ZOOOM

Free Software and AI openness:
Overcoming challenges in the licensing world

Lucas Lasota and Niharika Singhal

Free Software Foundation Europe

Contents

Executive Summary

The convergence of AI and Free Software is briefly analysed in this paper from the perspective of licensing, especially considering the necessity to promote openness respecting the historical principles of software freedom.

Some challenges in achieving openness in the licensing scheme of AI are highlighted. In particular, the uncoordinated growing proliferation of licences claiming to be "free and open source" but imposing extra limitations on software freedom may lead to licence incompatibility, a well-known challenge faced by the Free Software community.

This paper clarifies why using licences to address behavioural and usage restrictions may affect distribution of control over AI technologies, and how the complexities of non-free licences in multi-source software development can affect compliance efforts.

In conclusion, three recommendations are proposed:

- Preserving openness in AI by safeguarding the four freedoms of software. Restrictions on software freedom disable control, transparency and oversight over technology. This results in a negative impact on people's digital autonomy, distribution of power in the society and ultimately the democracy.

- Keeping licensing of AI technologies cohesive and interoperable with Free Software licences by avoiding licence proliferation, increasing legal interoperability and simplifying licence adoption.

- Encouraging engagement with civil society actors in initiatives aimed to make AI more open, accessible, transparent and auditable.

Introduction

Historically, Free Software (also known as Open Source Software) has been deeply connected with AI. The two concepts have evolved in parallel to converge in more explicit ways in recent years. Free Software is not only present in the tools used to develop AI, but also collaboration principles enabled by Free Software have played a key role in the building and flourishing of AI technologies. Freedom and openness promoted by the Free Software movement, which also aligns well with related areas, such as open access, open data, digital commons and free culture, carry the potential to make AI more human-centric by enabling access, transparency and distribution of power. The recent boom of development and adoption of AI has posed profound questions on how humans interact with technology, and openness of AI has been at its centre.

The legacy of the Free Software movement of democratising access and control over technology has attracted attention from companies and communities which want to be seen as "open" with regard to the products or technical environments that they promote. However, an increasing number of AI projects dealing with impactful technologies and calling themselves "open" do not grant users the four freedoms embedded in the Free Software definition. Besides, as a parallel phenomenon, due to concerns about the disruptive power of AI technologies, communities have departed from using exclusively Free Software licences on their projects to licensing solutions instead, prioritising restrictions on how software should be further used and distributed.

There is an urgent need for scrutinising the impact of AI on human relations in its diverse dimensions and manifestations. This is strongly supported by the FSFE also through its active engagement in the latest regulatory and legislative processes in the EU. However, the concept of openness of software embedded in the historical definitions of Free Software should not be forgotten in the current public debate.

This paper discusses some aspects of convergence of AI and Free Software in relation to licensing of the software, especially considering the necessity to promote openness in the context of historical principles of software freedom. We describe the challenges to achieving openness of AI from the licensing perspective. In particular, the uncoordinated and growing proliferation of licences claiming to be "free and open source" but imposing extra limitations on software freedom may lead to licence incompatibility, a well-known challenge to the Free Software community. We explain why using licences to address behavioural and usage restriction may affect distribution of control over AI technologies, and how the complexities of non-free licences in multi-source software development can affect compliance efforts. In conclusion, we propose a set of three recommendations: we ask for preserving the imperative of openness in AI, and keeping licensing of AI technologies cohesive and interoperable with Free Software licences. We also encourage engagement with civil society actors in initiatives aiming to make AI more open, accessible, transparent and auditable.

Openness as an Ethical Consideration

Openness, as a concept emphasising the accessibility of knowledge, technology and other resources, is central to the development, implementation and oversight of AI technologies. Free Software has been key in the history of AI.[1] The free sharing of knowledge, allied with the principles of common use of the underlying technologies enabling the AI boom, has been fostered by Free Software. The rich social, legal, policy and economic experience achieved at global scales by the Free Software movement has been crucial for the whole digital ecosystem where AI has flourished. Software freedom manifests the ideals of free exchange of knowledge, creative expression, collaborative development and respect for digital commons. Free Software has grown to become a foundational element not only of the economy in the digital age, but of the notion of democracy in the information society.

