Key recommendations for the forthcoming EC Guidelines on Article 50 AI Act and the Code of Practice for Transparent AI systems

1. **Encourage the use of mature technical solutions while preserving flexibility**: To promote trust and transparency in the context of AI-generated media, the Code should encourage the use of mature, widely deployed technical solutions for disclosing provenance information, where appropriate. Solutions developed collaboratively by global experts, tested across diverse use cases, and implemented at scale offer proven benefits in terms of technical robustness, interoperability, and security. Their adoption helps prevent fragmentation, accelerates deployment, and supports key regulatory objectives such as accountability, traceability, and user empowerment, especially in efforts to counter misinformation and ensure responsible AI development.

   We encourage relying on the following principles, among others, when assessing the suitability of technical solutions for compliance with Art. 50(2): strong security, easy availability, cross-platform interoperability, alignment with open metadata standards, and a mechanism that lets producers and editors communicate information directly to consumers. C2PA Content Credentials is a current example of a technical solution that meets these criteria for audio-visual media and text outputs in 'containerized' formats (e.g., documents and PDFs). At the same time, the field of verification and authentication for AI-generated content remains in active development, particularly for modalities such as text and audio. Before definitive recommendations can be made across all modalities, further research and innovation are needed. Accordingly, providers should retain the flexibility to select the most appropriate technical approach per modality, guided by risk-based assessments and evolving best practices.

   Technical considerations must also inform how disclosures are applied. For example, marking on AI-generated images should not obscure critical visual information, and marking on AI-generated text content should not alter or obscure the semantic meaning of phrases. Prescriptive requirements, particularly for modalities where techniques are still emerging, risk becoming outdated and may inadvertently hinder innovation or reduce the effectiveness of labeling systems, specially as authentication techniques continue to evolve alongside the technology itself.

   Flexibility is also essential to uphold accessibility. Applying uniform labeling techniques to all AI-assisted outputs could unintentionally disadvantage individuals who rely on accessibility technologies. It is important to distinguish between synthetically delivered human-authored content and content generated by AI. For example, individuals with disabilities may use Augmentative and Alternative Communication (AAC) or customized neural voices (CNV) to communicate. These tools are often the only means of verbal expression for many users. Marking human-authored, synthetically delivered communication with the same techniques as AI-generated content blurs a critical distinction and risks confusion, stigma, and unnecessary filtering. To avoid these harms, we urge that the Code includes a general

exemption for AI-assisted accessibility features, allowing for context-sensitive provenance solutions that preserve clarity for listeners while protecting the dignity and inclusion of the communities these technologies serve.

2. **Account for limitations in techniques across different modalities:** Both the guidelines and the Code of Practice should explicitly acknowledge that existing techniques for disclosing AI-generated content, while advancing rapidly, still face limitations in effectiveness, scalability, interoperability, and tamper-resistance. These limitations vary significantly across modalities. For example, more mature and robust solutions are available for applying provenance information to images, audio and video content, and those should be leveraged where appropriate. However, solutions for marking AI-generated text remain in early stages of development and require further investment, research, and standardization before they can be considered reliable or scalable.

   Recognizing these differences is essential to avoid one-size-fits-all mandates that could undermine the effectiveness of disclosure efforts. The Code should reflect the varying levels of maturity across modalities and support a flexible, risk-based approach that allows providers to adopt the most appropriate technical solutions as the field continues to evolve. This will help ensure that provenance and labeling practices remain both technically feasible and aligned with regulatory goals for transparency, accountability, and user trust.

3. **Seek alignment with international standards:** While industry works on standards and regimes to apply provenance information to AI-generated content, voluntary technical standards organizations, such as the World Standards Cooperation (an alliance of the IEC, ISO and ITU international standardization bodies), can help develop and test best practices, promoting consistency, interoperability, and collaboration across techniques. We encourage the AI Office to support such international efforts and use them as a basis for both the Code discussions as well as the development of any upcoming European harmonized standards. Ideally, companies should be able to leverage voluntary technical standards to comply with the EU AI Act, especially in the absence of harmonized standards.

4. **Adopt a risk-based approach in implementing Art. 50:** Disclosure obligations should be proportionate to the risks involved and the context in question. For example, indiscriminate marking or repetitive notifications to consumers in a gaming context would be detrimental, particularly for the entertainment sector. Recital 134 of the AI Act clarifies that deployers of AI systems generating deep fake content should clearly and distinguishably disclose that it has been artificially created or manipulated by labelling the output accordingly and disclosing its artificial origin. Regarding deep fake content that is part of evidently creative, satirical, artistic, fictional or analogous works or programs, Recital 134 clarifies that disclosure of the existence of such generated or manipulated deep fake content should occur 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. In line with Recital 134, we urge the AI Office to clarify that Art. 50(4) does

not apply to gaming. Publishing and distributing games and similar entertainment content retains material distinctions from deploying an AI system, and these two contexts should not be conflated. Not every game, or every user-generated-content feature within a game, relies on AI systems. These distinctions are critical for Art. 50(4), which is focused on deployers of AI systems, not publishers of entertainment content.

