GENERATIVE AI COMMONS

Practical Tools for Facilitating Openness in AI R&D:

The Model Openness Framework and OpenMDW License

Cailean Osborne, PhD

SciLifeLab AI Seminar Series

September 26, 2025

LF AI & DATA

Cailean Osborne, PhD

Senior Researcher,
Linux Foundation

Agenda

1. **Openness Challenges in AI**
2. Model Openness Framework
3. OpenMDW Licence
4. Discussion

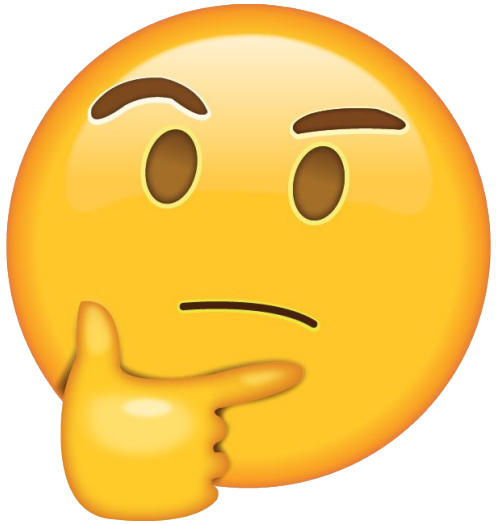
Openness Challenges in AI

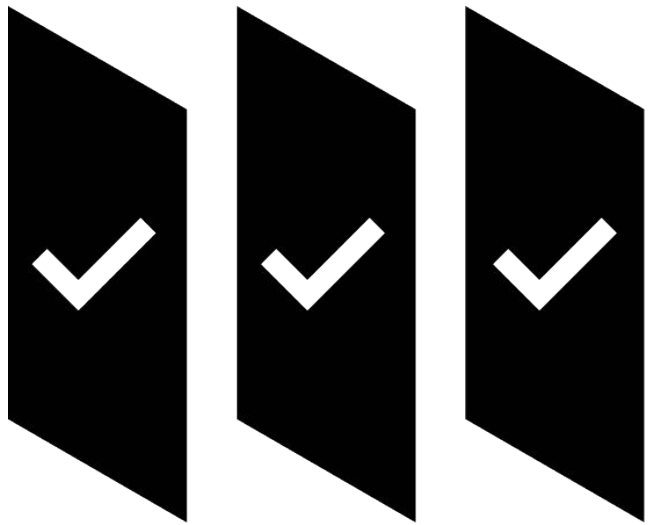
- Disagreement about **what is openness in AI**
- **Open-washing** (e.g. use of non-OSD-compliant licenses)
- Beyond weights, **few artifacts from model development** are released (e.g. training code, datasets)
- **Licensing issues** (e.g. OSS licenses, license-switching)
- **EU AI Act has exemptions** for AI systems and general-purpose AI models “released under a free and open source licence.” But which licenses are compliant?
- Not just an academic problem: **~5,000 models uploaded every day** to Hugging Face Hub 🤗

