Article 50 AI Act: Do the Transparency Provisions Improve Upon the Commission's Draft?

by **Nicolaj Feltes** *

Abstract: On April 21, 2021, the European Commission presented the first draft of the EU Artificial Intelligence Act, marking a significant step in Europe's regulatory approach to Artificial Intelligence (AI). The original proposal already included foundational transparency requirements, many of which are now formalised in Art. 50 of the Artificial Intelligence Act (hereinafter: AI Act). However, as AI technologies evolved rapidly – including the emergence of advanced tools like ChatGPT – the transparency obligations in Art. 50 AI Act were expanded to address new concerns around user awareness and content authenticity. Thus, notable additions such as labelling requirements for synthetic content and AI-generated texts were implemented in the final version of the AI Act.

In its finalised version, the AI Act specifies five distinct transparency obligations designed to enhance clarity and user protection across various AI applications. These obligations apply to interactive AI systems such as Chatbots (para. 1), AI systems for the creation of synthetic content (para. 2), systems for emotion recognition or biometric categorisation (para. 3), concerning AI-generated deep fake content (para. 4, subpara. 1), and AI-generated texts (para. 4, subpara. 2).

This article closely examines the transparency obligations, addressing potential issues of interpretation, practical challenges, and discusses whether the final version of the AI Act effectively addresses the problems present in the Commission's draft.

Keywords: AI Act, Transparency, Generative AI, Deep Fakes, DSA

Recommended citation: Nicolaj Feltes, Article 50 AI Act: Do the Transparency Provisions Improve Upon the Commission's Draft?, 16 (2025) JIPITEC 222 para 1.

A. Introduction

1 With its proposal for a Regulation on Artificial Intelligence in April 2021 (hereinafter: AI Act-COM)[1], the European Commission introduced a comprehensive set of transparency provisions aimed at regulating AI systems and addressing concerns related to user awareness, content authenticity, and potential misuse. These obligations included requirements to design and develop AI systems in a way that natural persons are informed when interacting with an AI system (Art. 52(1) AI Act-COM), to notify users exposed to emotion recognition systems or biometric categorisation systems (Art. 52(2) AI Act-COM), and to disclose deep fake content as artificially generated or manipulated (Art. 52(3) AI Act-COM). However, several shortcomings of these transparency provisions have been identified, particularly due to the use of undefined legal terms in the proposal[2] and concerns regarding the effectiveness of the

* The author is a Research Associate at the Digital Law Institute Trier (IRDT) and PhD candidate of JProf. Dr. Lea Katharina Kumkar.

1 Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts' COM(2021) 206 final.

2 See for example C. I., regarding the transparency obligation of Art. 50(1) AI Act (formally Art. 52(1) AI Act-COM), which uses several undefined terms such as "interaction" and "obvious".

obligations, especially those focused on AI system users.[3]

2 More than three years later, on July 12, 2024, the final version of the AI Act was published in the OJEU, incorporating several revisions of these provisions. In addition, as more potent AI tools – such as ChatGPT – have surfaced since the initial draft of the AI Act, new transparency obligations for synthetic content and AI-generated texts have been introduced. This raises the question of whether the revised provisions, together with the newly introduced obligations, effectively address the proposal's deficiencies. This article compares the transparency obligations in the Commission's draft with those in the final Act, critically evaluating whether the modifications effectively address these shortcomings. It also explores potential gaps in the regulatory framework, focusing on the scope of the obligations, their addressees, the associated requirements, legal consequences, and exceptions.

3 Moreover, the interplay between the AI Act and the Digital Services Act (DSA) is examined, highlighting potential synergies and conflicts, as undisclosed AI-content may have significant implications for users on online platforms. This is particularly pertinent in the case of unlabelled deep fakes and AI-generated news, which can facilitate the rapid spread of disinformation. Accordingly, it seems imperative to assess whether platform providers are obligated to remove content not labelled in compliance with the given provisions. Central to this discussion is whether such content qualifies as "illegal content" under the DSA and whether machine-readable markings prescribed by Art. 50(2) AI Act effectively support the implementation of risk mitigation measures under Art. 35(1) DSA. Finally, this article raises critical questions about the adequacy and enforceability of the AI Act's transparency obligations in mitigating the risks associated with rapidly evolving AI technologies.

B. Relevant Actors and Scope of the Obligations

4 Before examining the transparency obligations set out in Art. 50 AI Act, it is necessary to determine whether and to what extent the AI Act applies. This requires an analysis of its scope – including the material scope (which systems are covered by the obligations), the personal scope (who is subject to the obligations), and the territorial scope (in which geographical contexts the AI Act is applicable).

I. Material Scope

5 The AI Act primarily targets providers and deployers of AI systems. Art. 3(1) AI Act defines the term "AI system" comprising five main components. An AI system is "(1.) a machine-based system (2.) designed to operate with varying levels of autonomy and that (3.) may exhibit adaptiveness after deployment, and that, (4.) for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that (5.) can influence physical or virtual environments".[4] A key characteristic of such systems is the capability to infer.[5] According to Recital 12 AI Act, this capability refers to the process of obtaining outputs (e.g. predictions or content) "which can influence physical or virtual environments, and to a capability of AI systems to derive models or algorithms, or both, from inputs or data".[6]

6 The Commission's draft initially proposed a "technology-specific" concept, which faced significant criticism for diverging too much from the OECD's "technology-neutral" definition.[7] The definition of an AI system as provided in the final version of the AI Act aligns with this definition outlined by the OECD.[8] These adjustments made during the legislative process, however, were of little significance for the transparency obligations now outlined in Art. 50 AI Act. For instance, general adversarial networks (GANs), commonly used to create synthetic content and deep fakes, meet the criteria in both the Commission's draft and the final version of the AI Act.[9]

3 This concern has been mainly raised in the context of deep fake disclosure, as AI systems providers are more likely to implement disclosure solutions within the system itself, see C. IV. 3.

4 See Martina J. Block, 'A Critical Evaluation of Deepfake Regulation through the AI Act in the European Union' (2024) 4 EuCML 184, 185 f.

5 Recital 12, sentence 3 AI Act; for a more detailed analysis of the term "infer", see Alexander Steen, 'Ableitungen als wesentliche Fähigkeit von KI-Systemen nach der KI-VO' (2024) 1 KIR 2024, 7 ff.

6 Recital 12, sentence 4 AI Act.

7 Christiane Wendehorst in Martini/Wendehorst (ed.), KI-VO (2024), C.H. Beck, Art. 3 para 14 ff.

8 ibid para 17.

9 See Block, (n 4) 186; Additionally, as part of the "technology-specific" concept, the Commission's draft enumerates a list of covered technologies in Annex I. Notably, "deep learning" is highlighted in Annex I (a) AI Act-COM, which is primarily used in context of generative AI and deep fakes.

II. Personal Scope and Addressees

7 The personal scope of the AI Act includes both "providers" and "deployers" of AI systems, who are also the relevant actors subject to the respective obligations. The first two transparency obligations set out in Art. 50 AI Act pertain to the provider of the AI system, while the subsequent three fall under the responsibility of the deployer of the AI system.

