3Os and IP awareness raising for collaborative ecosystems

ZOOOM

Open Source Machine Learning

Ivo Emanuilov, IP & Tech Lawyer
Leuven Centre for IT & IP Law - imec

imec

KU LEUVEN CiTiP
CENTRE FOR IT & IP LAW

LIBRe
libreresearchgroup.org

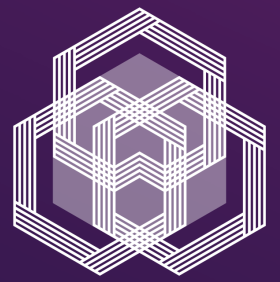
Funded by the European Union

Open Source AI: Enablement, Transparency and Reproducibility © 2023 by Ivo Emanuilov, Jutta Suksi is licensed under CC BY-SA 4.0. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/. DOI: 10.5281/zenodo.8418671

CC BY SA

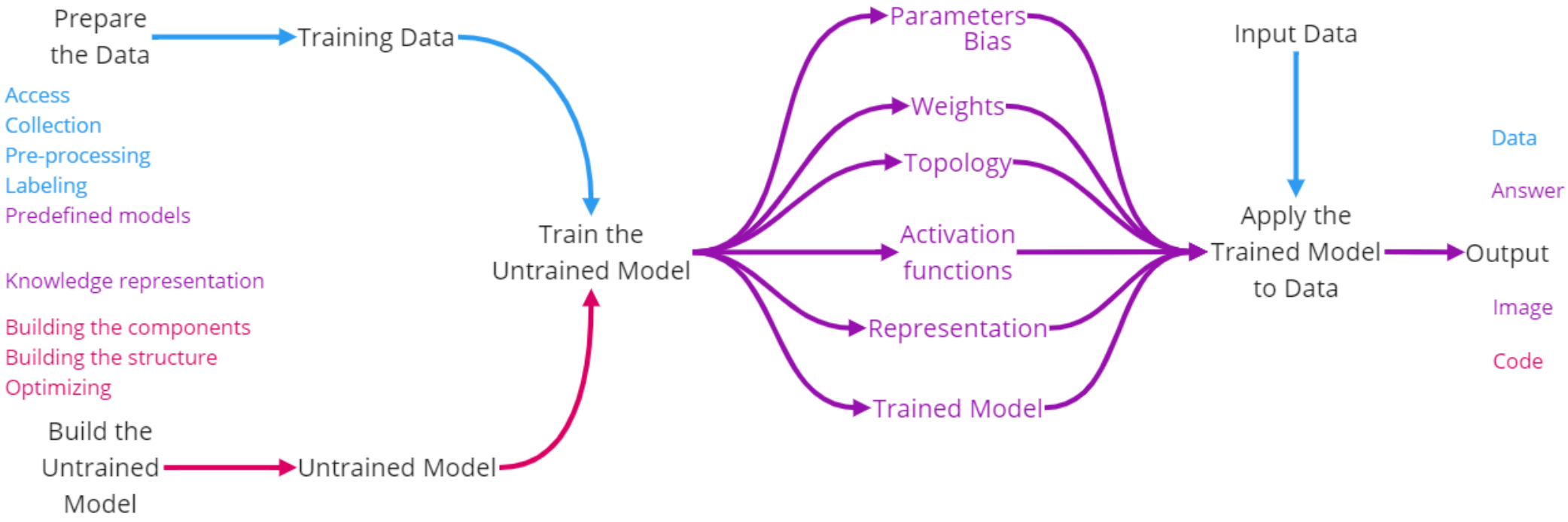
OUTLINE

- Building
- Deep dive in copyright and training data
- Layers in AI
- Deep dive in AI as hybrid IP
- Building open source AI
- Deep dive in the future of open source AI

