

Workshop

Introduction to Free Software Licensing in AI context

20.03.2023

Dr. Lucas Lasota · lucas.lasota@hu-berlin.de

Dr. Lucas Lasota, MA, PhD

- Senior Legal Project Manager, Legal Researcher and Lecturer
- Affiliation: Humboldt University and Free Software Foundation Europe

Birds



On Mars



Agenda

Why “Free Software”?

FS licensing compliance issues

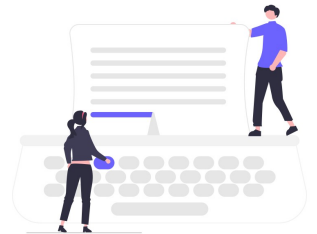
FS licensing and AI



This presentation does not configure neither legal nor business advice, but fits only educational and training purposes.

Why software licensing?

- Copyright applies by default: very strong protection
 - Injunctions, damages and criminal penalties
- Licenses are the main legal instrument to waive and grant rights
- *No license does not* configure public domain



Licensing

Full spectrum of rights management

Public Domain	Permissive Licenses (non-reciprocal)	Copyleft Licenses (reciprocal)	Proprietary Licenses	Trade Secret
All rights waived	<<< More rights granted to user		>>> More rights retained for owner	
				All rights retained



Software licensing models

```
graph TD; A[Software licensing models] --> B[Free and Open Source]; A --> C[Proprietary]; B --> D[Main purpose]; B --> E[Software freedom]; E --> F[. Reciprocal (copyleft)]; E --> G[. Non-reciprocal (permissive)]; C --> H[Main purpose]; H --> I[Limit ways to access and control source code];
```

Free and Open Source

Main purpose

Software freedom

- . Reciprocal (copyleft)

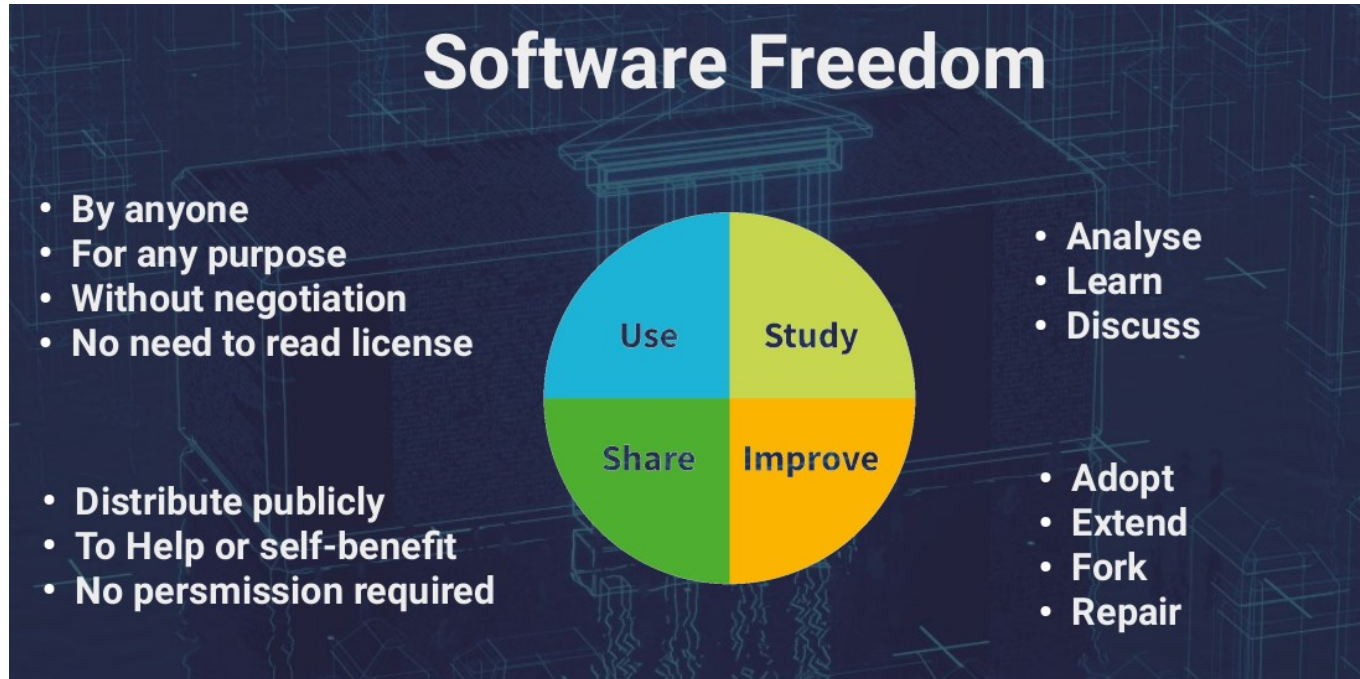
- . Non-reciprocal (permissive)

