CAN WE USE OPEN SOURCE LICENSES FOR ML/AI MODELS?

Mary Hardy
Senior Corporate Counsel in Open Source, Standards, and Open Machine Learning at Microsoft
MODEL LICENSES

https://huggingface.co/models, accessed 9/22/2023
OPEN SOURCE MACHINE LEARNING

Microsoft Bing Image Creator, “a data scientist handing an envelope to another data scientist, pixel art”
ML CORE COMPONENTS

MODEL
MODEL PARAMETERS
TRAINING DATASET
EVALUATION DATASET
TRAINING CODE?
WHAT IS A MODEL MADE OF?
WHAT ARE WEIGHTS MADE OF?

MODEL PARAMETERS
WHAT ARE DATASETS MADE OF?

TRAINING DATASET

EVALUATION DATASET
OPEN SOURCE LICENSE REVIEW
“...Work and such Derivative Works in Source or Object form.”
"...Work and such Derivative Works in Source or Object form."

Work = work of authorship, whether in Source or Object form, made available under the License

Derivative Works = any work, whether in Source or Object form, that is based on (or derived from) the Work

Source = the preferred form for making modifications

Object = any form resulting from mechanical transformation or translation of a Source form

Source examples: “...software source code, documentation source, and configuration files.”

Object examples: “...compiled object code, generated documentation...”
APACHE-2.0 [January 2004]

“...Work and such Derivative Works in Source or Object form.”

Work = work of authorship, whether in Source or Object form, made available under the License

Derivative Works = any work, whether in Source or Object form, that is based on (or derived from) the Work

Source = the preferred form for making modifications

Object = any form resulting from mechanical transformation or translation of a Source form

Source examples: “...software source code, documentation source, and configuration files.”

Object examples: “...compiled object code, generated documentation...”

MODEL
APACHE-2.0 [January 2004]

“...Work and such Derivative Works in Source or Object form.”

Work = work of authorship, whether in Source or Object form, made available under the License
Derivative Works = any work, whether in Source or Object form, that is based on (or derived from) the Work

Source = the preferred form for making modifications
Object = any form resulting from mechanical transformation or translation of a Source form

Source examples: “...software source code, documentation source, and configuration files.”
Object examples: “...compiled object code, generated documentation...”

MODEL
PARAMETERS
“...Work and such Derivative Works in Source or Object form.”

Work = work of authorship, whether in Source or Object form, made available under the License
Derivative Works = any work, whether in Source or Object form, that is based on (or derived from) the Work
Source = the preferred form for making modifications
Object = any form resulting from mechanical transformation or translation of a Source form

Source examples: “…software source code, documentation source, and configuration files.”
Object examples: “…compiled object code, generated documentation…”
MIT [1983? 1987?]
MIT [1983? 1987?]

Software = “…this software and associated documentation files (the “Software”),”
Software = “…this software and associated documentation files (the “Software”),”
Software = “...this software and associated documentation files (the “Software”),”

MODEL
PARAMETERS
Software = “...this software and associated documentation files (the “Software”),”
GPL-2.0 [June 1991]
GPL-2.0 [June 1991]

“...any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License...”

Program = any such program or work
GPL-2.0 [June 1991]

“...any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License...”

Program = any such program or work
GPL-2.0 [June 1991]

“...any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License...”

Program = any such program or work
GPL-2.0 [June 1991]

“...any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License...”

Program = any such program or work

Source code = preferred form of the work for making modifications to it.
"...any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License..."

Program = any such program or work

Source code = preferred form of the work for making modifications to it.
GPL-2.0 [June 1991]

“...any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License...”

Program = any such program or work

Source code = preferred form of the work for making modifications to it.

Distribute object code or executable form? “Accompany it with the complete corresponding machine-readable source code”
GPL-2.0 [June 1991]

“...any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License...”

Program = any such program or work

Source code = preferred form of the work for making modifications to it.

Distribute object code or executable form? “Accompany it with the complete corresponding machine-readable source code”

Complete source code = all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable.
GPL-3.0 [June 2007]

Program = any copyrightable work licensed under this License.
GPL-3.0 [June 2007]

Program = any copyrightable work licensed under this License.

