Transparency Obligations for All AI Systems:
Article 50 of the AI Act

Written by
Dr. Joan Barata Mir
November 2025

Contents

Executive Summary

The AI Act sets out in its Article 50, transparency obligations for providers and deployers of certain AI systems, including generative and interactive AI systems and deep fakes. The principle of transparency encompasses a series of obligations, imposed on different actors, in relation to the proper and timely disclosure of information about products and services directly or indirectly provided with the assistance of AI systems, as well as whether a user is interacting with a living being, or with an AI system imitating human or animal characteristics or presenting artificially created objects.

The European Commission's regulatory body for AI governance, the AI Office, has launched the drafting process of the Code of Practice on transparent generative AI systems. This Code of Practice is supposed to define clear compliance criteria for providers and deployers of generative AI systems to demonstrate compliance with the transparency obligations of the AI Act, which will become effective, in principle, on 2 August 2026. The drafting process started in November 2025 and is expected to take about 10 months, that is until 2 months before the legal provisions become enforceable. The AI Office aims at a multistakeholder process and has therefore invited providers of generative AI systems, providers of transparency techniques, associations of deployers of in-scope AI systems, civil society organisations, academic experts, and other relevant organisations, to express their interest to participate in the drawing-up.

This paper aims at supporting these efforts by providing a foundational understanding of Article 50, identifying potential overlaps with other existing legislations, and highlighting critical questions that should be considered as part of the code drafting. This paper has been elaborated with particular consideration for the overall simplification efforts and objectives set by the European Union, particularly when it comes to the regulation of emerging systems and technologies.

Article 50 obliges AI systems intended to interact directly with natural persons to be designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system. It also established that providers of AI systems shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.

Deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system. Finally, deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. Article 50 establishes that the mentioned information shall be provided to the natural persons concerned "in a clear and distinguishable manner at the latest at the time of the first interaction or exposure".

This paper underscores the need to use a flexible approach to such requirements, based on the context and technological options and changes. Furthermore, it also highlights the need for these obligations to be implemented respecting the principles of necessity and proportionality to avoid introducing unnecessary and disproportionate restrictions to a fundamental right such as freedom of expression.

1. Introduction

A very relevant principle related to the development and deployment of AI systems is that of transparency. This principle can be found in several international legal texts and standards dedicated to aligning the development and use of AI with the requirements of democracy, the rule of law, human rights, and certain fundamental ethical values.[1]

From an international standards perspective, the principle of transparency broadly encompasses a series of obligations, imposed on different actors, in relation to the proper and timely disclosure of information about products and services directly or indirectly provided with the assistance of AI systems, as well as whether a user is interacting with a living being, or with an AI system imitating human or animal characteristics or presenting artificially created objects.

While taking such standards as a general reference, this paper will consider a very specific aspect and regulation of the mentioned principle. The Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act)[2] contains a series of provisions directly or indirectly related to the notion of transparency. Specifically, transparency is defined in recital 27 which states that "transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights".

---

[1] See the UNESCO Recommendation on the Ethic of Artificial Intelligence (2022), available at https://unesdoc.unesco.org/ark:/48223/pf0000381137, the OCDE Recommendation of the Council on Artificial Intelligence (2023), available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, and the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024), available at https://rm.coe.int/1680afae3c. For further details on the diversity of international standards in the field of transparency see Gils, Thomas, A Detailed Analysis of Article 50 of the EU's Artificial Intelligence Act (June 14, 2024). Appeared in C. N.Pehlivan, N.Forgó and P.Valcke (eds.), *The EU Artificial Intelligence (AI) Act: A Commentary* (Kluwer Law International, 2025), 776-823, http://dx.doi.org/10.2139/ssrn.4865427

[2] https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

This paper will therefore focus on the general transparency regime applicable to any AI system in the cases and uses contemplated in Article 50 of the AI Act. It is particularly aimed at providing useful considerations to be considered during the upcoming drafting process of the Code of Practice on transparent generative AI systems, recently launched by the AI Office. This Code of Practice is supposed to define clear compliance criteria for providers and deployers of generative AI systems to demonstrate compliance with the transparency obligations of the AI Act. The AI Office aims for a multistakeholder process and has therefore invited providers of generative AI systems, providers of transparency techniques, associations of deployers of in-scope AI systems, civil society organisations, academic experts, and other relevant organisations, to express their interest to participate in the drawing-up[3].

