Volume 8, No. 6 | November–December 2025

RAIL

The Journal of Robotics, Artificial Intelligence & Law

fastcase FULL COURT PRESS

RAIL
The Journal of Robotics, Artificial Intelligence & Law
Volume 8, No. 6 | November–December 2025

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at $495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. For customer support, please contact Fastcase, Inc., 729 15th Street, NW, Suite 500, Washington, D.C. 20005, 202.999.4777 (phone), or email customer service at support@fastcase.com.

Publishing Staff
Publisher: David Nayer
Production Editor: Sharon D. Ray
Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

729 15th Street, NW, Suite 500, Washington, D.C. 20005
https://www.fastcase.com/

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 729 15th Street, NW, Suite 500, Washington, D.C. 20005.

Articles and Submissions

Direct editorial inquiries and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc., 26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@meyerowitzcommunications.com, 631.291.5541.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please contact:

David Nayer, Publisher, Full Court Press at david.nayer@vlex.com or at 202.999.4777

For questions or Sales and Customer Service:

Customer Service
Available 8 a.m.–8 p.m. Eastern Time
866.773.2782 (phone)
support@fastcase.com (email)

Sales
202.999.4777 (phone)
sales@fastcase.com (email)

ISSN 2575-5633 (print)
ISSN 2575-5617 (online)

Demystifying Hugging Face Licenses

Hemant Gupta*

In this article, the author details the legal ins-and-outs for the most popular licenses for artificial intelligence models. He notes that the license can determine whether you can publish derivative works, if your research can be commercialized later, and what obligations you have to the original model creators. The author reviews each of the major models and explains what they allow and what they do not, as well as how to find and understand license information.

The rapid advancement of artificial intelligence (AI) research has made model repositories like Hugging Face essential for researchers and developers. However, navigating the complex landscape of licensing can be challenging, especially when exploring models for research or commercial applications. This guide aims to demystify the various license types found on Hugging Face, helping researchers make informed decisions when using these powerful AI models.

Why Licenses Matter in AI Research

Before diving into specific license types, it is important to understand why licenses matter in the AI ecosystem. When integrating a Hugging Face model into your research or commercial project, conducting a thorough review of its licensing terms is essential to ensure compliance and protect your project from legal complications. Licenses dictate how models can be used, modified, and redistributed, with implications for both academic research and commercial applications.

The license attached to an AI model can significantly impact your research trajectory. It determines whether you can publish derivative works, if your research can be commercialized later, and what obligations you have to the original model creators. As AI research increasingly bridges academic and commercial contexts,

Robotics, Artificial Intelligence & Law / November–December 2025, Vol. 8, No. 6, pp. 395–406.

ISSN 2575-5633 (print) /ISSN 2575-5617 (online).

understanding these nuances becomes crucial for planning research that can navigate both worlds effectively.

The Hugging Face Licensing Landscape

The Hugging Face Hub hosts a diverse array of models with various licensing arrangements. A significant portion of models on Hugging Face lack explicit license information, while those that do carry licenses span a wide spectrum from traditional open source licenses to more specialized AI-specific frameworks. Major projects like Midjourney, BLOOM, and Llama often employ custom licensing terms that do not fit traditional open source definitions.

This mixed landscape creates challenges for researchers trying to determine which models they can safely use for different purposes. Unlicensed models present particular risks, as the absence of a license does not imply freedom to use. Conversely, it means all rights are reserved by default. For research that may eventually have commercial applications, this uncertainty can create significant complications down the road.

Let's explore the primary categories of licenses you will encounter on Hugging Face and what they mean for your research.

Traditional Open Source Licenses

Many Hugging Face models use conventional open source licenses originally designed for software. These licenses were created before the rise of modern AI and may not perfectly address all aspects of AI model usage, but they provide established frameworks that many researchers are already familiar with.

Apache 2.0

The Apache 2.0 license is one of the most popular licenses for AI models on Hugging Face. It is a permissive license that allows for commercial use, modification, and distribution of the model. Importantly for AI research, it includes explicit patent grants, which can be critical when working with technologies that may involve patented methods or algorithms.

Under Apache 2.0, you can use models for both academic and commercial research without significant restrictions. However, you must include appropriate attribution, noting the original source of the model. You must also clearly document any significant modifications you make to the original model. For researchers, this means keeping careful records of how you have fine-tuned or adapted models for your specific research questions.

The Hugging Face Transformers library uses the Apache 2.0 license, making it compatible with a wide range of research and commercial applications. Many models contributed directly by the Hugging Face team also use this license, creating a consistent licensing environment for core components of the ecosystem.

