Legal profiles of AI and copyright in Europe

Dr. Simona Lavagnini
attorney at law
LGV Avvocati
Milan - Italy

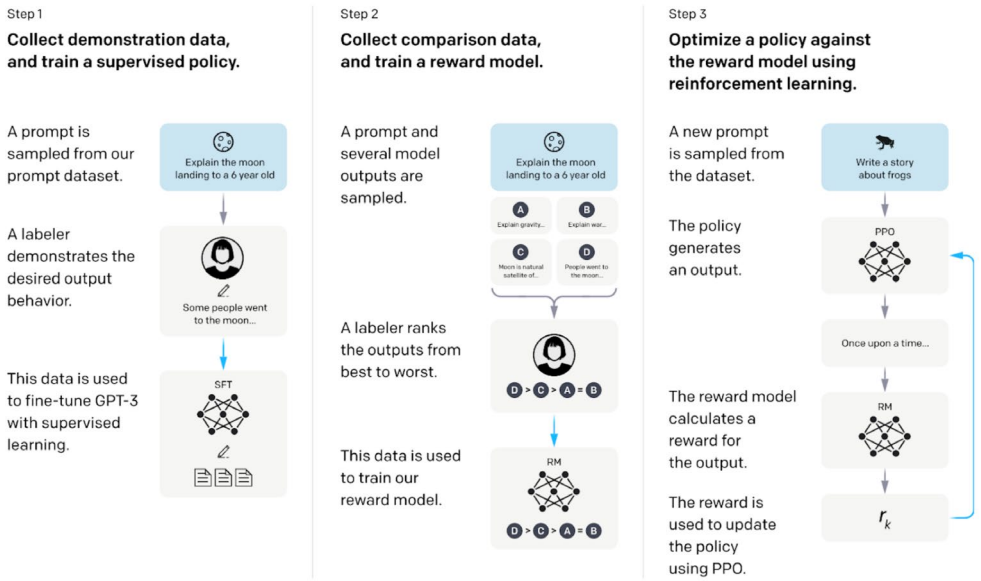
What is AI and AI General Purpose (from the AI Act)

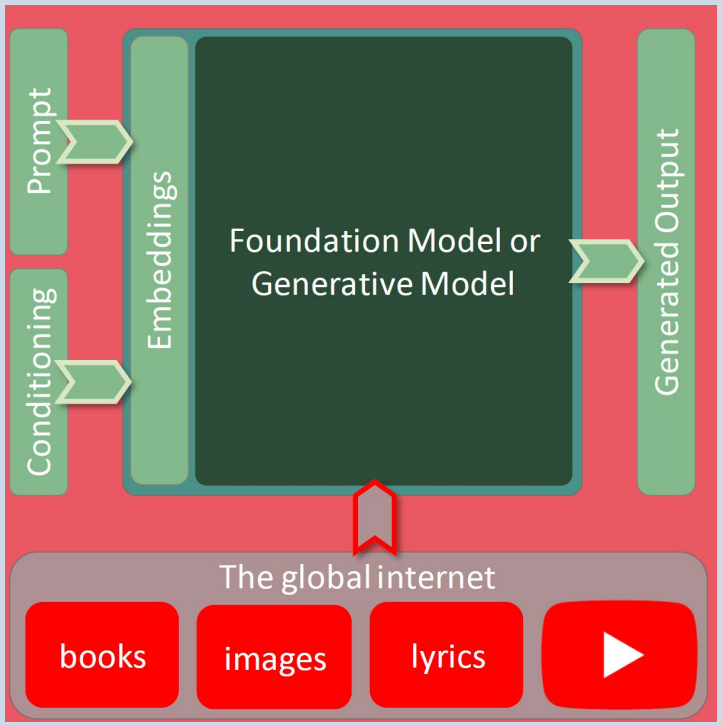
"AI system" means a machine-based system designed to operate with varying levels of **autonomy**, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to **generate** outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

"General-purpose AI system" means an AI system which is based on a general-purpose AI model, which has the capability to serve a variety of purposes, both for direct use as well as for integration into other AI systems.

LGV

How AI works


Step 1
Collect demonstration data, and train a supervised policy.
A prompt is sampled from our prompt dataset.
Explain the moon landing to a 6 year old
A labeler demonstrates the desired output behavior.
Some people went to the moon...
This data is used to fine-tune GPT-3 with supervised learning.
SFT
Step 2
Collect comparison data, and train a reward model.
A prompt and several model outputs are sampled.
Explain the moon landing to a 6 year old
A labeler ranks the outputs from best to worst.
D > C > A = B
This data is used to train our reward model.
RM
D > C > A = B
Step 3
Optimize a policy against the reward model using reinforcement learning.
A new prompt is sampled from the dataset.
Write a story about frogs
The policy generates an output.
PPO
Once upon a time...
The reward model calculates a reward for the output.
RM
The reward is used to update the policy using PPO.
r_k


Prompt
Conditioning
Embeddings
Foundation Model or Generative Model
Generated Output
The global internet
books
images
lyrics

