Responsible AI Licenses (RAIL): Here’s What You Need to Know

AJ STARITA, MAY 22, 2024

#Application Security

Responsible AI Licenses (RAIL) are a class of licenses created with the intention of preventing harmful or unethical uses of artificial intelligence while also allowing for the free and open sharing of models between those who intend to use and improve them for authorized purposes. Anyone can make their own version of RAIL for their model, and in doing so can create more or less restrictions than those detailed in the template licenses. The RAIL initiative exists to help people create ethical licenses, but they don’t govern each license. If you’ve got any AI license that includes restrictions that do the following:

- Limit behavior use
- Require these use restrictions to be passed on to any derivatives of the model

According to the RAIL initiative, you’ve got RAIL!

At the time of this writing, there are just under 50,000 models (and over 5,000 data sets) on Hugging Face that fall under some form of RAIL license[1]. To put that into perspective, that’s about 10,000 more than the number of models using the MIT license and a bit more than half of the number of models that fall under the Apache 2.0 license. Famous AI models covered under RAIL licenses include Stable Diffusion and BLOOM.

In this blog we’ll answer some common questions about RAIL licenses. The licenses themselves are dry, sure, but they’re perfectly readable by a non-lawyer so be sure to look through any specific licenses that cover models you want to use. Speaking of non-lawyers, we are not legal professionals and nothing in this blog should be construed as legal advice. If you need legal help, g the pros!

Are RAIL truly open source licenses?

Not really. Traditional open source licenses, whether permissive like the MIT license or copyleft like the GPL, are based strictly in copyright law. RAIL licenses combine copyright permissions with an agreement on how you can use the model, which is why you typically have to accept the RAIL terms on Hugging Face *before* you can even download the model—not something you would normally do when procuring open source software.

The RAIL Initiative-governed RAIL-A license comes in two parts, an end-user agreement and a source code license. Both reiterate each other's use restrictions. Some (like the creators and users of the licenses), certainly call it responsible, while others might call it paternalistic. Either way, use restrictions are not typical of open source licenses, which have generally been written with the philosophy that software users should enjoy the freedom to use software however they please, even if some might find those uses objectionable.

[1] Yes, yes, we know we're technically saying "licenses license" here. Go kick an ATM machine, pedant.

What kinds of uses are restricted by RAIL licenses?

Many of the things that you are not allowed to do within the terms of these licenses would already be forbidden by law in most countries, so there's some redundancy given you're not allowed to break the law regardless of what a license says. But you're generally also not permitted to use RAIL-licensed models to *administer* the law. Each of the RAIL family licenses varies somewhat on the exact restrictions here. For instance, RAIL-A gives a specific list of things you cannot do, whereas BigScience Open RAIL-M's criminal justice use restriction is much broader, stating that you are not permitted to "generate or disseminate information for the purpose to be used for administration of justice, law enforcement, immigration or asylum processes, such as predicting an individual will commit fraud/crime commitment (e.g., by text profiling, drawing causal relationships between assertions made in documents, indiscriminate and arbitrarily-targeted use)."

Another interesting specific restriction is that you are not permitted to use a model under the BigScience Open RAIL-M to "provide medical advice and medical results interpretation". This is probably intended to mean to a non-medical professional end user, but could potentially also restrict an application that would be used by a doctor. The RAIL-A license is clearer, and while it includes a restriction on providing medical diagnoses without human oversight, it's mostly concerned with the use of the model to deny insurance claims or coverage.

The vast majority of restrictions would seem acceptable to reasonable people but as with all legal things, it's all in the interpretation. Some examples from the BigScience version: you are not allowed to use the model to discriminate against people based on certain characteristics, harm or defame people, spread misinformation, make deep fakes, or generate or release PII in order to harm people.

Some restrictions might even preclude non-nefarious uses. The RAIL-A license, for instance, prohibits you from using a model to "[d]etect or infer aspects and/or features of an identity any person, such as name, family name, address, gender, sexual orientation, race, religion, age, location (at any geographical level), skin color, society or political affiliations, employment status and/or employment history, and health and medical conditions." It's not difficult to imagine antisocial applications for detecting things in that list, but there's plenty of prosocial applications that would be restricted as well.

Are RAIL licenses copyleft or permissive?

Neither. Both. It's complicated.

Strictly on the basis of code copyright and patent use, they are permissive—that is, if they're written to be. There's nothing in the template license nor any of the major RAIL licenses on Hugging Face that says you must distribute your derivative work or source code in any particular way. But be sure to check the particular license, because there's also nothing restricting someone from creating their own RAIL license and making the copyright assignments research-only or even copyleft.

On the basis of use, they are by design *not* permissive. In creating a derivative work, you may be able to change the copyright permissions, depending on the original license, but you must pass the use restriction parts of the agreement downstream to any user of your work or anyone who makes their own derivative work based on yours. In that way, the use restrictions spread a particular philosophy "virally" somewhat like a copyleft license does.

What’s in a RAIL name?

The capital letter(s) at the tail of RAIL license names are based on what is covered by the use restrictions in the license. The most common is “M” for “model”. The entire list is as follows:

D – data

A – application/executable

M – model

S – source code

By this convention, a license that restricts more than one aspect should be named by including a dash and the appropriate letters after “RAIL”. For example,. a license that restricts all four might be called “MyProject RAIL-DAMS”.

Another convention is to put “open” in the name if the copyright part of the agreement is like that of a traditional open source license. So your license that allows for royalty-free use but restricts how the data and application are used might be called “MyProject Open RAIL-DA”.

Which RAIL licenses are indexed on Hugging Face?

Hugging Face currently indexes projects by license under the following RAIL licenses:

- OpenRAIL license family
- BigScience OpenRAIL-M
- CreativeML OpenRAIL-M
- BigScience BLOOM RAIL
- BigCode Open RAIL-M v1
- Open Rail++-M License (also called CreativeML OpenRAIL++-M)

Some projects on Hugging Face use lesser known RAIL licenses but they are indexed under “other” licenses.

Parting words

Hopefully you found this guide to the world of RAIL licenses helpful. If there’s anything else about this license family that you think we should cover, reach out to us!

Meet The Author

AJ STARITA
AJ Starita is fascinated by the challenges and triumphs of cybersecurity and open source software. When not writing about technology, AJ can usually be found exploring nature or reading detective novels.

To All Articles