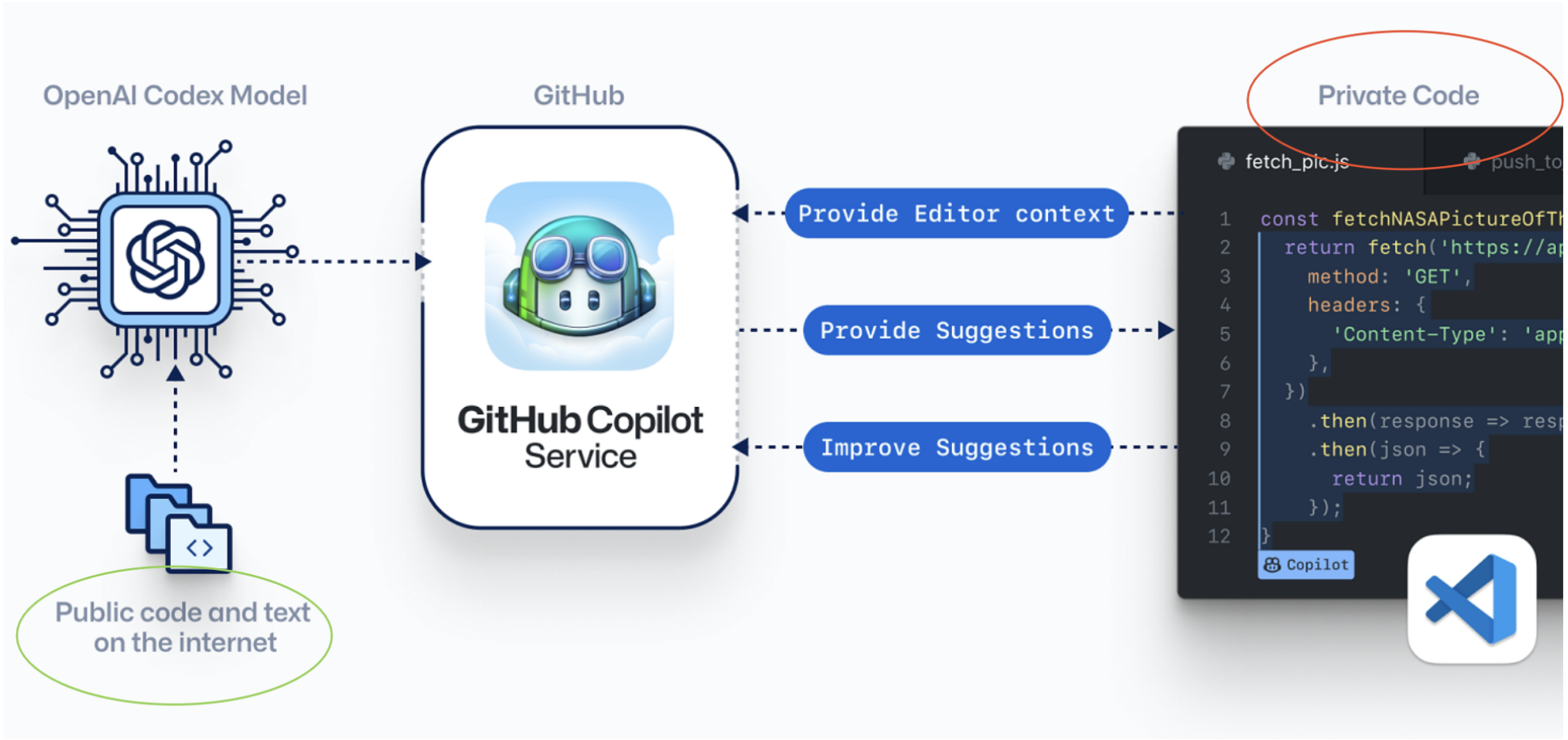


ZOOOM

AI Building Blocks



DEEP DIVE INTO COPYRIGHT AND TRAINING DATA



FROM CODE TO DATA TO CODE…

Article 4

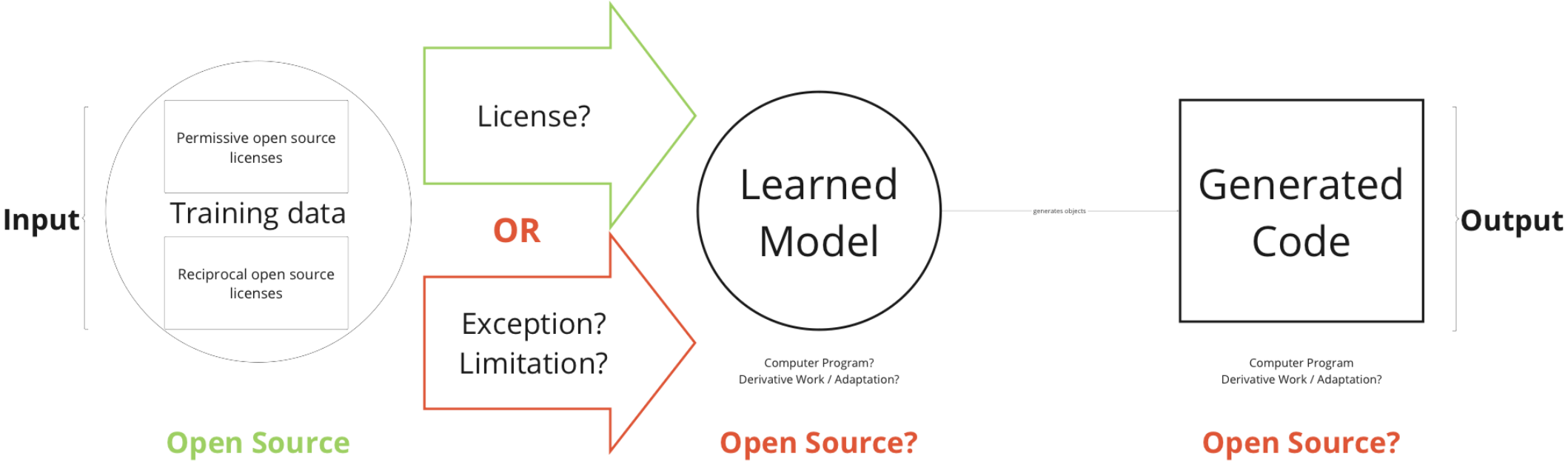
Exception or limitation for text and data mining

1. Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.

