

# Mori & Partners

Roppongi Hills North Tower 16F, 6-2-31 Roppongi, Minato-ku, Tokyo 106-0032 Japan  
TEL: 03-6831-9281 FAX: 03-6800-5542  
<https://mps-legal.com/>

## Hidden Licensing Risks in Training AI Models with Outputs from Open Models

July 14, 2026

### I. Introduction

In recent years, businesses have increasingly deployed AI models within their own environments in response to growing concerns over data security and privacy.<sup>1</sup> In particular, large language models (“LLMs”) operated within a company’s own infrastructure (commonly referred to as “local LLMs”) have emerged as a practical option for implementing generative AI in environments that handle highly sensitive information, such as those in the financial sector.

As the use of local LLMs continues to expand, businesses may increasingly seek to improve model performance by using the outputs of existing AI models as training data for other models.

In practice, however, when introducing an AI model, businesses do not always evaluate the model’s licensing terms with the future use case of training another AI model using its outputs

---

<sup>1</sup> Examples include the following:

- “MUFG deploys real-time anonymization across unstructured data,” Asian Banking & Finance, February 2026  
<https://asianbankingandfinance.net/banking-technology/news/mufg-deploys-real-time-anonymization-across-unstructured-data>
- “BNP Paribas provides its businesses with an LLM as a Service platform to accelerate the industrialization of generative AI use cases.” BNP Paribas Press Release, June 5, 2025  
<https://group.bnpparibas/en/press-release/bnp-paribas-provides-its-businesses-with-an-llm-as-a-service-platform-to-accelerate-the-industrialization-of-generative-ai-use-cases>
- “HSBC and Mistral AI join forces to accelerate AI adoption across global bank,” HSBC Media Release, Dec 1, 2025  
<https://www.hsbc.com/news-and-views/news/media-releases/2025/hsbc-and-mistral-ai-join-forces-to-accelerate-ai-adoption-across-global-bank>
- “The biggest AI shift is taking place in your employees’ bags,” Fast Company, May 5, 2026  
<https://www.fastcompany.com/91536554/the-biggest-ai-shift-is-taking-place-in-your-employees-bags>

specifically in mind. Nevertheless, if it is later discovered that such use violates the applicable license terms, the consequences may vary depending on both the terms governing the AI model that generated the training data and the manner in which the subsequently trained model is deployed. In some cases, the business may be unable to continue using the relevant training data and may be required to retrain the model from scratch. In others, achieving compliance may require amendments to the business's own terms of use and renewed user consent after the relevant service has already been launched. Such remedial measures can impose significant operational and legal burdens.

This article examines the licensing terms applicable to several major open models and considers the obligations that may arise where their outputs are used to train another AI model. It then discusses practical measures that businesses should consider during both the implementation and operational phases to appropriately manage AI model outputs and mitigate the associated legal risks.

In this article, the following terms are used as follows:

- “Open Model” means an AI model whose components (such as its weights, parameters, or source code), in whole or in part, have been made publicly available and whose use, modification, and redistribution by third parties are permitted within a specified scope.
- “AI System” means a machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs (such as predictions, content, recommendations, or decisions) that may influence physical or virtual environments.
- “AI Model” means a model incorporated into an AI system that is produced through machine learning using training data and generates predictions based on input data.

## **II. Models Covered by This Article and Obligations Arising from Their Use**

### **1. Models Covered by This Article**

This article examines the terms of use of the following models (collectively, the “Covered Models”):

- **Model:** Llama4  
**License:** Llama4 Community License Agreement  
**URL:** <https://www.llama.com/llama4/license/>
- **Model:** Gemma3

**License:** Gemma Terms of Use

**URL:** <https://ai.google.dev/gemma/terms>

- **Model:** stable-diffusion-xl-base-1.0 (the “SDXL”)

**License:** CreativeML Open RAIL++-M License<sup>2</sup>

**URL:** <https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0/blob/main/LICENSE.md>

## 2. Obligations That May Arise for AI Models Trained Using Outputs from Each Covered Model

### A. Llama4

If a licensee uses outputs from Llama4 to train another AI model and redistributes that model or otherwise makes it available for use, the licensee must place the name “Llama” at the beginning of that model’s name.<sup>3</sup>

---

<sup>2</sup> As a reference, the article numbers of the relevant provisions of the CreativeML Open RAIL++-M License are indicated below, numbered by the author based on the CreativeML Open RAIL-M License (<https://huggingface.co/spaces/CompVis/stable-diffusion-license>).

