

EVIDEN

Data & trained models licensing

Where are we today?

Yann Dietrich
Head of IP / Eviden
09/10/2023

© Eviden SAS

an atos business

How do we access to data for AI?



Open Data mechanisms

Data scratching

Foundation models

Open data mechanisms

Numerous sources available



Creative commons license

in a statement in August, 2023, CC General Counsel explained the dependency on CC licenses on how fair use will apply to AI -

<https://creativecommons.org/2023/08/18/understanding-cc-licenses-and-generative-ai/>



Open source license

Some datasets are made available under Open source licenses such as Apache, BSD, MIT – are they really adapted? Copyright license? Liability? Output?



Open Data

No unique license – each open data initiative has its own terms & conditions – in France, « Licence ouverte / Open license » -

<https://www.etalab.gouv.fr/licence-ouverte-open-licence//> - <https://www.etalab.gouv.fr/wp-content/uploads/2018/11/open-licence.pdf>

Data scratching

Not the magic recipe so far



No clarity in US about how fair use applies to AI

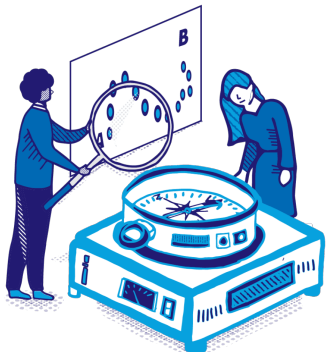
Unclear whether you can scratch data for AI training without a license for copyrighted materials – US Copyright office seeks comments and inputs

- <https://www.copyright.gov/newsnet/2023/1017.html>



Text & data mining exceptions in EU

The exception is not consistently transposed across Europe and we can expect massive opt-out by all copyright owners (and be mindful of website terms & conditions)



Not copyrighted but privacy ...

Copyright is not the only issue when scratching data – Privacy is also a concern – Personal data may be extracted out of a trained model –

Some scientists demonstrated that pictures of individuals can be extracted out of a LLM (e.g. stable diffusion) - <https://arxiv.org/pdf/2012.07805.pdf>

LLMs & trained models

Please read the terms & conditions – new issues are coming



Llama Community license agreement

A very open and free LLM accessible – with a non-compete restriction (cannot be used to improve any other LLMs) and also use restriction -

<https://github.com/facebookresearch/llama/blob/main/LICENSE>



Azure & OpenAI

Complex contractual licensing scheme through Microsoft terms & conditions with various interesting legal features



RESPONSIBLE AI
LICENSES

RAIL license

RAIL is an initiative to develop specific open licenses adapted to AI addressing especially the ethical considerations to AI -

<https://www.licenses.ai/ai-licenses>



One Ring to
Rule them all
One Ring to
find them
One Ring to
Bring them all
and in the darkness
Bind them

LORD OF THE RINGS

Initiatives to develop templates

Some efforts are done to facilitate data sharing

INTERNATIONAL DATA SPACES ASSOCIATION



A working group of IDSA members is working on legal template for data sharing



Licenses and tools to foster data sharing

THE LINUX FOUNDATION PROJECTS

LF AI & DATA

Develop and promoting the Community data license agreement (<https://cdla.dev/permissive-2-0/>)



A WG is working on data provisions contracts (<https://documents-dds-ny.un.org/doc/UNDOC/LTD/V23/064/75/PDF/V2306475.pdf?OpenElement>)



Working on various policy aspects and business principles



Working group on data sharing templates as contemplated by the Data act

Key points of attention

And also applicable in all your data&AI related agreements



Transparency & reliability (and the liability related)

Ensuring transparency & reliability of datasets will be a strong obligation on any AI developer under the AI act (or other législations) and some related liabilities



Ethical limitations

Proper management of such ethical restrictions are essential for compliance with all licenses especially when you are involved in certain activities – This is corresponding to high risks AI under the future AI act.



Outputs

Most of the template are making it clear that the owner of the data and/or of the trained models will not have any ownership or alike on the outputs. Important question also of the de minimis portion of the data (CDLA license)



Trade Secrets

The issue is coming with the data act and the ability of the data owners to seek for some contractual protection of their trade secrets.

A very recent template

EU model contractual AI Clauses to pilot in procurements of AI

<https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/eu-model-contractual-ai-clauses-pilot-procurements-ai>

Some examples of provisions

The risk management measures referred to in paragraph 2.2, point (c) shall be such that relevant residual risks associated with each hazard as well as the overall residual risk of the AI system is reasonably judged to be acceptable by the Supplier, provided that the AI System is used in accordance with the Intended Purpose or under conditions of Reasonably Foreseeable Misuse.

The Supplier ensures that the Data Sets used in the development of the AI System are relevant, representative, and to the best extent possible free of errors and be as complete as possible in view of the Intended Purpose

Next challenges : how do we develop objective criteria to assess compliance with such criteria?

