Legal 500
Country Comparative Guides

Hot Topic | Intellectual Property

AI and Deepfakes: EU and Italian Regulations

Contributor

Jacobacci Avvocati
and Jacobacci & Partners

For a full list of jurisdictional Q&As & hot topic articles visit legal500.com/guides/

AI and Deepfakes EU and Italian Regulations

The Oxford English dictionary defines the term “deepfake" as “*any of various media, esp. a video, that has been digitally manipulated to replace one person's likeness convincingly with that of another, often used maliciously to show someone doing something that he or she did not do*".

The manipulation of images, video, or audio to generate “seemingly" real content is not new, and certainly precedes the widespread application artificial intelligence technologies. Indeed, techniques for manipulation of images were developed in 1800s along with the birth of photography, while the advent of digital images allowed rapid and realistic editing by experienced users through the use of specialized software.

However, the recent, pervasive and incredibly rapid development of advanced artificial intelligence technologies, such as neural networks, has allowed incomparable accessibility and speed in generating deepfake content. These techniques are innumerable and often combined to achieve incredibly realistic audio and visual results.

The earliest neural network-based techniques used for the development of synthetic images were autoencoders representative. Such architectures involved the use of decoding neural networks that learn to reconstruct an image from compressed and simplified versions generated by an encoding network. By providing the decoder with new compressed images, the network would attempt to reconstruct the image on which it was trained, while retaining the features in the new image, allowing, for example, the features of a face to be applied to another person's body.

More recently, architectures based on generative adversarial networks (GANs) have been applied. GAN-type architectures use two neural networks trained in parallel to compete in a zero-sum game via an optimization loop in which the parameters of each neural network are modified based on the output of the other. A first network, called a “generator," is trained to generate synthetic samples, while a second, called a “discriminator," is trained to distinguish between synthetic and real samples. The generator resulting from this process can then be used to generate content that is really difficult to distinguish from the real thing, except by the discriminator with which it has been trained.

Developments in convolutional neural networks (CNNs) have also made it possible to more accurately identify the location of key features of a face, such as eyes, nose and mouth, helping to produce natural expressions and movements.

As a matter of fact, the advent of AI has made it not only deepfakes incredibly realistic, but it has also made it also easy for anyone to produce and disseminate deepfake content so that fake media environments are no longer limited to experts. Research reveals a staggering 550% increase in AI-manipulated photos between 2019 and 2023, demonstrating the alarming rate of accessibility and potential for misuse.

While originally confined mainly to pornographic content, in the last few months, many cases of deepfakes have been identified on the Internet in completely different fields. Just few examples: while Taylor Swift had to deal with a collection of sexually explicit AI-generated images of her that were published across several social media platforms, a fake image of Tom Hanks was used to advertise a dental plan. Meanwhile, another deepfake featured Mario Draghi promoting an absurd investment plan, convincing some 45 Italians that saw the video on Instagram to follow "his" advice.

The urgent call for laws regulating such harmful and invasive conduct has not yet seen results, even though the EU has taken an interesting approach to the problem, introducing the EU Artificial Intelligence Act (EU Regulation 2024/1689 dated 13 June 2024, the "EU AI Act"). The EU AI Act is the first of its kind worldwide and aims is to foster trustworthy AI in Europe and beyond, by addressing the risks of powerful and impactful AI models and ensuring that AI systems respect fundamental rights, safety, and ethical principles.

This course of action is perfectly in line with the pronunciation of the 2009 decision of the European Court of Human Rights holding that the right to the protection of a person's image is *'one of the essential components of personal development and presupposes the right to control the use of that image*'.

With regards to deepfakes, the EU AI Act does not ban them entirely, as one might have expected, and instead decided to require that AI creators act generally in a manner that is transparent. According to this framework, anyone who creates or uses a deepfake must disclose its artificial origin and provide information about the techniques that are used.

In this respect, the EU AI Act addresses the "*deepfakes problem*" through three key elements: (1) the definition of "deepfake"; (2) transparency obligations for AI providers and those who use AI, also known as deepfake creators; and (3) the objectives set forth in Recitals 132 to 137. More in detail:

1. "deep fake" is defined in Article 3(60) as "*AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful*".
2. The EU AI Act categorises AI systems based on potential risk (unacceptable, high, limited and minimal risk), imposing stricter regulations for higher risks. Deep fakes are commonly sorted into the limited risk category.

   "Limited risk" refers to the risks associated with a lack of transparency in AI usage. Article 50 of the AI Act introduces specific transparency obligations to ensure that humans are informed, when necessary, the goal of which is to foster trust. For instance, when using AI systems such as chatbots, humans should be made aware that they are interacting with a machine so they can make an informed decision to continue or step back. Providers will also have to ensure that AI-generated content is identifiable, and that AI-generated text that is published with the purpose to inform the public on matters of public interest must be labelled as artificially generated. This also applies to audio and video content constituting deep fakes.

3. Recital (134) underscores the previous point (2) and the importance of the transparency obligations provided in Article 50, anticipating that deployers of AI systems should be required to clearly disclose that the content has been artificially created or manipulated by labelling the AI output accordingly and disclosing its artificial origin.

   While the "default" placement of deepfakes into the limited risk category is certainly debatable in consideration of the harm they are capable of causing, the good news is that the AI Act also provides the theoretical possibility that deepfakes could fall into the "high-risk category" due to the potential for manipulation of political or electoral content. This refers to the use of deepfakes to disrupt the democratic process, by spreading political misinformation or impersonating candidates, with the goal of swaying public opinion and deceiving voters (see Article 6(2), stating that AI systems referred to in Annex III shall be considered to be high-risk, whereby Annex III (8)(b) refers to those that are intended "*to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda […]*").