Proprietary

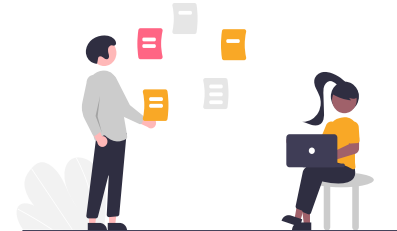
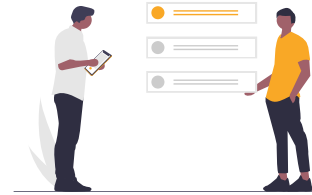
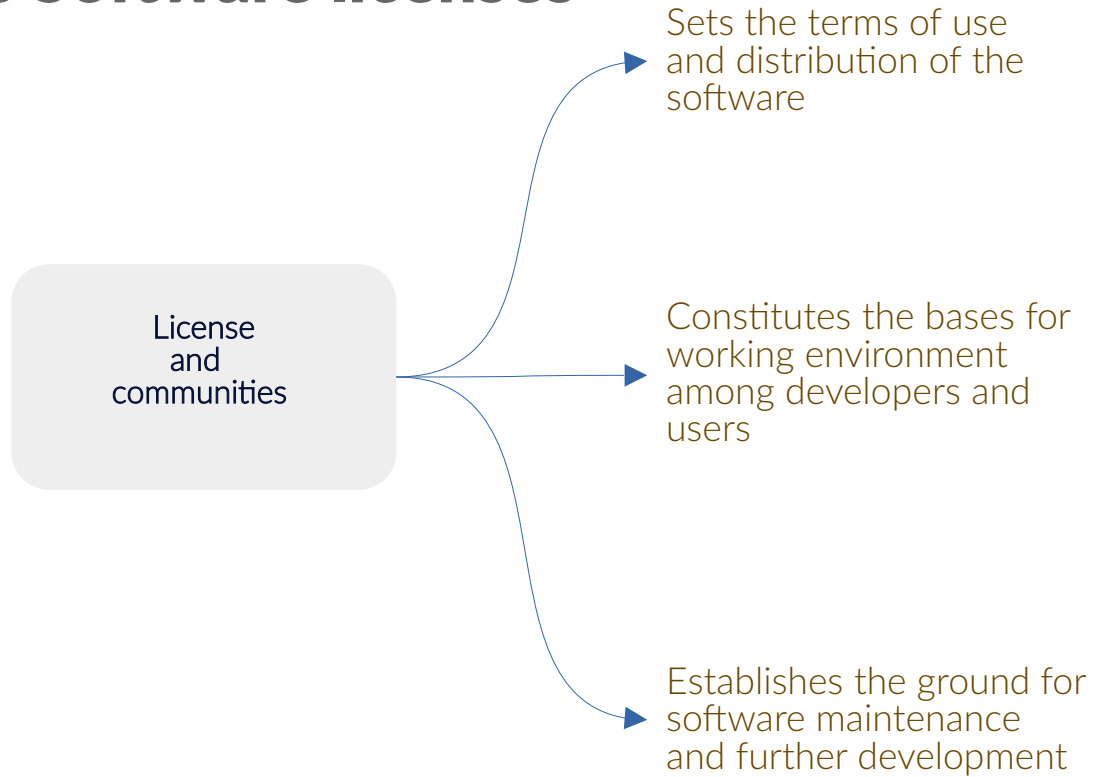
Main purpose

Limit ways to access and control source code

Free Software as an ethical movement



Social components of Free Software licenses



Free Software Licenses Types

Reciprocal (copyleft)

Main aspects

- . If you distribute, you must use the same license terms
- . If you distribute, you must provide source code



Non-reciprocal (permissive)