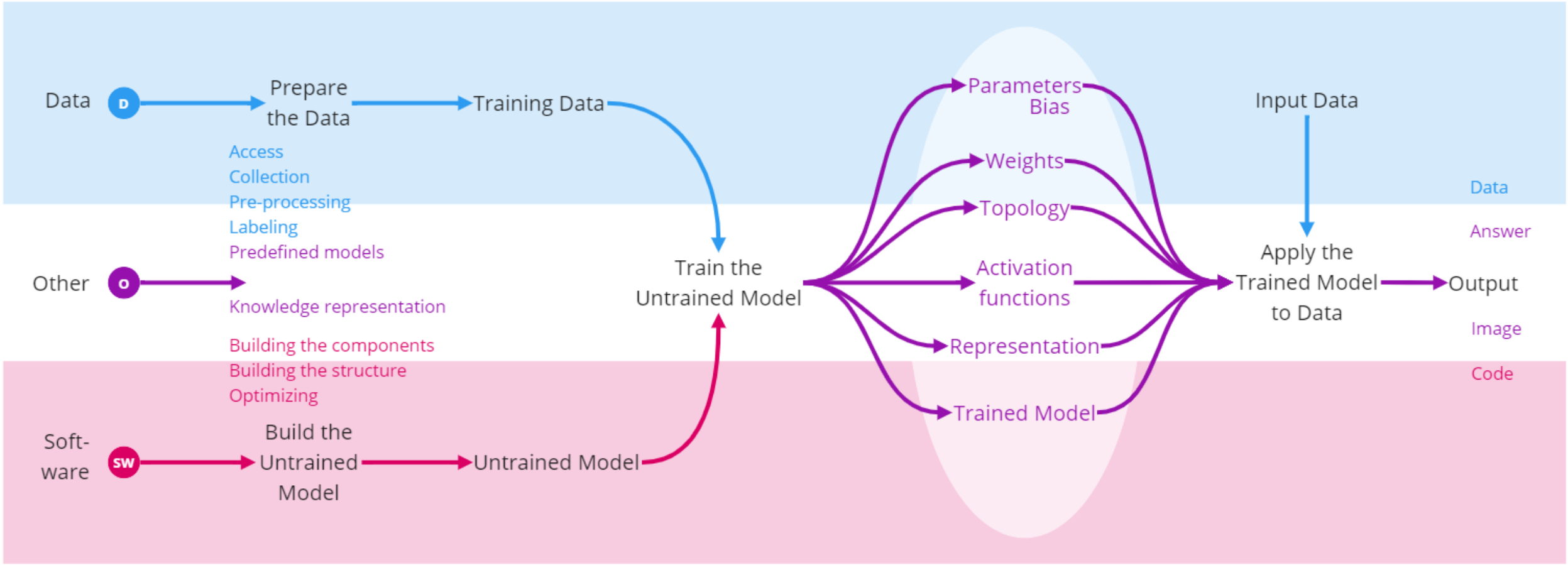
2. Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.

3. The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.

4. This Article shall not affect the application of Article 3 of this Directive.



LAYERS IN AI


Data
D
Prepare the Data
Training Data
Access
Collection
Pre-processing
Labeling
Predefined models
Other
O
Knowledge representation
Building the components
Building the structure
Optimizing
Soft-ware
SW
Build the Untrained Model
Untrained Model
Train the Untrained Model
Parameters Bias
Weights
Topology
Activation functions
Representation
Trained Model
Input Data
Apply the Trained Model to Data
Output
Data
Answer
Image
Code