Moreover, due to the never-ending evolution of technology, Article 7 has thought ahead, allowing the Commission to expand the list of high-risk AI systems in two ways. Firstly, by adding new areas or applications if the AI system is designed for uses listed in points 1–8 of Annex III and thus presents a comparable or higher level of risk to health, safety, or fundamental rights compared to current high-risk systems. Secondly, under Article 7(2), the Commission maintained the right to modify the classification of existing high-risk systems based on several factors that can be summarised by the history of harm that has been caused, impact scope and existing regulations.

As a consequence, at the EU level, while legislation already provides a possible development in consideration of the practical uses of AI and how it will impact European citizens, the only requirement at this stage is transparency. Frankly speaking, this seems insufficient to address the malicious potential of deepfakes, especially when the creator fails to inform the public, as (now) required by law.

Indeed, at the European level, it would have been helpful to mandate that software developers incorporate measures to prevent the generation of harmful deepfakes through their products and the use thereof.

With that said, Italian law provides a higher level of protection, even in the absence of ad hoc, "new" legislation. In fact, while some proposals for new AI regulation within Italy are already on the table, the "old" legal framework is perfectly capable of addressing the risks presented by deepfakes.

Under Italian law, deepfakes are viewed as reproductions of convincing image, audio and video content that constitute both a violation of the personality rights of the individual whose subjective characteristics and traits are replicated, as well as a violation of that individual's rights in their own image.

Indeed, according to Italian case law, the right to one's own image is among the so-called inviolable personality rights enshrined in Article 2 of the Italian Constitution, which include those rights that are concerned with essential aspects of a human personality.

In addition to the constitutional protection for personality rights, the right to one's image is regulated by the Italian Civil Code, Article 10, which defines the abuse of image rights as the unauthorized use of another's image and provides for compensation for damages from any third party that exposes or publishes the image of a person "*outside the cases in which the exposure or publication is permitted by the law*" or "*with prejudice to the decorum or reputation of the person themself*" (this provision can also be invoked by the heirs of the person depicted). In other words: an image may not be displayed or put on the market when that would be detrimental to the honor, reputation and decorum of the person portrayed.

This provision (Italian Civil Code, Article 10) is generally understood to be read in conjunction with Articles 96 and 97 of the Italian Copyright Law (Law No. 633/41), which contain specific regulations on portraiture that can be claimed in the event a person's image is exploited in an unauthorized manner.

In particular, Article 96 of the Italian Copyright Law introduces the so-called “principle of consent,” prohibiting the exposition, reproduction or use for profit of a reproduction of an individual's image without his/her consent. It is important to emphasize that by virtue of this principle, the right to one's own image is personal and inalienable. While the right to use the image may be sold or licensed, the image right itself is never transferred.

Indeed, even within the principle of consent, it is possible to define limits to the consent to the use of one's own image. Italian case law requires, however, that these limits are strictly confined to the circumstances of time, place and purpose for which the consent was given (objective limits), and with exclusive regard to the subject and/or subjects in whose favor the consent was given (subjective limits).

Article 97 of the Italian Copyright Law provides specific exceptions to the principle of consent set forth in Article 96, stating that the publication of an individual's image is in any case allowed:

- when justified by the notoriety or public stature of the person portrayed;
- when the reproduction of the image is related to facts, events, or ceremonies of public interest, or which were held in public;
- by necessity of justice or the police, or for scientific, educational or cultural purposes.

Italian case law has, however, regulated the exceptions set forth in Article 97. For instance, regarding the reproduction of images of a famous person, the majority position of the case law is that while notoriety is necessary for the exception to apply, it is not sufficient alone to justify the absence of consent: it is also necessary to demonstrate that the disclosure of the image meets the needs of public information.

At the same time, case law has also clarified that the reproduction of the image of a common person that is involved in acts of public interest, is permissible without his/her consent only if the picture is necessary to offer a better understanding of the accompanying article on the newspaper.

Indeed, in Italy, the use of the image of an individual without consent or outside the exceptions provided by law, or in any case to the detriment of that person's honor and reputation, constitutes a violation of that person's rights, giving rise to a claim for both economic and moral damages.

Returning to the issue of deepfakes, a deepfake creator in Italy would not benefit from any exemption provided by law, as they would have used a person's image without consent, to create content that is aimed at distorting reality, creating misinformation and/or causing damage to that person. Therefore, if a creator publishes deepfake content, the legal framework in Italy would allow the person depicted to seek an immediate injunction against the infringing use and, at the same time, to claim damages based on the consequences deriving from the exploitation of his/her rights to their personality and image.

For the sake of completeness, one must also consider that the unauthorized publication of deepfake images can also result in a violation of the right to protection of personal data according to art. 7 of the GDPR, and the implementing provisions within Italy, which also requires express consent to the use of a person's image.

To summarize, while the EU AI Act is breaking ground to provide an European-wide framework mandating transparency in the use of AI systems, including those generating deepfake content, Italy's national legislation already provides legal protections that can be applied to protect individuals from the risks associated with deepfake content. It is undeniable, of course, that more targeted and specific provisions – potentially including international cooperation on the matter, effective enforcement mechanisms at a transactional level, measures imposing legal consequences for individuals who create or disseminate deepfakes, especially with malicious intent, criminal penalties for those who create or facilitate the distribution of deepfakes, and so on – are needed to effectively address the legal and social implication of this technology, which can certainly be marvelous, but only if correctly used.

Contributors