MIT License

The MIT (Massachusetts Institute of Technology) license is even more permissive than Apache 2.0 and is extremely popular for data sets on Hugging Face, as well as being the second most common license for models. It allows for virtually any use with minimal restrictions, requiring only that the original copyright notice be preserved when redistributing the model or derivative works.

The simplicity of the MIT license makes it attractive for research projects where wide adoption is a priority. However, it lacks the explicit patent grants found in Apache 2.0, which can be a consideration for research with potential commercial applications. In practice, most researchers find MIT-licensed models suitable for almost any purpose, but consulting with intellectual property experts is advisable if patent concerns exist in your specific domain.

The MIT license's minimal restrictions make it ideal for exploratory research where you may not yet know all potential applications. It gives researchers maximum flexibility to pursue different directions as their work evolves, without worrying about license compatibility issues.

GPL

The GPL (GNU General Public License) is a "copyleft" license that differs significantly from permissive licenses like Apache 2.0 and MIT. When using GPL-licensed models, any derivative works you create must also be released under the GPL, effectively ensuring

that modifications remain open source. This can have significant implications for research that might eventually lead to proprietary applications.

While the GPL includes patent grants similar to Apache 2.0, its viral nature means that researchers planning to commercialize their work or integrate it with proprietary systems may need to carefully consider whether using GPL-licensed models aligns with their long-term goals. In academic contexts focused on open science, the GPL can help ensure that research remains accessible to the community.

The GPL is less common on Hugging Face than permissive licenses, but still appears in certain model families. It represents a philosophical approach to open source that prioritizes continued openness over maximum flexibility, which may align well with research projects committed to open science principles.

AI-Specific Licenses

As AI development has evolved, new license types designed specifically for AI models have emerged to address the unique considerations that traditional software licenses may not fully cover. These AI-specific licenses attempt to balance openness with responsible use concerns.

OpenRAIL

OpenRAIL (Open Responsible AI Licenses) licenses represent an emerging approach to AI licensing that combines open access principles with responsible use requirements. Developed with support from Hugging Face and inspired by initiatives such as BigScience and Creative Commons, OpenRAIL licenses could become for AI what Creative Commons became for content—a widespread community licensing tool.

The fundamental concept behind OpenRAIL is that openness alone is insufficient for responsible AI development. These licenses enable open access, use, and distribution of AI artifacts while requiring responsible use through behavioral restrictions. By April 2023, over 8,000 repositories on Hugging Face were using OpenRAIL licenses, making them the second most used category after permissive open source licenses.

OpenRAIL licenses include use-based restrictions clauses that give model creators better control over how their models are used. These clauses also act as a deterrent against potential misuse. Importantly, OpenRAIL licenses require downstream adoption of these use-based restrictions in derivative works, creating a copyleft-like effect for the ethical guidelines rather than just the code.

Common examples of OpenRAIL licenses include BigScience OpenRAIL-M, CreativeML OpenRAIL-M (used by Stable Diffusion), and BigScience BLOOM RAIL. Each variant maintains the core principles while adapting to specific model domains. For researchers, using models under these licenses means committing to ethical usage guidelines and propagating those commitments in any derivative work.

Unlike traditional open source licenses, RAIL licenses typically require explicit acceptance of terms before downloading a model. This reflects their dual nature as both copyright licenses and behavioral agreements. Researchers should review these terms carefully to ensure their intended uses align with the permitted applications.

Meta's Llama License

Meta's Llama models represent some of the most capable open access models available, but they come with a custom license that has sparked considerable discussion in the research community. Although Meta has described Llama as "open sourced" in marketing materials, the license does not qualify as an open source license according to the Open Source Initiative definition.

The Llama license (across versions from Llama 2 to the recent Llama 3.1 and 3.2) grants a nonexclusive, worldwide, nontransferable and royalty-free limited license to use, reproduce, distribute, copy, create derivative works of, and modify the Llama materials. However, it includes several notable restrictions that researchers should be aware of.

First, the license requires prominent attribution through phrases like "Built with Llama" on websites, user interfaces, or documentation. Any derivative model must include "Llama" at the beginning of its name, which can have implications for research branding and identity.

Second, the license prohibits using Llama models or their outputs to improve competing models. This restriction is particularly

significant for comparative research or work on model alignment and improvement techniques that might span multiple model families.

Third, organizations with more than 700 million monthly active users must request special permission from Meta to use the models, which could impact large-scale collaborative research involving major institutions.