AI AS HYBRID IP

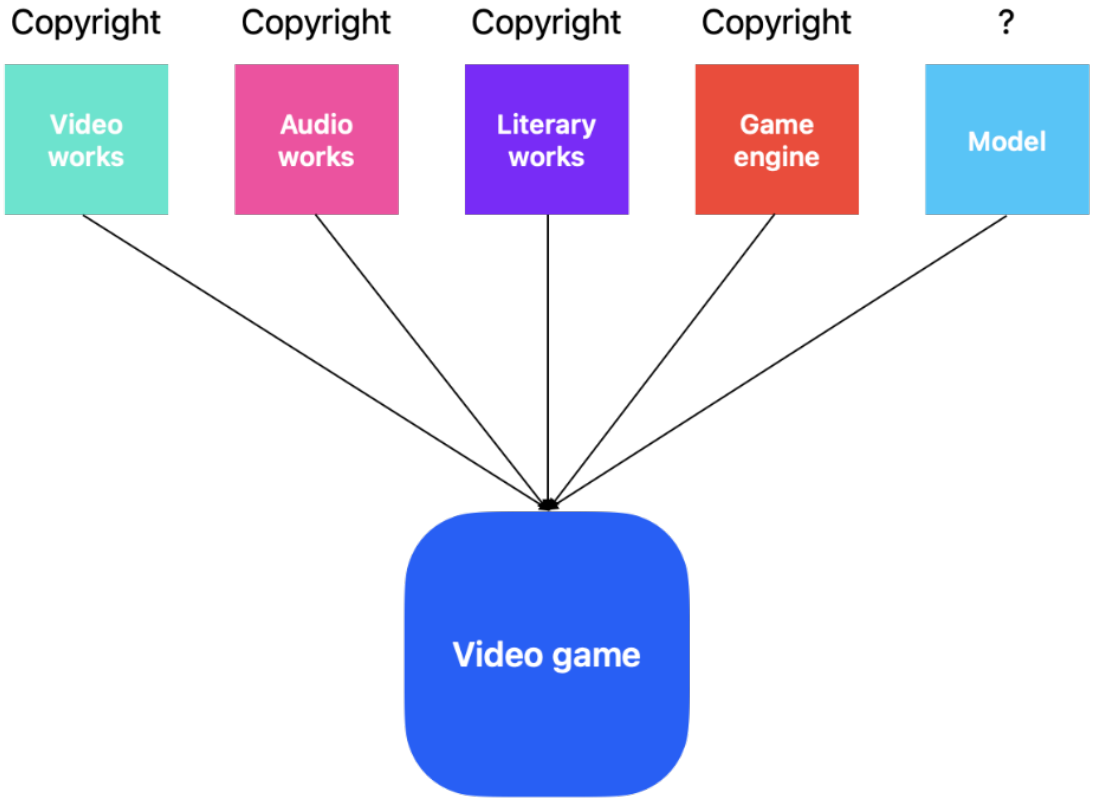
22. As regards the parts of a work, it should be borne in mind that there is nothing in Directive 2001/29 [InfoSoc Directive] indicating that those parts are to be treated any differently from the work as a whole. It follows that they are protected by copyright since, as such, they share the originality of the whole work (see Infopaq International, paragraph 38).

23. That finding is not weakened by the fact that Directive 2009/24 [Software Directive] constitutes a lex specialis in relation to Directive 2001/29 (see Case C-128/11 UsedSoft [2012] ECR, paragraph 56). In accordance with Article 1(1) thereof, the protection offered by Directive 2009/24 is limited to computer programs. As is apparent from the order for reference, videogames, such as those at issue in the main proceedings, constitute **complex matter** comprising not only a computer program but also graphic and sound elements, which, **although encrypted in computer language**, have a **unique creative value** which cannot be reduced to that encryption. In so far as the parts of a videogame, in this case, the graphic and sound elements, are part of its originality, they are protected, **together with the entire work**, by copyright in the context of the system established by Directive 2001/29.

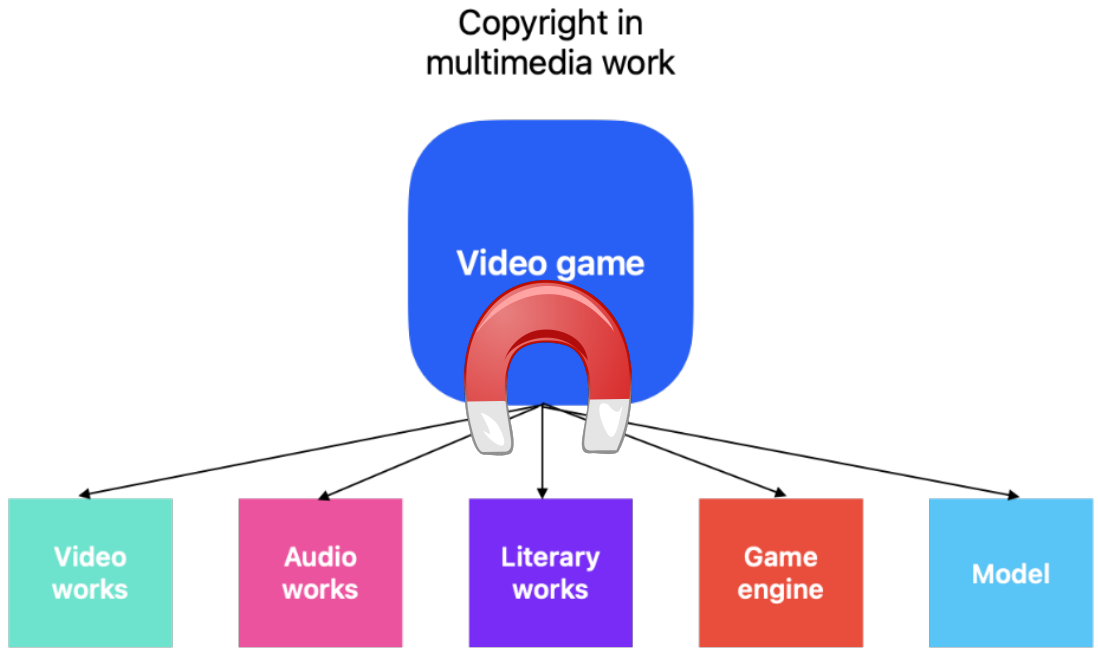
Case C-355/12, *Nintendo v Pc Box*, ECLI:EU:C:2014:25