Agenda

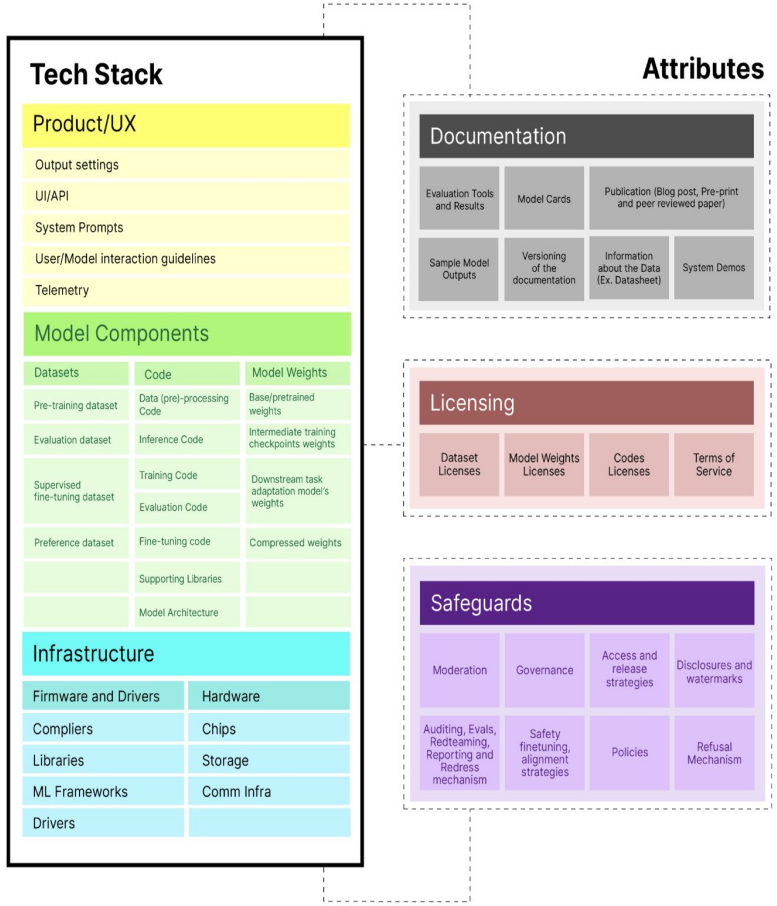
1. Openness Challenges in AI
2. **Model Openness Framework**
3. OpenMDW Licence
4. Discussion

What does it mean to be open in AI?



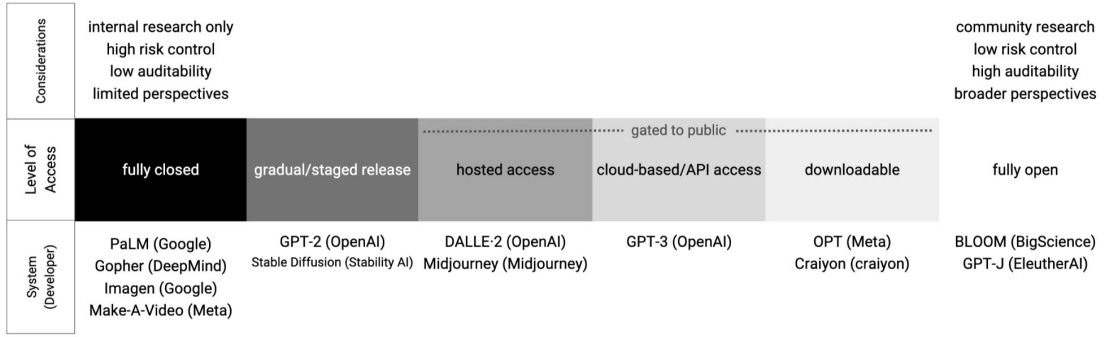


MODEL OPENNESS FRAMEWORK



Framework for Openness in Foundation Models – Columbia Convening on Openness & AI

Gradient of Generative AI Release / Access – Irene Solaiman



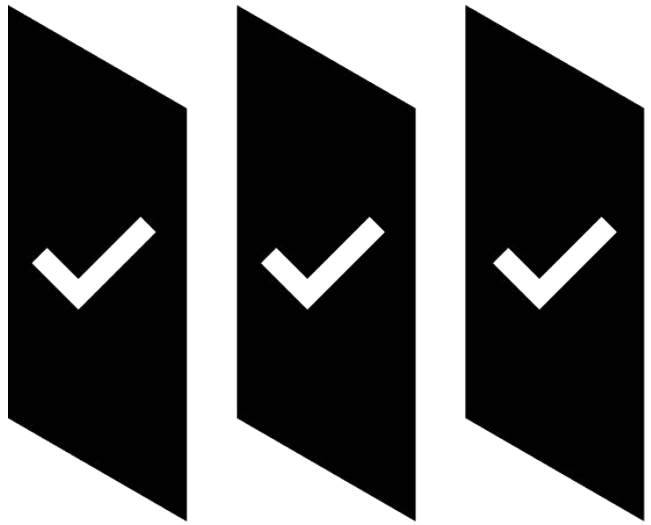
Ten Simple Rules for Good Model-Sharing Practices