This paper aims at supporting these efforts by providing a foundational understanding of Article 50, identifying potential overlaps with other existing legislations, and highlighting critical questions that should be considered as part of the code drafting.

It is also important to note that, as it will be shown in this paper, Article 50 needs to be interpreted with particular consideration of the indications contained in recitals 132 to 137, which include some clarifications regarding the scope and meaning of its certain provisions[4].

Article 50 will only be in force in August 2026. The drafting process of the already mentioned Code of Practice began in November 2025 and is expected to take about 10 months. It is therefore important to note that between the expected finalisation of

---

[3] https://digital-strategy.ec.europa.eu/en/news/participate-drawing-code-practice-transparent-generative-ai-systems

[4] These recitals provide a wide range of specific criteria for interpretation, including regarding the risks of reception and impersonation deriving from AI systems intended to interact with natural persons or to generate content (recital 132), the generation of synthetic content in relation with risks of misinformation and manipulation at scale, fraud, impersonation and consumer deception (recital 133), compliance with transparency obligations in relation with so-called deep fakes (recital 134), the drafting of codes of practice (recital 135), the relation of Article 50 with the provisions included in the so-called Digital Services Act (article 136), and the existence of other provisions establishing transparency obligations in other pieces of legislation (recital 137).

the drafting and the actual enforceability of the legal provisions there may only be a few weeks, which obviously represents a tight timeframe to adopt all the necessary measures, standards and practices to ensure compliance. It is important to note however that just a few weeks after the drafting process was started, the European Commission put forward a legal reform proposal affecting and amending the AI Act with the purpose of "the simplification of the implementation of harmonised rules on artificial intelligence", also known as the Digital Omnibus on AI[5].

Before starting with the detailed analysis of the provisions included in Article 50, it must be underscored that this is a legal provision that essentially articulates and frames the provision of information in order to protect certain manifestations of the public interest. It is however essential to bear in mind that the provision of information to natural persons on the basis of the requirements included in the mentioned article is not a complete panacea to address all the possible issues around the risks and challenges for humans in their interactions with AI systems and consumption of AI-generated content. Therefore legal and policy discussions around interpretation and enforcement of Article 50 of the AI Act must particularly take the following elements into consideration:

a) This is an area substantially conditioned by broad underlying factors including digital literacy, access to technology, and overall progress. Article 50 cannot thus be treated as or transformed into an instrument to address societal problems of trust, unhealthy communication environments or technological culture.

b) Any efforts linked to the interpretation and implementation of the provisions included in Article 50 must be translated into standards and rules with a sufficient degree of flexibility and adaptability to the evolving AI landscape. Excessively granular requirements might be counterproductive in terms of fostering the adoption and building public trust in AI.

c) Several of the legal measures considered in this paper may have an impact on relevant human rights. Therefore, the principles of necessity and proportionality must guide the adoption of any rules or standards. Obligations must be strictly needed for the protection of the interests as established by Article 50, must represent the lesser or technically less cumbersome degree of

[5] https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal

intervention, and ought to be appropriate to effectively and efficiently achieve the legal objectives.

2. Scope of Article 50

The obligations contemplated under Article 50 apply to several categories of AI systems, independently of their classification in terms of risks according to the provisions of the AI Act. The scope of Article 50 thus does not depend on whether an AI system qualifies as high-risk but on whether the specific use fulfils the criteria and circumstances established in this article. One of the main aims of this paper is to facilitate precisely the criteria for the identification of such elements, which in several cases are presented in a significantly ambiguous manner.

This being said, it is also important to bear in mind that due to the characteristics of the uses contemplated in this article, some of the provisions would apply to the developers of AI systems (i.e. those who develop an AI system or a general-purpose AI model or that have an AI system or a general-purpose AI model developed and place it on the market or put the AI system into service under their own name or trademark, whether for payment or free of charge) while others, as it will be shown, are to be understood as covering the deployment of such systems (i.e., any "natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity").

As it will also be shown, in some cases the obligations contemplated under Article 50 may overlap with or complement other similar transparency obligations applicable to online platforms, as contemplated by the Regulation (EU) 2022/2065 of the European Parliament, and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act, DSA)[6].

