July 24, 2023

Skadden

AInsights

If you have any questions regarding the matters discussed in this memorandum, please contact the following attorneys or call your regular Skadden contact.

Stuart D. Levi
Partner / New York
212.735.2750
stuart.levi@skadden.com

Mana Ghaemmaghami
Associate / New York
212.735.2594
mana.ghaemmaghami@skadden.com

MacKinzie M. Neal
Associate / New York
212.735.2856
mackinzie.neal@skadden.com

One Manhattan West
New York, NY 10001
212.735.3000

Online Terms of Use and the Training of AI Models

A key building block of artificial intelligence (AI) large language models (LLMs) is that they are trained on vast amounts of content and data. In many cases, this content and data is amassed by running bots or other automated programs that extract information from the web. For example, an earlier version of GPT (GPT-3) was trained in part through the use of filtered data from Common Crawl, an open, but unpermissioned, repository of data extracted through web crawling. Similar methods that programs may employ to extract data include "web scraping" or "bulk downloading." Importantly, nearly all of these programs are run without obtaining authorization to extract and use the content and data in this manner.

Companies are starting to appreciate the extent to which their website content and data is possibly being used in this way without their consent. For example, in April 2023, the *Washington Post* ran a feature that allowed readers to input a website URL and see if content provided at that URL was included in Google's C4 data set.[1] That data set contains the contents of 15 million websites that have been used to train LLMs.

While some companies may not take issue with their content and data being used without permission in this manner, a number of companies have started to push back on such uses of their data. Their motivations can range from a general unease with their data being used in ways they cannot control and without their permission, to the business argument that their data should not be monetized for free.

Companies and organizations are taking varied approaches to this issue. Some are entering into formal deals to permission use of their content, such as the recent deal between the Associated Press and OpenAI that allows OpenAI to train on Associated Press content for two years. Others have started to address this issue by putting their content behind paywalls or by implementing fees to access the interfaces that would be required to download large amounts of data. Still other companies have looked to whether they have arguments that such usage constitutes a violation of their copyright rights.

However, a growing trend among companies is to expressly prohibit use of their data for training of AI models through the terms and conditions of their website or service. Their strategy is to argue that use of their data in breach of the terms and conditions would

[1] Kevin Schaul, Szu Yu Chen and Nitasha Tiku, "Inside the Secret List of Websites That Make AI Like ChatGPT Sound Smart," *The Washington Post* (April 19, 2023).

potentially give rise to a breach of contract claim against the entity extracting the data. Given that such a claim would be based on contract law and not copyright law, the potential defendant would not have a fair use defense to such a claim.

The breach of contract argument is already starting to find its way into cases challenging the use of content in AI models. For example, in *Doe v. GitHub*, a putative class action by computer programmers challenging the unauthorized use of their computer code to train the Codex and Copilot AI-based code generation models, one of the plaintiffs' arguments is that the use of their code in this manner violated GitHub's Terms of Service since those terms prohibited use of their materials outside of GitHub.[2]

Companies seeking to rely on a website or service's terms of use to prohibit use of their data in connection with training AI models should be mindful of the attention courts have paid to the presentation of these terms when determining if they constitute binding and enforceable agreements. We summarize some of the key points to be aware of below.

Legal Standard for a Binding Contract

A contract is formed when parties manifest their mutual assent to the terms of the agreement. Generally, courts have held that for an online contract to be enforceable, the user must be on notice of the contract's terms and must unambiguously manifest assent through some type of affirmative action.

The notice requirement for online contracts can be satisfied if the user has actual knowledge of the terms or if the website or service provides reasonably conspicuous notice of the terms.[3] Courts generally look to the font size of the notice, the visibility of the notice in comparison to the surrounding text, and the overall design of the website to determine whether the provider of the site has taken steps to alert a reasonably prudent user of the terms.

Courts will also consider whether the website design draws the user's attention away from the link to the terms, so that it cannot be presumed the user saw the link. If a user must click a link in order to be directed to the terms, the link itself should be apparent, such as through the use of a contrasting font color or all capital letters; underscoring the link alone will likely not be sufficient to meet a presumption of notice. Similarly, requiring the user to hover their mouse over text to find the link will likely not be sufficient notice of the terms.

Overall, the onus is on website providers to put users on notice to the terms to which the providers wish the user to be bound since the providers control the design of their own sites.

