OUTLINE FOR THE PRESENTATION

- Phases of building AI
- Components of building AI
- Deep dive to copyright and training data
- Layers in AI
- Deep dive to AI as hybrid IP
- Building open source AI
- Deep dive to open source AI definition
- Evolution of the AI license scene
- Learnings from the AI license scene
- Adoption of AI licenses
- Regulatory requirements and impact
- Learning to understand hybrid technologies
PHASES OF BUILDING AI
PHASES OF BUILDING AI

Training Data

Train the Untrained Model

Untrained Model
PHASES OF BUILDING AI
PHASES OF BUILDING AI

Training Data

Train the Untrained Model

Untrained Model

Trained Model

Input Data

Apply the Trained Model to Data

www.zooom4u.eu
PHASES OF BUILDING AI

Training Data

Train the Untrained Model

Untrained Model

Trained Model

Input Data

Apply the Trained Model to Data

Output
COMPONENTS OF BUILDING AI

Let’s dig a bit deeper!
COMPONENTS OF BUILDING AI
COMPONENTS OF BUILDING AI

Prepare the Data → Training Data → Train the Untrained Model → Applied to Data → Output

Input Data → Apply the Trained Model to Data

Untrained Model → Trained Model

Additional steps:
- Access
- Collection
- Pre-processing
- Labeling
COMPONENTS OF BUILDING AI

Prepare the Data → Training Data
- Access
- Collection
- Pre-processing
- Labeling

Train the Untrained Model
- Building the components
- Building the structure
- Optimizing

Build the Untrained Model → Untrained Model

Input Data → Apply the Trained Model to Data → Output

Trained Model
COMPONENTS OF BUILDING AI

Prepare the Data
- Access
- Collection
- Pre-processing
- Labeling
- Predefined models

Training Data

Train the Untrained Model
- Knowledge representation
- Building the components
- Building the structure
- Optimizing

Untrained Model

Input Data
Apply the Trained Model to Data

Output

Trained Model
COMPONENTS OF BUILDING AI

- Prepare the Data
  - Access
  - Collection
  - Pre-processing
  - Labeling
  - Predefined models
- Knowledge representation
- Building the components
- Building the structure
- Optimizing
- Build the Untrained Model

Train the Untrained Model
- Parameters
  - Bias
- Weights
- Topology
- Activation functions
  - Representation
  - Trained Model

Input Data
- Apply the Trained Model to Data

Output
COMPONENTS OF BUILDING AI

1. Prepare the Data
   - Access
   - Collection
   - Pre-processing
   - Labeling
   - Predefined models

2. Build the Untrained Model
3. Train the Untrained Model
4. Input Data
5. Parameters Bias
6. Weights
7. Topology
8. Activation functions
9. Representation
10. Trained Model
11. Apply the Trained Model to Data
12. Output
   - Data
   - Answer
   - Image
   - Code
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

Let’s analyze different property layers involved in AI!
LAYERS IN AI

Prepare the Data → Training Data → Train the Untrained Model

Access
Collection
Pre-processing
Labeling
Predefined models

Knowledge representation
Building the components
Building the structure
Optimizing

Parameters Bias → Weights → Topology
Activation functions
Representation
Trained Model

Input Data → Apply the Trained Model to Data → Output
Data
Answer
Image
Code
LAYERS IN AI

Data → Prepare the Data → Training Data → Parameters Bias
  → Weights → Topology → Activation functions
  → Representation → Trained Model → Apply the Trained Model to Data

Input Data → Output
  Data → Answer
  Image → Code

Access
Collection
Pre-processing
Labeling
Predefined models
Knowledge representation
Building the components
Building the structure
Optimizing
Build the Untrained Model
LAYERS IN AI

Data → Prepare the Data → Training Data
- Access
- Collection
- Pre-processing
- Labeling
- Predefined models

Knowledge representation
- Building the components
- Building the structure
- Optimizing

Software → 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

www.zooom4u.eu
LAYERS IN AI

Data
- Prepare the Data
  - Access
  - Collection
  - Pre-processing
  - Labeling
  - Predefined models

Other
- Knowledge representation
- Building the components
- Building the structure
- Optimizing

Software
- 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

www.zooom4u.eu
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).

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.
CLOSED SUBJECT MATTER SYSTEMS / OPEN SUBJECT MATTER SYSTEMS

Closed subject matter systems

Copyright
Video works

Copyright
Audio works

Copyright
Literary works

Copyright
Game engine

Copyright
Model

Video game

Open subject matter systems

Copyright in multimedia work

Video game

Video works

Audio works

Literary works

Game engine

Model
BUILDING OPEN SOURCE AI

How to build similar dynamics to open source AI as we have for open source software?
BUILDING OPEN SOURCE AI
BUILDING OPEN SOURCE AI
BUILDING OPEN SOURCE AI
Deep dive into the elements of an open source AI definition

- Transparency
  - data
  - databases
  - algorithms
  - neural networks
  - architecture

- Enablement
  - model card
  - disclosure
  - collaboration policy

- Reproducibility
  - continuous reproducibility
Evolvement of the AI License Scene

- Montreal data license (2019):
  - Taxonomy for licensing of data in AI and ML
  - Focus on data: the relationship between 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
Evolvement of the AI License Scene

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 License Scene

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)
LEARNINGS FROM THE LICENSE SCENE TO OPEN SOURCE AI
LEARNINGS FROM THE LICENSE SCENE TO OPEN SOURCE AI
LEARNINGS FROM THE LICENSE SCENE TO OPEN SOURCE AI
LEARNINGS FROM THE LICENSE SCENE TO OPEN SOURCE AI
LEARNINGS FROM THE LICENSE SCENE TO OPEN SOURCE AI
LEARNINGS FROM THE LICENSE SCENE TO OPEN SOURCE AI

Data
- Montreal data license
- Open weights permissive license
- BLOOM RAIL License 1.0
- FOSS licenses
- OpenRAIL-DAMS

Other
- Knowledge representation
- Building the components
- Building the structure
- Optimizing
- Predefined models

Softwae
- Access
- Collection
- Pre-processing
- Labeling

Input Data
- Parameters
- Bias
- Weights
- Topology
- Activation functions
- Representation
- Trained Model

Output
- Data
- Answer
- Image
- Code

www.zooom4u.eu
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.”
# Regulatory Requirements and Impact

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

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

<table>
<thead>
<tr>
<th>Draft EU AI Act Requirements</th>
<th>GPT-4</th>
<th>Cohere</th>
<th>Stable Diffusion v2</th>
<th>Claude</th>
<th>PaLM 2</th>
<th>BLOOM</th>
<th>LLaMA</th>
<th>Jurassic-2</th>
<th>Luminous</th>
<th>GPT-NeoX</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sources</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>22</td>
</tr>
<tr>
<td>Data governance</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>19</td>
</tr>
<tr>
<td>Copyrighted data</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>7</td>
</tr>
<tr>
<td>Compute</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>17</td>
</tr>
<tr>
<td>Energy</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>16</td>
</tr>
<tr>
<td>Capabilities &amp; limitations</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>27</td>
</tr>
<tr>
<td>Risks &amp; mitigations</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>16</td>
</tr>
<tr>
<td>Evaluations</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>16</td>
</tr>
<tr>
<td>Testing</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>21</td>
</tr>
<tr>
<td>Machine-generated content</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>9</td>
</tr>
<tr>
<td>Member states</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>9</td>
</tr>
<tr>
<td>Downstream documentation</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>☐ ☐ ☐</td>
<td>24</td>
</tr>
</tbody>
</table>


@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}}
Learning to understand 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!
TIME FOR DISCUSSION!