• **CreativeML Open RAIL++-M License**

*Section III: CONDITIONS OF USAGE, DISTRIBUTION AND REDISTRIBUTION*

*4. Distribution and Redistribution. You may host for Third Party remote access purposes (e.g. software-as-a-service), reproduce and distribute copies of the Model or Derivatives of the Model thereof in any medium, with or without modifications, provided that You meet the following conditions:*

- *Use-based restrictions as referenced in paragraph 5 MUST be included as an enforceable provision by You in any type of legal agreement (e.g. a license) governing the use and/or distribution of the Model or Derivatives of the Model, and You shall give notice to subsequent users You Distribute to, that the Model or Derivatives of the Model are subject to paragraph 5. This provision does not apply to the use of Complementary Material.*
- *You must give any Third Party recipients of the Model or Derivatives of the Model a copy of this License;*
- *You must cause any modified files to carry prominent notices stating that You changed the files;*
- *You must retain all copyright, patent, trademark, and attribution notices excluding those notices that do not pertain to any part of the Model, Derivatives of the Model.*
- *You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions - respecting paragraph 4.a. - for use, reproduction, or Distribution of Your modifications, or for any such Derivatives of the Model as a whole, provided Your use, reproduction, and Distribution of the Model otherwise complies with the conditions stated in this License.*

*5. Use-based restrictions. The restrictions set forth in Attachment A are considered Use-based restrictions. Therefore You cannot use the Model and the Derivatives of the Model for the specified restricted uses. You may use the Model subject to this License, including only for lawful purposes and in accordance with the License. Use may include creating any content with, finetuning, updating, running, training, evaluating and/or reparametrizing the Model. You shall require all of Your users who use the Model or a Derivative of the Model to comply with the terms of this paragraph (paragraph 5).*

<sup>3</sup> Section 1.b.i of the Llama4 Community License Agreement.

Unlike the two licenses discussed below, however, Llama4's FAQ suggests that if a licensee uses only Llama4 outputs to train another AI model, without including all or part of Llama4 itself or its documentation, that resulting AI model does not fall within Llama4's "derivative works."<sup>4</sup>

## B. Gemma3<sup>5</sup>

If a licensee uses outputs from Gemma3 to train another AI model so that it behaves similarly to Gemma3, the resulting AI model qualifies as a "Model Derivatives"<sup>6</sup> of Gemma3.

When reproducing or "Distributing" a "Model Derivatives," the following obligations apply:

- (a) An obligation to include, as legally enforceable provisions in the relevant agreement, the use restrictions set out in Section 3.2 (i.e., restrictions prohibiting uses under the Gemma Prohibited Use Policy or unlawful uses), and to notify users of the relevant "Model Derivatives" that they are required to comply with those use restrictions.<sup>7</sup>
- (b) An obligation to provide a copy of the Gemma Terms of Use to third-party recipients of the "Model Derivatives."<sup>8</sup>
- (c) An obligation to include, on modified files, a notice stating that such files have been modified.<sup>9</sup>

---

<sup>4</sup> Meta Platforms, Inc., "Troubleshooting & FAQ", <https://www.llama.com/faq/-Legal>

Q : What are Llama's attribution requirements?

A : If a developer uses a Llama 3.1, Llama 3.2, Llama 3.3, or Llama 4 model, such as Llama 3.1-405B, to create or train another AI model, for example by generating a synthetic dataset that is then used to train another AI model, then that developer must include "Llama" at the beginning of such AI model's name if it is distributed. This case is different from the case above, where Llama models or materials are being distributed or a product or service is built on or incorporates Llama (a derivative work), since the AI model in this case may not be based on or incorporate Llama. Consequently, a statement that the AI model is "Built with Llama" is not required; however, "Llama" should be included at the start of the AI model's name since a Llama model was used to create, train, fine tune, or otherwise improve the AI model.

<sup>5</sup> Gemma4, released on March 31, 2026, is licensed not under the Gemma Terms of Use but under Apache License 2.0.

Release date: <https://ai.google.dev/gemma/docs/releases?hl=ja>

Source: <https://ai.google.dev/gemma/terms>

<sup>6</sup> Section 1.1(e) of the Gemma Terms of Use.

<sup>7</sup> Sections 3.1, item 1 and 3.2 of the Gemma Terms of Use.

<sup>8</sup> Section 3.1, item 2 of the Gemma Terms of Use.

<sup>9</sup> Section 3.1, item 3 of the Gemma Terms of Use.