Last but not least, it is also necessary, for proper and appropriate interpretation of the obligations established in the mentioned article, to keep in mind the specific determinations in terms of scope contemplated by article 2, and the definitions and categorisations included in article 3 of the AI Act. In this sense and without prejudice to the fact that the Code of Practice must particularly focus on clarifying and providing criteria for the interpretation and enforcement of the different provisions included in Article 50, it shall also provide adequate criteria to avoid any ambiguity when it comes to the application of such provisions in relation to the

[6] https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

general scope requirements established under article 2, as well as in light of the definitions provided by article 3.

3. Specific Transparency Obligations

3.1. Obligations Vis-à-Vis Interactive AI Systems (Article 50.1)

Main Provisions

According to paragraph 1 of Article 50, natural persons who directly interact with AI systems must be informed of such interaction. This obligation affects the design and development of the systems, even though this may become particularly relevant vis-à-vis the effective deployment of the systems.

This obligation has two exceptions:

a) Cases where such interaction is "obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use".

b) Cases of interactions with AI systems "authorised by law to detect, prevent, investigate or prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties". An exception to such exemption would apply to systems available for the public to report a criminal offence.

According to recital 132, when implementing the disclosure obligation, "the characteristics of natural persons belonging to vulnerable groups due to their age or disability should be taken into account to the extent the AI system is intended to interact with those groups as well". It also "should be provided in accessible formats for persons with disabilities".

Main Obligations

The obligations included in these provisions must be read in light of the core legal goal of tackling "specific risks of impersonation or deception irrespective of whether they qualify as high-risk or not" (recital 132).

They can be described as follows:

a) The information obligation applies to interactions of natural persons with AI systems only.

b) Being informed about the interaction with AI systems can be conceptualised and legally defined as a right of users.

c) An additional element to be particularly considered is the fact that the law circumscribes the scope of the obligation to AI systems "intended to interact directly with natural persons". This element of purpose, as well as the requirement of direct interaction, must therefore be particularly taken into consideration when it comes to the interpretation and enforcement of these provisions.

Areas for further clarification

There are a few areas where some open questions or ambiguities can be detected. There also are some key recommendations to highlight when it comes to promote a reasonable and appropriate interpretation and implementation:

a) **Intention and risk of deception.** As mentioned, the law does not define when AI systems are "intended to interact directly with natural persons", even though this clearly excludes other types of interactions, particularly those between AI systems with no human intermediation or intervention. It is important to highlight the connection between this type of interaction with the main objective of this provision, as stated in recital 132, which is to tackle "risks of impersonation or deception". Therefore, it is very important for the future Code of Practice to facilitate the criteria for an interpretation, firstly, that limits the scope to systems where the user provides a concrete input and receives a subsequent response/output triggered by the former, and secondly, identifying the specific conditions and circumstances where such interaction creates the mentioned risks. Consequently, non-conversational interactions or input/output sequences where there is no synchronicity or includes other factors such as additional human interactions must not be considered to be under the scope of the law. Examples of the former include chatbots, conversational virtual assistants, or online AI interactive games, while email autoresponders, content suggestions or classifiers, spam filters, fraud

detection systems, or cybersecurity threat detection would not fulfil the mentioned legal requirements.

b) **No predetermined transparency solutions.** The specific fulfillment of this legal obligation shall affect the design and development of AI systems. There is no single, pre-determined and universal technical solution in this sense, in as much as information is clearly conveyed to users immediately prior to their first interaction. Therefore, different transparency solutions are possible, including labelling or proper disclosures as part of the terms of service. Furthermore, paragraph 5 of Article 50 particularly refers to the need to provide information in a "clear and distinguishable manner" (as it will be discussed later). Connected with this, recital 132 requires that when implementing that obligation, "the characteristics of natural persons belonging to vulnerable groups due to their age or disability should be taken into account to the extent the AI system is intended to interact with those groups as well". In any case, further clarification by the Code of Practice may be needed in order to properly, and at the same time, flexibly define clarity and distinguishability, as well as possible cases where the need to avoid possible deceptions might require providing information about specific characteristics of the disclosed interaction.

