"*Artificial Intelligence and Intellectual Property*"

Steering AI: Legal Challenges and Ethical Standards from an Engineering Prospective

prof. avv. Francesca Ferrari – DIECO – Insubria University

TECHNOLOGY AND INNOVATION

Technological breakthroughs have drastically changed the face of societies and economies. They have also changed how innovators generate new technologies and creators produce original works.

AI can be viewed

- as just a tool that enhances human capabilities, similar to the tools that have emerged in the past or

- as fundamentally new and different compared to past technologies in that it diminishes – or perhaps even replaces – the human ingenuity that so far has been at the root of every innovative and creative endeavor.

IP AND INNOVATION

The protection of intellectual property (IP) rights is at the heart of the incentive structure that underlies innovative and creative activities.

1. **Incentive Structure**

2. **Promoting Competition and Progress**

3. **Economic Growth**

4. **Cultural and Artistic Expression**

5. **International Trade and Collaboration**

6. **Balancing Rights and Public Interest**.

In conclusion, the protection of intellectual property rights is fundamental to fostering innovation, creativity, and economic growth. By providing incentives for investment, promoting competition, and facilitating cultural expression, strong IP protection contributes to a thriving and dynamic society.

AI & IP

AI poses tricky questions to the IP system, such as:

Should an AI-generated invention qualify for patent protection?

Should copyright protection apply to AI-generated novels?

Should it be legal to extract images from the Internet to train an AI algorithm that produces original works?

AI & IP AND ECONOMIC CONSIDERATIONS

The cited legal questions have important economic ramifications.

History has shown that seemingly small changes to IP laws can have profound effects on evolving innovation ecosystems.

Adjusting the IP system in the face of new technologies thus requires careful consideration.

Before changing any legal framework, it makes sense to first think about what is really changing, how any change affects the incentive structure for innovation and review available evidence.

NEED FOR RULES? WHAT RULES?

Artificial intelligence is now talked about on a daily basis: from industry to trade and stock markets, from medicine to robotics, from means of transport to information and communication systems, these are all areas in which intelligent computer systems are increasingly being applied.

New rules are often invoked to regulate a subject that raises ethical as well as theoretical and practical problems.

What rules are really necessary for artificial intelligence?

NEED FOR RULES?

Artificial intelligence is scary: the development and diffusion of the technology has been preceded by decades of cultural expressions, from literature to film, that have made it a myth:

2001: A Space Odyssey

Blade Runner

Matrix.

The same phenomenon has occurred in the field of literary production:

the work of Isaak Asimov dates back to the 1930s and was even mentioned in an act of the European Union, the European Parliament resolution of 16 February 2017 with recommendations to the Commission concerning civil law rules on robotics.

► One of the ways of reacting to fear is to invoke regulation and call for intervention by the legislature!

AI ACT

- contains a single definition of artificial intelligence:

'it is a machine-based system designed to operate with varying levels of autonomy and 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'.

- establishes a solid risk methodology for defining artificial intelligence systems and provides for four risk levels according to which AI applications should be categorised, which should consequently be subject to different levels of monitoring.

But …. How did we get to February 2024 ands the compromise agreement?

AI ACT

The EU Commission's proposal for the regulation dates back to 21 April 2021, that the Council subsequently unanimously adopted the general approach on the proposal on 6 December 2022, a few days before the advent of GPT chat with all the problems that not so much and not only Chat GP, but generative artificial intelligence in general poses.

► This is the reason why, although the general approach dates back to 6 December 2022 and the vote in the plenary of the European Parliament to 14 June 2023, the final approval of the compromise text of the regulation was only achieved on February 2, 2024.

IP ISSUES POSED BY AI

Intellectual property issues arise from:

❑ the data used to train the algorithms of generative artificial intelligence;

❑ the possible recognition of intellectual property rights to artificial intelligence systems and

❑ the protection of works generated by or with generative artificial intelligence systems.

INPUT AND … LITIGATION

Lawsuits concerning the allegedly unlawful use of training data are multiplying:

Getty Images vs. Stability AI,

Andersen vs. Stability AI

Doe vs. GitHub

New York Times vs. OpenAI and Microsoft.

► generative artificial intelligence systems would be trained with copyright-protected works for the production of outputs and all this would be done '*without consent, without credit, and without compensation*' for authors and owners of intellectual property rights.

INPUT … EVIDENCE ISSUE

in general, there is no evidence of counterfeiting and plagiarism in works produced by artificial intelligence systems as the latter certainly draw useful information from the data used to train them, but do not draw from a single source in producing an output.

INPUT: VIOLATION OF IP RIGHTS OR LEGITIMATE CONDUCT?