(d) Except where provided through a “Hosted Service,”<sup>10</sup> an obligation to include the following statement in the “Notice” file for all “Distributions”<sup>11</sup>:

*“Gemma is provided under and subject to the Gemma Terms of Use found at [ai.google.dev/gemma/terms](https://ai.google.dev/gemma/terms).”*

A point of particular importance is that “Distribution” includes not only the redistribution of Gemma3 or its Model Derivatives themselves, but also making their functionality available through APIs, web services, or other electronic or remote means.<sup>12</sup>

Accordingly, if a licensee uses outputs from Gemma3 to train another AI model so that it behaves similarly to Gemma3, the provider must also comply with the obligations described in items (a) through (c) above when it makes the relevant model’s functions available as a hosted service through APIs, web services, or other electronic or remote means.

### **C. stable-diffusion-xl-base-1.0**

If a person uses outputs from SDXL to train another AI model so that it behaves similarly to SDXL, the resulting AI model qualifies as a “Derivatives of the Model” of SDXL.<sup>13</sup>

When reproducing or “Distributing”<sup>14</sup> such Derivatives of the Model to Third Parties,<sup>15</sup> the following obligations apply<sup>16</sup>:

- (a) An obligation to include, as legally enforceable provisions in the relevant agreement, the use restrictions set out in Attachment A<sup>17</sup>, and to notify users of the relevant Derivatives of the Model that they are required to comply with those use restrictions.
- (b) An obligation to provide a copy of the CreativeML Open RAIL++-M License to Third Parties receiving the Derivatives of the Model.

---

<sup>10</sup> Under section 1.1(b) of the Gemma Terms of Use, “Hosted Service” means providing or making Gemma or its functionality available as a hosted service via API, web access, or any other electronic or remote means.

<sup>11</sup> Section 3.1, item 4 of the Gemma Terms of Use.

<sup>12</sup> Section 1.1(b) of the Gemma Terms of Use.

<sup>13</sup> Section I.1 of the CreativeML Open RAIL++-M License, as numbered by the author.

<sup>14</sup> Strictly speaking, the text states “host for Third Party remote access purposes (e.g. software-as-a-service)” as a condition; however, because the following paragraph states, “You shall give notice to subsequent users You Distribute to, that the Model or Derivatives of the Model are subject to paragraph 5,” on the premise that such hosting falls within “Distribution,” the author has supplemented the original wording in light of the intended meaning of that text.

<sup>15</sup> This refers to a natural person or legal entity not under common control with either the licensor or the licensee.

<sup>16</sup> Section III.4 of the CreativeML Open RAIL++-M License, as numbered by the author.

<sup>17</sup> Broadly speaking, this sets forth prohibitions on unlawful use and infringement of third-party rights.

- (c) An obligation to include, on modified files, a notice stating that such files have been modified.
- (d) An obligation to retain all copyright, patent, trademark, and attribution notices associated with any part of SDXL or the Derivatives of the Model.

As with Gemma3, a point that requires attention when a party provides services using SDXL or its Derivatives of the Model is that “Distribution” covers not only redistribution of SDXL or its Derivatives of the Model themselves, but also the act of making their functions available through APIs, web services, and similar means.<sup>18</sup>

Accordingly, if a licensee uses outputs from SDXL to train another AI model so that it behaves similarly to SDXL, the provider must also comply with the obligations described in items (a) through (d) above when it makes the relevant model’s functions available through APIs, web services, or similar means.

#### **D. Summary of Terms of Use Relating to Outputs from the Covered Models**

Under the Llama4 license, the relevant obligation is relatively straightforward: the licensee must begin the name of the AI model with “Llama.”

By contrast, where outputs from Gemma3 or SDXL are used to train another AI model, the applicable obligations may vary depending on how the resulting model is deployed and the manner in which the related service is provided.

If the provision of the service constitutes “Distribution” under the applicable terms of use, the provider must, among other things, appropriately incorporate into the service’s terms of use any applicable use restrictions set out in the Gemma3 or SDXL license terms. Where the service has already been launched, complying with these requirements may, in some cases, require the provider to obtain renewed consent from users.

Accordingly, the specific compliance measures required, and the ease with which they can be implemented, will depend on how the relevant AI model is used.

### **III. Risks Associated with Using AI Model Outputs for Training and Practical Responses**

---

<sup>18</sup> Section I.1 of the CreativeML Open RAIL++-M License, as numbered by the author.

## 1. Preliminary Considerations

As discussed in Section II.2 above, even where an AI model is an Open Model, using its outputs to train another AI model may give rise to certain obligations with respect to the resulting model.