1. Provider

8 Under Art. 3(3) AI Act, a *"provider"* is a natural or legal person, public authority, agency, or other body that develops an AI system or general-purpose AI model or has one developed. Additionally, the system or model has to be placed on the market[10] or put into service[11] under its own name or trademark, regardless of whether for payment or free of charge. In many cases, the term "provider" is synonymous with the software developer.[12] For instance, in the case of ChatGPT or Dall-E, OpenAI would be considered the "provider" of the AI system.[13]

2. Deployer

9 The three other transparency obligations pertain to the deployer of the AI system. Art. 3(4) AI Act defines the term *"deployer"* as a *"natural or legal person, public authority, agency or other body using an AI system under its authority"*. It is noteworthy that in the event that these subjects use the AI system "*in the course of a personal non-professional activity"* they are not considered to be a "deployer".[14] Thus, users who privately operate AI systems are not subject to the transparency obligations outlined in Art. 50(3) and (4) AI Act, as these obligations apply exclusively to deployers. Strong arguments favour a narrow interpretation of this exclusion, drawing parallels to the restrictive application of the household exemption in the General Data Protection Regulation (hereinafter: GDPR).[15] In this context, only "personal activities" are covered by the exception, suggesting that only natural persons can invoke this exception.[16]

10 Within the scope of Art. 50 AI Act, it is questionable what requirements should be placed on such "personal non-professional activities". In the context of the transparency obligations, this is particularly relevant with regard to the disclosure obligation for deep fakes outlined in Art. 50(4) subpara. 1 AI Act, as such content is typically disseminated via the internet, raising the question of whether such dissemination can still be considered a "personal non-professional activity". The exception under Art. 3(4) AI Act is contingent on whether the "use" of the AI system occurs within the context of a personal activity. Strictly speaking, the system itself is exclusively being operated during the creation of the deep fake – but not during the utilization of the output (such as the dissemination of the deep fake). This raises the question of whether, in cases where the creation of a deep fake occurs within the private sphere of an individual, the intention to subsequently disseminate the content constitutes a decisive criterion for assessing whether the activity falls outside the scope of a purely personal non-professional activity.

11 This would result in substantial evidentiary challenges, as the intention to disseminate content is typically difficult to prove. Accordingly, all essential steps – from the input of the input data to the utilization of the output of the system – must take place within the control of the user.[17] Contrarily, the mere use of an output, without prior operation of the generative AI system is not sufficient to qualify the disseminator as a deployer.[18]

12 Furthermore, the term "deployer" was originally referred to as "user" in the AI Act-COM.[19] The definition itself, however, has remained unchanged in the final version of the AI Act. The term "deployer" was introduced by the Parliament in response to

10 See Art. 3(9) AI Act.

11 See Art. 3(11) AI Act.

12 Mireille M. Caruana and Roxanne Meilak Borg in Sammut, Mifsud (ed.), The EU Internal Market in the Next Decade – Quo Vadis?, (Brill 2025) 108, 124. <https://library.oapen.org/bitstream/handle/20.500.12657/98980/9789004712119_webready_content_text.pdf?sequence=1#page=119> accessed 24 April 2025.

13 OpenAI has also acknowledged this, see <https://openai.com/global-affairs/a-primer-on-the-eu-ai-act/?utm_source=chatgpt.com> accessed 24 April 2025.

14 See Art. 3(4) AI Act; Art. 2(10) AI Act reiterates this in a contradictory manner: it excludes *"deployers"* utilising AI systems in the course of such a private activity from the scope of the AI Act. However, in accordance with Art. 3(4) AI Act, natural persons using such systems for private purposes cannot conceptually be classified as deployers in the first place, see Wendehorst in Martini/Wendehorst (ed.), (n 7) Art. 2 para 93.

15 Viktoria Kraetzig, 'Deliktsschutz gegen KI-Abbilder – Teil 1: Täuschende Deepfakes' (2024) 3 CR 207, 210.

16 LeaKatharinaKumkar/MoritzGriesel,'Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO' (2024) 4 KIR 117, 121; this aligns with the originally intended clarification in Art. 2(10) AI Act, which explicitly states that only natural persons can invoke the exception.

17 Wendehorst in Martini/Wendehorst (ed.), (n 7) Art. 3 para 83.

18 ibid.

19 See Art. 3(4) AI Act-COM.

repeated criticism that the term "user" could be misleadingly interpreted as referring only to the "end user."[20]

III. Territorial Scope

13 Art. 2(1) AI Act establishes the territorial scope of the AI Act. Art. 2(1) (a) AI Act mandates a "market place-principle"[21] for providers of AI systems: The AI Act applies to providers placing their AI systems (or general-purpose AI models) on the (EU-) market or putting such systems into service in the Union. This principle applies irrespective of whether the provider is established or located within the Union or in a third country.

14 Conversely, Art. 2(1) (b) AI Act prescribes a "principle of establishment" for deployers, signifying that the AI Act is applicable only to deployers who are either established or located within the Union.[22] For other provisions, such as the transparency requirement for synthetic and deep fake content, it must be taken into account that the recipient – the viewer of the content – interacts solely with the output generated by the AI system, rather than the AI system itself. The legislator has addressed this by stipulating in Art. 2(1) (c) AI Act that the regulation applies if the output is situated within the Union, regardless of the provider's or deployer's location.

C. Obligations

15 Article 50 AI Act encompasses five distinct transparency obligations. Recital 132 explains that these obligations, set out in paragraphs 1 to 4, are motivated by the fact that certain AI systems pose a particular risk with regard to identity fraud or deception. This is especially true for AI systems that interact with natural persons or generate content, regardless of whether these AI systems are classified as high-risk or not. Accordingly, Art. 50(6) AI Act underscores that the transparency obligations in paragraphs 1 to 4 do not alter or replace requirements and obligations for high-risk AI systems outlined in the AI Act. Additionally, these transparency obligations operate without prejudice to other transparency requirements imposed by Union or national laws for deployers of AI systems.[23]

16 Moreover, under Art. 50(5) AI Act, these transparency obligations must be presented to the affected natural persons "in a clear and distinguishable manner", at the latest by the time of their first interaction or exposure to the AI system. This information must also comply with applicable accessibility requirements, ensuring that it is accessible to all individuals as required by the AI Act.[24]

I. Art. 50(1) AI Act: Transparency for Chatbots and Interactive AI Systems

17 Art. 50(1) AI Act imposes an obligation on providers of AI systems intended to directly interact with natural persons to ensure that these systems are designed and developed in a manner that informs the individuals in question that they are interacting with an AI system.

1. Systems Intended for Direct Interaction

18 This requirement specifically applies to providers of AI systems designed for direct human interaction. The AI Act does not provide a definition of *"interaction"*. Therefore, its common linguistic meaning should be applied, which denotes reciprocal and interrelated actions, specifically through tactile, auditory, or visual influence.[25]

19 Furthermore, the system must be specifically intended for *"direct"* interaction. Notably, the

20 Christiane Wendehorst, 'The Proposal for an Artificial Intelligence Act COM(2021)206 from a Consumer Policy Perspective' (Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz, 14.12.2021), 11 <https://www.sozialministerium.at/dam/sozialministeriumat/Anlagen/Themen/Konsumentenschutz/Konsumentenpolitik/The-Proposal-for-an-Artificial-Intelligence-Act-COM2021-206-from-a-Consumer-Policy-Perspective_dec2021__pdfUA.pdf> accessed 24 April 2025; Wendehorst in Martini/Wendehorst (ed.), (n 7) Art. 3 para 81.

21 Similarly, the GDPR stipulates a "market place-principle" in Art. 3(2) GDPR, requiring controllers and processors from third countries to comply with the GDPR if they target individuals in the EU, see Gerrit Hornung in Spiecker gen. Döhmann/Papakonstantinou/Hornung/De Hert (ed.), General Data Protection Regulation (2023), Beck/Hart/Nomos, Art. 3 para 32.

22 This principle also resembles the "establishment principle" found in Art. 3(1) GDPR, which ensures that the GDPR is applicable to controllers or processors established in the EU, ibid, Art. 3 para 13.