1. Define what you mean by "model"
2. Involve the community in informing and promoting model-sharing practices
3. Acknowledge diverse contributions
4. Provide accessible documentation for the appropriate audience
5. Embrace FAIR principles for sharing models
6. Publicly recognize and reward research software engineers
7. Deploy user-friendly tools for collaborative modelling practices
8. Influence publishers to promote good model-sharing practices
9. Break down silos
10. Don't wait for perfection when sharing models

10 Simple Rules – Kherroubi Garcia et al.

open source initiative®

Open Source AI Definition – OSI



MODEL OPENNESS FRAMEWORK

Completeness vs. Openness

- **Completeness:** In open science, completeness refers to providing comprehensive and well-documented information for all artifacts, ensuring that each element is thorough, self-contained, and meaningfully usable without requiring additional context or resources.

- **Openness**: Binary property indicating whether a particular artifact is licensed under an open license or not. An artifact is considered *open* if and only if it is distributed under a license that grants users the rights to freely access, use, modify, and share the artifact. If an artifact is not licensed under an open license, it is not *open*.

Breakdown of Machine Learning Models

CODE


Model Architecture


Preprocessing Code

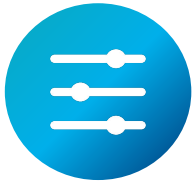

Training Code


Inference Code


Evaluation Code


Libraries & Tools

DATA


Model Parameters


Training Datasets


Evaluation Data


Sample Model Outputs


Model Metadata


MOF Config File

DOCS


Model Card


Data Card


Research Paper


Technical Report


Evaluation Results

Class	Usage	Artifacts
Class I Open Science		
Class II Open Tooling		
Class III Open Model		

Class	Usage	Artifacts
Class I Open Science		
Class II Open Tooling		
Class III Open Model	• Unrestricted usage (access, use, modify, redistribute) • Create a product or service • Finetune and align • Model optimizations	

Class	Usage	Artifacts
Class I Open Science		
Class II Open Tooling		
Class III Open Model	• Unrestricted usage (access, use, modify, redistribute) • Create a product or service • Finetune and align • Model optimizations	• Model Architecture • Model Parameters (final checkpoint) • Technical Report • Evaluation Results • Model Card • Data Card • Sample Model Outputs (optional)

Class	Usage	Artifacts
Class I Open Science		
Class II Open Tooling	• Understand training process • Validate benchmark claims • Inference optimizations	
Class III Open Model	• Unrestricted usage (access, use, modify, redistribute) • Create a product or service • Finetune and align • Model optimizations	• Model Architecture • Model Parameters (final checkpoint) • Technical Report • Evaluation Results • Model Card • Data Card • Sample Model Outputs (optional)

Class	Usage	Artifacts
Class I Open Science		
Class II Open Tooling	• Understand training process • Validate benchmark claims • Inference optimizations	• Training Code • Inference Code • Evaluation Code • Evaluation Data • Supporting Libraries & Tools (optional) • All Class III Components
Class III Open Model	• Unrestricted usage (access, use, modify, redistribute) • Create a product or service • Finetune and align • Model optimizations	• Model Architecture • Model Parameters (final checkpoint) • Technical Report • Evaluation Results • Model Card • Data Card • Sample Model Outputs (optional)

Class	Usage	Artifacts
Class I Open Science	• End to end analysis and auditing • Reproduction of a similar model • Data exploration and experimentation	
Class II Open Tooling	• Understand training process • Validate benchmark claims • Inference optimizations	• Training Code • Inference Code • Evaluation Code • Evaluation Data • Supporting Libraries & Tools (optional) • All Class III Components
Class III Open Model	• Unrestricted usage (access, use, modify, redistribute) • Create a product or service • Finetune and align • Model optimizations	• Model Architecture • Model Parameters (final checkpoint) • Technical Report • Evaluation Results • Model Card • Data Card • Sample Model Outputs (optional)