c) **Exception for cases where interaction is "obvious".** The exception applicable to cases where the interaction with an AI system becomes "obvious" (sic) is broad and requires further analysis and elaboration. Similarly to the requirement analysed in the previous paragraph, this is a dynamic standard, very much connected to the evolution of technology and AI digital literacy. The provision particularly establishes as a standard to determine obviousness "the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use". Therefore, any further elaboration or development of this standard needs to meet the requirements of necessity, adaptability and flexibility. It must be particularly reiterated that any obligation in this area must be consistent and necessary in relation to the core goal of avoiding users' deception, and thus is not applicable to cases where such risk would be clearly absent. More detailed standards must consider specific factors such as the nature and characteristics of the interface, nature and characteristics of the provided output, the type of user and context of use, as well as the output's

attributes (for example, a very speedy and comprehensive output clearly points at the use of AI).

d) **Investigating and prosecuting offences.** Regarding law enforcement purposes, there are still some open issues for further clarification, particularly when it comes to the need for, and characteristics of, a specific and separate criminal law provision establishing in detail the circumstances, competent bodies, and safeguards applicable to such types of use.

3.2. Generation of Synthetic Content (Article 50.2)

Main Provisions

Paragraph 2 of Article 50 establishes that:

a) AI systems that generate synthetic audio, image, video, or text content "shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated".

b) Providers must guarantee that "technical marking solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of the various types of content, the costs of implementation and the generally acknowledged state of the art".

The AI Act also contemplates two exemptions:

a) AI systems that perform an assistive function for standard editing or do not substantially alter the input data or its meaning.

b) Systems authorised by law to detect, prevent, investigate, or prosecute criminal offences.

Main Obligations

It shall be noted that these provisions apply to a wide variety of content and different technical solutions to generate artificial content, particularly within the area of what is commonly known as generative AI. This would include well known systems in

fields like voice generating (Apple's Siri), image generating (Dall-E), or text generating (ChatGPT). Recital 133 refers to the fact that AI systems can generate large quantities of synthetic content that becomes "increasingly hard for humans to distinguish from human-generated and authentic content". For this reason, the main obligation included in this paragraph revolves around the notion of "marking".

In order to properly determine the scope of these obligations, it is necessary to consider the indications contained in some of the recitals of the Act, including the following:

a) This obligation applies to providers who are required to embed technical solutions that not only enable the marking in a machine-readable format but also the detection that the output has been generated or manipulated by an AI system and not a human. Recital 135 also refers to obligations "regarding the detection and labelling of artificially generated or manipulated content."

b) According to recital 133, technical measures would include "watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints or other techniques, as may be appropriate". It also specifies that these measures "can be implemented at the level of the AI system or at the level of the AI model, including general-purpose AI models generating content, thereby facilitating fulfilment of this obligation by the downstream provider of the AI system."

c) Recital 133 also emphasises the need to consider "the specificities and the limitations of the different types of content and the relevant technological and market developments in the field, as reflected in the generally acknowledged state of the art."

d) Recital 136 clarifies that obligations particularly apply "as regards the obligations of providers of very large online platforms (VLOPs) or very large online search engines (VLOSEs) to identify and mitigate systemic risks that may arise from the dissemination of content that has been artificially generated or manipulated, in particular the risk of the actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, including through disinformation." This refers to articles 34 and 35 of the DSA, as it will be further developed.

a) It is important to stress, in any case, that the legal obligation in question refers only to the need to provide tools for the proper differentiation between synthetic (i.e., AI generated or manipulated) and "authentic" content, the latter being understood as content that corresponds to "real" audio, image, video or text content. This consideration or differentiation must, in any case, be separated from other assessments regarding trustworthiness or veracity of content. Even though technical measures mentioned above can also help when it comes to the second type of assessments, this is not the objective of the commented Article 50. It shall not be forgotten that the mentioned measures may complement other different mechanisms generally dedicated to promoting trustworthy information, such as community notes, trusted flagging, fact-checking, digital literacy initiatives, and other verification or monitoring systems, as well as the overall improvement of the information ecosystem.