---

[1] Vaughan-Nichols, S. (2023). Open source is actually the cradle of artificial intelligence. Here's why. Zdnet. Available at: https://www.zdnet.com/article/why-open-source-is-the-cradle-of-artificial-intelligence/

Historically, to promote inclusivity as opposed to exclusiveness in the software realm, Free Software communities have developed the concept of openness in the curated definitions provided by the Free Software Foundation (FSF) [2] and the Open Source Initiative (OSI).[3] "Openness" is defined in terms of freedom to use the software, study and modify the source code, and the possibility of sharing and reusing the program. Traditionally these freedoms have been given by copyright licences defining the grants and reservation of rights. For instance, the so-called reciprocal (copyleft) licences have been instrumental in sustaining the freedom of software by promoting the continuity of the licence choice, and avoiding any licence changes in the future. The FSF [4] and the OSI [5] maintain lists of reviewed and approved licences.

Use
Free Software can be used for any purpose and is free of restrictions such as licence expiry or geographic limitations.

Study
Free Software and its code can be studied by anyone, without non-disclosure agreements or similar restrictions.

Share
Free Software can be shared and copied at virtually no cost.

Improve
Free Software can be modified by anyone, and these improvements can be shared publicly.
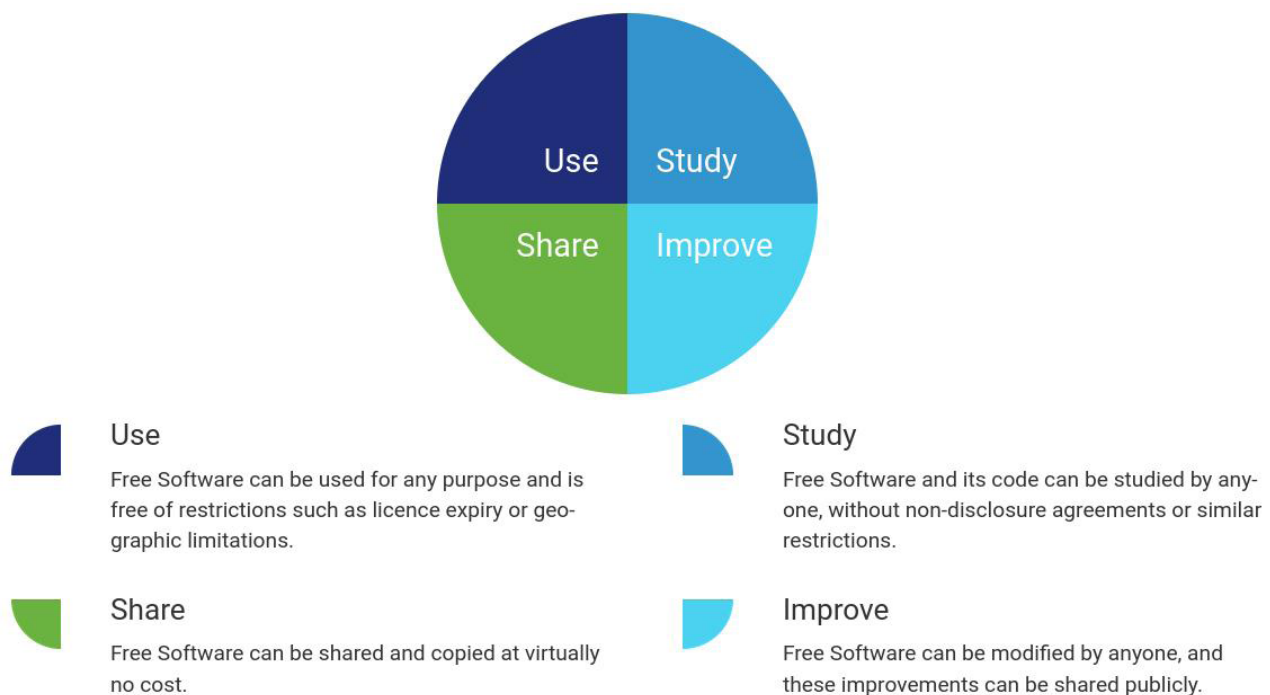
Figure 1. The Free Software definition refers to the four freedoms of the software