23 Art. 50(6) AI Act.

24 Art. 50(5) AI Act.

25 David Bomhard/Marieke Merkle, 'Europäische KI-Verordnung' (2021) 6 RDi 276, para 35; Marieke Merkle, 'Transparenz nach der KI-Verordnung – von der Blackbox zum Open-Book?' (2024) 9 RDi 414, para 32.

Commission's draft did not differentiate between direct and indirect interactions.[26] This distinction in the final version of the AI Act raises questions about the types of interactions the legislator intended to address. For instance, automated recommendation systems that simply analyse user data to offer personalised suggestions likely fall outside the scope of Art. 50(1) AI Act, as they lack "direct" interaction.[27] In these cases, the provider is therefore not required to disclose that it is an AI system. Conversely, the transparency requirement primarily applies to systems like chatbots, which clearly engage in "direct" interaction.[28]

2. Designed and Developed to Inform Natural Persons about the AI Interaction

20 As a legal consequence, the provider shall ensure that the system is designed and developed to inform the natural persons concerned that they are interacting with an AI system. This indicates that para. 1 does not directly obligate the provider to inform affected individuals. Rather, it requires the provider to ensure the technical provision of this information by design.[29]

21 According to Art. 50(1) AI Act, affected individuals must only be informed "that" they are interacting with an AI system. This suggests that the information provided is limited to the mere question whether they are interacting with an AI system. Additional information – e.g. information concerning the operation of the system – must not be provided upon the persons concerned.[30]

22 Pursuant to Recital 132, providers shall take into account characteristics of natural persons belonging to vulnerable groups – particularly those affected by age or disability – when fulfilling their obligations, given the AI system is intended to interact with those groups. This notice should not be understood to mean that a "greater" amount of information must be provided to these individuals. Instead, it aims to ensure that the respective group can receive and comprehend the information.[31] For a technically savvy audience, an information such as "I am your virtual assistant" would suffice, whereas an older audience is likely to understand the information only if it is explicitly disclosed as, e.g., "You are interacting with an AI system."

3. Exceptions

23 According to Art. 50(1) AI Act, the transparency obligation does not apply when it is *"obvious"* – taking into account the circumstances and context of the use – that the individuals concerned are interacting with an AI system. The AI Act-COM originally lacked clarity on when interactions with AI systems are *"obvious"*.[32] This was regrettable, as the providers of AI systems could interpret this undefined term in very broad ways.[33] An amendment proposed by the Czech Presidency (of the Council of the European Union) introduced the notion of a *"reasonably well-informed, observant, and circumspect"* natural person as a benchmark for this determination – a standard consistent with consumer protection law and used by the European Court of Justice (ECJ) since the 1990s.[34] Further points of reference or illustrative cases are not provided in this context. Likely referenced here are aspects such as predefined response options and the instantaneous appearance of responses.[35] Predefined response options differ significantly from the open-ended, free-form responses typical

26 See Art. 52(1) AI Act-COM.

27 Philipp Roos/Johanna Voget, 'Transparenzpflichten für die Nutzung von KI auf Online-Marktplätzen' (2024) 10 RDi 487, 490 f.; also see Thomas Gils, 'A Detailed Analysis of Article 50 of the EU's Artificial Intelligence Act', 9 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4865427> accessed 24 April 2025.

28 The European legislator had already aimed primarily at regulating Chatbots during the drafting of the Commission's proposal, see COM(2021) 206, Explanatory Memorandum, 1.1.

29 Merkle, (n 25) para 31; the focus on the requirement of systems to be "designed and developed" in a certain way likely stems from the increased responsibility imposed upon providers. Similarly, the obligations set out in Arts. 13-15 AI Act for providers of high-risk AI systems also require the system to be "designed and developed" in a certain way. Critical: Gils, (n 27) 9, who argues that the "deployment-phase" is particularly crucial in this context.

30 Lea Katharina Kumkar in Hilgendorf/Roth-Isigkeit (ed.), 'Die neue Verordnung der EU zur künstlichen Intelligenz' (2023), C.H. Beck, § 6 para 38.

31 See Gianclaudio Malgieri and Maria-Lucia Rebrean, 'Vulnerability in the AI Act: Building an interpretation', 23 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5058591> accessed 24 April 2025.

32 Christoph Engelmann/Nico Brunotte/Hanna Lütkens, 'Die Regulierung von Legal Tech durch die KI-Verordnung' (2021) 7 RDi 317 para 22; Frauke Rostalski/Erik Weiss, 'Der KI-Verordnungsentwurf der Europäischen Kommission' (2021) 4 ZfDR, 329, 351.

33 ibid.

34 See <https://artificialintelligenceact.eu/wp-content/uploads/2022/09/AIA-CZ-3rd-Proposal-23-Sept.pdf> accessed 24 April 2025; See for example Case C-210/96 *Gut Springenheide und Tusky v Oberkreisdirektor des Kreises Steinfurt* [1998] ECLI:EU:C:1998:369, para 31.

35 Kilian Georg Wolf, 'Chatbots als KI-Systeme mit besonderen Transparenzpflichten nach Art. 52 KI-Verordnung' (2022) DSRITB, 601, 613.

of conversations with human interlocutors. This difference can manifest itself in different ways: Some AI systems include buttons or clickable options to help the user navigate, for instance, multiple-choice menus. In addition, many chatbots provide almost instantaneous responses to user input. Typical delays that occur in human conversations, such as when the conversation partner is typing or thinking, are absent. This lack of delay is often considered an indicator that an AI system is involved.[36]

24 In many cases, chatbots are already labelled as *"[company name] bot"* by design.[37] In such instances, it would be obvious that one is interacting with an AI system.[38] Here again, it should be noted that this standard changes if, for example, a disadvantaged group is part of the target audience of the chatbots.[39]

25 Lastly, the obligation does not apply to AI systems legally authorised for the detection, prevention, investigation, or prosecution of criminal offenses, if appropriate safeguards for the rights and freedoms of third parties are in place, unless such systems are made available to the public for reporting a crime.[40]

II. Art. 50(2) AI Act: Synthetic Content

26 Art. 50(2) AI Act imposes a transparency obligation for providers of AI systems generating synthetic audio, image, video and text content. Providers shall ensure that system outputs are marked in a machine-readable format, allowing such content to be detectable as artificially generated or manipulated. Paragraph 2 seeks, on the one hand, to ensure transparency regarding the authenticity of the content, specifically addressing whether events depicted in AI-generated photos or videos might mistakenly be perceived as *"real"*.[41] On the other hand, the provision clarifies whether content is human-made or AI-generated (e.g., whether the design of a logo or a music piece is AI-generated).[42]

1. AI Systems Generating Synthetic Content

27 Subject to the marking obligation are AI systems generating synthetic audio, image, video or text content. The AI Act does not provide a definition of the term "synthetic". Rather, this term should be equated with the term(s) "AI-generated" as the content should be distinguished from human-made content.[43] Accordingly, (almost) every output of such an AI system is subject to this marking requirement.

28 Paradoxically, paragraph 2 requires outputs to be marked as "artificially generated" or "manipulated", yet it only applies to systems that generate synthetic content.[44] Neither the term "artificially generated" nor "artificially manipulated" are defined in the AI Act. Linguistically, the phrasing suggests that "artificially generated" refers to content created entirely by AI, whereas "artificially manipulated" pertains to the modification of pre-existing content through AI.[45] The aforementioned stipulation, however, reflects a legislative imprecision. Correctly, paragraph 2 should encompass both artificially generated and artificially manipulated content, as this would align with its intended scope – to ensure that all content shaped by AI, whether through generation or manipulation, can be distinguished from purely human-made material.[46] The further wording of paragraph 2, which explicitly provides an exception for cases where the system does not substantially "alter" input data, does not support a differing interpretation.[47] This indicates that the provision is meant to also address AI manipulated content.

