

4 August 2026

EU Finalises Transparency Rules for AI-Generated Content

On 20 July 2026, the European Commission (the “**Commission**”) published its finalised guidelines (the “**Guidelines**”)¹ on the implementation of the transparency obligations for certain AI systems under Article 50 of the EU Artificial Intelligence Act (the “**AI Act**”). Together with the Code of Practice on Transparency of AI-Generated Content (the “**Code**”)², the Guidelines complete the Commission’s guidance framework for the Article 50 AI Act transparency obligations which broadly take effect from 2 August 2026. This alert summarises the key points of the Guidelines and the Code and highlights the practical considerations for businesses developing or deploying generative AI.

Key Takeaways

1. **People must know when AI is involved.** Article 50 requires transparency about AI, it does not restrict what AI can be used for. Providers must tell users when they are dealing with an AI system and mark AI-generated content so it can be detected as such, and deployers must flag more sensitive uses by giving notice where emotion recognition or biometric categorisation is used and labelling deepfakes and AI-generated text published on matters of public interest.
2. **Voluntary in name, important in practice.** The Guidelines and Code are not binding (and becoming a Code signatory is optional, although the Commission announced that, as of 31 July, around 190 organisations have signed the Code)³, but signing the Code and compliance with the Guidelines is likely the clearest route to demonstrating compliance. Non-signatories to the Code will need to justify the adequacy of their own measures to market surveillance authorities.
3. **No single technical fix for marking / labelling.** The Commission has kept its advice on the technical methods to comply with various marking and labelling obligations under Article 50 very high level. Although there is useful guidance on how, for example, the technical compliance steps for labelling may differ between text, audio and video outputs from an AI system, the Commission is clear that it has not identified any single marking or labelling technique that currently meets the AI Act’s standards, meaning that a multi-layered approach (e.g. metadata, watermarking etc.) is required.

Transparency Guidelines

Broadly, Article 50 AI Act requires providers and deployers of AI systems (for most obligations limited to certain types of AI systems) to comply with certain information obligations for those that interact with their system and the output of those systems, with a primary aim of ensuring that individuals know when they are dealing with AI or AI-generated content. These apply to AI systems rather than AI models and so encompass application layer providers and other AI-enabled tech operators that use third-party AI models. The obligations apply from 2 August 2026 for most AI systems covered by Article 50, subject

¹ Available here: <https://digital-strategy.ec.europa.eu/en/library/guidelines-transparency-obligations-providers-and-deployers-ai-systems>

² Available here: <https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content>

³ Available here: <https://digital-strategy.ec.europa.eu/en/news/strong-backing-code-practice-transparency-ai-generated-content>

to a deferral (as a result of the Digital Omnibus⁴) to 2 December 2026 for obligations under Article 50(2) for marking and detection of generative systems already on the market and to 2 February 2027 for obligations under Article 50(2) for watermark-detection interoperability. The Guidelines and the Code add several points of practice for interpreting the Article 50 requirements.