The diverse parameters governing the choice for Free Software are multi-faceted. The engagement occurs in many ways (e.g., as users or makers/contributors, or both) and for different reasons. Strategic and competitive advantages are relevant, but in some cases not necessarily economic in nature. For instance, they can involve the potential for growth (both personal and collective), technological innovation, practical needs, personal and collective motivations (e.g. freedom of choice, digital autonomy and sovereignty, personal and social development), and ethical values (e.g. reciprocity, altruism, democratisation of knowledge, social justice, promotion of digital commons).

---

[2] Stallman, R. (2021). What is Free Software? GNU Operating System. Version 1.169. Available at: https://www.gnu.org/philosophy/free-sw.html

[3] OSI (2023). The Open Source Definition. Available at: https://opensource.org/osd/

[4] FSF (2023). Various Licenses and Comments about Them. Available at: https://www.gnu.org/licenses/license-list.html

[5] OSI (2022). Open Source Licenses by Category. Available at: https://opensource.org/licenses-old/category/

Autonomy

Free Software helps to develop and maintain tailored software that suits your needs, not just the vendor's business model.

Share & Copy

A Free Software licence allows a limitless number of installations to be run, without paying extra.

Reuse Code

Free Software provides the freedom to reuse the code for other projects.

Competition

Free Software resists monopolisation and enhances competition.

Collaboration

Free Software can be shared and used in a non-exclusive way by everyone – serving the public good.

No Lock-in

Free Software licences reinforce independence from vendors and provide more choice in service providers.

Innovation

A Free Software licence encourages innovation for your software.

Security

Free Software allows for independent security checks that help close security holes faster.

Figure 2. Besides economical, strategic and operational factors, there are a multitude of ethical reasons for engaging with Free Software

Free Software positively contributes to AI accessibility and transparency, making it less discriminatory.[6] Accessibility for AI can be understood as making it reusable, so that everyone may tinker with it, improve it and use it for their own purposes. A Free Software licence can be instrumental for reusing code elements in an AI system. AI transparency can be subdivided into openness and interpretability. In this context, openness is defined as the right to be informed about the AI software, and interpretability can be defined as being able to understand how the input is processed so that one can identify the factors taken into account to make predictions, and their relative importance. Free Software can help with that as well. Last but not least, Free Software can improve AI systems by minimising any discriminatory effects. This is possible as an AI system released under a Free Software licence can be used, inspected and verified by everyone to determine whether it is free of potentially harmful discriminatory elements more easily than if it were proprietary. Discriminatory biases based on sensitive characteristics such as gender, ethnicity, religion, disabilities, or sexual orientation can be mitigated by Free Software. Moreover, this synergises with AI transparency, as a transparent AI application facilitates the understanding of the factors considered for making predictions. While doing so is desirable, releasing an AI application as Free Software does not make it completely non-discriminatory. However, Free Software makes it easier to evaluate and audit AI systems for any discriminatory effects.

---

[6] Lequertier, V. (2021). Controlling technology at the age of Artificial Intelligence: a Free Software perspective. Free Software Foundation Europe. Available at: https://fsfe.org/freesoftware/artificial-intelligence.html

Challenges for Keeping AI Open

The recent AI revolution is based on the enormous capacity of current data processing used for specific machine and deep learning techniques. Data may be consumed, transformed, and incorporated into AI models in ways that are different from how software and other creative content are generally used. There may be difficulties when data subject to commonly used Free Software or even well-established Creative Commons licences are used as input to train AI-powered computer models. Traditional licences for software and content might not apply in expected ways to open data. In recent years, several strategies have been developed by different communities to overcome such limitations. Some adopt a more pragmatic approach, by applying different licences for code and data. Others prefer to implement additional behavioural restrictions by means of licensing, and there are still efforts in trying to define "Free & Open Source AI". These processes are not necessarily coordinated, which can create challenges to guaranteeing openness in AI.