The collection of data for training purposes presents critical issues: while the use of public databases is obviously lawful, generalized scraping falls into a grey line, which could be covered by the Text and Data Mining exceptions regime under Art. 70 *ter* and *quater of* the Copyright Act (introduced by Legislative Decree no. 771 of 8 November 2021 implementing the Directive 'On Copyright and Related Rights in the Digital Single Market') or in the very broad notion of '*fair use*' in the United States, but it could also well qualify as unlawful use of copyrighted works.

TEXT AND DATA MINING (TDM) EXCEPTION

The Copyright Directive 2019/790/EU introduced the text and data mining (TDM) exceptions, which are regulated in Articles 3 (Text and data mining for the purposes of scientific research) and 4 (Exception or limitation for text and data mining). TDM is defined in Article 2 of the Copyright Directive as "*any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.*"

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790

TDM IN ITALIAN LEGAL SYSTEM

At the domestic level, these articles have been transposed, respectively, with the introduction into the Copyright Law of Articles 70-*ter* - which deals only with extraction for scientific purposes by research organizations and cultural heritage protection institutions - and 70-*quarter* - which allows the extraction of text and data in general, by anyone, even for mere profit.

Among the two TDM exceptions regulated by the European directive, the second one, which allows mining also for profit, deserves particular attention.

► Article 70-quater of the Copyright Law in fact exempts any text and data mining activity that is carried out on the intellectual work, including software or database protected by a related right, regardless of the purpose or the qualification of the person performing it.

ART. 70 *QUATER* ITALIAN COPYRIGHT LAW (633/1941)

It requires that:

- the person had legitimate access to the content for the purpose of text and data mining; and

- the owner of the copyright and related rights and/ or the owner of the database have not expressly reserved the extraction of text and data (opt out mechanism), thus bringing TDM's activities under its exclusive control.

THE RESERVATION AS PER ART. 70 *QUATER*

The liberalizing scope of the opt-out mechanism granted by Article 70-*quater* depends on the manner in which the reservation is made by the rights holder.

Article 4, para. 3 of the Copyright Directive requires that the reservation be expressed "*in an appropriate manner, such as machine readable means in the case of content made publicly available online.*"

This provision requires that the reservation statement be readable in an automated manner when the work to which it relates is made available to the public on the internet.

But how at domestic level the reservation should be made? The Italian lawmaker nothing stated on this issue in the context of article 70 quarter and thus the ambiguity remains.

TDM EXCEPTION AND AI TRAINING

Are we sure that this exception is applicable to the use of IP protected data for training Generative artificial intelligence algorithms and systems?

Many scholars and nowadays many courts too pointed out that Gen-AI training and TDM are distinct activities and thus the exception for the second does not apply to the first.

► TDM is the process of "transforming unstructured text into a structured format to identify meaningful patterns and new insights." This generally involves the scraping and combining of data from digital sources into a tabular dataset format, and thereby focuses on extracting and analyzing existing information.

► AI training is the application of a "model's mathematical framework to a sample dataset whose data points serve as the basis for the model's future real-world predictions.

TDM EXCEPTIONS APPLICABLE OR NOT?

if we deem that TDM exception:

➢ applies to the use done my Gen AI systems of data protected by IP rights, we still need to know exactly how the reservation should be made;

➢ does not apply the AI training with data protected by IP rights has to be qualified as unlawful use of copyrighted works.

OUTPUT AND … LITIGATION

The other side of the controversial relationship between generative artificial intelligence and copyright is to determine whether the output generated by a generative artificial intelligence system can be protected through intellectual property.

Patentable? DABUS case ('*Device for Autonomous Bootstrapping of Unified Sentience*')

Protected through copyright? Thaler tried again to obtain protection for the output of generative artificial intelligence systems in the field of copyright, claiming that his Creativity Machine would be the exclusive author of a two-dimensional work of art entitled '*A Recent Entrance to Paradise*'.

THE MONKEY'S SELFIE

Between 2011 and 2018, a series of disputes took place about the copyright status of selfies taken by a monkey using equipment belonging to a British wildlife photographer. The disputes involved Wikipedia and a blog which have hosted the images following their publication in newspapers in July 2011 over the photographer's objections that he held the copyright, and the association People for the Ethical Treatments of Animals, who have argued that the copyright should be assigned to the macaque.

The photographer argued that he haa a valid copyright claim because he engineered the situation that resulted in the picture. The refusal to remove the pictures from its library by Wikipedia and the blog was instead based on the understanding that copyright is held by the creator, that a non-human creator hold copyright, and that the images should thus be in public domain.

In December 2014, the United States Copyright Office stated that works created by non human, such as a photograph taken by a monkey, are not copyrightable.

MILAN COURT AND ECJ

The same conclusion had been reached, among others:

by the Court of Milan in the Bansky case

as well as

by the Court of Justice of the European Union (CJEU) declaring on various occasions, particularly in its landmark *Infopaq* decision (C-5/08 *Infopaq International A/S* v *Danske Dagbaldes Forening*), that copyright only applies to original works, and that originality must reflect the "author's own intellectual creation".