CLOSED SUBJECT MATTER SYSTEMS / OPEN SUBJECT MATTER SYSTEMS

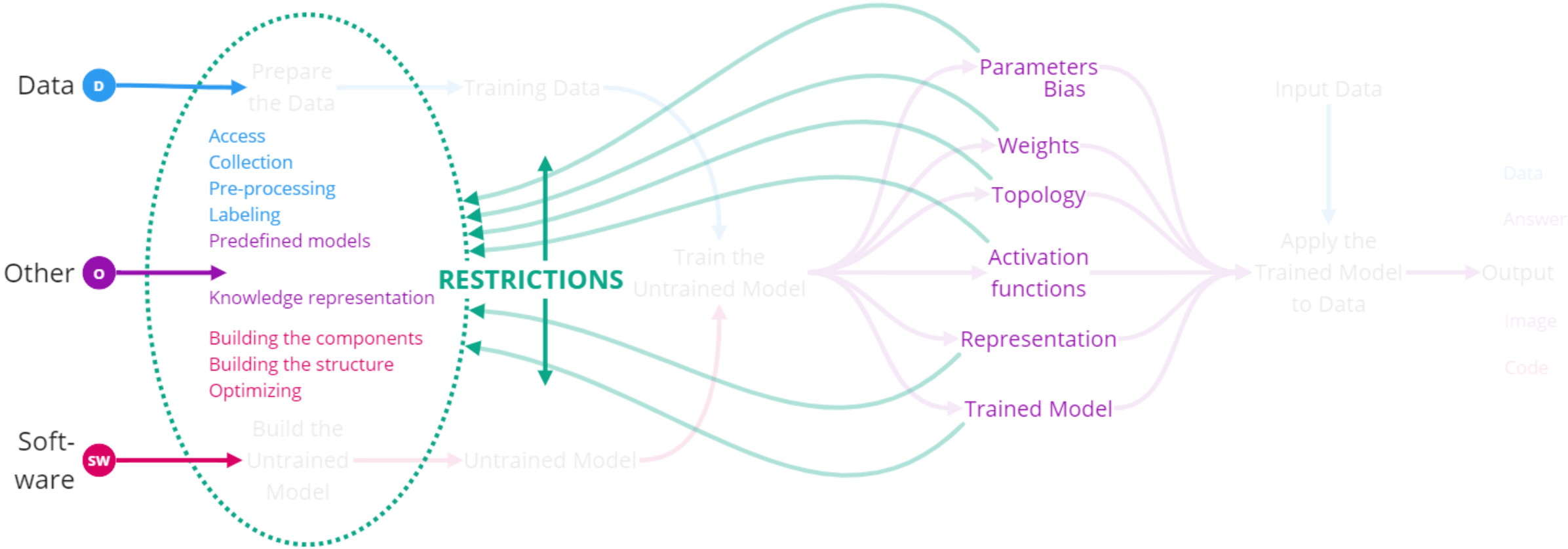
Closed subject matter systems



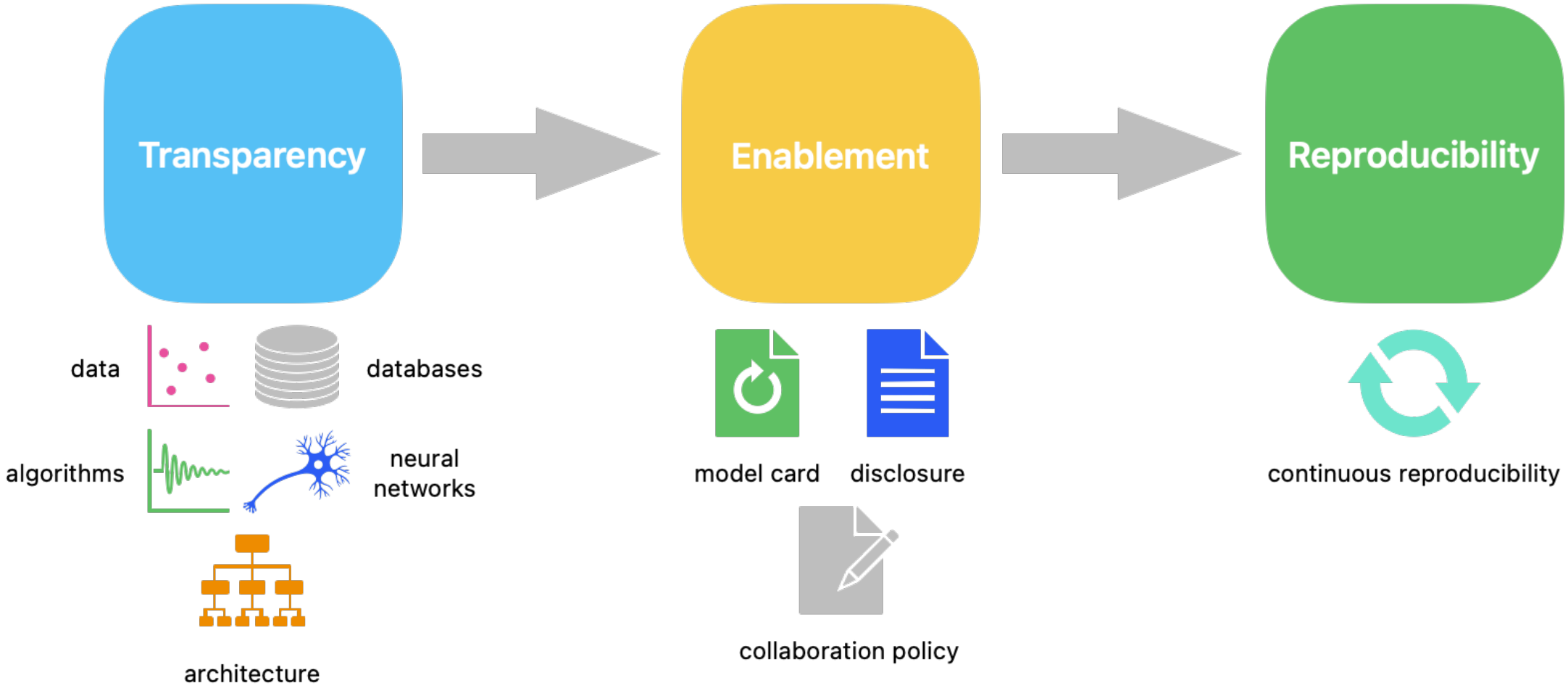
Open subject matter systems
Unitary approach



BUILDING OPEN SOURCE AI



Deep dive into the elements of an Open Source AI definition



EVOLUTION OF THE AI LICENSE SCENE

- Montreal data license (2019):
  - Taxonomy for licensing of data in AI and ML
  - Focus on data: the relationship betwen data and the models, not on openness
- Big Science BLOOM RAIL license 1.0 (2022)
  - Model specific
  - Focus on restrictions: open and permissive character while striving for responsible use of the model
- Open weights permissive license (2023):
  - Definition and license for open source compatible licensing of neural network weights (NNWs)
  - Focus on human-machine interactions: underlines the differences between open source software and NNWs with focus on openness