Source code = preferred form of the work for making modifications to it.
GPL-3.0 [June 2007]

Program = any copyrightable work licensed under this License.

Source code = preferred form of the work for making modifications to it.

Modify = “...to copy from or adapt all or part of the work in a fashion requiring copyright permission...”
GPL-3.0 [June 2007]

Program = any copyrightable work licensed under this License.

Source code = preferred form of the work for making modifications to it.

Modify = “...to copy from or adapt all or part of the work in a fashion requiring copyright permission...”

Distribute object code? “...also convey the machine-readable Corresponding Source under the terms of this License.”
GPL-3.0 [June 2007]

Program = any copyrightable work licensed under this License.

Source code = preferred form of the work for making modifications to it.

Modify = “…to copy from or adapt all or part of the work in a fashion requiring copyright permission…”

Distribute object code? “…also convey the machine-readable Corresponding Source under the terms of this License.”

Corresponding Source = all the source code needed to generate, install, and run the object code and to modify the work, including scripts to control those activities.
GPL-3.0 [June 2007]

Program = any copyrightable work licensed under this License.

Source code = preferred form of the work for making modifications to it.

Modify = “…to copy from or adapt all or part of the work in a fashion requiring copyright permission…”

Distribute object code? “…also convey the machine-readable Corresponding Source under the terms of this License.”

Corresponding Source = all the source code needed to generate, install, and run the object code and to modify the work, including scripts to control those activities.
AFL-3.0 [2005]
"… any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice…”
“... any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice...”

Source code = preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work
AFL-3.0 [2005]

“... any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice...”

Source code = preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work
“... any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice...”

Source code = preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work.
“... any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice...”

Source code = preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work

Distribute object code? “...provide a machine-readable copy of the Source Code of the Original Work”
"... any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice...”

Source code = preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work

Distribute object code? “...provide a machine-readable copy of the Source Code of the Original Work”
“… any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice…”

Source code = preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work

Distribute object code? “…provide a machine-readable copy of the Source Code of the Original Work”

Network hook: “made available as an application intended for use over a network”
“… any original work of authorship (the “Original Work”) whose owner (the “Licensor”) has placed the following licensing notice…”

Source code = preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work.

Distribute object code? “…provide a machine-readable copy of the Source Code of the Original Work”

Network hook: “made available as an application intended for use over a network”

“Right to Use. You may use the Original Work in all ways not otherwise restricted or conditioned by this License or by law, and Licensor promises not to interfere with or be responsible for such uses by You.”

MODEL
MODEL PARAMETERS
TRAINING DATASET
EVALUATION DATASET
TRAINING CODE?
DO WE NEED A NEW OPEN SOURCE LICENSE FOR AI/ML COMPONENTS?

LICENSE PROLIFERATION BAD
DO WE NEED A NEW OPEN SOURCE LICENSE FOR AI/ML COMPONENTS?

LICENSE PROLIFERATION BAD

FALCON EXAMPLE

• Falcon 180B TII License Version 1.0
  [September 2023]
“Source” form shall mean the preferred form for making modifications, including but not limited to software source code, training datasets used for training or fine tuning a machine learning model or artificial intelligence model, documentation source, and configuration files.

“Object” form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, a trained and/or fine-tuned machine learning model or artificial intelligence model, generated documentation, and conversions to other media types.
“...use-based restrictions incorporating the Acceptable Use Policy in the manner set out in Section 5 of this license, and which do not otherwise conflict with the Acceptable Use Policy, must be included as enforceable provisions ...”

“The Acceptable Use Policy may be updated from time to time. You should monitor the web address at which the Acceptable Use Policy is hosted... “
MODEL/WEIGHTS CASE STUDY

Model, training code, inference code under MIT license;

Weights under non-open source CC Attribution-NonCommercial-NoDerivatives 4.0 International

“B. produce and reproduce, but not Share, Adapted Material for NonCommercial purposes only.”
THANK YOU