Proskauer

3 Items To Check When Evaluating AI Terms And Conditions

Law360 on **October 10, 2023**

A deluge of new generative artificial intelligence tools has hit the market, with companies ranging from startups to tech giants rolling out new products.

In the large language model space alone, we have seen OpenAI's GPT-4, Meta's LLaMA, Anthropic's Claude 2, Microsoft's Bing AI, and others.

A proliferation of tools has meant a proliferation of terms and conditions.

Many popular tools have both a free version and a paid version, with each subject to different terms, and several providers also have "enterprise" grade tools available to the largest customers.

For lawyers, law firms and other businesses looking to trial generative AI, the number of options can be daunting.

This article sets out three key items to check when evaluating a generative AI tool's terms and conditions.

Although determining which tool is right for a particular user is a complex question that requires an analysis of terms and conditions in their entirety — not to mention nonlegal considerations like pricing and technical capabilities — the below items can provide prospective customers with a starting place, as well as bellwether to help spot terms and conditions that are more or less aggressive than the market standard.

Training Rights

Along with new generative AI tools, 2023 has been a year of new generative AI lawsuits.

OpenAI and other generative AI providers currently face claims alleging unauthorized and improper use of plaintiffs' proprietary data as generative AI model training material, with claims variously based on copyright, contract and privacy law.

And lawsuits aren't the only way that generative AI providers have lately faced increased scrutiny over how and where they obtain training data to develop their generative AI products.

For example, in April, the popular social media platform Reddit announced a plan to begin charging for access to the platform's application programming interface, which is generally how generative AI providers import data into their models — i.e., Reddit has decided that user posts shouldn't be given away for free to generative AI providers whose products might undermine the popularity of their platform.

Other popular news publishers and media companies have taken defense technical measures — e.g., amending robots.txt files to disallow certain bots — to block OpenAI's GPTBot web crawler from ingesting news stories for future training purposes.

On top of these new hurdles, the Federal Trade Commission is reportedly looking into OpenAI's collection of user data — among other issues, such as publication of false information and potentially anti-competitive practices surrounding generative AI.

In light of these challenges, many of which pertain to the training of generative AI models, it is perhaps not surprising that some generative AI providers have revised their tools' terms to reassure users about how user data may — or more precisely, may not be — used.

For example, Microsoft Corp. has updated its default commercial terms for its Azure OpenAI service, which provides licensed access to OpenAI's GPT models, to explicitly state that user inputs are not used for training, and GitHub Inc. has done the same for its generative AI coding tool, Copilot.

OpenAI has made a similar update to its template enterprise agreement. Even Anthropic PBC, which is the provider of ChatGPT competitor Claude — the newest player on the scene whose terms assert a broad right to use user data to develop new products and services — explicitly excludes model training.

On the other side of the coin, Zoom faced backlash this summer over asserting a broad right to turn user data into training material — a position it eventually walked back.

The upshot is that it is now off-market for a generative AI provider to assert a right to use customer data for training purposes without at least providing an opt-out mechanism. The biggest providers have abandoned this position, but as generative AI companies proliferate, customers should be watchful.

For lawyers, this means checking a tool's terms of use to make sure they include a no-training commitment or opt-out — whether the lawyer is evaluating the tool for use in his or her own practice or conducting diligence on the tool for a client.

If it turns out the generative AI provider might use inputs for training, the lawyer will need to make sure guardrails are in place to ensure use of the tool does not result is unintentional leakage of proprietary or sensitive information.

Use Restrictions

For many businesses, one of the most attractive aspects of generative AI in general and large language models in particular is use-case flexibility.

Unlike most machine learning technology, which traditionally has been designed and deployed for a particular job or small set of jobs — e.g., speech-to-text, facial recognition — many of the newest large language models are capable of a surprising range of tasks, from responding to customer queries in chat interfaces to fixing bugs in software code.

However, just because a generative AI tool is technically capable of a function does not mean that function is permitted under the tool's terms of use.

Virtually all generative AI providers prohibit at least some uses of their tools. Some of these restrictions are predictable and generally unobjectionable, such as prohibitions on uses that violate applicable laws. However, many terms go beyond such obvious bans, and it is here where end users need to be careful.