EVOLUTION OF THE AI LICENSING LANDSCAPE

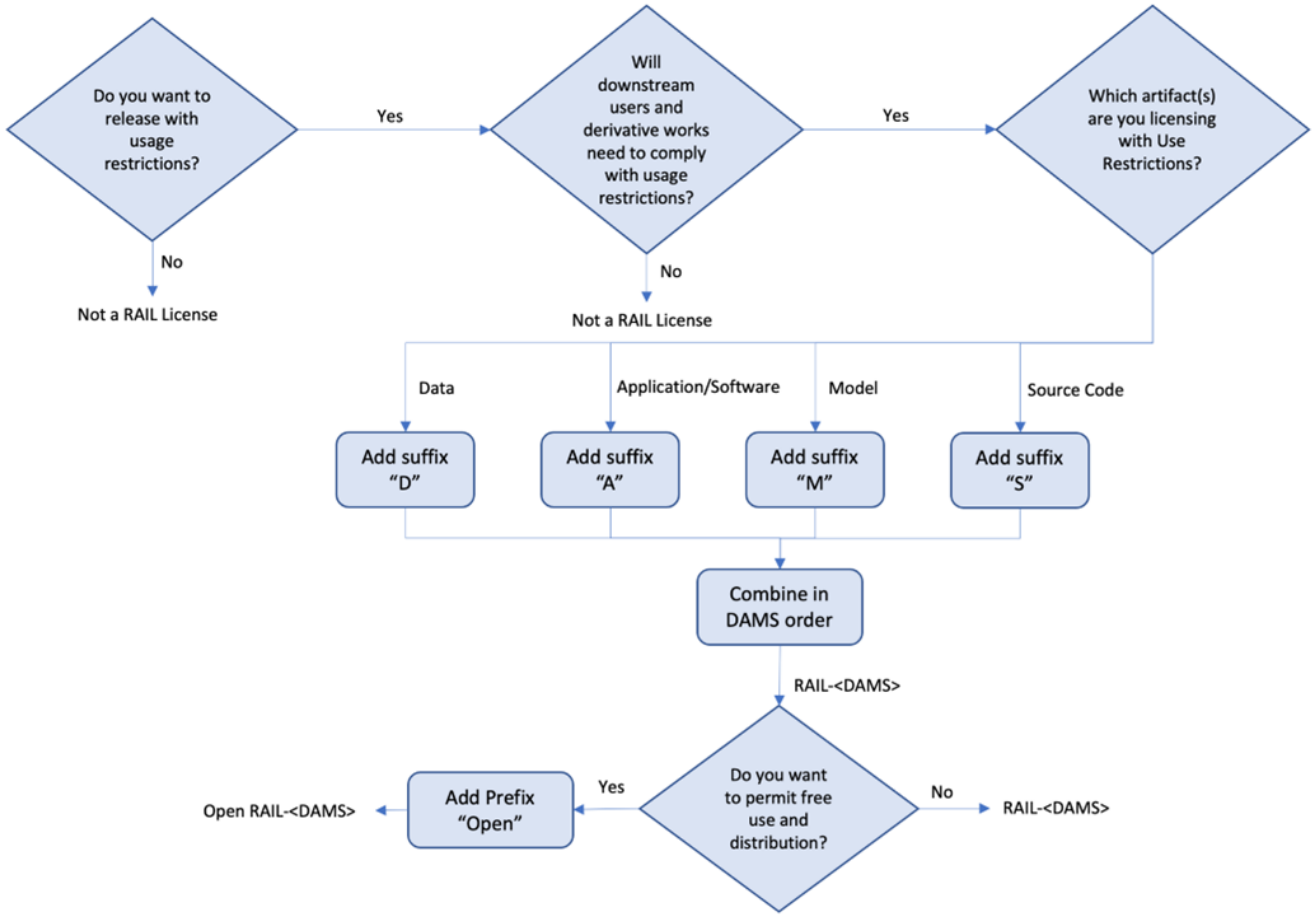
OpenRAIL-DAMS licenses (2022)

- A family of licenses aiming to prevent irresponsible and harmful applications
- Focus on the restrictions and combinations of data, applications, models and source code
  - Data license
  - Application/Executable license
  - Model license
  - Source code license

