VENABLE LLP

Contracting for AI Model Training: Key Considerations for Customer Data Rights

June 05, 2026

Leah A. Druckerman, Emma R. W. Blaser and Rob Hartwell

As laws and market expectations regarding AI continue to evolve, so do contracting considerations associated with AI-enabled products and services. While contract language permitting vendors to use customer data for "improvement" and "development" of services has been a routine software-as-a-service (SaaS) negotiation point, the proliferation of "AI-enabled" products and services has brought new focus to these provisions from customers and vendors.

What Is AI Model Training, and Why Does It Matter?

AI models are "trained" by ingesting sample data resembling the data that the model will analyze in practice. Training data is typically "memorized" by the model, which can then reproduce similar or identical content when prompted. New generative AI models are trained on, and therefore can potentially reproduce, copyrighted works, software code, and confidential or personal data. For example, there are reports of generative AI models reproducing portions of copyrighted works in response to user prompts.

The potential for a model to "regurgitate" confidential or personal training data drives many customers of AI tools to scrutinize any use of their data for model training. However, a prohibition may not always be the best approach for those customers seeking to obtain the full value of a tool and for AI vendors seeking to improve their services. There are technical, contractual, and compliance considerations that may help unlock that value for both customers and AI vendors.

Key Considerations for AI Training Data Provisions

Align on Permitted Uses for Customer Data

Customers and vendors should align on goals for the data processing permitted under the contract. For customers, language that permits a vendor to process customer data to "improve," "build," or "enhance" services can authorize model training and may enhance the services they receive. However, some customers may wish to limit these permitted uses or include express provisions regarding the use of customer data for training if concerns about proprietary, personal, or confidential information

are paramount.

For vendors, transparency and education are key in negotiating AI training rights and building customer trust. Carefully tailoring permitted use and training data language and identifying benefits of model training can help customers understand the impact of training restrictions.

Consider the Risks and the Added Value of Training

It is possible to provide AI-enabled services without training on customer data. In such cases, a pre-trained model is deployed that does not "train" on user inputs or outputs. The result is a model that does not memorize customer data but also does not improve over time. Users of AI tools should assess whether these pre-trained models provide the right fit for their needs and consider whether allowing training may provide better value for their organization.

Foundational vs. Private Model Training

Some vendors may offer the use of either a "foundational" or "private" model. "Foundational" models are public, shared, or used across customers, but may carry a higher risk of "regurgitating" training data to third parties.

Training of "private" instances of models on customer data does not typically carry these same risks because the customer's data is only used to train the customer's "private" copy of the model, not a model used by others.

This private model structure allows customers to benefit from model training while reducing risk of disclosure of training data, but potentially without the benefit of the broader improvements in the foundational model. As a result, it is common for many enterprise AI providers to offer a "private" version of their services, though this may come at a premium.

Privacy Impacts of Personal Data Training

Allowing a vendor to train a foundational AI model with customer personal data may result in the vendor acting as a "controller" instead of a "processor." Engagement of vendors that are controllers can impose additional compliance obligations, including requiring privacy policy updates, and could even result in "sales" of personal data, triggering opt-out and assessment obligations under U.S. state privacy laws.

Deidentification or Aggregation Is Not a Silver Bullet

We often see terms that permit the training of foundational AI models with only "deidentified" or "aggregated" data. While this can address some customer concerns, training data can be "deidentified" and still contain confidential or valuable information. Terms outlining more specific requirements for deidentification or aggregation of training data can help mitigate some of these risks.

What Should Companies Consider When Negotiating AI Model Training Contract Terms?

Contracting around training data is evolving rapidly as new products and use cases come to market, so the above is not an exhaustive list of considerations. Whether you are an AI customer, AI vendor, or both, updating contracting playbooks and educating your teams on the ins and outs of AI contracting can help shorten negotiations and mitigate emerging risks.

If you have questions about AI contracting, please reach out to Venable's Privacy and Data Security Group for assistance in building or negotiating AI contract terms that work for your organization and use cases. If you'd like to learn more about emerging AI and data privacy issues, please check out our recent webinar.