2. Marking in a Machine-Readable Format

29 Firstly, according to Art. 50(2) AI Act, providers are legally required to mark synthetic content in a machine-readable format. This marking obligation is, in a sense, specified in the recitals:

36 ibid.

37 For instance, Pizza Hut utilizes its 'Pizza Hut Chatbot', while H&M employs the 'H&M AI' bot, and Sephora features the 'Sephora Virtual Artist' for personalized customer experiences, see Anuj Kumar, Nimit Gupta, Gautam Bapat, 'Who is making the decisions? How retail managers can use the power of ChatGPT' (2024) 3 Journal of Business Strategy 161, 167 <https://www.emerald.com/insight/content/doi/10.1108/jbs-04-2023-0067/full/html> accessed 24 April 2025.

38 Maximilian Becker, 'Generative KI und Deepfakes in der KI-VO' (2024) 6 CR 2024, 353 para 48.

39 See C. I. 2.

40 Art. 50(1) AI Act.

41 Additionally, many of the affected contents are likely to also fall under the deep fake provision of paragraph 4. In that case, the transparency requirements apply cumulatively.

42 Recital 133, sentence 1 AI Act.

43 Angelica Fernandez, '"Deep fakes": disentangling terms in the proposed EU Artificial Intelligence Act' (2021) 2 UFITA 392, 413; Mario Martini in Martini/Wendehorst (ed.), (n 7) Art. 50 para 62.

44 Block, (n 4) 188.

45 See Lea Katharina Kumkar, 'Deepfakes – Risiken und Regulierung im europäischen Verordnungsentwurf für künstliche Intelligenz' (2023) 10 supplement 1 K&R 32, 35.

46 See Recital 133 sentence 1 AI Act.

47 Gils, (n 27) 17.

Providers must embed technical solutions to enable marking in a machine-readable format. Recital 133 sentence 4 AI Act gives a few examples for available techniques and methods to be used, namely watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods and fingerprints. It is emphasised that multiple labelling methods can also be combined.[48] The use of combined marking methods may even be essential, as for example metadata-based techniques can be easily bypassed by screenshots or automatic metadata removal on online platforms, rendering them ineffective.[49]

30 According to Art. 50(2) AI Act, the providers "shall ensure that 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, implementation costs and generally acknowledged state of the art, as may be reflected in relevant technical standards".

31 The "effectiveness" of the chosen solution extends beyond its technical functionality, as the other requirements (such as "robustness", "reliability" or the "generally acknowledged state of the art") already cover this aspect.[50] Rather, this criterion requires that the technical solutions practically enable a clear distinction between AI-generated and human-created content. For example, effectiveness is particularly not ensured if it is disproportionately difficult to decode the machine-readable marking.

32 The requirement to account for the specifications and limitations of the various content types implies that the watermark must be appropriately aligned with the nature of the respective content. For instance, it would be incongruous to apply an auditory watermark to an AI-generated image.

3. Detectable as Artificially Generated or Manipulated

33 Secondly, according to Art. 50(2) AI Act, providers must ensure that the content is "*detectable as artificially generated or manipulated*". The understanding of cumulative recognisability of the artificial origin ("marked in a machine-readable format *and* detectable as artificially generated or manipulated") suggests that a visible label for humans must be provided in addition to the machine-readable marking.[51] However, the stipulation that the mark must be *"detectable"* implies that the artificial origin of the content does not need to be visible to the human eye without the aid of additional tools.[52] Recital 135 further mentions (the access of) "detection mechanisms" if essential for enabling the public to effectively distinguish AI content. The fact that specific mechanisms are required for detection suggests that the obligation is intended to facilitate detection by technical systems, subsequently enabling recipients to be informed about the artificial origin of the content through these mechanisms.[53] Therefore, the two-tiered marking approach should rather be understood as a duty that distinguishes between the implementation of a machine-readable marking itself and the specific information to be disclosed when tracing the marking.

34 Pursuant to Art. 50(7) AI Act, the Commission is requested to facilitate the drawing up of codes of practice to ensure the effective implementation of the obligation to detect AI-generated content, including the access to detection mechanisms.[54] Additionally, other actors – such as the providers of very large online platforms (hereinafter: VLOPs) or very large online search engines (hereinafter: VLOSEs) in the sense of the Digital Services Act – are taken into account for embedding algorithmic detection mechanisms.[55]

4. Exceptions

35 Three exceptions apply to the marking obligation under Art. 50(2) AI Act. Firstly, according to para. 2, the obligation shall not apply to the extent the AI systems perform an "assistive function for standard editing". This could be the case if the AI system is used "as a tool for an essentially human product".[56] A straightforward example of such a fundamentally human-driven action is the automatic recognition of objects in photos to enable the segmentation of individual objects.[57]

36 Furthermore, the obligation set out in para. 2 shall not apply to the extent the AI system in question does not substantially alter the input data provided by the deployer or the semantics thereof. It is probable that this exception will cover programs

48 Recital 133, sentence 4 AI Act.

49 Becker, (n 38) para 55; Ramak Molavi Vasse'i, 'Watermarking von KI-generierten Inhalten als regulatorisches Instrument' (2024) 9 RDi 406, para 16.

50 Molavi Vasse'i, (n 49) para 31.

51 Kumkar/Griesel, (n 16) 124.

52 Block, (n 4) 188.

53 Molavi Vasse'i, (n 49) para 10.

54 Recital 135 AI Act.

55 European Commission, C/2024/3014, subsection 3.3. (40) (d).

56 Becker, (n 38), para 57.

57 ibid, Becker cites the example of the object isolation function on iPhones.

designed for spelling and grammar verification.[58] In addition, the scope may extend to systems that focus on optimizing image quality, translating text, reducing noise from recordings and converting file formats.

37 Lastly, a similar exemption to para. 1 is provided for instances where the system is legally authorised for crime investigation or prevention purposes.

5. Issues

38 A particular issue is that the provision does not require the deployer to specify which part of the content is AI-generated. Accordingly, it remains unclear whether only an insignificant part of the content was artificially generated or manipulated, or whether the content was predominantly or entirely AI-generated. Since no further information regarding the nature or extent of the generation or manipulation of the content is disclosed, the labelling remains largely uninformative.[59]

39 It is not yet possible to determine whether machine-readable markings effectively contribute to combating disinformation and ensuring transparency regarding whether content is AI-generated or human-made. A key factor will be whether social networks such as Facebook, Instagram, or X implement technical solutions that enable the identification of such content,[60] as users are unlikely to independently verify the origin of every piece of content they encounter.

III. Art. 50(3) AI Act: Emotion Recognition Systems and Biometric Categorisation Systems

40 According to Art. 50(3) AI Act, "deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system". In contrast to the Commission's draft, paragraph 3 now declaratively states that the deployers of these systems must process personal data in accordance with the GDPR and the EU Data Protection Regulation (EU-DPR), as well as the Data Protection Law Enforcement Directive (DP-LED). In this context, Art. 2(7) AI Act already clarifies that the AI Act and the aforementioned regulations and directive apply concurrently.

1. Additional regulations

41 The transparency obligation for emotion recognition systems and biometric categorisation systems is part of a broader set of regulations targeting such AI systems. For example, in certain cases, these systems are prohibited entirely: Art. 5(f) AI Act explicitly prohibits systems used to infer emotions of a natural person in a workplace or educational institution. Pursuant to Art. 5(g) AI Act, the same applies to biometric categorisation systems used to deduce or infer sensitive information such as the race, political opinions or sexual orientation of the affected individuals.