PROTECTION FOR AI GENERATED WORKS? UK…

In the same orbit is the question of the protection of AI generated works

►to exclude a priori that a work generated by algorithms can be the subject of privative rights entails a significant legal loophole, with serious social and economic consequences.

Some legal systems do not seem to be entirely resistant giving authorship to the programmer.

- In UK copyright law, section 9(3) of the Copyright, Designs and Patents Act (CDPA), states: "*In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.*" Furthermore, section 178 of the CDPA defines a computer-generated work as one that "*is generated by computer in circumstances such that there is no human author of the work*". .

- Similar provisions are contained in the copyright laws of Hong Kong (SAR), India, Ireland and New Zealand.

PROTECTION FOR AI GENERATED WORKS? EU AND ITALIAN SUPREME COURT (ORDER OF 16 JANUARY 2023, NO. 1107)

European Parliament Resolution of 20 October 2020 'On intellectual property rights for the development of artificial intelligence technologies' reiterated that, in cases where artificial intelligence is only used as a tool to assist an author in the creative process, the existing copyright framework remains applicable and also that works generated by artificial intelligence systems must find forms of legal protection under intellectual property rights in order to encourage investments in the context of artificial intelligence itself.

This approach seems to have been shared by our Court of Cassation in the context of the dispute between architect Carla Biancheri and RAI. The architect had created through a software system the work *'The Scent of the Night'*, a digital fractal flower used - without any recognition of rights - for the set design of the 2016 Sanremo Festival.

SUPREME COURT OBITER DICTUM IN ORDER OF 16 JANUARY 2023, NO. 1107

the use of an algorithm in the creative process of an image cannot be considered sufficient to deny the creative nature of an original work, requiring only a rigorous assessment of the level of creativity, aimed at verifying whether and to what extent the use of the algorithm has absorbed the creative elaboration of the artist who has used it.

AI ACT COMPROMISE AGREEMENT AND COPYRIGHT

The compromise agreement states that

- providers of general purpose AI models need to put in place a policy to respect Union copyright law as well as

- make publicly available a sufficient detailed summary about the content used for training of the general purpose AI model, based on a template provided by the AI Office.

▶ While taking into due account the need to protect trade secrets and confidential business information, this summary should be generally comprehensive in its scope instead of technically detailed to facilitate parties with legitimate interests, including copyright holders, to exercise and enforce their rights under Union law, for example by listing the main data collections or sets that went into training the model, such as large private or public databases or data archives, and by providing a narrative explanation about other data sources used."

MACHINE UNLEARNING?

Our society has moved in recent years from a millenary culture of recording information to a new culture of erasing it. And in fact, while until recently it was difficult to collect information and ensure that it survived, today most information is born in digital format and is therefore already recorded, which also means that it accumulates by default.

There are two ways in which digital information can be removed:

we can delete it to make it unavailable or

we can block it to make it unaccessible.

The distinction is fundamental from a conceptual, technical as well as cost perspective: blocking information is certainly cheaper and simpler and can be reversed more easily, just think to

- the Court of Justice's historic 2014 decision on the right to be forgotten: in that context, relevant information was not deleted but made inaccessible;

- the Cambridge Analitica scandal of 2018 as a result of which the UK Information Commissioner's Office requested Cambridge Analitica to delete the data it held.

MACHINE UNLEARNING?

Once the model has been trained, if there is a privacy or intellectual property issue, one has to ask whether there is another solution than:

- Delete the template completely

- Removing unwanted data from the training set and then re-training the model (in machine unlearning culture, this is referred to as exact unlearning)

- Blocking information completely.

MACHINE UNLEARNING?

One possibility seems to be the machine unlearning, a kind of selective and potentially incremental partial amnesia that takes the form of a reversal of the process that led a model to learn to provide some specific information.

Machine unlearning is an issue in its infancy and not without its difficulties.

Some argue that it is also an area that is only well defined at an algorithmic level, and certainly, purely by way of example, no reference to this technique is to be found in the context of the decision of the Garante per la protezione dati personali (Italian Data Protection Authority), which, in March 2023, ordered with immediate effect the provisional restriction of the processing of Italian users' data against OpenAI, noting the lack of information to users and all interested parties whose data are collected, but also the absence of a legal basis justifying the massive collection and storage of personal data.

PRINCIPLES NOT RULES

The very existence of a technique such as Machine Unlearning, which is selective and, therefore, does not lead to Hal being switched off, but rather to him forgetting certain specific data, leads, in my view, to believe that not rules, but principles are needed in order to regulate the relationship between artificial intelligence and intellectual property and, therefore, what is needed is not the law of horses, but, at most, that of animals.

Thanks for your attention

francesca.ferrari@uninsubria.it

francesca.ferrari@esteri.it