Licensing Data and Software

Some communities have followed the path of setting up licences for AI systems that cover software and data elements separately. Combinations of software and data are common but until recently data had hardly been a subject of licensing discussions. This has changed with big data analytics and AI. Inspired by the success of Free Software, the drafters of the first data licences have tried to apply the fundamental principles enshrined in the Free Software definition to data as well. For instance, the Linux Foundation's Community Data License Agreement (CDLA)[7] provides a grant of rights for recipients of data to use, share, and modify the data for any purpose. It also permitted using the results from analysed data to create AI and machine learning models, without any obligation to share the data. It was launched with two initial types: a non-reciprocal (permissive) variant, with attribution-style obligations, and a reciprocal (share-alike) variant, with a reciprocal commitment regarding resharing the raw data. Code elements remain under a Free Software licence. In other cases, some companies and communities prefer to let the AI use assemblages of code and data under a single licence, namely the licence applicable to the software. In any case, despite the blurry zone between data and code, licensing hybrid combinations of software and data has been a practical approach adopted by some.

---

[7] LF (2021). Community Data License Agreement: Collaborative licenses to enable access, sharing and use of data openly among individuals and organisations. Available at: https://cdla.dev/
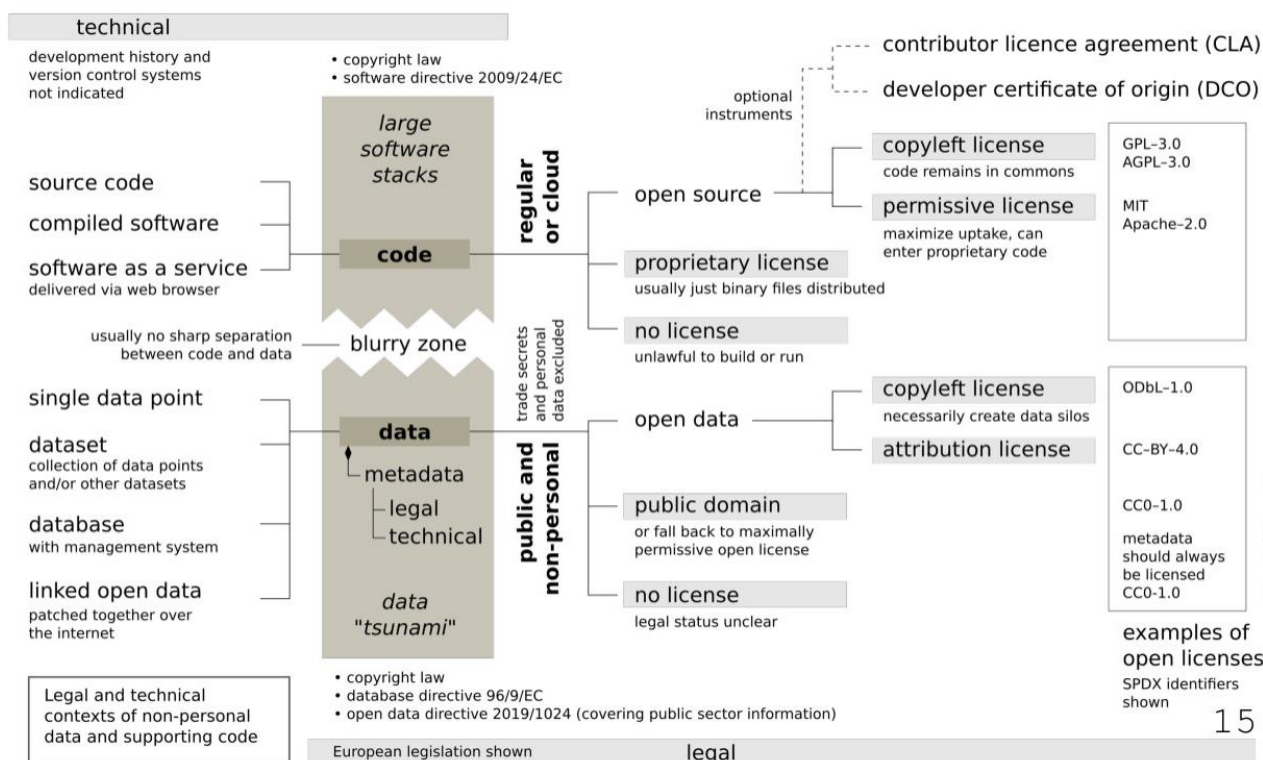
Figure 3. Notwithstanding the blurry zone between software and data, there is a vast variety of legal instruments for licensing code and data from the perspective of EU law.[8]