A very recent template

Society for Computers & Law (SCL) AI Group Artificial Intelligence Contractual Clauses

<https://www.scl.org/files/download/2158-AI%20Clauses%20Project%20-%20October%202023.pdf>

Some examples of provisions

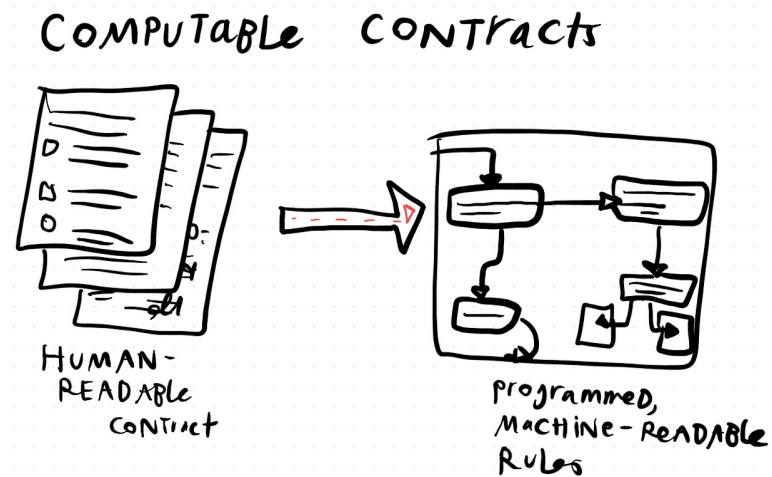
Circuit Breaker provision - S1.9 The Supplier reserves the right to deploy a circuit-breaker capable of [interrupting and stopping] the AI System. The Supplier shall be entitled to deploy the circuit-breaker in any circumstances where the Supplier, acting reasonably, considers it is necessary to do so or to comply with applicable law. The Customer shall not be entitled to any compensation or refunds in any circumstances where the Supplier deploys the circuit-breaker in accordance with this clause S1.9.

S4.8 The Supplier warrants that it has taken reasonable steps to mitigate the risk of the AI system causing [material] discrimination against a class of individuals on the basis of one or more of the protected characteristics set out in the Equality Act 2010.

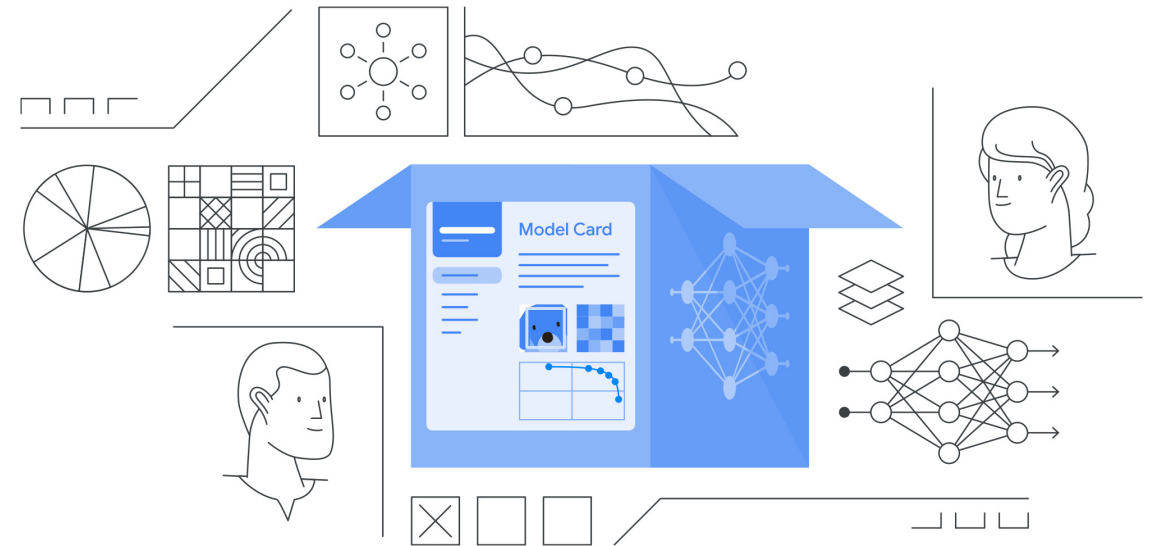
Interestingly, this document is showing pro-customer / pro-supplier views

And some « technical » trends

And also applicable in all your data&AI related agreements



Computer contracts & automated compliance monitoring



Data cards for data traceability and transparency

EVIDEN

Confidential information owned by Eviden SAS, to be used by the recipient only.
This document, or any part of it, may not be reproduced, copied, circulated
and/or distributed nor quoted without prior written approval from Eviden SAS.

© Eviden SAS