Main aspect

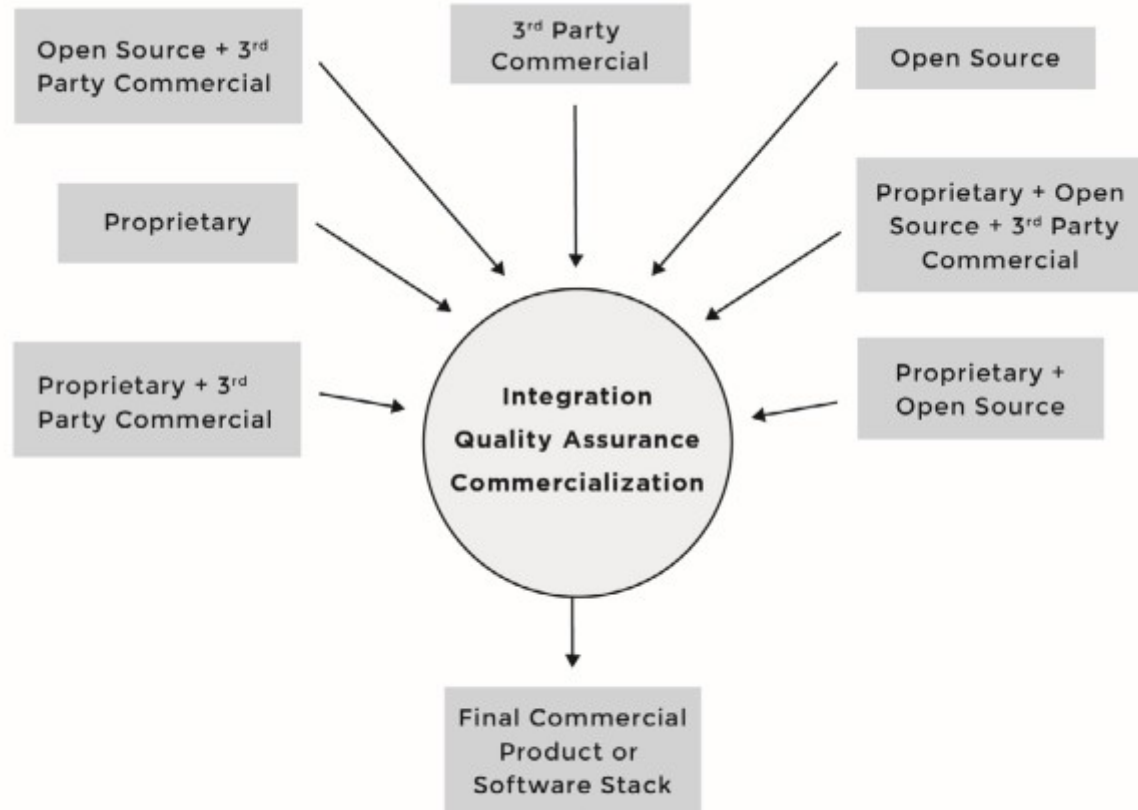
- . If you distribute, you must provide license/copyright notice



BSD



Free Software License Compliance



Source: Haddad, Ibrahim et al. Open Source Compliance in the Enterprise. The Linux Foundation, 2018, p. 18.