Imposing Additional Behavioural Restrictions by Licensing

Discussions on AI ethics, trustworthiness and responsibility are crucial for a more human-centric AI that respects human rights and democracy. The FSFE strongly supports and fosters ethical, policy and legal efforts in establishing such frameworks through appropriate regulatory and legislative solutions by directly engaging with legislative and regulatory processes in the EU.

The concerns of the impact of AI on human relations in its different dimensions and manifestations have motivated communities to depart from using Free Software licences exclusively on their projects. Instead, they have developed new types of licensing solutions which impose restrictions on the openness of software. The prioritisation of other imperatives against openness has led some communities to introduce limitations in licences on how software should be further used and distributed. For instance, the "BigScience RAIL License", which has the goal 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", includes use-based restrictions.[9] Establishing further behavioural restrictions on AI openness

---

8   Robbie Morrison (2021). Open Source Software and Open Data: Open Licensing of Software and Data for Public Policy Analysis and for Collaborative Research — Release 03. Available at: https://zenodo.org/record/4537157. CC-BY-4.0.

9   Ferrandis, C. et al. (2022). The BigScience RAIL License. Available at: https://bigscience.huggingface.co/blog/the-bigscience-rail-license

creates additional challenges for the Free Software movement mainly because of the complexity and broad interpretations of ethical values across jurisdictions. It further complicates the alignment of licences with restrictive characteristics based on purported ethical considerations which may create incompatibilities with the four freedoms of software.

Particularly concerning is the popularisation of the term "open" in relation to AI technologies when there is a discrepancy between what is implied and the real terms and conditions under which such systems are actually licensed. The dissonance between the "open source" claims of AI systems and the restrictions imposed by licences casts shadows over the definitions of Free Software and Open Source Software. Such licensing schemes emphasise restrictions related to fields of endeavour, behaviour, community management and commercial practices. This practice has now also spilled over to creation of suo moto ethics codes on AI, leading to creation of licences with restrictive characteristics. For instance, Meta's LLaMA model, a language learning model, which forms the basis of many chatbots and other systems, was originally licensed under GPL v3 and is available for research purposes, but cannot be used commercially. In the same vein, LLaMA-2 is proclaimed as an "open source model" by Meta; however, it imposes restrictions that forbids its use to train other language models, and it requires a special licence from Meta for use involving a large number of users.[10]

Ethical considerations may have a significantly broad and wide spectrum of interpretations, especially judging from a strict legal perspective. These licences relying on ethical clauses that may involve different interpretations should not cause overstepping the several current legislative processes, or substituting for the upcoming regulations. Licensing restrictions on the distribution and modification of the software may impinge on software freedom and cause obstruction to its advancement and to a more distributed control over AI:

- **Barriers against use and reuse:** The conditions of the licences with behavioural restrictions raise ambiguities in their implementation. For example, clauses prohibiting defamation, disparagement or harassment of others; or terms related to non-discrimination against individuals or groups based on characteristics or categories. The vague terminologies may result in an overarching broad prohibition of AI use. This disallows larger collaborative efforts and hinders downstream application and integration.