For most academic researchers, these restrictions will not present immediate obstacles, but they do need to be considered when planning the trajectory of research that might eventually scale or transition to commercial applications. The license terms have evolved across Llama versions, so researchers should always check the specific terms for the version they are using.

Creative Commons Licenses

Some Hugging Face content uses Creative Commons (CC) licenses, which were originally designed for creative works rather than software or AI models. These licenses include variations such as CC BY (Attribution), CC BY-SA (Attribution-ShareAlike), and CC BY-NC (Attribution-NonCommercial).

The Open Source Initiative does not consider CC licenses to be open source licenses. They are generally not recommended for software or AI models because they were not written with code distribution in mind, creating legal ambiguity about whether they cover the source code, the model weights, or just the outputs.

For researchers, CC licenses on models present particular challenges because of this ambiguity. The CC-NC (NonCommercial) variant especially can create uncertainty about what constitutes commercial use in research contexts. Does publishing in a journal with a commercial publisher count? What about research funded by industry grants? These questions lack clear answers under CC licensing frameworks applied to AI models.

When encountering CC-licensed models on Hugging Face, researchers should proceed with caution and consider consulting legal experts if the research has potential commercial dimensions or involves partnerships with industry. In purely academic contexts with no commercial element, these licenses generally permit research use with appropriate attribution.

How to Find and Understand License Information

Navigating license information on Hugging Face requires knowing where to look and how to interpret what you find. The platform has made efforts to standardize license information, but inconsistencies still exist across the thousands of hosted models.

The primary place to check for license information is the model card, which is essentially the README.md file in the repository. Hugging Face encourages model publishers to include license information in the metadata section of this file. This standardized location makes it easier to quickly assess whether a model's license aligns with your research needs.

Many repositories also include a separate LICENSE file with the complete license text. This is particularly important for custom licenses or when specific modifications have been made to standard licenses. Reading the full license text is always recommended for models central to your research, as summarized license descriptions may not capture important nuances.

For models that originated outside of Hugging Face, it is wise to check the original repository or publication for authoritative license information. Sometimes models are re-uploaded to Hugging Face with incomplete license information, creating potential compliance risks.

If license information is unclear or missing entirely, the safest approach is to contact the model creators directly for clarification. Most researchers in the AI community are responsive to licensing questions, as they want their work to be used appropriately. Hugging Face also provides mechanisms to reach out to model publishers through the platform.

A practical tip for researchers is to document license information at the time you download or begin using a model. License terms occasionally change, and having records of the specific terms under which you began your research can be valuable if questions arise later.

Key Considerations for Researchers

When selecting models for your research, several license-related factors should inform your decision-making process to ensure both legal compliance and alignment with your research goals.

Commercial Use Permissions

Understanding whether a license permits commercial use is essential, even for academic researchers. Research often spans boundaries between academic and commercial contexts through industry collaborations, start-up spin-offs, or technology transfer. A model perfectly suitable for publication may create complications if your research leads to commercial applications later.

Some licenses, like MIT and Apache 2.0, place virtually no restrictions on commercial use. Others, like CC-NC or certain custom licenses, explicitly prohibit commercial applications. Still others, like the Llama license, allow commercial use but impose restrictions based on organization size or other factors. Carefully consider your research's potential trajectory before committing to models with commercial use limitations.

Modification Rights and Obligations

AI research frequently involves fine-tuning, adapting, or extending existing models. All licenses on Hugging Face permit some level of modification, but they differ in what obligations attach to those modifications. Under permissive licenses like MIT and Apache 2.0, you have few obligations beyond attribution. Under copyleft licenses like GPL, you must release modifications under the same license terms.

AI-specific licenses often have unique requirements for modifications. OpenRAIL licenses require maintaining ethical use restrictions in derivative works. The Llama license requires naming conventions for derivative models and prohibits using modifications to improve competing models. Consider how these obligations align with your research publication plans and collaboration strategies.

Attribution Requirements

All licenses on Hugging Face require some form of attribution, but the specifics vary considerably. MIT and Apache 2.0 licenses have relatively straightforward attribution requirements, typically satisfied by preserving copyright notices and acknowledging the original model in publications.

AI-specific licenses often have more detailed attribution requirements. The Llama license, for instance, requires prominently displaying "Built with Llama" and including "Llama" in derivative model names. These requirements may affect how you present your research and could have implications for building a distinct research identity if your work heavily modifies an existing model.

Patent Considerations

Patent protection is particularly relevant for AI research, as the field involves many potentially patented techniques. Licenses differ significantly in how they address patents. Apache 2.0 includes explicit patent grants, providing some protection against patent claims from contributors. MIT does not address patents explicitly, potentially leaving researchers more exposed to patent risks.