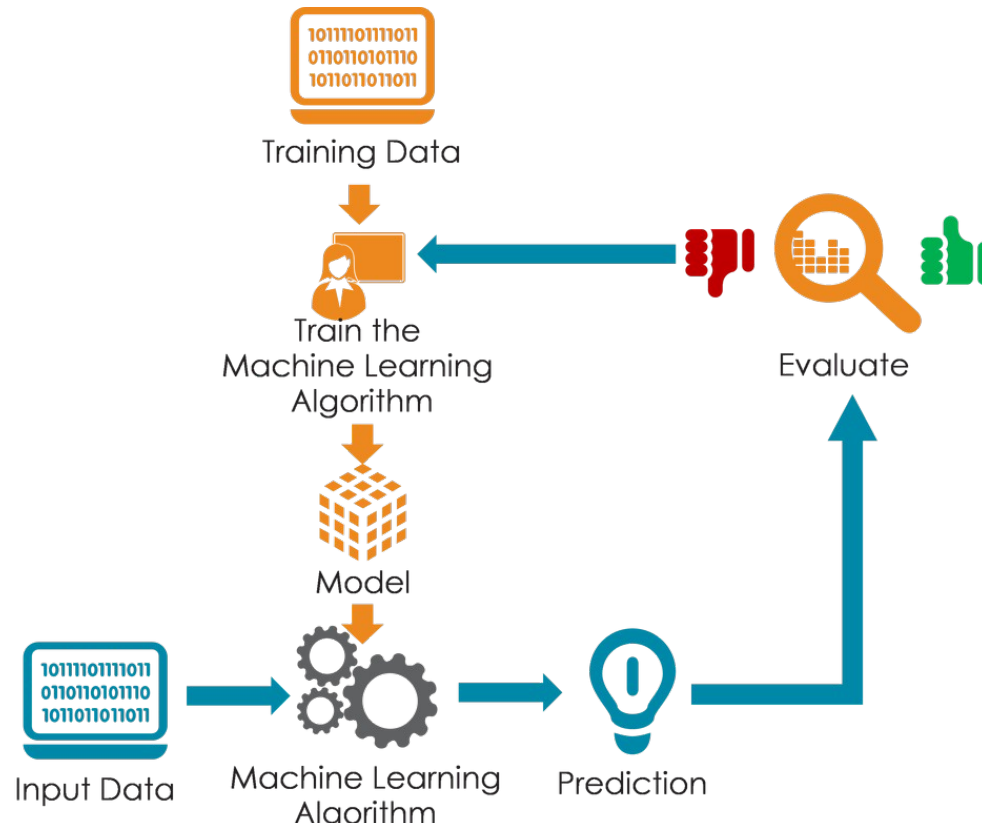
Compliance issues related to Free Software

Problem type	Solution
Inbound license obligations	License scan and mapping License compatibility workflow
Outbound license obligations	License compliance workflow (source code provision, licenses notices, etc)
Patents and trade secrets (NDA) involved?	IP management workflow
Project's trademarks?	Trademark policy
Exporting or transferring technology of dual-use items?	Export control compliance workflow
Managing external contributions	Contribution policy Contribution licensing agreements (CLA, DCO, employment issues)

AI

AI models and licensing

Source: What is a training data set in Machine Learning and rules to select them? Cogito, 2018.



Element	Type of license
Training data (labelled or not)	Data license
Software to be trained (model)	Software license
Input data (real life raw data)	Data license?
Trained software	Software license

Content & Data Licensing



You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Audio/video/images

Documentation

Data and metadata

PS: Copyright exceptions for **data mining**
(Directive EU 2019/70)

- Non profit (research, cultural)
- Any other usage if not expressly reserved by right-holders
- Exceptions override contrary contractual provisions

Community Data License Agreement - Permissive - Version 2.0

This is the Community Data License Agreement - Permissive, Version 2.0 (the "agreement"). Data Provider(s) and Data Recipient(s) agree as follows:

1. Provision of the Data

- 1.1. A Data Recipient may use, modify, and share the Data made available by Data Provider(s) under this agreement if that Data Recipient follows the terms of this agreement.
- 1.2. This agreement does not impose any restriction on a Data Recipient's use, modification, or sharing of any portions of the Data that are in the public domain or that may be used, modified, or shared under any other legal exception or limitation.

2. Conditions for Sharing Data

- 2.1. A Data Recipient may share Data, with or without modifications, so long as the Data Recipient makes available the text of this agreement with the shared Data.

3. No Restrictions on Results

- 3.1. This agreement does not impose any restriction or obligations with respect to the use, modification, or sharing of Results.

4. No Warranty; Limitation of Liability

- 4.1. All Data Recipients receive the Data subject to the following terms:

THE DATA IS PROVIDED ON AN "AS IS" BASIS, WITHOUT REPRESENTATIONS, WARRANTIES OR CONDITIONS OF ANY KIND, EITHER EXPRESS OR IMPLIED INCLUDING, WITHOUT LIMITATION, ANY WARRANTIES OR CONDITIONS OF TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