42 Additionally, AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on the interference of those attributes or characteristics, are considered high-risk AI systems pursuant to Art. 6(2) AI Act and Annex III, 1(b) AI Act.[61] Likewise, systems intended to be used for emotion recognition are classified as "high-risk" pursuant to Annex III, 1(c) AI Act. If an emotion recognition system or a biometric categorisation system is classified as a high-risk AI system, the transparency requirements set out in Art. 50(3) AI Act must be fulfilled cumulatively.[62]

2. Emotion Recognition Systems and Biometric Categorisation Systems

43 Art. 50(3) AI Act obliges providers of "emotion recognition systems" and "biometric categorisation systems" to "inform the natural persons exposed thereto of the operation of the system". Art. 3(39) AI Act defines the term "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". Recital 18 further clarifies this notion by distinguishing between recognised emotions and intentions – such as happiness, sadness, anger and surprise – and explicitly excluding physical states like pain or fatigue. In practical terms, however, the threshold between emotions and physical states is likely to be

58 Block, (n 4) 189.

59 Molavi Vasse'i, (n 49) para 56.

60 E.g., "meta" labels content identified by its systems as AI-generated, as well as content that users themselves declare to be AI-generated, cf. <https://www.meta.com/en-gb/help/artificial-intelligence/How-ai-generated-images-in-ads-are-identified-and-labeled-on-Meta/> accessed 24 April 2025.

61 See Annex III, 1. b) and c) AI Act.

62 See Art. 50(6) AI Act.

difficult to determine.[63]

44 Under Art. 3(40) AI Act, a "biometric categorisation systems" is defined as an AI system "for the purpose of assigning natural persons to specific categories on the basis of their biometric data". Recital 16 AI Act specifies that these categories may encompass aspects such as "sex, age, [...] religion, membership of a national minority, sexual or political orientation". However, an AI system is not considered a "biometric categorisation system" if its categorisation function is merely ancillary to another commercial service and is strictly necessary for objective technical reasons. Recital 16 further clarifies that such a feature is deemed purely ancillary only if it cannot, for objective technical reasons, function independently of the principal service and if its integration is not intended to circumvent the requirements outlined in the AI Act.

45 Lastly, Art. 3(34) AI Act specifies the term "biometric data" as "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, such as facial images or dactyloscopic data". Pursuant to Recital 14 AI Act, the notion of "biometric data" should understood in line with its definitions in Art. 4(14) GDPR, Art. 3(18) EU-DPR and Art. 3(13) DP-LED.

46 Despite this intention to align these definitions, the term provided in the AI Act differs from that of the GDPR, as Art. 4(14) GDPR expressly subsumes only those data under the term "biometric data" that allow for the unique identification of a person. The AI Act, however, does not incorporate this requirement.[64] Interestingly, the definition presented in the Commission's draft aligns precisely with the definition provided in the GDPR.[65]

47 The identical wording of the definition found in the Commission's draft – particularly the phrase "allow or confirm the unique identification of that natural person" – has been subject to criticism in literature.[66] This phrasing results in only "strong" biometric features, such as fingerprints, being covered by the definition.[67] For emotion recognition, so-called "weak" biometric features (such as body shape or voice) are crucial; for biometric categorization, "soft" biometric features, such as those associated with a specific age, gender, or skin colour, are relevant.[68] As a result, the final version of the AI Act deliberately omits the phrase "allow or confirm the unique identification of that natural person" as found in the GDPR to tailor the concept of "biometric data" to emotion recognition systems and biometric categorisation systems. This change is to be welcomed.

3. Inform the Person Exposed of the Operation of the System

48 Unlike Art. 50(1) AI Act – which requires the AI system to be designed in a way that informs the individual affected *"that"* they are interacting with an AI system – paragraph 3 requires the deployer to inform the person exposed *"of the operation of the system"*. This comparison implies that the obligation in paragraph 3 goes beyond a mere notification that an individual is exposed to such a system.[69] This alludes to the deployer not only informing the person exposed about "whether" they are using such a system but also about the specific manner of use – the "how" of the system.[70] Accordingly, essential parameters, based on which the system makes a decision – such as which feature (e.g. voice or facial structure) the system evaluates – must be disclosed.[71]

4. Exceptions

49 Art. 50(3) AI Act also includes an exception for systems authorised by law for the detection, prevention and investigation of criminal offences. Unlike other similar exemptions set out in Art. 50 AI Act, the exception in para. 3 sentence 2 does not include systems that are legally authorised to *"prosecute"* criminal offences.[72] This implies that the use of such

63 Mario Martini in Martini/Wendehorst (ed.), (n 7) Art. 3 para 279.

64 ibid, Art. 50 para 84.

65 See Art. 3(33) AI Act-COM.

66 Wendehorst, (n 20) 93 ff.; Wendehorst in Martini/Wendehorst (ed.), (n 7) Art. 3 para 232.

67 Wendehorst in Martini/Wendehorst (ed.), (n 7) Art. 3 para 232.

68 ibid para 240 f.

69 Kumkar in Hilgendorf/Roth-Isigkeit, (n 30) § 6 para 53; critical: Gils (n 27) 20, who notes that the phrasing "of the operation of the system" is ambiguous. It may either refer to the individual being exposed to the system "in operation" – in which case it suffices to inform the person of their exposure to an emotion recognition or biometric categorisation system – or to the person being informed "about the operation", which would require informing the individual about how the system functions.

70 ibid.

71 Martini in Martini/Wendehorst (ed.), (n 7) Art. 50 para 89.

72 The transparency obligations in paras. 1 and 4 (including subparas. 1 and 2) all contain exemptions related to the detection, prevention, investigation, and prosecution of criminal offences. Moreover, the Commission's draft also included an exception for emotion recognition systems and biometric identification systems legally authorised to prosecute criminal offences, see Art. 52(2) sentence 2 AI

AI systems for crime prosecution purposes is not outright prohibited but is in authorised cases always subject to the transparency obligation.[73]

50 Additionally, the exemption declaratively mentions that the AI system also has to be used in accordance with Union law. Likely, this addition was included to highlight the significance of the associated fundamental rights, as some applications of these systems are even prohibited or classified as high-risk.[74]

IV. Art. 50(4) subpara. 1 AI Act: Disclosure Obligation for "Deep Fakes"

51 Art. 50(4) subpara. 1 AI Act establishes a disclosure obligation for deep fake content. Under this provision, deployers of AI systems that generate or manipulate image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated.

1. AI Systems Generating or Manipulating Content Constituting a Deep Fake

52 Art. 50(4) subpara. 1 AI Act pertains to deployers of AI systems that generate or manipulate image, audio or video content constituting a deep fake. As explained regarding the transparency obligation under Art. 50(2) AI Act, the obligation covers both fully AI-generated but also merely modified content.[75] However, in contrast to paragraph 2, Art. 50(4) subpara. 1 AI Act does not extend to text-based content.

53 Furthermore, the content must qualify as a deep fake. Art. 3(60) 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"*. Accordingly, Art. 3(60) AI Act provides an extensive list of potential "targets" for deep fakes: In addition to persons, objects, places, entities, or events can also be the subject of a deep fake. This expansive list contradicts the common understanding that primarily humans can be subject of deep fakes, ultimately leading to numerous overlaps with Art. 50(2) AI Act.[76]

54 The deep fake definition no longer explicitly requires an *"appreciable resemblance"* to the mentioned subjects like Art. 52(3) AI Act-COM did.[77] At first glance, it appears that the legislator is broadening the definition of deep fakes to include content that is similar to the given subjects, rather than requiring near-identical resemblance. Recital 134, however, still mentions that the content has to *"appreciably resemble"* these subjects. This raises confusion and appears to be a regulatory imprecision with no practical impact.[78]

55 It is important to note that Art. 3(60) AI Act explicitly refers to "a person" perceiving the content as authentic or truthful.[79] In particular, there is no reference to a specific benchmark such as "a reasonably well-informed, observant, and circumspect natural person".[80] This suggests that a different standard is intended here. Overly stringent standards should not apply here to ensure that the regulatory purpose is not undermined. Since the spectrum of potential recipients includes both technically skilled and unskilled individuals, the question of whether image artefacts alone can disrupt the impression of authenticity should be critically considered.[81] Rather, it should suffice if the content appears authentic or truthful at a cursory glance to an average recipient.[82]

2. Disclosure Obligation

56 Deployers must *"disclose"* that the deep fake content is *"artificially generated or manipulated"*. Recital 134 further mandates that deployers must "clearly and distinguishably disclose"[83] AI-generated content by "labelling the AI output accordingly" and "disclosing its artificial origin".