Class	Usage	Artifacts
Class I Open Science	• End to end analysis and auditing • Reproduction of a similar model • Data exploration and experimentation	• Research Paper • Datasets • Data Preprocessing Code • Model Parameters (intermediate checkpoints) • Model Metadata (optional) • All Class II Components
Class II Open Tooling	• Understand training process • Validate benchmark claims • Inference optimizations	• Training Code • Inference Code • Evaluation Code • Evaluation Data • Supporting Libraries & Tools (optional) • All Class III Components
Class III Open Model	• Unrestricted usage (access, use, modify, redistribute) • Create a product or service • Finetune and align • Model optimizations	• Model Architecture • Model Parameters (final checkpoint) • Technical Report • Evaluation Results • Model Card • Data Card • Sample Model Outputs (optional)

Recommended Licenses

COMPONENT	DOMAIN	CONTENT TYPE	RECOMMENDED LICENSE
Datasets	Data	Data	Preferred: CDLA-Permissive-2.0
Data Preprocessing Code	Data	Code	Acceptable: OSI-approved
Model Architecture	Model	Code	Acceptable: OSI-approved
Model Parameters	Model	Data	Recommended: OpenMDW
Model Metadata	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: Permissive Open Data Licenses
Training Code	Model	Code	Acceptable: OSI-approved
Inference Code	Model	Code	Acceptable: OSI-approved
Evaluation Code	Model	Code	Acceptable: OSI-approved
Evaluation Data	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: Permissive Open Data Licenses
Evaluation Results	Model	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Supporting libraries and Tools	Model	Code	Acceptable: OSI-approved
Model Card	Model	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Data Card	Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Technical Report	Model & Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Research Paper	Model & Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Sample Model Outputs	Model	Data or Code	Unlicensed

Challenges with the MOF

- Complicated to implement for model producers
  - Content-appropriate license needed to attribute to each artifact
  - LICENSE files distributed over different locations in repo
  - Need to submit model to MOT to generate MOF.JSON file

- Difficult to understand for model consumers
  - Most consumers do not understand license implications
  - Complexity could lead to unnoticed license changes for an artifact

Agenda

1. Openness Challenges in AI
2. Model Openness Framework
3. **OpenMDW Licence**
4. Discussion

Licensing Challenges with Open Models

- Open source software licenses were designed for software
- ML models are more complex and don't directly map to software components
- Model weights cannot be copyrighted or patented
- Limited rulings if training on copyrighted data can be considered fair use
- In some case, models with restrictive licenses are being fine-tuned and license converted to permissive
- Many publicly available foundation models available under restrictive licenses

Common Open vs. Proprietary/Restrictive Licenses

Open	Proprietary/Restrictive
Apache 2.0 (Pythia, MPT, Mis/xtral, RWKV, Gemma*)	Open RAIL (BLOOM/Stable Diffusion)
MIT (GPT-2, Dolly)	Llama Community License Agreement
BSD-3-Clause	DBRX Model License (DBRX)
CDLA-Permissive-2.0	TII License (Falcon)
CC-BY-SA-4.0 (StableLM-Alpha)	Google Gemma AUP (Gemma weights)

OpenMDW in a Nutshell 🥜

1️⃣ First permissive license designed for ML models

📦 Global license for models + code, data, and documentation

🪜 Reuses MIT, CDLA-Permissive-2.0, CC-BY-4.0

🎛️ Compatible with other licenses used for code, data, docs

🔐 Grants rights under copyright, patent, database, and trade secret laws

🧾 Patent litigation protection and attribution

🖨️ No restrictions on model outputs

🇪🇺 Complies with EU AI Act def of "free and open source license"