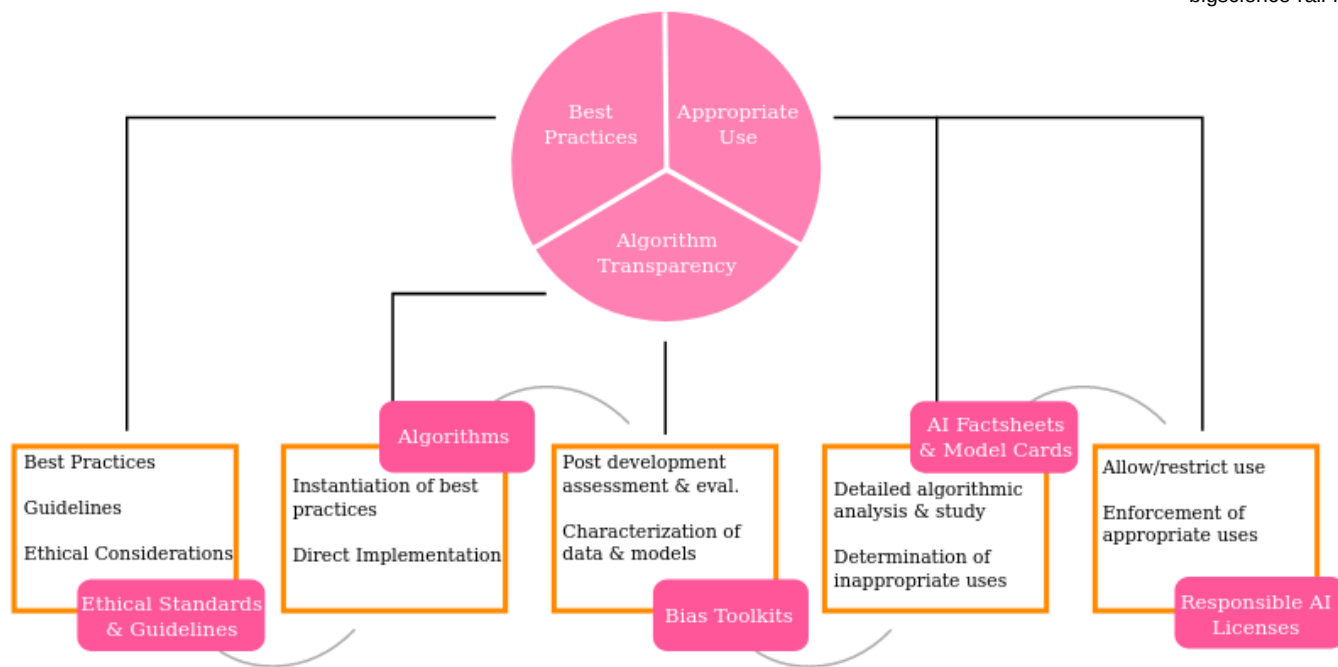
NO DATA PROVIDER SHALL HAVE ANY LIABILITY FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING WITHOUT LIMITATION LOST PROFITS), HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE DATA OR RESULTS, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

5. Definitions

- 5.1. "Data" means the material received by a Data Recipient under this agreement.
- 5.2. "Data Provider" means any person who is the source of Data provided under this agreement and in reliance on a Data Recipient's agreement to its terms.
- 5.3. "Data Recipient" means any person who receives Data directly or indirectly from a Data Provider and agrees to the terms of this agreement.
- 5.4. "Results" means any outcome obtained by computational analysis of Data, including for example machine learning models and models' insights.

BigScience RAIL License model

Source: The BigScience RAIL License 2022.
Available at:
<https://bigscience.huggingface.co/blog/the-bigscience-rail-license>



The goal of our approach to licensing is to support AI researchers who may be concerned about the possible inappropriate use of their models and would still like to share their work for advancing science. As a result, we opted to design an open and permissive license that also includes use-based restrictions. Although, the Apache 2.0 license was applicable to resources used to develop the **Model**, the licensing conditions have been modified for the access and distribution of the **Model**. This has been done to further BigScience's aims of promoting not just open-access to its artifacts, but also a responsible use of these artifacts. We include some FAQs to help answer some questions that we often faced during the development of our license.

Further resources

Bibliography on FOSS compliance

https://www.zotero.org/groups/4870654/foss_compliance/library

FSFE's License Questions Team

<https://fsfe.org/activities/licence-questions/licence-questions.html>

Take aways

AI systems may involve several data and software elements that are licensed differently. These licenses operate differently and have diverse scope of application.

Free software licenses are the main instrument for safeguarding the “four freedoms”: use, study, improve, share. They come in two flavors: reciprocal and non-reciprocal.

License compliance issues are complex and involve not only copyright but other legal and operational elements (e.g. GitHub Copilot case). Establishing a FS compliance workflow is key.



Thank you!

Discussion warm up

How Free Software helps improving AI systems?

20.03.2023

Dr. Lucas Lasota · lucas.lasota@hu-berlin.de

Illustrations kindly provided by Undraw