For research in areas with active patent landscapes, such as specific machine learning optimization techniques or application domains like healthcare, consulting with intellectual property experts may be advisable when using models with licenses that lack explicit patent provisions.

Ethical Use Restrictions

The emergence of AI-specific licenses like OpenRAIL reflects growing concern about potential misuse of powerful AI models. These licenses include specific restrictions on applications considered harmful or unethical. While most academic research easily falls within ethical boundaries, researchers working in sensitive areas like security, biology, or content generation should carefully review any ethical use restrictions.

These restrictions can sometimes be subjectively worded, creating interpretation challenges. When in doubt about whether your research falls within permitted uses, reaching out to model creators for clarification is the most prudent approach.

Best Practices for Researchers

Navigating the complex licensing landscape of Hugging Face models requires thoughtful practices. Following are expanded recommendations for researchers working with these models.

Document License Compliance Throughout Your Research

Maintaining comprehensive records of the models you use and their associated licenses is essential for both current compliance and future flexibility. Create a standardized way to document this information in your research materials, including specific version information and the date you accessed the model. This documentation proves valuable if questions arise about compliance or if you need to evaluate whether new research directions remain compatible with your existing model licenses.

Consider creating a license inventory document for each research project, detailing all models used, their licenses, key restrictions, and compliance notes. This practice is particularly important for collaborative research where multiple team members may incorporate different models.

Implement Attribution Best Practices

Proper attribution is a universal requirement across all licenses, but implementing it well requires attention to detail. Develop consistent practices for attributing models in your research outputs, including papers, presentations, and code repositories. Include specific model versions and access dates, not just model names, as licenses and models can change over time.

For models with specific attribution requirements, like the Llama family, ensure these requirements are satisfied across all relevant materials. When using multiple models with different attribution requirements, create a comprehensive attribution section that satisfies all applicable licenses simultaneously.

Ensure License Compatibility Across Your Technology Stack

Research often involves combining multiple models or integrating models with other software components. When doing this, verify that the licenses of all components are compatible with each other and with your intended use cases. License incompatibilities can emerge in subtle ways, particularly when combining permissively licensed components with those under copyleft or custom licenses.

Create a visual map of your technology stack with license information for each component to identify potential conflicts early. Pay particular attention to the flow of data and code between components under different licenses, as this is where compatibility issues typically arise.

Plan for License Transitions

Research evolves, and so do licensing needs. What begins as purely academic research may develop commercial potential, requiring reevaluation of licensing choices. Develop contingency plans for how you would handle license transitions if your research changes direction. This might include having alternative models identified with more flexible licenses that could substitute for restricted ones if needed.

For long-term research programs, schedule periodic license reviews to ensure continued compliance as both your research and the licensing landscape evolve. This is particularly important for rapidly developing areas where new models with different licensing terms regularly emerge.

Consider the Licensing Impact on Research Reproducibility

The reproducibility of AI research depends partly on other researchers' ability to access and use the same models you employed. When selecting models, consider whether their licenses might impede reproducibility. Models requiring explicit acceptance of terms or with commercial restrictions might create barriers for other researchers trying to verify or build upon your work.

In publications, clearly document any license-related steps required for reproducibility, such as acceptance of terms or attribution requirements. Where possible, provide alternative approaches using more openly licensed models as a fallback for reproducibility.

Contribute to Licensing Clarity in the Community

As a researcher, you can help improve the licensing clarity of the Hugging Face ecosystem. When publishing your own models, choose clear licenses and document them thoroughly. When using

models with unclear licensing, reach out to creators for clarification and encourage them to update their documentation.

Consider sharing your license compliance approaches and documentation templates with the community to help establish best practices. As AI licensing continues to evolve, collaborative development of community standards will benefit all researchers.

Conclusion

Understanding Hugging Face licenses is crucial for researchers navigating the AI ecosystem. By familiarizing yourself with the various license types and their implications, you can make informed decisions about which models to use in your research while ensuring compliance with legal requirements and respecting the intentions of model creators.

What remains constant in the developing field of AI licensing is the importance of transparency, responsibility, and respect for the intellectual property that drives innovation in AI research. By adopting thoughtful licensing practices, researchers can help build a sustainable ecosystem that promotes both innovation and responsible use of increasingly powerful AI technologies.

Note

* The author, an attorney with Epstein Becker & Green, P.C., may be contacted at hgupta@ebglaw.com.