Areas for Further Clarification

Based on the above, it is also important to outline the following:

a) **Non-human generated content as the main target.** Obligations considered in this section cover the use of AI systems to both generate completely new synthetic content (for example, from a prompt) or create content based on the manipulation or alteration of existing images, sounds, etc. The main objective of these legal constraints is to achieve a proper understanding among users about the nature and origin of content they are accessing, particularly when the use of AI is not evident and thus misleading. Therefore, even though the law does not require in these cases the intention to mislead or misrepresent, specific standards will need to focus on the core objective of such obligations, which is to guarantee that users properly identify non-human-generated content.

b) **Technical principles and solutions.** Issues around standardisation, interoperability, adaptation to different types of content, proportionality, and human readability remain very much open and subject to evolving technical progress in these areas. In particular, it is very important to bear in mind, from a technical point of view, the inherent tension between a solution's robustness (being hard to remove) and its interoperability (being universally detectable). Therefore, any additional standard established in this area via

codes of practice or guidelines must acknowledge this trade-off, recognizing that fully open detection can compromise the robustness and security of a marking technique. Other relevant technical principles to bear in mind in this area include technical feasibility and efficiency (also to avoid disproportionate costs), as well as future proofing since technology is still quite nascent and constantly evolving.

c) **Content provenance solutions.** Technical collaborative solutions, such as those in the area of content provenance, to mark the nature and source of a piece of content as it is created or edited, may be helpful to fulfil these legal obligations, even though it is important to bear in mind that they do not aim at tackling broader issues of content veracity. A specific example in this area would be the Coalition for Content Provenance and Authenticity (C2PA)[7], whose provenance metadata standard ("content credentials") has a fast-growing community of contributors and implementers including, but not limited to, technology companies like Google, OpenAI, and Microsoft, media companies like the BBC and CBC, camera companies like Sony, and more. A combination of different metadata techniques that includes robust digital watermarking may provide a significantly balanced solution to the trade-offs between reliability, robustness and interoperability.

d) All the above shall be read without prejudice to other relevant and connected legal issues:

d.1) *Information about a creator's identity is personal data.* It is therefore necessary to establish standards that while guaranteeing content provenance also provide proper safeguards against unnecessary disclosure of personal data, particularly in contexts where there is an overriding interest (or even a fundamental right) to protect the identity of content creators and speakers.

d.2) *Content provenance standards must not be established as adjudicators or indicators of copyright ownership*. It must also be taken into account that in any case, copyrights and usage rights are dynamic and thus subject to changes.

[7] https://c2pa.org

d.3) *Provenance shall not be in any case articulated as a mechanism for rights management*. This area shall be tackled by different legal provisions and separated technical tools.

e) **Scope and interpretation of exceptions.** As mentioned already, obligations included in these provisions are exempted in cases of AI systems that perform an assistive function for standard editing, or introduce non-material alterations. Even though the law is once again very broad, any further standard must be based on the principle of necessary exclusion of cases where the output of synthetic content consists of non-substantive modifications that do not alter any essential element of the original content and do not generate manipulation or disinformation risks. This would include, amongst other elements, trivial or accessory content elements (captions, icons and similar additions). When it comes to the specific reference to "standard editing", it must be understood as referring to the most common editing processes for text, audio and images respectively, such as spelling correction, grammar checks, tone adjustment, or auto-complete functions, cropping, resizing, filtering, or adjustments to brightness, contrast, saturation, and colour-grading, as well as standard audio production techniques like background noise removal, reverb, compression, or volume adjustments.

In any case, it is very important that the Code of Practice provides very clear and detailed criteria for the proper enforcement of these provisions. Since AI is currently used to perform a significant number of assistive functions or to support content modification in ways that are completely unrelated to the public interest objectives guiding this article, any ambiguity or uncertainty may trigger the unnecessary labelling of an important amount of content. Apart from the consequences in terms of use of resources, the potential extensive labelling of most of the content we access online may simply deprive the commented legal provisions of any meaning and effectiveness as well as may deteriorate public trust in ways that will be described in other sections of this paper.

3.3. Emotion Recognition Systems (Article 50.3)

Main Provisions

Article 50.3 establishes the obligation for deployers of an emotion recognition system or a biometric categorisation system to inform the natural persons exposed to the operation of the system. It also indicates the obligation to process the personal data in accordance with relevant EU legislation in this field.