Moreover, businesses increasingly rely on multiple AI models (including SaaS-based AI services) for different purposes. In this context, providers of SaaS-based AI services commonly prohibit or restrict the use of their outputs to develop AI models or services that may compete with their own offerings.<sup>19</sup>

Given the above, whenever a business uses the outputs of one AI model to train another AI model, it should first identify and review the terms of use applicable to those outputs.

## 2. Risks in the Event of Non-Compliance

If a licensee fails to take the measures described above and, as a result, violates the terms of use applicable to those outputs, two principal risks may arise:

- A) The risk of receiving cease-and-desist demands, warning letters, claims for damages, or similar assertions from the rights holder of the relevant AI model.
- B) The risk of being perceived as having inadequate compliance practices.

The first risk may result not only in the costs associated with responding to a dispute, but also in additional costs for implementing an alternative AI model and retraining or improving that model, depending on the outcome of the dispute. In addition, because many providers of AI models, including Open Models, are based outside Japan, a company may also face legal proceedings, including litigation, in foreign jurisdictions.

The second risk extends beyond the ordinary reputational harm that may result in a decline in users of the company's services. It may also give rise to practical disadvantages, such as a reduction in contributions from the developer community to open-source software that the company has made publicly available.

---

<sup>19</sup> See, for example:

- “Terms of Use” (ChatGPT), Using our Services  
<https://openai.com/policies/row-terms-of-use/>
- “Consumer Terms of Service” (Claude), Section 3.2  
<https://www.anthropic.com/legal/consumer-terms>

### **3. Practical Responses**

To comply with the terms of use applicable to AI model outputs and mitigate the risks discussed above, businesses should consider implementing the following measures:

- (a) When selecting an AI model, review the terms applicable to that model in light of the anticipated future uses of its outputs.
- (b) When maintaining logs of inputs to and outputs from an AI model, also record the identity of the AI model that generated those outputs.
- (c) Where an AI model is trained using a company's internal data, determine whether the service from which the relevant data was obtained involved another AI model. If so, establish operational procedures to ensure that the applicable terms of use of that AI model are reviewed, and complied with.

Among these measures, item (a) is particularly important at the initial implementation stage. By selecting an AI model whose terms of use permit broader use of its outputs, a business may be able to avoid additional compliance measures at a later stage, including those described in item (c).

In practice, however, responsibility for model selection, data management (including the management of output logs), and license compliance is often divided among different business units or personnel. As a result, individual departments may find it difficult to implement these measures independently. Businesses should therefore recognize these issues at an organizational level and establish an internal governance framework appropriate to their own operational structure.

Even where AI models have already been introduced without implementing the measures described above, it may still be possible to avoid or mitigate the associated risks, depending on the manner in which those models have been used. For example, a business may review the terms of use applicable to the AI models currently in use when considering future model improvements, or verify the source of outputs included in its training data and take any additional steps necessary to achieve compliance with the applicable license terms.

Ultimately, businesses should maintain an accurate understanding of the terms of use applicable to the AI models they use and establish appropriate internal processes to ensure ongoing compliance.

## **IV. Conclusion**

This article has examined practical approaches to managing the terms of use applicable where AI model outputs are used for training. In practice, however, uniformly managing the obligations applicable to particular outputs may prove difficult, particularly where multiple AI models are used in combination or where human intervention modifies the outputs. Businesses should therefore establish governance and management frameworks that reflect the specific characteristics of each use case, in addition to implementing the measures discussed in this article.

Identifying, at the outset, the need to comply with the terms of use governing the use of AI model outputs for training enables businesses to consider a broader range of options, including the selection of AI models, while reducing future operational burdens. Accordingly, when introducing AI models, businesses should ensure that relevant internal stakeholders coordinate from the earliest stages (including model selection) and assess, from the perspective of the AI model's entire lifecycle within the organization, how compliance with the applicable terms of use can be maintained and what governance framework should be established to support that compliance.

**Further contact:**



**Shinichiro ("Shin") Mori**  
MANAGING PARTNER

E: [shinichiro.mori@mps-legal.com](mailto:shinichiro.mori@mps-legal.com)



**Takamasa Tsumurai**  
ASSOCIATE

E: [takamasa.tsumurai@mps-legal.com](mailto:takamasa.tsumurai@mps-legal.com)

**About Mori & Partners**

Mori & Partners is a Tokyo-based boutique law firm focusing on cross-border corporate, technology, gaming, fintech, and regulatory matters. The firm advises foreign companies entering the Japanese market on entity structuring, regulatory licensing, and ongoing compliance. For inquiries, please contact us at [mps-legal.com](mailto:mps-legal.com).

*This article is provided for general informational purposes only and does not constitute legal advice. Readers should seek specific legal counsel in relation to their particular circumstances.*