1. **Interaction with AI (Article 50(1)).** Providers of any AI system intended to interact directly with natural persons (irrespective of whether general purpose AI or not, and including chatbots, agents and voice assistants) must tell users they are dealing with AI, unless obvious to a reasonably well-informed, observant and circumspect user or for AI systems used to detect, prevent, investigate or prosecute criminal offences. The Guidelines provide further specific guidance that: (i) AI agents (including as part of multi-agent architecture) must disclose both their artificial nature as well as the person on whose behalf they are acting; (ii) disclosure buried in terms and conditions, or readable only by machines, or as a generic statement (e.g. “assist”, “this system uses LLMs”) is not sufficient; and (iii) the “obvious interaction” exception will be construed narrowly and only apply where the AI is obvious to a reasonably well-informed user (such as a coding assistant used by professional developers), with general awareness that AI exists being insufficient and a higher bar to take advantage of the exception where the reasonably foreseeable audience may include children, elderly people or those with disabilities.
2. **Marking of AI Content (Article 50(2)).** Providers of AI systems (expressly including general purpose AI systems) that generate synthetic audio, image, video or text content must ensure that: (i) outputs are marked in a machine readable format that is detectable as AI-generated or manipulated; and (ii) their technical solutions are effective, interoperable, robust and reliable taking into account various factors, in each case, excluding AI systems that perform an assistive function for standard editing or do not substantially alter the input data, and AI systems used to detect, prevent, investigate or prosecute criminal offences. The Guidelines note that Article 50 applies to AI systems generally, not only general purpose AI (as noted in the AI Act itself), so narrow, single-purpose tools such as voice cloners are potentially caught. However, the Guidelines also confirm that certain AI systems and outputs fall outside Article 50(2), such as systems that merely reproduce or rearrange existing content (e.g. music playlists or recommender systems). The Guidelines also identify specific categories that fall outside of scope: (a) content that falls outside Article 50(2) altogether on the basis that there is no risk of deception, manipulation or to ensure integrity and trust in the information, and that is not perceptible to people or not intended to be interpreted, verified or acted upon by natural persons, including short outputs (e.g. single words, image captions and data labels), source code, machine-to-machine communications, and interim outputs in closed loop environments in industrial and product development workflows; and (b) what amounts to an assistive function for standard editing or does not substantially alter the input, such as grammar and spell-checking, format conversion and minor cropping.
3. **Emotion recognition and biometric categorisation (Article 50(3)).** Deployers of emotion recognition or biometric categorisation AI systems (other than those used under law to detect, prevent, or investigate criminal offences) must disclose to those exposed that the system is operating. The Guidelines make clear that this disclosure must be at the latest on first exposure, whether in writing, orally or through the use of standardised icons. The Guidelines emphasise that notice does not make the use lawful, and these systems may often also be prohibited activities under Article 5 or the high-risk activities under Article 6, and the AI Act does not give any leeway on compliance with separate privacy laws.
4. **Labelling deepfakes and public-interest text (Article 50(4)).** Deployers of AI systems (other than those used under law to detect, prevent, investigate or prosecute criminal offences) that: (i) generate or manipulate image, audio or video content that is a deepfake must disclose that the content has been AI-generated or manipulated, provided that this obligation for artistic / creative works is limited to disclosure of existence of such generation or manipulation in a manner that does not hamper the display or enjoyment of the work; or (ii) generate or manipulate text which is published for the purpose of informing the public on matters of public interest must disclose that the text has been AI-generated or manipulated, provided that this obligation does not apply where the AI-generated text has undergone a process of human review or editorial control by a natural or legal person with responsibility. The Guidelines provide clarity that:

A. the provider’s machine-readable marking obligation does not discharge the deployer’s duty to label;

⁴ For Paul, Weiss commentary on the Digital Omnibus, see <https://www.paulweiss.com/insights/client-memos/next-stop-on-the-digital-omnibus-ai-act-reforms-and-new-commission-guidance>

-
- B. the test for whether content is a deepfake is objective, with intent to deceive not required and photorealism not decisive. The subject need only be plausibly real, which catches AI-generated or de-aged actors and digital replicas but leaves out wholly fantastical content, or content no audience would take as genuine; and
 - C. the trigger for whether text is published for the purpose of informing on public interest matters should be construed broadly, covering matters open to public debate such as politics, public health, consumer safety, the environment, financial, and scientific or cultural developments, but not AI-generated fiction, product descriptions or a chatbot reply seen only by the user who asked. The construction can also turn on the publisher's purpose, not solely the content. The exception to this for editorial responsibility has a high bar, with the editing needing to be by someone with relevant competence, and spell-checking or cursory sign-off not sufficient and any substantive AI edit after editing removing the exception.
5. **Horizontal requirements (Article 50(5)).** For all obligations, Article 50 states that the information must be provided in a clear and distinguishable manner and given at the latest on first interaction or exposure. The Guidelines: (i) give examples of first interaction or exposure, such as launching a conversation with a chatbot or at the start of a deepfake video; (ii) confirm that the information must be perceivable in practice, so it cannot be buried in terms or menus, and a single up-front notice may not suffice where people may not see the start of the content (e.g., on a podcast or video); and (iii) confirm that content generated (or, in the case of Article 50(4), published) prior to 2 August 2026 does not need to be retroactively labelled.

Code of Practice

The Code is a voluntary tool to assist providers and deployers of generative AI systems who become signatories in demonstrating compliance with Articles 50(2), (4) and (5) of the AI Act. The Code is split into two sections, one for providers and one for deployers, with each commitment signposted as mandatory ("will"), recommended ("encouraged") or optional ("may"). Adherence has not yet been stated to be conclusive evidence of compliance, but following the positive adequacy assessment by the Commission and the AI Board, it is now the most straightforward, EU-wide route to demonstrating compliance (and the Commission has taken the view on the earlier GPAI Code of Practice that becoming a signatory is the best route to compliance). Non-signatories must instead show their own measures are adequate, typically through a gap analysis against the Code, and can expect closer regulatory scrutiny and more information requests.