The AI Act defines an emotion recognition system as an AI system "for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data", whereas biometric categorisation systems consist of an AI system "for the purpose of assigning natural persons to specific categories on the basis of their biometric data, unless it is ancillary to another commercial service and strictly necessary for objective technical reasons" (article 3.39 and 3.40).

Main Obligations

The obligation to inform needs to meet the general requirements established in Article 50.5, and particularly the fact that it must take place "in a clear and distinguishable manner" with particular consideration of the characteristics of the AI systems involved and the implications of their use in terms of fundamental rights. On the other hand, the obligation regarding data protection shall be essentially seen as a reminder of the general legal applicable regime.

It is important to emphasise that this obligation should apply to AI systems that are used specifically for the purpose of emotion recognition or biometric categorisation (e.g., a system designed to analyse customer emotions in a retail space).

Areas for Further Clarification

Once again, a main challenge deriving from these provisions would be to define the exact scope of such obligation, considering the wording of the mentioned article.

This is an area where the legislator has also exempted systems "permitted by law to detect, prevent or investigate criminal offences". The provision also warns that such practice must be "subject to appropriate safeguards for the rights and freedoms of

third parties, and in accordance with Union law". Even though this is a matter beyond the scope of the present paper, it is very important to insist in this context on the need to articulate adequate safeguards to preserve the principle of presumption of innocence and defence rights, for example.

3.4. Deep Fakes (Article 50.4)

Main Provisions

Article 50.4 requires deployers of an AI system that generates or manipulates image, audio, or video content constituting a deep fake to disclose that the deep fake content has been artificially generated or manipulated.

The AI Act defines deep fakes 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" (article 3.60). It therefore cumulatively refers to the generation or manipulation of content via AI, the existence of a resemblance with existing elements of reality, and the fact that this content would falsely appear to be authentic or truthful. Recital 134 establishes that deployers must "clearly and distinguishably disclose that the content has been artificially created or manipulated by labelling the AI output accordingly and disclosing its artificial origin".

Article 50.4 also establishes that this obligation "shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offence".

Furthermore, in cases where the content "forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations set out in this paragraph are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work". Recital 134 connects this specific regime with the need to protect the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the EU Charter of Fundamental Rights.

This provision also applies to deployers of AI systems that generate or manipulate text "which is published with the purpose of informing the public on matters of

public interest shall disclose that the text has been artificially generated or manipulated", with two exceptions:

a) Where the use is authorised by law "to detect, prevent, investigate or prosecute criminal offences".

b) Where the AI-generated content "has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content".

Main Obligations

The obligations contained in the provisions mentioned above can be summarised as follows:

a) Deployers of AI systems that generate or manipulate image, audio, or video content constituting a deep fake must disclose that the deep fake content has been artificially generated or manipulated, except for cases where the use is authorised by law to detect, prevent, investigate or prosecute criminal offence.

b) In cases where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work.

c) Deployers of AI systems that generate or manipulate text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated, "except for cases where the use is authorised by law to detect, prevent, investigate or prosecute criminal offences or where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content".

c) The obligations mentioned above for deployers must be clearly distinguished and separated from those applicable to providers according to Article 50.2 and in the terms already described in this paper.

d) According to article 35.1 of the DSA, very large online platforms (VLOPs) and very large online search engines (VLOSEs) have the obligation to put in place reasonable, proportionate and effective measures that mitigate the systemic risks that may generate according to article 34 DSA. These measures may include, where applicable, prominent markings. This is thus an additional legal framework that indicates possible best practices for platforms that at the same time may also have the legal consideration of AI deployers. In addition to this, the European Commission issued guidelines precisely under the DSA for VLOPs and VLOSEs to mitigate risks to elections, including those posed by deep fakes. They also indicate possible good practices to assess and mitigate specific risks linked to AI, for example by clearly labelling content generated by AI (such as deep fakes), adapting their terms and conditions accordingly, and enforcing them adequately[8]. In any case, these considerations around new obligations for online platforms and search engines shall not be understood in the sense of altering the applicable intermediary liability exemptions legal regime where these actors simply host third-party content created using AI systems. In these cases, legal liability would still rest in the user generating or posting the content.