- **Hurdles to adaptation or improvement:** Unlike the direct licensing model provided by Free Software licences, some of the restrictive licences follow the sub-licensing model. In this case, downstream users are not bound to contribute back to the commons, being allowed to re-license the model under their own licence provided that they propagate the use restrictions as provided in the restrictive licences. This can create challenges for the sustainability of the software, as the adaptations are no longer free.[11]

---

[10] See, Sec 2. Additional Commercial Terms, Llama 2 Version Release Date: July 18, 2023, available at: https://ai.meta.com/llama/license/ and Llama 2 Acceptable Use Policy, available at: https://ai.meta.com/llama/use-policy/

[11] Albers, E. (2021). On the Sustainability of Free Software. Free Software Foundation Europe. Available at: https://fsfe.org/freesoftware/sustainability/sustainability.html

For instance, some AI models provide a text-to-image foundation model which can potentially create accurate medical images leading to better clinical healthcare. However, the use of restrictive licences has created barriers to adoption by preventing users from generating images for medical advice or medical results interpretation

- **Hindering control over technology:** Some companies provide access to several components of the AI models, such as the database and source code of the training models, but restrict the purposes of their reuse. Some exercise tight control over APIs. Others use licences that make models available only for research purposes but prohibit their commercial use. The consequence is an increased loss of users' control over technology, lock-ins and dependency on the providers.

- **Weakening oversight and transparency:** Although a proprietary AI model can be transparent, Free Software facilitates transparency by making auditing and inspection easier. While some types of data might be too sensitive to be released under a free licence, statistical properties of the data can still be published.[12] The lack of openness hampers oversight over key aspects of AI systems and how data is being processed. Scholars have also highlighted a range of different approaches to ethics-based auditing of AI that already exists. Functionality audits focus on the rationale behind the decision, code audits entail reviewing the source code, and impact audits investigate the effects of an algorithm's outputs.[13] As model evaluation and standardised risk assessment procedures are being rapidly developed,[14] the use of Free Software licences enables such code audits by allowing any user to not only review the source code but also improve it. Free Software also aids in detection of bugs and identification of security issues, and may lead to better explainability of AI systems. Governments should in this regard be particularly wary of these developments, especially where public funding is involved, as they hinder openness, transparency and collaborative efforts for innovation in society.

Defining Free & Open Source AI

A more comprehensive issue relates to a formal definition for the meaning of "Free and Open Source AI". While AI technologies are contemporary to the Free Software movement itself, it is only in the last decade, with advancements in deep learning and large language models, that public and policy attention has exponentially grown in this domain. However, applying the Free Software definition directly to AI may not render the same result, as AI involves a myriad

---

[12] Lequertier, V. (2021). Controlling technology at the age of Artificial Intelligence: a Free Software perspective. Free Software Foundation Europe. Available at: https://fsfe.org/freesoftware/artificial-intelligence.html

[13] Mökander, J. and Floridi, L. (2021). Ethics-Based Auditing to Develop Trustworthy AI. Minds and Machines 31, 323. Available at: https://link.springer.com/article/10.1007/s11023-021-09557-8

[14] Constanza-Chock, S., Raji, I. and Buolamwini, J. (2022). 'Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem', ACM Conference on Fairness, Accountability and Transparency, p. 1571-1583.

of other elements that are not software. The OSI has undertaken the initiative to provide a definition departing from the traditional four freedoms of software, which aims to cover the wide spectrum of AI technologies.[15] At the same time, the OSI focuses on openness of the AI and does not try to define how to develop and deploy trustworthy or responsible AI, notwithstanding the support garnered through appropriate governmental regulation on these issues.[16]

The OSI's process has involved a broad spectrum of stakeholders. Emanuilov and Suksi, leading partners of the ZOOOM initiative, have collaborated with the process by proposing a definition of "Open Source AI", by presenting suggestions that include access to models, weights, biases, algorithms or training and testing data and promote the four freedoms of software under a three-criteria scheme prioritising data transparency, disclosure of model details and verification by reproducibility of the models.[17]

Licence Proliferation and Operational Hurdles

Since its inception, Free Software has expanded from an idealistic movement to becoming a key aspect of today's digital infrastructure. As a result of these developments, many software projects, including those of large companies developing proprietary programmes for commercial use, have started to incorporate Free Software components into their platforms and software stacks, resulting in the emergence of software projects that fall under the categorization of the "multi-source development model" involving a large quantity of different proprietary code, software under different types of Free Software licences, data, information and other types of content under diverse data licences. All of that must be incorporated and integrated in a cohesive and compatible way to satisfy licensing compliance. Depending on the size of the project, it may have licensors or contributors (copyright holders) in the hundreds or even thousands.