Comparison of licenses

Aspect	OpenMDW	MIT	Apache 2.0	OpenRAIL
Version	Dated April 23, 2025	-	2.0	Dated August 22, 2022
Scope	**Models, Software, Data, Documentation**	**Software-only**	**Software-only**	**Models**
Length	**Short**	**Very short**	**Moderate**	**Long**
OSD Conformant	**Targeting OSD conformance**	**Yes**	**Yes**	**No**
MOF Compatible	**Yes (for all components)**	**Yes (for source code)**	**Yes (for source code)**	**No (not permissive license)**
Patent Grant	Yes	No	Yes	Yes
Attribution Requirement	Yes	Yes	Yes	Yes
Modification (of distribution)	Allowed	Allowed	Allowed	Allowed
Distribution	Allowed	Allowed	Allowed	Allowed
Warranty	Disclaimed	Disclaimed	Disclaimed	Disclaimed
Liability	Limited	Limited	Limited	Limited
Specific AI Model Disclaimer	**Yes**	**No**	**No**	**Yes**
Data Disclaimer	**Yes**	**No**	**No**	**Data not licensed under this license**
Indemnification Clause	No	No	No	Optional for redistributors
Commercial Use	Allowed	Allowed	Allowed	Allowed
Patent Litigation Clause	**Yes**	**No**	**Yes**	**Yes**
Contribution Definition	Not included	Not included	Included	Included
Notice Preservation	Yes	Required	Required	Required
Use-based Restrictions	**No**	**No**	**No**	**Yes**
Output Rights	**No restrictions**	**Not specified**	**Not specified**	**Licensor claims no rights, user is accountable**
Updates and Runtime Restrictions	Not specified	Not specified	Not specified	Licensor reserves right to restrict usage and update
Ethical Considerations	Not specified	Not specified	Not specified	Explicit restrictions on harmful uses

One more reason why model licensing matters:

EU AI Act

AI Act Exemptions for AI Systems and GPAI Models if...

- "**Does not apply to AI systems or AI models**, including their output, specifically developed and put into service **for the sole purpose of scientific research and development."** (Article 2:6)
- **"Does not apply to AI systems released under free and open-source licences**, unless they are placed on the market or put into service as high-risk AI systems or an AI system that falls under Art 5 or 50." (Article 2:12)
- **"Shall not apply to providers of AI models that are released under a free and open-source licence** that allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, information on the model architecture, and information on model usage, are made publicly available. This exception shall not apply to GPAI models with systemic risks." (Article 53:2)

AI Act guide for open source AI developers

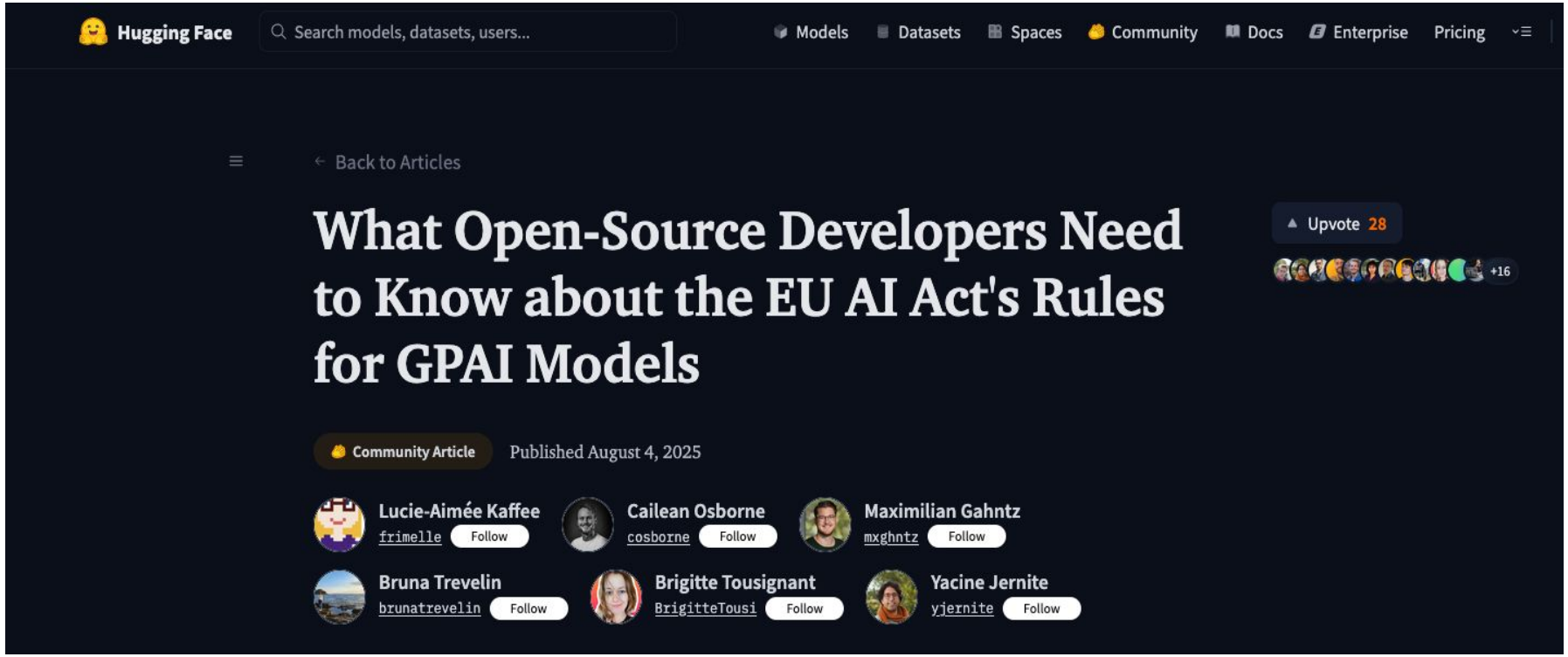
Home / Blog / What Open Source Developers Need to Know about the EU AI Act