Areas for Further Clarification

The complex set of obligations, specifications and exceptions established under Article 50.4 creates some uncertainties as well as the need to properly investigate and determine their definition and scope:

a) **Definition: resemblance and false appearance.** The definition of deep fake, as already mentioned, covers AI-generated or manipulated image, audio or video content that fulfils two main conditions: resemblance to existing persons, objects, places, entities or events, and falsely appearing to a person to be authentic or truthful resemblance is not necessarily the same as identicality, which might mean that deep fakes could also consist of cases of presenting content that depicts persons, objects, places, entities or events that, as such, do not or did not actually exist although they would still appear to an average user as "realistic" (for example, the creation of a virtual person or a video showing an artificially created event). However, the element of false appearance is fundamental in terms of legal interpretation. In connection

[8] https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1707

with the already mentioned recital 133, this false appearance only becomes legally relevant in as much as it generates "risks of misinformation and manipulation at scale, fraud, impersonation and consumer deception". Therefore, both the AI Act regime and the demand to apply the principles of necessity and proportionality determine the exclusion from the scope of Article 50.4 (and 50.2) cases where content creation merely results in material that, as such, may not present deceptive risks, such as the depiction of generic objects or persons (an artificial landscape with kids running around), unrealistic content (someone walking on Saturn's rings), neutral alterations (including dramatic clouds in an advertising picture), minor manipulations or edits (as mentioned vis-à-vis Article 50.2), and generally content that per se does not raise the risks mentioned above (presenting the image of a newly published book on an artificially created colourful background).

b) **Expressive activities.** In connection with the last element above, it is also important to mention the specific rules applicable to cases where the content forms part of an "evidently artistic, creative, satirical, fictional or analogous work or programme". Once again, the legislator presents as "evident" something that may not be so in many possible scenarios. As recital 134 rightfully points out, this is also an area with relevant implications when it comes to freedom of expression, and particularly vis-à-vis the capacity of individuals, organisations, and political entities to engage in a significantly protected form of speech such as political speech. In other words, the law appears to be particularly cautious when it comes to preserving from unnecessary restrictions certain types of speech that relate to human dignity and creativity, and thus the development of individual personality. For all these reasons, the reference to all these categories must be interpreted in a broad sense in order to embrace any form of human artistic, fictional and creative work, even if commercially oriented (for example, in the case of advertisements).

c) **Potential impact on freedom of expression.** Based on the above, these specific obligations must be implemented while respecting the principles of necessity and proportionality. Otherwise, we might not only be contradicting the core objectives of these provisions but also introducing unnecessary and disproportionate restrictions to a fundamental right such as freedom of expression. Furthermore, it has already been explained that recital 134 establishes, when it comes to the content mentioned in the previous

paragraph, that the disclosure obligation must be fulfilled in an appropriate manner that "does not hamper the display or enjoyment of the work, including its normal exploitation and use, while maintaining the utility and quality of the work". Any further standard in this area must therefore be based on two main principles: flexibility (to adapt to the specific characteristics, format and intended impact of the creative work in question) and lesser degree of necessary obstruction (thus generally avoiding prominent, on-content marks).

d) **Human review and editorial control of publications.** As it has already been mentioned, AI systems that generate or manipulate text which is published with the purpose of informing the public on matters of public interest are also subject to transparency requirements where relevant exceptions apply. Article 50 refers to the existence of a process of human review or editorial control and the existence of an identifiable natural or legal person holding editorial responsibility for the publication of the content. In this sense, publication should be understood as any piece of content, or organised pieces of content, made available to the public after following an internal editorial process of prior consideration and approval (instead of private and informal user-generated comments or posts), matters of public interest as publications on non-private or individual issues, thus referring to a variety of topics besides advertising (culture, politics, health, fashion, economy) and targeting a general public; as well as human review or control as any human editorial intervention or approval prior to publication. In any case, the future Code of Practice must contain specific indications when it comes to the interpretation of such concepts.

e) **Counterproductive effects of labelling.** As a final reflection it is important to note that in some contexts a visible label (for example, "AI-generated content") might be wrongly interpreted as equal to a malicious intention to mislead or inversely, that non-labelled content may be inauthentic[9]. This may