One area where restrictions are common — and potentially problematic — is development of new products and services. Many providers do not want their tools to be used to build competitive technology, and so terms and conditions often restrict use of the applicable provider's services accordingly.

The ubiquity of these sorts of restrictions shows that the market has congealed to some degree on this issue, such that most businesses — and lawyers — likely will not get far in a negotiation with a generative AI provider if they attempt to reject the concept of a competitive use restriction outright.

However, the specific language of these restrictions is critical and can vary significantly, from narrow prohibitions referring specifically and narrowly to training large language models that directly compete with the provider's model, to general prohibitions on developing or improving similar products or services of any kind.

While restrictions of the former type generally would not be a problem for most end users — unless of course they are intending to build their own generative AI models — the latter can trip up businesses precisely because the providers' generative AI tools have so many possible functionalities, such that many of the end users' products and services into which the providers tools might be integrated could be considered similar.

For lawyers, this means it is critical to review all use restrictions — but especially competitive prohibitions — before a client or the lawyer themselves begins using a particular generative AI tool, to ensure that the end user is on the same page with the provider about how exactly the tool may and may not be used.

Of course, when evaluating a tool's terms on behalf of a client, this will also require investigating the client's planned use cases for the tool and other plans in the generative AI space, making sure the client understands the boundary between acceptable and prohibited use, and ensuring guardrails are in place to prevent out-of-bounds usage in the future.

Where competition prohibitions for a given tool are broad and vague and the provider is inflexible, some users might consider looking at other options.

Responsibility for Outputs

At this point, it is well established that even the most advanced generative AI tools can produce outputs with a litany of flaws, from made-up facts, or hallucinations, to elements copied verbatim from training data, aka memorization — and that these flaws can cause harm and sometimes lead to lawsuits.

Sometimes, these flaws are easily spotted through basic diligence and removed before they do any harm, but this will not always be practicable.

For example, it may not be feasible for an end user to detect that a generative AI tool has generated output that infringes a third party's intellectual property — until the owner of that IP comes knocking.

At the same time, most generative AI providers are not in a position to police every individual output themselves, and so far, no one has developed a foolproof filter or other technical solution — at least not publicly. The upshot is that is responsibility for outputs is often one of the most hotly negotiated subjects in generative AI contracts, since the risks can be difficult to mitigate.

Unsurprisingly, most generative AI providers disclaim all responsibility for their tools' outputs, such that customers must use the outputs at their own risk.

However, Microsoft recently partially broke from this trend with its Copilot Copyright Commitment, assuring paid customers that they can "use Microsoft's Copilot services and the output they generate without worrying about copyright claims."[1]

Microsoft Copilot actually isn't the only tool to provide this sort of assurance, but it nevertheless may presage a market shift in response to customer concerns, similar to the shift on model training rights discussed above — but note that this copyright commitment does not address other output issues like hallucinations.

In any case, lawyers evaluating generative AI tool options for self or client use should consider output responsibility generally and the potential for harm to third parties specifically — particularly keeping an eye out for generative AI provider warnings and disclaimers in terms of use — and should make sure that internal guardrails are appropriately strong where providers do not offer output assurances or otherwise use their tool terms to make the generative AI end user responsible for third-party harm.

Otherwise, the end user could find themselves holding the bag if the generative AI tool goes off the rails, as in the infamous case of the **so-called ChatGPT lawyer.**

Conclusion

Determining whether a given generative AI tool is right for a particular business — or for use in legal practice — requires an analysis of the tool's terms and conditions in their entirety, as well as nonlegal considerations like pricing and technical capabilities.

That said, training rights, use restrictions and output responsibility will virtually always be relevant considerations, and where a generative AI provider stands on these issues may help prospective customers and their counsels evaluate that provider's terms in the context of the broader generative AI market.

[1] Microsoft Copilot's commitment actually extends to over types of IP as well, but does not cover non-IP claims and includes certain conditions, such as the user having sufficient rights to their inputs. See https://www.microsoft.com/en-us/licensing/news/microsoft-copilot-copyright-commitment.

Reproduced with permission. Originally published October 10, 2023, "3 Items To Check When Evaluating AI Terms And Conditions," Law360.

Related Professionals

- **Peter J. Cramer**
  Associate