

Building AI and Machine Learning Technologies: Data Licensing Tips and Traps

Data is the fuel for software development, and developers use historical data from existing products to train algorithms and build AI and machine learning models. Companies are well aware of privacy and regulatory restrictions on data use, but often do not consider the potential impact of data use restrictions on intellectual property ownership and use rights.

By Anna Remis

In the race to innovate, developers must consider how data use could impact ownership of their technology. Data is the fuel for software development, and developers use historical data from existing products to train algorithms and build AI and machine learning models. Companies are well aware of privacy and regulatory restrictions on data use, but often do not consider the potential impact of data use restrictions on intellectual property ownership and use rights.

In the United States, contracts are the most important source of rights and restrictions for data use, laying out between companies what is acceptable data use and what isn't. However, copyright law does provide limited protection for compilations of data, so companies should still analyze copyright infringement principles and defenses — as well as other potential claims — when dependence on third-party data is central to the technology development process

If you use third-party data, the question of whether a data owner could assert rights in your technology always requires a review of contractual restrictions. Get-

ting clear contractual permissions is a best practice.

Let's break this down with a hypothetical example. You are a fintech start-up company developing a tool to assess customers' financial planning readiness. You plan to offer this tool to financial advisors on a white-label basis as a lead-generation tool so they can market to customers whose assessment reveals they have a need for help. You are relying on several important data sources when you develop your tool, including *historical* customer behavior data from a prior app that your company built to help customers develop savings habits, *commercially licensed* anonymized datasets purchase for research and development purposes from a large financial firm, and *pilot program* data from a strategic partner testing new features and functionality under development. Setting aside the privacy and regulatory restrictions, you also need to analyze whether the large financial firm, or the strategic partner for the pilot program, could claim any rights in your newly developed tool based on the data they provided, whether



Credit: peshkov/stock.adobe.com

under copyright law or as a result of the contracts you signed with them.

Here are a few tips on data licensing you might consider to be sure you're aware of your rights to the data — and what potential pitfalls might be lurking in the shadows if you use the data.

Data Licensing Tips for Developers of Technology Products and Platforms:

- Carefully define in your contract the technology that each company is contributing and who owns improvements and derivative works. Third-party datasets could be considered intellectual property that another company is contributing. Is there any risk under the contract that the

product you develop could be considered an improvement or derivative work of their dataset, or jointly developed?

- Check whether there are obligations to share insights back to the data licensor. Remember, data is a valuable form of currency, and these insights may be part of the price you pay for the data, particularly where the initial data was “free” or discounted. If you agreed to share insights exclusively with any data licensor, that exclusivity could become an encumbrance on your technology.
- Negotiate for rights to use data beyond the immediate services offered. If rights are limited to aggregated or anonymized data, will that meet your needs? If you need rights to use data for R&D, then rights only to support and maintain the current platform are insufficient. If data is licensed in connection with a pilot program, can the data be used beyond the pilot and/or for the benefit of other strategic partners?
- Distinguish — and make sure you own — data that relates to the way the products and services that you develop operate, including data about use of the products and services, system performance data, and transactions on the platform. If you do not have full rights to product or platform data, it may limit the value of your product or platform and rights to make future developments.
- If rights are limited to providing and improving a product or service or for “internal business purposes,” carefully assess whether new proposed use cases are permitted. You may not be able to use the data as broadly as you had hoped.

On the other hand, if you are using a product or service, you should scrutinize broad requests to leverage your data that extend beyond the product or service provided. Are you retaining the rights you need to monetize the data for future use cases? Are you receiving adequate value in exchange for data use rights? If you are giving broad rights in aggregated and anonymized data, consider whether the recipient could reverse engineer the data to derive commercially sensitive information.

Balancing Licensing Risks Against Innovation Imperatives

What if you do not have a perfect dataset available, with clear data use rights? Companies may feel pressure to conduct exploratory research and modeling to determine what products or services might be possible, before obtaining clearer rights. Proceed cautiously, as exploratory analytics could breach contracts, infringe another company’s intellectual property or give rise to other claims.

In this situation, companies should press their development teams to consider at the outset: Are there more economical datasets you could obtain for exploratory purposes? Can your algorithms run with synthetic datasets and still yield useful results? Can skunkworks projects be done in an isolated development environment so as not to contaminate products and services already in production? Waiting until the development process is further along could create problems that are expensive, difficult, or even impossible to fix later.

Restrictions on data use could form a basis for a data owner to claim rights in a licensee’s technology applications. Companies engaged in data-driven technology development should adopt thoughtful strategies towards data usage, with contractual strategies in the spotlight alongside regulatory and ethical considerations.

Anna Remis is a partner in the Technology and Life Sciences Transactions practice at the law firm of Sidley Austin LLP in San Francisco. She counsels technology companies from start-ups to public companies on a broad range of intellectual property issues, including IP and technology development, ownership, use, acquisition, sale, licensing, and protection across industries. She can be reached at aremis@sidley.com.

This article has been prepared for informational purposes only and does not constitute legal advice. This information is not intended to create, and the receipt of it does not constitute, a lawyer-client relationship. Readers should not act upon this without seeking advice from professional advisers. The content therein does not reflect the views of the firm.

[This article appeared in Cybersecurity Law & Strategy, an ALM publication for privacy and security professionals, Chief Information Security Officers, Chief Information Officers, Chief Technology Officers, Corporate Counsel, Internet and Tech Practitioners, In-House Counsel. Visit \[the website to learn more\]\(#\).](#)