57 This requirement implies that only the artificial origin needs to be disclosed, without necessarily labelling the content explicitly as a "deep fake". However, the

Act-COM.

73 *Argumentum a contrario* Art. 5(1)(g) AI Act.

74 See C. III. 1.

75 See C. II. 1.

76 However, if both provision are applicable, the obligations must simply be fulfilled cumulatively, see fn 42.

77 Gils, (n 27) 21.

78 Kristof Meding, Christoph Sorge, 'What constitutes a Deep Fake? The blurry line between legitimate processing and manipulation under the EU AI Act', 5 <https://arxiv.org/abs/2412.09961> accessed 24 April 2025; Łabuz, 'Deep fakes and the Artificial Intelligence Act – An important signal or a missed opportunity?' (2024) 4 Policy & Internet 783, 787 <https://onlinelibrary.wiley.com/doi/full/10.1002/poi3.406> accessed 24 April 2025; Gils (n 27) 21.

79 Block, (n 4) 189.

80 ibid.

81 Kumkar/Griesel, (n 16), 120.

82 ibid.

83 Also cf. Art.50(5) AI Act.

term "disclosure" remains somewhat ambiguous:[84] Neither Art. 50(4) subpara. 1 AI Act nor Recital 134 provide further details on the specific manner in which the disclosure should be implemented (e.g., whether the labelling must be affixed directly to the medium itself or if a notice in the caption suffices). An *argumentum a contrario* regarding the exception in sentence 3 can be drawn, suggesting that the deep fake must be directly labelled. Within the scope of this exemption, the legislator has aimed to limit the disclosure obligation in a manner that does not impair the presentation or enjoyment of the work. If disclosure in the caption were sufficient to meet the requirements of Art. 50(4) subpara. 1 AI Act, this exemption would be redundant.[85]

3. Does Art. 50(2) AI Act Effectively Complement Deep Fake Disclosure?

58 Unlike Art. 50(2) AI Act, the regulation set out in Art. 50(4) subpara. 1 AI Act is aimed at deployers. This issue was heavily criticized in the Commission's draft, as such disclosure requirements are for one technically much easier for the provider who could, for example, embed the necessary disclosures directly into the software code.[86] Additionally, provider obligations prove ineffective when malicious actors are the users of such systems.[87] Specifically, under Art. 50(4) subpara. 1 AI Act, if the provider harbours malicious intent, they are unlikely to label their deep fakes before dissemination. Imposing the disclosure obligation on the provider would at least require malicious actors to either establish their own generative AI systems or modify existing AI systems to eliminate the labelling applied by the provider. In either case, there would be at least some technical barrier to overcome, necessitating at least a certain level of effort from these actors. While this requirement may not deter professional disinformants, it does present an obstacle for everyday users attempting to disseminate disinformation on social media from their home computers.

59 One reason the obligation under Art. 50(4) subpara. 1 AI Act may not apply to providers is that a substantial share of deep fakes is created using general-purpose generative AI systems such as DALL-E or Midjourney.[88] Imposing a labelling requirement on providers would mean they must track every piece of content their systems create to assess whether it qualifies as a deep fake under Art. 3(60) AI Act and then disclose its artificial origin. This would, in effect, turn the transparency obligation for disclosing deep fakes into a moderation duty for providers.[89]

60 These issues appear to have been addressed, as the disclosure obligation is now supplemented by the aforementioned marking requirement for synthetic content (Art. 50(2) AI Act). As previously discussed, the recognition of these machine-readable markings requires the aid of detection mechanisms.[90] For instance, if a deployer circulates a deep fake without labelling it (as permitted under Art. 50(4) subpara. 1 AI Act), identifying its artificial origin still depends either on the recipient himself verifying its authenticity or the social media platform offering detection solutions.

61 Given the sheer volume of content circulating on social media, expecting every user to consistently verify the authenticity of content is highly unrealistic.[91] Consequently, the effectiveness of the transparency obligation under paragraph 2 relies heavily on social media platform providers independently applying visible labels to such content. However, such a platform provider focused framework seems to align with the legislator's objectives.[92]

84 cf. Lea Katharina Kumkar/Julian Philipp Rapp, 'Deepfakes' (2022) 3 ZfDR 199, 224; Becker, (n 38) para 73.

85 Kumkar/Griesel, (n 16) 122.

86 Mario Martini/Jonas Botta, 'Der Staat und das Metaversum', (2023) 12 Supplement MMR 887, 895; Kumkar/Rapp, (n 84) 224; Tobias Hinderks, 'Die Kennzeichnungspflicht von Deepfakes' (2022) 2 ZUM 110, 112; in a similar vein, Veale and Zuiderveen Borgesius note that, when such obligations are placed on deployers, enforcement becomes particularly challenging, see Michael Veale/Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 4 CRi, 97, 108.

87 Differently, Georg Borges, 'Die europäische KI-Verordnung (AI Act) Teil 3 – Transparenzpflichten, Durchsetzung, Gesamtbewertung' (2024) 10 CR 663 para 59, who argues that service providers (such as in the case of ChatGPT) should be regarded as the deployers of the AI system rather than the "users".

88 On the capability of such systems to generate highly realistic image, video, and text content, see Zhengyuan Jiang/Jinghuai Zhang/Neil Zhenqjang Gong, 'Evading Watermark based Detection of AI-Generated Content', (2023) ACM Conference on Computer and Communications Security 1168 <https://dl.acm.org/doi/10.1145/3576915.3623189> accessed 24 April 2025.

89 For a proposal on the transfer of content moderation obligations of the DSA to providers of large GenAI models, see Philipp Hacker/Andreas Engel/Marco Mauer, 'Regulating ChatGPT and other Large Generative AI Models' FAccT'23, 1112 , 1120 <https://dl.acm.org/doi/10.1145/3593013.3594067> accessed 24 April 2025.

90 See C. II. 3.

91 See C. II. 5.

92 See for example E. I.; the European legislator foresees that platform providers should take risk mitigation measures based on the labelling of synthetic content in machine-

4. Exceptions

62 Art. 50(4) subpara. 1 sentence 3 AI Act provides limitation: If the content in question forms part of an "evidently artistic, creative, satirical, fictional, or analogous work or programme", the disclosure provision is limited to disclosing "the existence of such generated or manipulated content in a manner that does not impede the display or enjoyment of the work."

63 In this regard, Recital 136 AI Act clarifies that compliance with Art. 50(4) subpara. 1 AI Act should not be interpreted as indicating that the use of the AI system or its output impedes the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights (hereinafter: CFR). This is particularly relevant if the limitation specified in Art. 50(4) subpara. 1 sentence 3 AI Act is applicable.

