

# The Open Source AI Definition:

Carlo Piana – **Array**

EOLE 2024

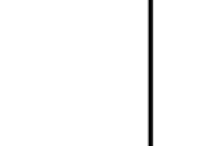
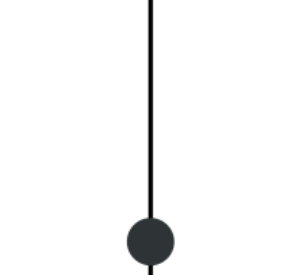
“Navigating the New EU Regulatory Landscape in Open Source”







WHY?



# Why “Why”?

- Open *source* → Source → code
- AI *models* do not have “source” (systems do, yet impact == 0)
- Open source **source** would mean nothing, we need open **models**



# Why Open Source, Act 1

- Some models already on the market, *soi-disant* “Open source” (or open-source)
- OSI **stewards** the definition of Open Source in software (OSD) and the very **name**
- Letting AI models in the wild calling themselves “Open Source” would **dilute** the term
- Especially if they are **far** from being Open Source, whatever it means:
  - EG, **Field of endeavour limitation** or **Business model limitations** (See Meta’s Llama)



# Why Open Source, Act 2

- Open Source is a **means to an end**
- The end is **frictionless, permissionless innovation**
- The characteristic of Open Source allowing it are the **Four Freedoms**
- The Four Freedoms are allowed, in software by two things:
  - Availability of source code (OSD #2)
  - License permissions (mainly copyright)
- Open Source is therefore a **synecdoche** (a part for the whole)



# Why Open Source, Act 3: Gran Finale

Open Source AI systems and models are expressly regulated in the **AI Act**  
(badly, alas!)



# Open Source: **General exclusion:**

- The AI Act “**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 as an AI system that falls under Article 5 [prohibited practices] or 50 [Certain AI systems, like chatbots and deep fakes] (art 2.12)





# [A] Definition

- **Software and data, including models**, released under a free and open-source licence that allows them to be openly shared and where users can freely access, use, modify and redistribute them or modified versions thereof [...] (Recital 102).
- The licence should be considered to be free and open-source also when it allows users to run, copy, distribute, study, change and improve **software and data**, including models under the condition that the original provider of the model is credited, the identical or comparable terms of distribution are respected (Recital 102).



# General exclusions for Open Source AI models

- **No obligation to comply** with requirements targeting the responsibilities along the AI value chain **for third parties** which make accessible to the public public tools, services, processes, or AI components other than general-purpose AI models **when those tools, services, processes, or AI components are made accessible under a free and open licence** (Art. 25, par. 4, Recital 89)
- Developers are encouraged to implement widely adopted documentation practices



# Obligations for providers of general-purpose AI models (artt. 53-54)

- **not applicable to providers of AI models released under a free and open licence** that allows for the access, usage, modification, and distribution of the model, and whose parameters, **including the weights**, the **information on the model architecture**, and the information on model usage, **are made publicly available**



# Question

What IP right controls

1. model weights;  
and
2. model  
architecture?



# Rude awakening for Open Source People

And lawyers, mainly:

## Licenses are not central any more



# Everybody back to basics

- OSD: 9.5 out of 10 rules are on licensing
- 0.5 (#2) is about source code availability
- nothing of this makes sense anymore in AI
- Bonus point: No, Training data is **not** "AI source code"



# How

- We started with “AI Deep Dive”
- Planned to create a process, to collect consensus, and then start the actual collaborative process
- Then hell broke loose
- We hastened going public and proactively collect voices others than the unusual suspects
- Main issue: **training data**



# Long Story Short

- One year of process
- Lost one Board Member
- Lost sleep (alas, not weight, me)
- Lost funding (but Sloan shelled \$0.25M)
- Announced Oct 28 2024
- Same day, I stepped down from Chair position





# The Open Source AI Definition

- Preamble
- Definition – The Four Freedoms: Use, Study, Modify, Share
- Preferred form to make **modifications** for machine learning systems
- Open Source Models, Open Source Weights



# Adoption



We have several endorsement, open criticism.

We are endorsed by Eleuther AI and LLM360, for instance. Software Heritage, Eclipse, Mozilla and many others have also endorsed.



# Thank You For Your Attention

 <https://opensource.org>

  Carlo Piana



This work is licensed under a [Creative Commons - Attribution - ShareAlike 4.0](#)  
Presentation made using [Reveal.js](#) and a [Markdown](#) workflow with [reveal-md](#)