Key Points for Copyright Issues

- Artificial intelligence systems are based on training in which a large set of data (inputs), chosen by the system developer, is initially provided to the system.
- the input can be a work of art (music, video, photographs, etc.)
- the input is corrupted by the system, i.e. decomposed, adding 'noise'.
- the AI must reconstruct the dataset, reversing the corruption process. The goal is to recreate the initial input (work), but since the method used by the system is digital and statistical, the system reconstructs the digital inputs by inserting one piece of data after another according to the probability of the relationship between the data, so that the final result can never be identical to the original before the corruption (only similar, or of the same type).
- this training creates the model (which identifies the relationships between the elements that make up the input).
- once the model is generated, the AI will be able to produce new content, without using any specific original content, by applying the notions acquired through the analysis of the datasets originally provided, which were identified by linking each content to its description in natural language.
- AI systems do not keep copies of the input, or even parts. Rather, they are trained to determine - in a general way - which element tends statistically to be in a certain position in the work. During the training there is however reproduction of the work of art.

What impact of AI on the cultural industry?

- Generative artificial intelligence systems represent a potential revolution for the industry
- They are both an opportunity and a risk
- Examples:

Easy creation by everybody of lyrics and music

Realization of deepfakes (e.g. use of images and sounds in which a performance by a physical person artist is emulated, creating a song and having it performed with the voice and image of the artist himself - example the Drake's deepfake).

Some risks and opportunities

- **Reconfiguration of the cultural industry and copyright landscape** (which actors and which exclusive rights - emergence of personality rights).
- **Impact on cultural diversity** (recommender systems direct users towards certain types of content, there is standardization and a risk of discrimination).
- **Access to data and monopolization** (some companies hold large databases on users and provide access only in part to other players).

AI Act

- It is the first comprehensive legislation of its kind
- It is a regulation and therefore directly and immediately applicable in Europe by national and EU courts, although further action will be needed at EU and local level
- Provides for new governance structures (European AI Office - European AI Board - Advisory Forum - Scientific Panel)
- What relationship with national copyright laws?

The primary objective of the AI Act

Key provision for copyright in the AI Act:

- Art. 53 regarding the conditions under which the use of protected works is lawful, and transparency regarding the content used
- Art. 50 regarding transparency to the public that the content is generated by the IA
- There are no provisions for the protection of outputs from the point of view of Union law

"The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter of Fundamental Rights, including democracy, the rule of law and environmental protection, against the harmful effects of artificial intelligence systems (AI systems) in the Union, and to support innovation".

The discipline of AI in Europe

- 1 August 2024 - entry into force of the EU AI Act (Regulation 1689/2024).
- 2 February 2025 - banned artificial intelligence practices must be withdrawn from the market (unacceptable risks such as technologies to manipulate people's behavior, biometric surveillance, etc.).
- 2 May 2025: codes of conduct between developers, companies and industry associations and the European Union will be ready to ensure the inclusion of AI Act principles in AI development (environmental and social sustainability, training and literacy, ethical principles).
- 2 August 2025: Gpai must comply with the new rules (content tagged in a machine-readable system and recognizable as generated by an AI; deepfakes tagged; users made aware if they are interacting with a chatbot). A governance structure will have to be set up (AI Office, European Committee for Artificial Intelligence, national market surveillance authorities, etc.).
- 2 February 2026: the EU Commission will adopt an implementing act with detailed provisions establishing a template for the post-market monitoring plan and the list of elements to be included in the plan.
- 2 August 2026: all AI Act regulations become applicable, including the requirements for high-risk systems defined in Annex III (list of high-risk use cases). Member States shall ensure that their competent authorities have established at least one operational AI regulatory testing environment at national level.
- 2 August 2027: the obligations for high-risk systems defined in Annex I (list of EU harmonization rules) apply

Highlights of the AI Act on the lawful use of data/works by AI

- Training as text and data mining is an exception, therefore permissible under certain conditions - yes for scientific research, no for commercial activities if there has been an opt out by the rights holders
- There remains the problem of the three steps test for some types of generative AI
- There remains the problem of the opt-out and its compatibility with the Berne Convention and the ban on formalities for protection

Extraction of protected works to instruct AI

TDM is defined in Art. 2 of the Copyright Directive as "*any automated analysis technique aimed at analyzing texts and data in digital format with the purpose of generating information including, but not limited to, patterns, trends and correlations*". Relating provisions have been transposed at national level (see for instance Articles 70-ter and 70-quater Italian Copyright Law).

- Providers must comply with Union law on copyright and related rights, in particular to identify and respect reservations of rights expressed by right-holders pursuant to Article 4(3) of Directive (EU) 2019/790.

- Any supplier placing a general-purpose AI model on the Union market must comply with this obligation, **regardless of the jurisdiction in which the relevant copyright acts underlying the training of such general-purpose IA models take place**.