---

[15] OSI (2022). Deep Dive AI: Discover the Future of Open Source. Available at: https://deepdive.opensource.org/

[16] OSI (2023). The Open Source AI Definition – draft v. 0.0.4. Available at: https://opensource.org/deepdive/drafts/

[17] Emanuilov, I., Suksi, J. (2023). Open Source AI: Building Blocks for a Definition. ZOOOM Policy Briefs #1. Available at: https://zooom4u.eu/index.php/2023/12/14/zooom-policy-brief-1-open-source-ai-building-blocks-for-a-definition/

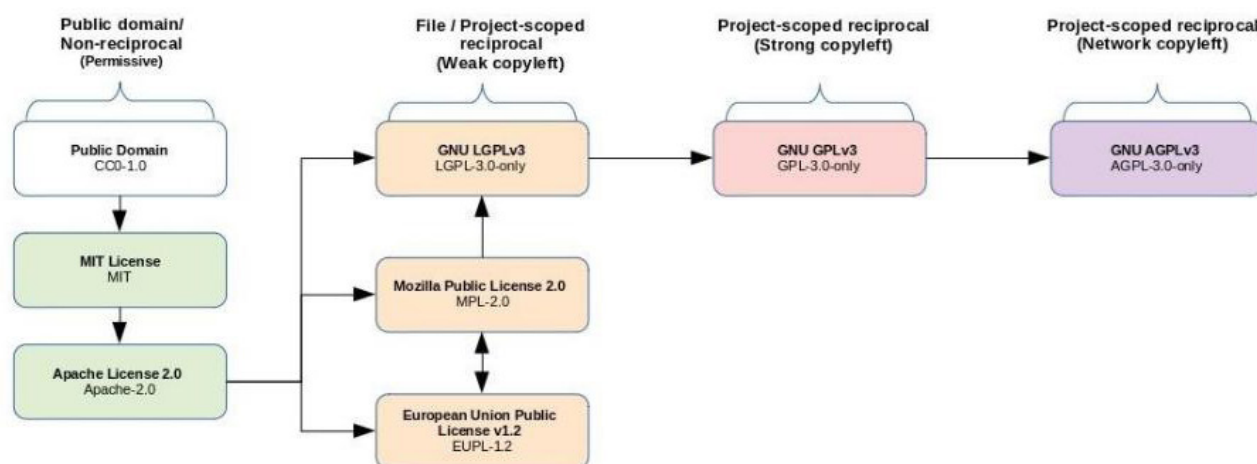## COMMON FOSS LICENSES COMPATIBILITY CHART

Public domain/ Non-reciprocal (Permissive)	File / Project-scoped reciprocal (Weak copyleft)	Project-scoped reciprocal (Strong copyleft)	Project-scoped reciprocal (Network copyleft)
Public Domain CC0-1.0	GNU LGPLv3 LGPL-3.0-only	GNU GPLv3 GPL-3.0-only	GNU AGPLv3 AGPL-3.0-only
MIT License MIT	Mozilla Public License 2.0 MPL-2.0		
Apache License 2.0 Apache-2.0	European Union Public License v1.2 EUPL-1.2		

Figure 4. Depending on the size of the project, and the number of different licences involved, licence compatibility can be challenging.[18]

Achieving licensing compliance in a software project is key to avoiding liabilities, to assure the long-term sustainability of the project, and it is also important for a healthy licensing ecosystem among software projects in the Free Software communities. While proprietary licences are fundamentally incompatible with each other, the Free Software movement has made efforts to avoid licence proliferation, to increase legal interoperability and to simplify licence adoption. The existing well-known licences are standardised, well-documented and have withstood complex legal issues. The compatibility between the existing well-known licences can also be ascertained by developers or any potential user with the help of available sources and materials.[19] Free Software licences do impose conditions on exercise of the license, but these are not limited as to the type of use, location of use or number of copies, as how its occurs with proprietary schemes and as is also reflected in the licences with additional restrictive characteristics as discussed above. Therefore, compared to proprietary software licences, including the licences with additional restrictive characteristics in question, Free Software licences are easier to comply with.