64 The EU Commission's draft included a similar exception in Art. 52(3) AI Act-COM.[93] Under the AI Act-COM, deep fakes whose use is necessary for the exercise of the rights to freedom of expression and freedom of the arts or science, as guaranteed by the CFR, were entirely exempt from the disclosure requirement. However, it is commendable that the exemption in the final version of the AI Act does not provide such a complete exemption from disclosure. Such exceptions ultimately create uncertainty for the recipient regarding whether the content in question is truly a deep fake, as AI system deployers could simply invoke the exception.[94] Accordingly, it is to be welcomed that a complete exemption from this disclosure obligation is no longer possible in this context.

65 In the final version, the ambiguity of when the work or programme of which the deep fake forms part is *"evidently"* artistic, creative, or satirical (etc.) raises questions about the applicable standard for assessment.[95] In comparison, the transparency obligation set out in Art. 50(1) AI Act uses the linguistically (almost) identical term *"obvious"*, which is further defined by the benchmark of a "reasonably well-informed, observant, and circumspect natural person". The fact that Art. 50(4) subpara. 1 AI Act avoids the term "obvious" suggests that a different standard is intended.[96]

66 Lastly, the second sentence of Art. 50(4) subpara. 1 AI Act further provides for an exception in cases where the use is authorized by law to detect, prevent, investigate or prosecute criminal offences. Here, the preceding discussions to the other exceptions apply.

V. Art. 50(4) subpara. 2 AI Act: Disclosure of AI-Texts

67 Art. 50(4) subpara. 2 AI Act introduces a new disclosure requirement for AI-generated text, which was not included in the original Commission's draft. Pursuant to this obligation, deployers of AI systems that generate or manipulate texts published with the purpose of informing the public on matters of public interest shall disclose that the text is AI-generated.

68 A key point to note is that the general requirement for marking synthetic content under Art. 50(2) AI Act also applies to AI-generated texts. Conversely, the term "deep fake" in Art. 3(60) AI Act – and, by extension, the deep fake disclosure obligation stated in Art. 50(4) subpara. 1 AI Act – excludes AI texts.

69 The primary reason lies in the nature of resemblance and its applicability to textual data. Unlike image, audio or video content – which can directly mimic the physical or sensory attributes of a specific entity (e.g. the appearance of a person) – text does not inherently "resemble" any such tangible or sensory reality.[97] However, although AI texts may not offer the same realistic illusion as other types of content, they can still serve as effective tools for disseminating misinformation. This is particularly concerning in the scope of automated journalism, where AI systems generate news articles or reports.

1. Text Published with the Purpose of Informing the Public on Matters of Public Interest

70 The transparency obligation set out in Art. 50(4) subpara. 2 AI Act pertains to deployers of AI systems generating or manipulating text content, provided the text is published with the purpose of informing the public on matters of public interest.

71 Notably, the obligation exclusively applies to texts that are "published". A linguistic understanding of the term "published" implies that the deployer must intentionally make the text accessible to the public. This aligns with the second requirement as

readable format.

93 See Art. 52 para. 3 subpara. 2 AI Act-COM.

94 Kumkar/Rapp, (n 84) 224.

95 Also see Gils (n 27), 25, who argues that the assessment of whether the work is "evidently" creative or satirical is highly subjective.

96 Paradoxically, the German language version of the AI Act uses the term "offensichtlich" for both cases. However, the use of two different terms in other language versions suggests that these terms do not share the same benchmark.

97 See Łabuz, (n 78) 787.

the obligation only concerns texts published with the specific intention of informing the "public". This term presupposes that the text is intended to be addressed to more than a limited number of people. In internet-related scenarios, particularly in the case of automated journalism as mentioned initially, this requirement is generally met.

72 Lastly, the text must address "matters of public interests". Recitals 7 and 8 AI Act give examples of "matters of public interest" as "health, safety and (the protection of) fundamental rights".[98] Beyond the aforementioned enumeration, the term "matters of public interest" may also include political, social, economic and cultural matters, with the key indicator being their relevance to public discourse and their relevance for the public opinion formation.[99]

2. Disclosure Obligation

73 Pursuant to Art. 50(4) subpara. 2 AI Act, where AI-generated or manipulated text is published with the purpose of informing the public on matters of public interest, the deployer "shall disclose that the content has been artificially generated or manipulated". In a similar manner to subpara. 1, subpara. 2 fails to provide any clarification regarding the method of "disclosure". Specifically, it remains unclear whether the text must be highlighted, for instance, through bold lettering, colour emphasis, or specific placement.[100] As with the other transparency obligations, the information must generally be provided to the affected party in a "clear and distinguishable" manner (Art. 50(5) AI Act).

3. Exceptions

74 The transparency obligation laid down in para. 4, subpara. 2, is exempted in the same way as the other obligations when the use is authorised by law for the purpose of detection, prevention, investigation or prosecution of criminal offences.

75 In addition, if the "content has undergone a process of human review or editorial control and a natural or legal person holds editorial responsibility for the publication of the content" the provider must not disclose the artificial origin of the text.

76 The extent to which human review or editorial control must occur is not directly evident from the wording of the Art. 50(4) subpara. 2 AI Act. In this context, it should be noted that the purpose of the transparency obligation is not solely to reveal whether an article was authored by a human or an AI system. Rather, its aim is to prevent the spread of misinformation, which could proliferate on a large scale if the substantial volume of automatically AI-generated news content were left unchecked.[101] The purpose of the subpara. 2 is of paramount importance when determining the extent of editorial control. Therefore, the human reviewer or editorial controller must ensure that no misinformation is disseminated through these AI-generated texts. A mere review of spelling errors and grammar is insufficient in this regard.[102]

D. Penalties

77 In the event of non-compliance with these transparency obligations, Art. 99(4) (g) AI Act provides for administrative fines. These fines may be up to € 15 million or, if the offender is an undertaking, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher.

78 In addition, according to Art. 99(1) AI Act, the Member States shall lay down rules on penalties and other enforcement measures. These shall include warnings and non-monetary measures, applicable to infringements of the AI Act by operators. Moreover, the Member States shall take all measures necessary to ensure that they are properly and effectively implemented, thereby taking into account the guidelines issued by the Commission pursuant to Art. 96 AI Act.[103] Due to the absence of measures by the Member States, nothing further can be said regarding the sanctions and measures provided here.

E. Interplay with the Digital Services Act

79 Transparency obligations set forth in Art. 50 AI Act have various implications for the Digital Services Act (hereinafter: DSA).

I. Risk Mitigation Measures

80 Both Recital 120 and (the almost identical) Recital 136 AI Act emphasize the risk of actual or foreseeable negative effects on democratic

98 Gils, (n 27) 26.
99 Martini in Martini/Wendehorst (ed.), (n 7) Art. 50 para 113.
100 Gils, (n 27) 26.
101 ibid para 114.
102 Gils (n 27) 27.
103 Art. 99(1) AI Act.

processes, civil discourse, and electoral integrity, including disinformation. Moreover, these Recitals underscore that the obligations established in the AI Act for enabling detection and disclosure of artificial origin (referring to both Art. 50(2) AI Act and Art. 50(4) AI Act) are essential for the effective implementation of the DSA. These obligations hold particular significance for providers of VLOPs and VLOSEs, as they relate to the risk mitigation measures mandated in Art. 35(1) DSA.

