

Publications

Key Contract Terms and Conditions for AI Products and Services, Part 1 – Data Ownership and Licensing

Related Attorneys

Craig R. Auge

Related Services

Copyrights

Data Strategy, Privacy and Security

Intellectual Property

Patents

Technology Transactions

Trade Secrets

Related Industries

Colleges and Universities

Financial Institutions

Health Care

Insurance

Manufacturing

Private Equity

Restaurants, Food and Beverage

Retail and Consumer Products

CLIENT ALERT | 8.29.2023

*This is the first installment on key contract terms and conditions for AI products and services and is more from a customer-side vantage. A **second installment** will cover commitments, disclaimers, risk allocation, regulatory, and privacy considerations. Scope note – this is **not** about AI preparing contract forms.*

Artificial intelligence (AI) is pervasive in current news.

While popular media has just recently latched on to promoting “powered by AI” in a manner similar to media promotions of “the Cloud” a few years ago or “the Web” a couple decades ago, AI as a discipline can trace its roots to the 1950s, and generative AI has seen major breakthroughs over the last few years. With this history and current momentum, the far-reaching impact of generative AI technologies in products and services for nearly all industries is undeniable.

AI leverages computers and machines to mimic problem-solving and decision-making capabilities of the human mind.

When entering contracts for AI products and services, providers and customers should apply well-known legal concepts to lesser-known AI elements. Parties experienced with SaaS agreements will find some familiar landscape, and indeed AI solutions are often provided by SaaS.

But they also should be prepared for unique issues and more complexity – and potential risks and liabilities – associated with an array of types and sources of data that train, fuel, guide, emanate from, and modify generative AI-based models, solutions, and systems (AI Solutions).

For convenience, “provider” references in this Client Alert could be the actual developers of the foundational model for an AI Solution or application developers that build on that foundational model or otherwise make it available (e.g., as a part of larger offerings from Microsoft or Oracle, which contracts with customers).

Data Types and Sources

- **Training the AI Solution**

The algorithm that is a part of the AI Solution is exposed by the provider to massive data sets to “train” it, typically before being made available to customers. *Think of it as “initial training data.”*

Subsequently, the initial training data is often fine-tuned, improved, and optimized by more data added as a new layer, sourced from data contributed by the provider, by the customer, by both of them jointly, or by or through a third party from which they have obtained it. *Think of it as “enhanced training data.”*

- **Prompting the AI Solution**

Customers provide the AI Solution with prompts, instructions, queries, data (including, for example, Internet-of-Things device data), or other input. *Think of it as “input prompts.”* It is possible that input prompts become a part of enhanced training data.

- **Generating Outputs from the AI Solution**

In response to input prompts, the AI Solution generates responsive output – such as other data, text, images, video, audio, new code, or other materials or content. *Think of it as “output.”* Output could become a part of enhanced training data.

Practice Tips

*AI Solutions are often “black boxes” on how they were constructed, the sources of data, and the flow of data. If the contract does not adequately describe, at least at a higher level or in ways that impact key rights and obligations, then the provider and the customer should have an initial joint session to “whiteboard” it. Skipping this step and proceeding directly to negotiating the contract leads to business parties not fully grasping the proposed AI Solution, frustrating and unproductive drafting turns, or, at worst, inaccurate or missing terms and conditions.

*Definitions matter a lot.

Data Ownership and Licenses

As a matter of U.S. copyright law, raw data as facts are not copyrightable, but database compilations may be protectable by copyright. However, as a matter of contract (and “as between the parties”), providers and customers regularly declare ownership of certain data and convey license rights to the other, often with a limited license scope specifying rights to use for a certain time, to display, to modify or create derivative works, or to distribute – and then for permitted described purposes.

Customers accustomed to simply declaring ownership of their data provided to, for example, a SaaS solution should quickly realize that data treatment for an AI Solution requires more finesse. And while such customers may be familiar with terms that permit a provider to use or re-purpose anonymized and aggregated versions of their customer-provided data, that too is not enough. Best practices include

moving beyond the word “own” to spell out restrictive parameters for use cases.

- **Training Data**

Less clear than input prompts (below), initial training data could be owned by the provider, by third parties (e.g., obtained by the provider web-scraping the internet), or, if in an enterprise- or more closed-type of AI Solution, by the customer. Such initial training data could also have been contributed pursuant to a license.

Enhanced training data ownership is more challenging, if inclusive of customer input prompts or output contributed back to the AI Solution by the customer (or other customers). In such instance, best practice would be including some type of express but limited grant from the customer(s) to the provider, such as whether it can be used just for that contributing customer or other customers under certain conditions.

- **Input Prompts**

Customers can likely succeed in declaring ownership of input prompts. If the input prompt was obtained by the customer from a third party, by license or otherwise, they need to ensure they have rights to input it.

In such instance, a provider may get an implied or express license to use the input prompt for customer’s use case – but that provider may ask for a broader scope, such as refining the AI Solution for the betterment of other customers too (some of which could be competitors). Providers may also have a default provision that input prompts can train the AI Solution, but give customers opt-out rights.

- **Output**

Stickiest of all, output stretches concepts like derivative works and may require express distinctions between and among underlying training data and input prompts and the AI Solution itself as to ownership and licensed rights.

A customer may, for example, want to assert ownership of that portion attributable directly to its input prompt that is not otherwise training data.

Humans using generative AI may be able to create a copyrightable work (how much human control being a “line” that is still being drawn), but the U.S. Copyright Office has taken the position that the AI Solution itself cannot be the “author” of a copyrighted work. Regardless, prudent parties will stipulate in their contract what output is owned or licensed, as between them.

Providers may also stipulate that output for one customer may be the same or nearly the same as output resulting from another customer’s use case.

- **Assignments for Output**

Customers can ask for assignments, but providers may resist more traditional express assignments to customers of ownership in output, citing the complexities of output data or the amalgamated nature

of output content to be assigned and lack of clarity on copyright and patent treatment.

Providers may instead be more inclined to disavow their own ownership in output (as carefully defined) or concede by acknowledgement that, as between them and customers, customers own.

- **Confidentiality Provisions**

Traditional confidentiality provisions alone are insufficient to cover the types of data treatment described above, but, thoughtfully applied, they can buttress particular positions on ownership and licensing scopes of use.

Customers may, for example, want their input prompts or output to be a part of confidential information (as defined in the contract and with non-disclosure and non-use restrictions) so as to protect their trade secrets. Whether that is appropriate, however, depends on the AI Solution model and use case. Moreover, standard exceptions would exclude information rightfully received from a third party and independently developed information.

Practice Tip

*Avoid a pre-determined stance that “ownership” is in all cases better than “licensing.” Ownership can sometimes invite additional possible liabilities, and often licensing (or sub-licensing) is all that can or needs to be done. Obtaining ownership through negotiation may also be a hollow victory, to the extent the owned output, for example, is only usable with the AI Solution or has a short lifespan of value.

When swimming in the waters of AI products and services, be sure to consult your Vorys attorney.

Next installment on key contract terms and conditions for AI products and services – commitments, disclaimers, risk allocation, regulatory, and privacy considerations.