[9] See for example the conclusions of some studies conducted in these areas in Chloe Wittenberg, Ziv Epstein, Gabrielle Péloquin-Skulski, Adam J Berinsky, David G Rand, Labeling AI-generated media online, *PNAS Nexus*, Volume 4, Issue 6, June 2025, pgaf170, https://doi.org/10.1093/pnasnexus/pgaf170

erode trust in certain sources or information or types of content[10] and alter the conditions for equal access to information and the free formation of the public opinion it could also particularly deter or harm the credibility and value of the use of AI-generated synthetic content for purposes of political activism, engagement, criticism or even campaigning. Additionally, excessive and unjustified marking may also create saturation and information fatigue, thus defeating the purpose of the disclosure. All this being said, this is probably an area where more analysis, discussion among stakeholders, as well as empirical evidence is needed. In any case, it is important to implement solutions that take into proper consideration these elements and impact credibility in order to properly fulfill and respect the objectives of the legislator when establishing these obligations.

[10] For further elaborations on such sensitive societal impacts see Baek, T. H., Kim, J., & Kim, J. H. (2024). Effect of disclosing AI-generated content on prosocial advertising evaluation. *International Journal of Advertising*, 1–22. https://doi.org/10.1080/02650487.2024.2401319, Sacha Altay, Fabrizio Gilardi, People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation, *PNAS Nexus*, Volume 3, Issue 10, October 2024, page 403, https://doi.org/10.1093/pnasnexus/pgae403, and Oliver Schilke, Martin Reimann, The transparency dilemma: How AI disclosure erodes trust, *Organizational Behavior and Human Decision Processes*, Volume 188, 2025, 104405, https://doi.org/10.1016/j.obhdp.2025.104405 .

4. Concluding Reflections

As a first final remark, it shall be highlighted that Article 50.5 establishes that, as per the disclosure requirements included in the four previous paragraphs, relevant information shall be provided to the natural persons concerned "in a clear and distinguishable manner at the latest at the time of the first interaction or exposure". The previous sections have already presented the specific interpretation challenges associated with the different disclosure obligations applicable to each of the uses. It is relevant in any case to underscore the need to use a flexible approach to such requirements, based on the context and technological options and changes. While there may not be one single standardised solution, what remains relevant is to protect comprehensibility (clarity) and distinguishability in terms of access and presentation.

Secondly, and from a broader perspective, it is important to note once again that any development in terms of further rules and standards adopted by the competent bodies in development of the provisions of the Act must not establish or prescribe narrow and specific technical solutions but determine general and flexible frameworks capable of guaranteeing the respect and promotion of the core values guiding the provisions included in Article 50. In addition to this, such core values must also be promoted using policy and normative instruments beyond the AI Act, including the support to AI literacy and other measures to counter and prevent misinformation and disinformation.

Lastly, and despite the limited and technical scope of Article 50, its interpretation and enforcement must in any case be guided by a series of very relevant principles, connected to basic fundamental rights: protection of human expressive activities, necessary and proportionate interventions, and adoption of flexible and adaptable solutions with proper consideration of their effectiveness and positive societal impact.

ANNEX: Article 50 AI Act

1) 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 point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate or prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, unless those systems are available for the public to report a criminal offence.

2) Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards. This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof, or where authorised by law to detect, prevent, investigate or prosecute criminal offences.

3) Deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system, and shall process the personal data in accordance with Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, as applicable. This obligation shall not apply to AI systems used for biometric categorisation and emotion recognition, which are permitted by law to detect, prevent or investigate criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, and in accordance with Union law.

4) Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offence. Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations set out in this paragraph are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work.

   Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offences or where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content.

5) The information referred to in paragraphs 1 to 4 shall be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure. The information shall conform to the applicable accessibility requirements.

6) Paragraphs 1 to 4 shall not affect the requirements and obligations set out in Chapter III, and shall be without prejudice to other transparency obligations laid down in Union or national law for deployers of AI systems.

7) The AI Office shall encourage and facilitate the drawing up of codes of practice at Union level to facilitate the effective implementation of the obligations regarding the detection and labelling of artificially generated or manipulated content. The Commission may adopt implementing acts to approve those codes of practice in accordance with the procedure laid down in Article 56 (6). If it deems the code is not adequate, the Commission may adopt an implementing act specifying common rules for the implementation of those obligations in accordance with the examination procedure laid down in Article 98(2).