving legal landscape surrounding generative models, both model providers and model customers will be better equipped to successfully navigate the complexities of these transactions and enter into agreements that make sense.

Aaron P. Rubin is a partner and chair of the technology transactions group at Morrison Foerster LLP.

Heather Whitney is an associate at the firm.

Disclosure: Morrison Foerster represents Kashtanova.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] https://blog.ericgoldman.org/?s=scraping&submit=Search.

[2] https://www.law360.com/articles/1598495/ai-art-cos-want-out-of-artists-collage-tool-copyright-case.

[3] https://www.law.cornell.edu/uscode/text/35/102.

[4] https://www.law.cornell.edu/wex/one-year_rule.

[5] https://www.law360.com/articles/1599643.

[6] https://www.law360.com/articles/1599643/justices-reject-call-to-let-ai-be-inventor-on-patents.

[7] https://www.pacermonitor.com/public/case/44763624/THALER_v_PERLMUTTER_et_al.

[8] https://copyright.gov/docs/zarya-of-the-dawn.pdf.

https://www.law360.com/articles/1589396/artist-seeks-copyright-of-ai-artwork-that-uses-own-drawing.

[9] https://www.law.cornell.edu/uscode/text/47/230.

[10] https://www.supremecourt.gov/oral_arguments/audio/2022/21-1333.

[11] https://www.law.cornell.edu/definitions/uscode.php?width=840&height=800&iframe=true&def_id=47-USC-1900800046-1237841278&term_occur=999&term_src=title:47:chapter:5:subchapter:II:part:I:section:230