Providers

1. **Multi-layered marking techniques.** As no single technique currently ensures compliance with Article 50(2), the Code requires providers to mark AI outputs (audio, image, video and text) with at least two machine-readable techniques, such as digitally signed, tamper-evident metadata plus imperceptible watermarking, and to prohibit users from stripping metadata. A single technique however will be sufficient in two instances: (i) free-form text, which cannot carry metadata; and (ii) generative systems built into closed physical products where technical measures keep the output inside the product, so it cannot be shared online. The Code also makes clear that text under 200 tokens (or approximately 150 words, although the Code notes that as watermarking technology advances, that threshold may decrease) is exempt from the watermarking requirement.
2. **Mechanism to detect AI Content.** The Code, going beyond the express marking obligations of Article 50(2), requires providers to make available a mechanism to detect AI-generated or manipulated content. The mechanism must in principle be free of charge, although providers with fewer than one million monthly users whose detection solution incurs substantial costs may charge a reasonable and proportionate fee for high-volume requests from a single user. Regulators, law enforcement, media, fact-checkers, researchers and civil society must always have free, unrestricted access. Forensic detection of content whose marking has been stripped is optional for now, as available tools do not yet meet the AI Act's quality requirements.
3. **Interoperability of watermark detection mechanisms.** By 2 February 2027, providers must make their watermark-detection mechanisms interoperable, so content can be verified without running each provider's detector in turn. The Code gives examples of an industry-standard query-routing method, an embedded readable signpost or a shared provider-agnostic solution.

Deployers

1. **Labelling and the EU icon.** The Code sets out how deployers can apply an Article 50(4)-compliant label where the provider's marking does not satisfy the deployer's obligation. Labels must be immediately recognisable without user

interaction, remain visible for a sufficient duration, be embedded directly into the content where possible, and be perceivable at the time of first exposure. The Commission has finalised three EU icons, for fully AI-generated content, for AI-modified content, and a basic icon supplemented by an interactive second layer, in each case, which deployers may use where visual disclosure is possible. Audible disclaimers are required where visual disclosure is not possible, such as for audio-only content. The “editorial” exception requires documented workflows, not solely assertions that such review took place.

2. **Internal processes.** Deployers must also adopt proportionate internal processes, awareness and literacy measures, and review and feedback mechanisms for their labelling of deepfakes and text publications.

Comment

The Code and the Guidelines together give businesses the first workable baseline for Article 50 compliance, but they also confirm that the technology to support compliance with certain aspects (in particular the marking and labelling obligations of Articles 50(2) and 50(4)) is not yet up to the task (or at least that the Commission does not have a clear view on what that technology is and how it can be standardised). The Code openly accepts that no single technique satisfies the AI Act’s marking requirements, so a multi-layered approach will be the working standard for now. With enforcement live from 2 August 2026 (subject to limited deferrals to 2 December 2026 for generative systems already on the market, with watermark-detection interoperability required by 2 February 2027) and fines of up to the greater of EUR 15 million and 3% of worldwide turnover, businesses that have not yet mapped their generative AI systems against Article 50 should, as a priority, identify whether they are caught as a provider, a deployer or both, and assess their technical capabilities to comply.

This memorandum is not intended to provide legal advice, and no legal or business decision should be based on its content. Questions concerning issues addressed in this memorandum should be directed to:

Jonathan H. Ashtor

+1-212-373-3823

jashtor@paulweiss.com

John P. Carlin

+1-202-223-7372

jcarlin@paulweiss.com

Katherine B. Forrest

+1-212-373-3195

kforrest@paulweiss.com

Nicole Kar

+44-20-7601-8657

nkar@paulweiss.com

Henrik Morch

+32-2-884-0802

hmorch@paulweiss.com

John Patten

+44-20-7367-1684

jpatten@paulweiss.com

Georgina Hoy

+44-20-7601-8743

ghoy@paulweiss.com

Alex Zapalowski

+44-20-7367-1697

azapalowski@paulweiss.com

Associates Edmund Berney and Scott Caravello contributed to this Client Memorandum.