28 MIN READ

What Open Source Developers Need to Know about the EU AI Act

CAILEAN OSBORNE | 03 APRIL 2025

AI Act guide for open source GPAI developers



Agenda

1. Openness Challenges in AI
2. Model Openness Framework
3. OpenMDW Licence
4. **Discussion**

Thanks for listening!

OpenMDW License Agreement v1

Guide to EU AI Act (General)

Model Openness Framework

Guide to EU AI Act (GPAIM)

OpenMDW: Deeper Dive

OpenMDW: Paragraph 1

*"By exercising rights granted to you under this agreement, **you accept and agree** to its terms."*

Because the Model Materials include some content which is likely uncopyrightable (e.g. parameters), some assumptions of open source software licenses might not apply.

This sentence makes it clear that exercising rights granted by OpenMDW is an act of *accepting* the OpenMDW agreement terms.

OpenMDW: Paragraph 2

*"As used in this agreement, "**Model Materials**" means the materials provided to you under this agreement, consisting of: (1) one or more machine learning **models** (including architecture and parameters); and (2) **all related artifacts** (including associated data, documentation and software) **that are provided to you hereunder**."*

The Model Materials — the licensed "stuff" — includes the model itself, as well as any related artifacts if they are also provided in the distribution under OpenMDW.

OpenMDW: Paragraph 3

"Subject to your compliance with this agreement, ***permission is hereby granted****, free of charge,* ***to deal in the Model Materials without restriction****, including under all copyright, patent, database, and trade secret rights included or embodied therein."*

If you comply with OpenMDW's terms, then you can use the Model Materials without any restrictions. This explicitly includes where it would otherwise require permission under copyright, patent, database, or trade secret rights.

OpenMDW: Paragraph 4

*"**If you distribute** any portion of the Model Materials, **you shall retain** in your distribution (1) **a copy of this agreement**, and (2) **all copyright notices and other notices** of origin included in the Model Materials that are applicable to your distribution."*

Redistributions must retain the OpenMDW agreement text, together with any applicable attribution notices.

OpenMDW: Paragraph 5

*"If you file, maintain, or voluntarily participate in **a lawsuit** against any person or entity **asserting that the Model Materials** directly or indirectly **infringe any patent**, **then all rights and grants made to you hereunder are terminated**, unless that lawsuit was in response to a corresponding lawsuit first brought against you."*

If a recipient files a lawsuit alleging patent infringement by the Model Materials, then that recipient can no longer benefit from the rights granted by OpenMDW for those Model Materials.

OpenMDW: Paragraph 6

*"This agreement **does not impose any restrictions** or obligations with respect to **any use, modification, or sharing of any outputs generated** by using the Model Materials."*

OpenMDW expressly confirms that outputs generated by using the Model Materials are not subject to restrictions or obligations.

For instance, the attribution requirement for redistributing the *Model Materials* would not apply for redistributing their *outputs*.

OpenMDW: Paragraphs 7-9

*[Three paragraphs of **ALL-CAPS LEGALESE**]*

OpenMDW makes clear the intent that providers of the Model Materials disclaim all warranties and liabilities arising from the recipients' use, with particular reference to aspects relating to AI models and their components and uses.