EVOLUTION OF THE AI LICENSING LANDSCAPE

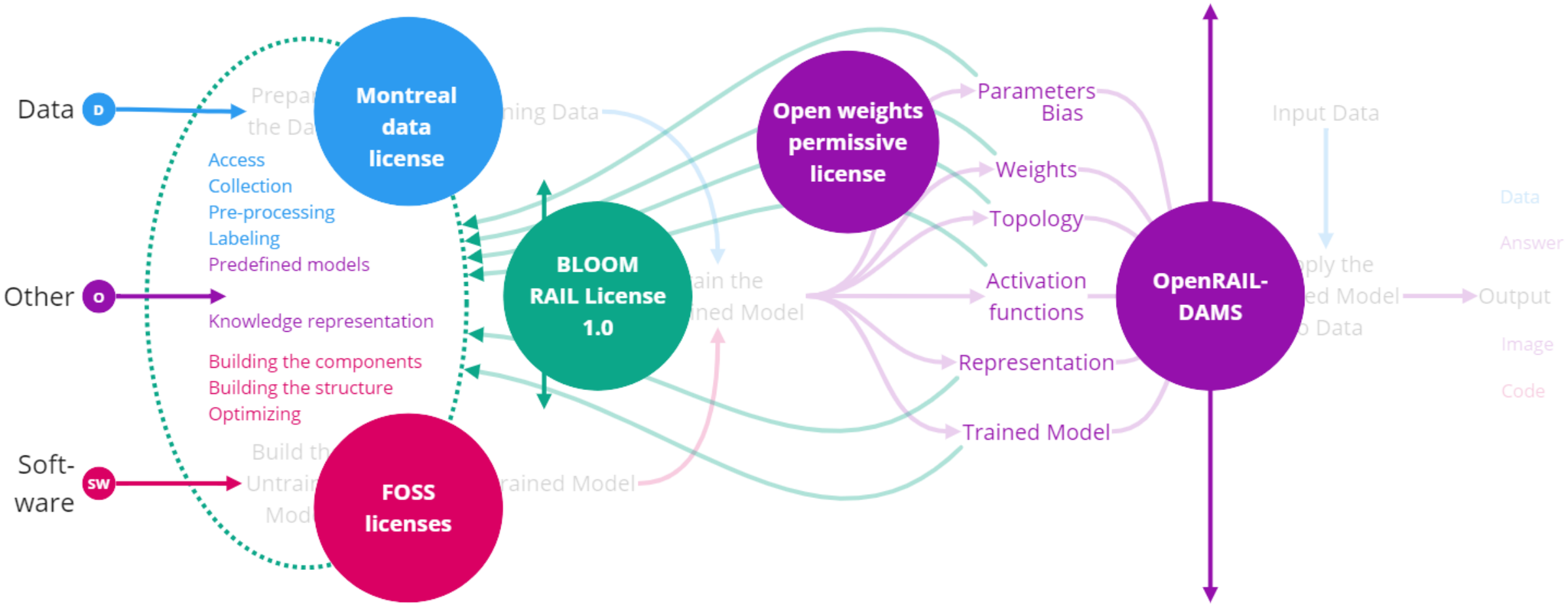
OpenRAIL-DAMS

Choosing the combination:



The OpenRAIL flow chart aids the selection and naming of a license for an ML project. Danish Contractor, Carlos Muñoz Ferrandis, Jenny Lee, & Daniel Mcduff. (2022, August)

WHAT DO WE KNOW…



ADOPTION OF AI LICENSES

Adoption of RAIL licenses:

"According to a recent paper that analyzed licenses attached to models on the HuggingFace hub, between September 2022 and January 2023, Open RAIL licenses have overtaken all other categories of restrictive open source licenses, and are now the second most used category after permissive open source software licenses."

Growth and Adoption of RAIL Licenses — Responsible AI Licenses (RAIL)

REGULATORY REQUIREMENTS AND IMPACT

Stanford Research on Foundation Models

Grading foundation model providers' compliance with the draft EU AI Act:

Grading Foundation Model Providers' Compliance with the Draft EU AI Act

Source: Stanford Research on Foundation Models (CRFM), Institute for Human-Centered Artificial Intelligence (HAI)

Draft AI Act Requirements	GPT-4	Cohere Command	Stable Diffusion v2	Claude	PaLM 2	BLOOM	LLaMA	Jurassic-2	Luminous	GPT-NeoX	Totals
Data sources	● ○ ○ ○	● ● ● ○	● ● ● ●	○ ○ ○ ○	● ● ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	22
Data governance	● ● ○ ○	● ● ● ○	● ● ○ ○	○ ○ ○ ○	● ● ● ○	● ● ● ●	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ○	19
Copyrighted data	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	7
Compute	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	● ○ ○ ○	● ● ● ●	17
Energy	○ ○ ○ ○	● ○ ○ ○	● ● ● ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	16
Capabilities & limitations	● ● ● ●	● ● ● ○	● ● ● ●	● ○ ○ ○	● ● ● ●	● ● ● ○	● ● ○ ○	● ● ○ ○	● ○ ○ ○	● ● ● ○	27