- This is necessary to ensure a level playing field among providers of general purposes AI models, as no provider should be able to gain a competitive advantage in the Union market by applying less stringent copyright rules than in the Union.

Highlights of the AI Act with regard to transparency regarding the content used

- AI model providers should develop and make available to the public a sufficiently detailed **summary** of the content used for training. While duly taking into account the need to **protect trade secrets** and confidential business information, this summary should be **broad and general in scope**, rather than detailed in technical terms, in order to facilitate parties in exercising and enforcing their rights under Union law, for example by listing the main collections or datasets that have been included in the model training, and providing a description of the other data sources used.

- The AI Office will provide a **template** for the summary, which should be simple and effective and allow the provider to provide the requested summary in descriptive form. The AI Office should check whether the provider has fulfilled these obligations without verifying or assessing the training data in terms of copyright compliance.

- The Regulation is without prejudice to the application of copyright rules under Union law

Highlights of the AI Act with regard to transparency regarding the interaction

- Providers shall ensure that AI systems intended to **interact directly** with natural persons are designed and developed in such a way that the natural persons concerned are **informed** that they are interacting with an AI system, unless this is **obvious** from the perspective of a reasonably informed, alert and prudent natural person, taking into account the circumstances and context of use.

- Providers **shall ensure that the results of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated**. Vendors shall ensure that their technical solutions are effective, interoperable, robust and reliable to the extent technically feasible, taking into account the specificities of different content, implementation costs and state of the art. **This obligation does not apply if AI systems perform a standard editing assistance function or do not substantially alter the input data provided by the user or their semantics**.

- The Regulation is without prejudice to the application of copyright rules under Union law

LGV

Article 50 (3 of 4)

4. **Anyone using an AI system that generates or manipulates image, audio or video content that constitutes a deep fake must declare that the content has been artificially generated or manipulated**. This obligation does not apply if the use is authorised by law to detect, prevent, investigate or prosecute crimes. If the content is part of an obviously artistic, creative, satirical, fictional or similar work or programme, the transparency obligations in this paragraph are limited to the disclosure of the existence of such generated or manipulated content in an appropriate manner that does not obstruct the viewing or enjoyment of the work. Persons who use an IA system that generates or manipulates a published text for the purpose of informing the public about matters of public interest are required to disclose that the text has been artificially generated or manipulated. This obligation does not apply when the use is authorised by law to detect, prevent, investigate or prosecute criminal offences or when the AI-generated content has undergone a process of human review or editorial control and when a natural or legal person has editorial responsibility for the publication of the content.

In summary regarding the transparency (labelling) of AI contents

- Outputs that are audio video or text material must be labelled by the suppliers.
- The so-called deep fakes must be recognizable as such and this is an obligation that also falls on the users of AI.

AI and case law in Europe (1 of 2)

- **ITALY - January 2023 Supreme Court**

The use of AI (creation of image) does not exclude protectability but creativity must be scrutinized more rigorously.

- **CZECH REPUBLIC - April 2024 Municipal Court of Prague**

Images created using AI tools cannot be protected by copyright as their author is not a natural person. Individuals who publish images created by AI tools cannot prevent others from copying and exploiting those images by invoking copyright law. The decision indicates that in the legal sphere of the EU, the outputs created by AI tools only will not be able to rely on copyright protection to prevent their copying by third parties.

AI and case law in Europe (2 of 2)

- GERMANY - September 2024 Hamburg District Court

A photographer brought a against a non-profit organization (LAION) that created a dataset for training AI models. According to the claimant's allegations, LAION had infringed his copyright by reproducing one of his images without permission.

LAION created the dataset by aggregating publicly available images and their corresponding textual descriptions. The dataset was made publicly available for free and could be used to train AI models. During the creation process, LAION's software analyzed the images to ensure that the descriptions matched the image content. It was not disputed that LAION had downloaded a copy of Kneschke's image that was available in low resolution and watermarked.

The Court of Hamburg dismissed the lawsuit ruling out a copyright infringement as LAION's reproduction of the image was covered by the TDM exception for scientific research under Article 3 of the Digital Single Market (DSM) Directive

GDPR (Regulation on personal data in EU)

- In March 2023 there was a temporary ban of ChatGPT in Italy and investigation commenced by the Italian Data Protection Authority (IDPA). It was the first case of this type in the world.

- At the end of 2024 IDPA issued a 15 million euro fine holding that:

- OpenAI trained ChatGPT with users' personal data without first identifying a proper legal basis for the activity, as required under GDPR;

- OpenAI also failed to notify about a data breach;

- OpenAI did not provide proper age verification mechanisms for users under age 13.

- In addition to the fine, OpenAI must also conduct a six-month public education campaign. OpenAI intends to appeal the decision…meanwhile the IDPA on Jan. 30, 2025 suspended DeepSeek in Italy.

LGV

Thank you for your attention!

Dr. Simona Lavagnini
Attoney at law

LGV AVVOCATI
Via Privata Cesare Battisti 2
20122 Milano
www.lgvavvocati.it
+39 02 540 6971