The traditional formulations of Free Software licences may not be sufficient when implementing AI algorithms to guarantee inspectability and replicability of the AI systems. Hence, the emergence of new licences and licensing schemes should be an inevitable natural progression. New developments and stewardship of dedicated licences for AI technologies are welcomed and desired. We plead, nevertheless, for a cautionary approach in creating new licences, and utmost consideration for interoperability with Free Software licences, so that accessibility, reusability and sustainability of AI systems benefit.

---

[18] Chart based on David Wheeler, The Free-Libre / Open Source Software (FLOSS) License Slide <https://dwheeler.com/essays/floss-license-slide.html>. CC-BY-SA 3.0. Modifications of the content: The title and classification of the licenses was altered; new licenses were added, such as EUPL and AGPL.

[19] See for example FSF's compatibility information at: https://www.gnu.org/licenses/license-list.en.html

Conclusion & Recommendations

AI has become a reality in people's lives. The complex questions on the individual and social impacts of AI have sparked debates on different aspects of this technology, and a key element asks what it means for an AI to be Free (as in Freedom). The urgency and necessity of having AI systems be available has forced communities to follow different strategies regarding licensing of the AI software and data elements, culminating in recent efforts to achieve an overarching definition in ways similar to the Free Software movement forty years ago. There are many challenges in such processes. The different conceptions of the terminology on what is "open" in regard to AI are not of particular help. Incompatibility among licences also represents a serious issue for the long-term sustainability of projects, as well as for the reusability of the involved technologies. Therefore, we make three recommendations as contributions towards the engagement in the ZOOOM initiative.

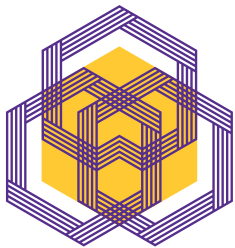## #1 - Preserve openness in AI

The convergence of AI and Free Software depends on keeping technology open. "Free Software" and "Open Source" are curated terms, taking with them a rich 40-year long history of success in democratising control over software. They specifically refer to the four freedoms of software: use, study, share and improve. The exponential growth and massive use of AI in these few years have led to the emergence of systems claiming to be "free" or "open source" that do not grant users the four freedoms. The dissonance between the marketing pitch of such projects and the restriction on software freedom disables control, transparency and oversight over technology. The negative impact on people's digital autonomy, distribution of power in the society and ultimately on democracy is clear.

2 - Keep AI licensing simple and coherent with software freedoms

The convergence of AI and Free Software depends on keeping technology open. "Free Software" and "Open Source" are curated terms, taking with them a rich 40-year long history of success in democratising control over software. They specifically refer to the four freedoms of software: use, study, share and improve. The exponential growth and massive use of AI in these few years have led to the emergence of systems claiming to be "free" or "open source" that do not grant users the four freedoms. The dissonance between the marketing pitch of such projects and the restriction on software freedom disables control, transparency and oversight over technology. The negative impact on people's digital autonomy, distribution of power in the society and ultimately on democracy is clear.

3 - Engage with civil society actors and initiatives

Discussions on AI ethics, trustworthiness and responsibility are crucial for a more human-centric AI that respects human rights and democracy. We strongly support and foster ethical, policy and legal efforts in establishing such frameworks through appropriate regulatory and legislative solutions. We ask, however, for a cautionary approach to new licences that can negatively affect Free Software, and utmost consideration for interoperability with Free Software licences. Engaging with civil society actors in initiatives aimed to make AI more open, accessible, transparent and auditable is a great way to contribute to digital commons and preserve a human-centric control over technology.

www.zooom4u.eu