Risks & mitigations	● ● ● ○	● ● ○ ○	● ○ ○ ○	● ○ ○ ○	● ● ● ○	● ● ○ ○	● ○ ○ ○	● ● ○ ○	○ ○ ○ ○	● ○ ○ ○	16
Evaluations	● ● ● ●	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ● ○	● ● ○ ○	○ ○ ○ ○	● ○ ○ ○	● ○ ○ ○	15
Testing	● ● ● ○	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ○ ○	○ ○ ○ ○	● ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	10
Machine-generated content	● ● ● ○	● ● ● ○	○ ○ ○ ○	● ● ● ○	● ● ● ○	● ● ● ○	○ ○ ○ ○	● ● ● ○	● ○ ○ ○	● ● ○ ○	21
Member states	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ○ ○ ○	○ ○ ○ ○	9
Downstream documentation	● ● ● ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	● ● ● ●	● ● ● ●	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ○	24
Totals	25 / 48	23 / 48	22 / 48	7 / 48	27 / 48	36 / 48	21 / 48	8 / 48	5 / 48	29 / 48	

@misc{bommasani2023eu-ai-act, author = {Rishi Bommasani and Kevin Klyman and Daniel Zhang and Percy Liang}, title = {Do Foundation Model Providers Comply with the EU AI Act?}, url = {https://crfm.stanford.edu/2023/06/15/eu-ai-act.html}, year = {2023}}

LICENSING HYBRID TECHNOLOGIES

ZOOOM – 3Os and IP awareness raising for collaborative ecosystems (zooom4u.eu)

- Focusing on hybrids of open source software, open hardware and open data
- Emphasis on 4 emerging technologies: AI, Blockchain, Quantum, Robotics
- Bridging the understanding between legal aspects and business aspects
- Practical tools to be identified and a toolbox developed
- Stay tuned!

LIBRe
libreresearchgroup.org

ZOOOM

THANKS FOR LISTENING!

Open Source AI: Enablement, Transparency and Reproducibility © 2023 by Ivo Emanuilov, Jutta Suksi is licensed under CC BY-SA 4.0. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/. DOI: 10.5281/zenodo.8418671

imec

KU LEUVEN CiTiP
CENTRE FOR IT & IP LAW