With respect to the "manifestation of assent" requirement, courts often start with the basic principle from the Restatement of Contracts that "[t]he conduct of a party is not effective as a manifestation of his assent unless he intends to engage in the conduct and knows or has reason to know that the other party may infer from his conduct that he assents."[4]

In the case of online contracts, this means looking at the actions, if any, that the user was required to take to signify their assent to the contract. According to the Ninth Circuit, for example, a website "must explicitly notify a user of the legal significance of the action she must take to enter into a contractual agreement."[5] Whether a user has manifested assent will depend on the firm of online agreement used:

- **Browsewrap agreements.** "Browsewrap" agreements refer to online agreements that seek to bind the user to terms because they appear on a webpage (typically through a link at the bottom of the webpage along with other links) and require no further action showing that the user has read or agreed to the terms. The term is derived from the idea that companies are seeking to bind users simply through their act of browsing the website. Courts have been reluctant to enforce browsewrap agreements because "there is no assurance that the user was put on notice as to the existence or content of the terms" or that they manifested acceptance of those terms.[6]
- **Clickwrap agreements.** "Clickwrap" agreements refer to online agreements where the terms are presented to the user though a clear link, often through a pop-up window, and the user can only proceed to use the website or service by clicking an "I agree" button or checking an "I agree" box. Courts are most likely to enforce these terms because they represent the clearest and most direct manifestation of assent.
- **Scrollwrap agreements.** "Scrollwrap" agreements are generally seen as a subset of clickwrap agreements in that they not only require the user to click on an "I agree" button, but also require the user to scroll through the terms before they can even access that button. These agreements are similarly likely to be deemed enforceable.

[2] Complaint at 45, *Doe v. Github* (N.D. Cal. 2022) (No. 3:22-cv-06823-KAW). See our May 23, 2023, client alert "Ruling on Motion To Dismiss Sheds Light on Intellectual Property Issues in Artificial Intelligence."

[3] *Berman v Freedom Financial Network, LLC*, 30 F.4th 849 (9th Cir. 2022) at 855 (finding that terms were not binding where the link appeared only in "tiny gray font considerably smaller than the font used in the surrounding website elements barely visible to the naked eye.").

[4] Restatement (Second) of Contracts §19(2) (1981).

[5] *Berman*, 30 F.4th 849 at 858.

[6] *See, e.g., Gaker v. Citizens Disability, LLC*, No. 20-CV-110310AK, 2023 WL 1777460, 6 (D. Mass. Feb. 6, 2023), citing *Kauders v. Uber Techs.*, 159 N.E.3d 1033, 1054 (2021); *Berman* at 1178.

Online Terms of Use and the Training of AI Models

- **Hybridwrap agreements and sign-in agreements.** "Hybrid" agreements or "sign-in" agreements typically refer to online agreements where the terms are clearly presented and some action is required to proceed, but the required action does not clearly manifest assent to those terms. For example, the user might be presented with a link to the terms and then must click "submit" or "continue" to proceed to the next screen, but the language employed does not make explicit that clicking one of those buttons means the user is agreeing to the terms. Where it may not be clear to the user that clicking "continue" or proceeding to "sign in" to the service means they are assenting to the terms — including instances where the "continue" button and the link to the terms are in close proximity — such agreements may not be enforceable against the user.[7] Importantly, including language such as "I understand and agree to the terms and conditions" with a "continue" button may not be sufficient where that text is in a small font and light color.[8] In other cases, however, a sign-in agreement can be enforceable where a user is clearly and explicitly put on notice that by creating an account or proceeding to use the service they are agreeing to the terms of use (*e.g.*, "By opening an account, you are agreeing to our terms of use").[9] In general, the key inquiry here, as with all online agreements, is whether a "reasonably prudent" user would be deemed to be on inquiry notice of the terms of use.

In the foregoing cases, especially with respect to hybrid and sign-in agreements, some courts will look at the context to ascertain whether a consumer would "expect" to be bound to contract terms (*e.g.*, a user would expect to be bound by terms when they sign up for a subscription as opposed to when they engage in a one-off transaction or activity, or try a service before signing up for it).

[7] *Berman* 30 F.4th 849 at 858, citing *Nguyen v. Barnes & Noble, Inc.* 763 F.3d 1171, 1179 (9th Cir. 2014).

[8] *Id.*

[9] *Meyer v. UberTechnologies, Inc.*, 868 F.3d 66, 74 (2d Cir. 2017).

Summer associate **Abby Rubinshteyn** contributed to this article.

Key Points

The use of publicly-accessible website data and content to train AI models has brought renewed focus on the enforceability of online agreements, as companies wish to use their terms and conditions as a mechanism to protect use of their data in unauthorized ways.

Companies looking to rely on provisions in their online agreements to bar content or data on their websites or services from being used to train AI models should be careful to present those agreements in a manner such that they will be deemed enforceable. As noted, a simple link at the bottom of a website that does not require any action by the user to manifest assent may not be enforceable. In such cases, it may be challenging for the company to claim that the unauthorized use of its content to train an AI model was a breach of those terms.

Companies that use clickwrap agreements will be better positioned to argue that use of their content violates an agreement with the user, but even they will want to make sure they have phrased the clickwrap prompt as a clear manifestation of assent to the terms of use, and present the link to those terms in a clear manner. While in the case of bots and web scraping tools there is no human clicking on an "I agree" button — the tool is programmed to find such prompts and do so automatically — a company with an enforceable clickwrap agreement will likely have a strong argument that the developer of the bot or tool violated the enforceable online agreement, and that its data was used in violation of those terms.

The enforceability of online terms of use is almost always a fact-specific inquiry. However, court decisions in this area have provided companies with useful guideposts with respect to the type of notice and manifestation of assent that is required to determine whether a contract may be deemed enforceable.