81 Online platforms (and online search engines) with over 45 million active users in the EU are designated as VLOPs and VLOSEs pursuant to Art. 33(4) DSA. According to Art. 34(1) DSA, providers of such VLOPs and VLOSEs shall diligently identify, analyse and assess any systemic risk stemming from their platform. These systemic risks may include, for instance, the dissemination of illegal content through their platforms or (actual or foreseeable) negative effects on the exercise of fundamental rights as well as on civic discourse and electoral processes.[104] Consequently, Art. 35 DSA requires such providers to put in place reasonable, proportionate and effective risk mitigation measures. In particular, Art. 35(1) sentence 2 (k) DSA outlines a risk mitigation measure to "prominently label" deep fake content or deep fake-like[105] content. In principle, however, providers of such platforms are not bound to a specific risk mitigation measure, but may freely choose between several measures, provided that these are found to be reasonable, proportionate, and effective.[106]

82 Additionally, the Commission has issued guidelines for providers of VLOPs and VLOSE on the mitigation of systemic risks for electoral processes pursuant to Art. 35(3) DSA.[107] These guidelines include specific risk mitigation measures linked to generative AI.[108] In addition to measures like labelling deep fakes[109], platform providers shall take measures such as adapting their content moderation processes to detect AI content marked in accordance with Art. 50(2) AI Act.[110] This enables VLOP- (and VLOSE-) providers to effectively search for AI-generated content as part of their moderation duties and, where necessary, filter out problematic content.

83 The guidelines were primarily developed in connection with the 2024 European Parliament elections but are intended to remain applicable beyond these elections, particularly concerning threats to electoral processes.[111] Nevertheless, the platform provider is granted a degree of discretion, similar to that of the risk mitigation measures mentioned in Art. 35(1) sentence 2 DSA. As a result, there is, in practice, no obligation to enforce these measures.

II. Unlabelled Content as Illegal Content in the Context of the DSA?

84 Furthermore, the DSA introduces a notice-and-action mechanism for service providers such as Facebook, Instagram or X. Generally, hosting service providers are not liable for illegal content uploaded by their recipients, provided they have no actual knowledge of such content.[112] Contrarily, if service providers become aware of illegal content, they are obligated to "expeditiously" remove it; otherwise, they might be liable for the respective content.[113] As part of the notice-and-action mechanism, users can report content they consider illegal to hosting service providers. According to Art. 16(6) DSA, service providers shall process these notices and take their decisions in respect of the information to which the notices relate, in a timely, diligent, non-arbitrarily and objective manner.

85 The basis for this decision is a generally broad definition of the term "illegal content". Pursuant to Art. 3(h) DSA, "any information that, in itself or in relation to an activity [...] not in compliance with Union law or the law of any Member State [...] irrespective of the precise subject matter or nature of that law" is considered "illegal content".

86 Recital 136 AI Act, however, emphasises that violations of the transparency obligations established in Art. 50 AI Act should not affect the assessment of the legality of the relevant content. That assessment "should be performed solely with reference to the rules governing the legality of the content".[114] In this context, "rules governing the legality of content" should be interpreted as referring to regulations that pertain to the "expressive content" of the given material.[115] Accordingly, content is classified

104 See Art. 34(1) DSA.

105 Unlike Art. 50(4) subpara. 1 AI Act, Art. 35(1) sentence 2 (k) DSA does not require the content to be artificially generated or manipulated. This initially seems contradictory, as Art. 50(2) AI Act also discloses only the artificial origin of the content. Nevertheless, it simplifies the labeling process for VLOPs in that the vast majority of content covered by Art. 35(1) sentence 2 (k) DSA is likely to be AI-generated as well.

106 This is already implied by the wording of Art. 35(1) sentence 2 DSA: "Such measures may include, [where applicable]".

107 C/2024/3014.

108 ibid subsection 3.3.

109 ibid (40) (b).

110 ibid (40) (d).

111 ibid subsection 1.1. (3).

112 Art. 6(1) (a) DSA.

113 Art. 6(1) (b) DSA.

114 Recital 136, sentence 4 AI Act.

115 Lennart Laude/Andreas Daum, 'KI als neues

as illegal under the DSA depending on whether it infringes personal rights, for example, or is defamatory, insulting or libellous.

87 In the context of deep fake content, this classification presents a notable weakness: A violation of the transparency obligation set out in Art. 50(2) and (4) AI Act does not automatically result in the content being removed.[116] Instead, if such content is not inherently illegal due to the lack of appropriate disclosure or marking, an individual assessment of its legality must be carried out.

88 This process carries significant risks. While this legal assessment is ongoing, the content may continue to circulate and be accessible, potentially causing irreversible damage – especially in cases involving manipulated media, such as deep fakes, where the rapid spread of misinformation or harmful content can have far-reaching consequences.

89 For example, if a deep fake is not disclosed in accordance with Art. 50(4) subpara. 1 AI Act and is simultaneously suspected of being defamatory towards the person depicted, the service provider is not automatically obligated to remove the deep fake due to the lack of proper disclosure. Rather, the content may remain on the platform until an assessment determines whether it defames the person portrayed. For these reasons, the absence of an immediate removal mechanism for unlabelled synthetic (deep fake) content highlights a critical gap in the regulatory framework.

90 Since the Commission's draft did not yet contain a marking obligation for synthetic content – and thus no illegality attached to such content – the final version of the AI Act could have marked a pivotal advancement in moderating disinformative content. Instead, this development represents a marked legislative regression: The AI Act-COM did not have a Recital corresponding to Recital 136 AI Act, thus, unlabelled deep fake content was considered "illegal" in the context of the DSA.[117] As a result, the AI Act's stance inadvertently weakens enforcement against potentially harmful AI-generated content by deprioritizing transparency violations and their consequences.

F. Final Evaluation

91 Compared to the Commission's draft, the transparency obligations were revised and refined during the legislative process, albeit mostly just in detail. Many of the adjustments made serve a clarifying purpose, such as the implementation of the term "reasonably well-informed, observant, and circumspect natural person" as standard for assessing obviousness under paragraph 1. In addition, existing transparency obligations have been optimised. For instance, the disclosure obligation for deep fakes no longer includes any exceptions for the exercise of certain fundamental rights. This was necessary, as such selective exceptions in the context of transparency obligations would undermine the intended transparency.

92 The most notable addition, however, is the new marking requirement for synthetic content under Art. 50(2) AI Act, which serves to support the disclosure obligation for deep fakes as well as the (also newly introduced) obligation to disclose AI-generated texts. Furthermore, this marking obligation synergises with the risk mitigation measures for providers of VLOP and VLOSE outlined in the DSA as these providers will be able to easily trace the machine-readable marking, allowing for a straightforward detection of AI content. This ease of AI content detection assists platform providers in specific risk measures such as the labelling of deep fakes in accordance with Art. 35(1) sentence 2 (k) DSA.

93 Nevertheless, the European legislator has missed a crucial step in the fight against disinformation by failing to classify unmarked content as "illegal content" within the meaning of the DSA. Accordingly, platform providers are not required to remove the content in cases of mere non-compliance with the transparency obligations set out in Art. 50(2) and (4) AI Act. This is particularly problematic concerning the spread of false information. In this context, deep fakes pose a significant threat because they can spread globally via the internet within seconds. This risk could have been mitigated – at least in part – by classifying unlabelled deep fakes as "illegal content" under the DSA and thereby holding platform providers accountable.

94 Overall, the adjustments compared to the Commission's draft are to be welcomed. However, the transparency obligations can only partially address the underlying risks. This is partly due to a fundamentally flawed approach, such as the assumption that malicious or uninformed actors will voluntarily disclose deep fakes as AI-generated. A more effective strategy would be to rely on trustworthy actors, such as AI system providers,

Wahlkampfinstrument' (Verfassungsblog, 3 May 2024) <https://verfassungsblog.de/ki-als-neues-wahlkampfinstrument/> accessed 24 April 2025.

116 Kumkar/Griesel, (n 16) 125.

117 See Kalbhenn, 'Designvorgaben für Chatbots, Deepfakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung' (2021) 8/9 ZUM 663, 671.

and to establish comprehensive regulations, such as mandatory content moderation for such materials, akin to the provisions of the DSA.