

Open Source Licensing Considerations for Artificial Intelligence Application

By Vlad Teplitskiy, Damien Howard and Fatima Kouadio

April 23, 2025

In January 2025, DeepSeek captured the headlines with the release of the company's artificial intelligence (AI) R1 model that, for at least certain use cases, appears to match the performance of ChatGPT at a fraction of the cost. Media coverage of the R1 model has mostly focused on the model's performance and low cost. In many ways, however, the success of the R1 model is also due to DeepSeek's decision to give away the R1 AI model for free. Rather than mimicking competitors that account for 80% of the global generative AI market and mainly follow closed-source approaches that guard the secrecy of their models, DeepSeek has chosen to freely distribute the R1 model for the public to use, modify and redistribute. Being a free model, R1 has captured the attention of developers, startups, and companies looking to build their own AI products. However, the cost-free nature of open source AI software does not necessarily equate to automatically being risk-free. As is explained below, a company should carefully consider the risks of using open-source AI models.



(l-r) Vlad Teplitskiy, Damien Howard, and Fatima Kouadio, with Knobbe Martens.

Open Source Licensing

Open source software (OSS) refers to source code that anyone can freely use, modify and distribute. The goals of OSS are unrestricted access and collaboration. OSS is provided under a variety of different licenses that seek to ensure fulfillment of these goals. Users of OSS code must comply with the license terms if they wish to use, modify and distribute the code or works based on the code.

According to a 2022 study from the Linux Foundation, 70-90% of any given software code includes open-source components. For AI applications, two-thirds of large language models (LLMs) released in 2023 were open source.

With the expanding availability of high-quality and free OSS, it is almost guaranteed that any technology company developing AI applications will be using this invaluable resource. While OSS may not come with a hefty price tag, the type of licenses used can cause varying levels of potential risk for using OSS AI models.

Generally, OSS licenses for software code that makes up AI models fall into three different categories: permissive, reciprocal (or weak copy-left), and restrictive (or copy-left). Permissive licenses allow the code to be copied, modified and distributed without limitation so long as you include the copyright notice in your software. Examples of these licenses include BSD, MIT, and Apache 2.0.

Reciprocal licenses add the requirement that any modified code must be distributed under the same license. Examples of these licenses include MPL, which is used by the Firefox web browser and EPL.

Restrictive licenses further add the requirement that any modified code and, in many cases, proprietary code that uses OSS code under such license, inherits the same license terms. Practically, restrictive OSS code can “taint” proprietary code and require mandatory source code distribution of the proprietary code. Examples of these licenses include GPL, which is used by the Linux operating system, and AGPL.

To function properly, an AI model based on the model code must be trained or refined with training data. While training datasets are often freely available to the public, certain uses of the datasets may be restricted by a license. Thus, in addition to considering the license of the software code that makes up the AI model, it is

important to consider the license of the training data used to train the AI model.

Broadly, licenses for open uses of training data fall into four different categories: attribution, share-alike, no derivative works and non-commercial. Attribution licenses, such as CC BY and ODC-By, are similar to permissive licenses for code in that they require attribution to the creator of the data. Share-alike licenses, such as CC BY-SA and ODbL, also require that all modifications be distributed under the same license and are often similar to a reciprocal license for software code. However, in some cases, share-alike licenses may also be restrictive depending on the scope of modification to the data and the use of the data. Unlike share-alike licenses, no derivative works licenses prohibit distribution of any modifications. An example of such licenses is CC BY-ND. In addition to attribution and any applicable share-alike or no derivative works requirements, non-commercial licenses prohibit distribution of modifications for commercial purposes. Examples of such licenses include CC BY-NC, CC BY-NC-SA, and CC-BY-NC-ND.

Lastly, trained AI models may be freely available to the public under the terms of one or more foregoing licenses for software code and training data.

Use of Open Source AI Models in Proprietary Applications

When choosing among freely-available AI models for developing its applications, a company should consider the licensing scheme for one or more of the model code, training datasets, or the trained model.

Choosing model code licensed under a permissive license would be the least risky. Similarly, selecting

model code licensed under a reciprocal license without making any modifications to the source would reduce the risk.

For commercial applications, it is important to try to avoid any licenses for training data or trained models that prohibit commercial use. While the applicability of some non-commercial use restrictions (such as those in CC BY-NC, CC BY-NC-SA, and CC-BY-NC-ND licenses) for using data to train an AI model has not been settled by the courts, it would be less risky to not use training data licensed under such licenses to avoid a legal dispute. In addition, certain training datasets are licensed under custom open-source non-commercial licenses, such as the license for the InsightFace facial recognition model that provides: “The training data containing the annotation (and the models trained with these data) are available for non-commercial research purposes only.” Courts may be less inclined to condone commercial use in face of such clear prohibition.

For similar reasons, it is advisable to try to avoid any share-alike licenses for trained data or trained models as well as any no derivative works licenses in case modifications are made.

For example, DeepSeek has licensed its R1 model under the well-known permissive MIT license. The trained R1 model is licensed under a custom license that permits commercial use and

does not include any share-alike or no derivative works restrictions. DeepSeek’s custom license contains several use restrictions, including restrictions against any military use, any use that would adversely impact an individual’s legal rights, or any use that would otherwise create or modify a binding, enforceable legal obligation. Thus, to promote adoption of its model, DeepSeek has chosen a permissive licensing scheme for the model.

While it may be tempting to select any freely available AI model software for development of proprietary AI applications, it is important to remember that freely available does not necessarily mean free of legal obligations and risks. Care must be taken to analyze, with the help of a qualified lawyer, all licenses under which AI model software is licensed to consider and mitigate potential risks. It is best to perform this assessment as early as possible during the development of the proprietary applications.

Vlad Teplitskiy and **Damien Howard** are partners at intellectual property firm *Knobbe Martens*. Their practices cover strategic patent portfolio development and management, open source software licensing, and emerging technologies. **Fatima Kouadio** is an associate at the firm who assists clients with a variety of intellectual property matters with a focus on patent litigation.