5. **Apply Art. 50(4)'s labelling requirements in line with the Act's legal text**: Under the AI Act, the legal responsibility to visibly label deep fake content and text published for the purpose of informing the public on matters of public interest lies with deployers (i.e., users) of AI systems, rather than providers. The Code should therefore align with the legal text and ensure that primary responsibility for implementation of Art. 50(4) rests squarely with downstream deployers. Providers may still offer optional labelling tools at customers' request, but any "one-size-fits-all" approach will inevitably fall short across the full range of deployment scenarios because system providers are technically and practically incapable of discerning *ex ante* the deployment context, which is essential for determining when Art. 50(4)'s triggers are met. For example, upstream system providers have no way of knowing independently, at the time of system development, whether text that system produces will be used "with the purpose of informing the public on matters of public interest." In practice, whether a text is used for a matter of public interest often depends on where it is published, such as whether it is published on a (social media) platform. The platform may have the AI system built in, or the content may be generated by a separate AI system and then uploaded. The Code should distinguish among these scenarios and ensure that the obligation to determine when the public interest trigger is met, and therefore when the disclosure obligation attaches, lies with the system deployer.

6. **Distinguish software code from semantic text content:** We ask the AI Office to confirm that AI-generated software code produced by code completion systems falls outside the scope of Article 50(2). Code completion systems are AI systems that offer AI-generated software code suggestions based on what a software developer is typing in their integrated development environment (IDE). While software code is expressed through text characters, it is fundamentally different from semantic text content. Its function is to give deterministic instructions to a computer, not to convey meaning to humans. Its use does not create risks for the trust in and integrity of the information ecosystem (see BeckOK KI-Recht/Lauber-Rönsberg KI-VO Art. 50 n. 26). Text content should be defined as human language conveying semantic meaning through text characters, distinct from software code.

7. **Clarify types of AI systems, features or components in scope of Art. 50.** Across Art. 50, clearer guidance is needed on what qualifies as an AI system, as opposed to AI features or components. For example, it is uncertain whether a standalone AI feature embedded in an otherwise non-AI service, such as an AI-assisted chart designer functionality embedded in a non-AI-based spreadsheet editor, would count as an AI system and therefore fall within Art. 50's remit. In the context of AI coding assistance, a product may include AI functionalities with different characteristics,

e.g. code completion, code suggestions based on text prompts, chat assistance in the IDE, coding agents etc. For such integrated AI features, it is unclear if it would be sufficient to inform a user upon initial sign-up.

8. **Adopt a proportionate approach that differentiates material from immaterial modifications to input data.** Art. 50(2) specifies that the marking obligation does not apply to AI systems that do not substantially alter the input data provided by the deployer or the semantics thereof, yet it gives no guidance on what counts as such a non-substantial alteration. Forthcoming guidelines and Code discussions should therefore define material versus immaterial modifications, so it is clear which changes fall outside Art. 50(2). This clarification is most critical for text content: with short-form AI-generated text, especially outputs under 30 words, the efficacy of provenance or labeling tools drops sharply, making enforcement impractical and potentially misleading. Exempting these brief outputs would focus regulatory effort where technical reliability and user impact are greatest and avoid overreach that could undermine trust in labeling systems.

9. **Acknowledge limitations to applying Art. 50 to open-source AI systems.** C2PA manifests are created, signed, and attached to assets (e.g., audio-visual media) after they have been generated. Thus, system developers are equipped to add such manifests when they are hosting the system for inferencing. In other cases, it will be more appropriate for the downstream users of open-source systems, who generate content or host the system for others to do so, to add this type of disclosure. We note that while an open-source system could use C2PA API calls to generate and validate manifests, this would come with security concerns. As with all open systems, this would open the possibility of hackers trying to modify the open-source system to interfere with its communication with C2PA API servers, potentially reducing the difficulty of creating successful attacks. One such attack could entail creating a secondary memory buffer that has a modified copy of the content, and sending that to the C2PA server, so the C2PA manifest will be created based on content that's different from the one that the manifest will be attached to by the open-source AI content generator. Additional challenges exist for disclosure for open-source systems via watermarking. A problem with open-source watermarking is that all the details of the watermarking algorithm are broadly known, making it much easier for hackers to implement removal attacks. While forgery attacks (forging a valid watermark on media that was not originally marked) are typically very difficult without algorithmic knowledge, open sourcing may make it easier for a hacker to figure out a forging approach. Disclosure via methods that are easy to attack should be avoided, as it will decrease the reliability of provenance information available in the ecosystem.

10. **Clarify overlap with the Digital Services Act.** There is a degree of overlap between Article 50 AI Act and the DSA. Under Article 35 of the DSA, very large online platforms and very large online search engines must implement reasonable, proportionate, and effective mitigation measures tailored to the specific systemic risks they face. These may include steps such as labelling of deep fake content, which closely mirrors the labelling requirement for deep fakes under the AI Act. The European Commission has

also issued DSA guidelines on systemic electoral risks, calling for generative AI content to be detectable, and for deep fake content to be clearly labelled or prominently marked. Finally, deep fake content in scope of Article 50(4) AI Act may potentially also qualify as “illegal content” under Article 3(h) of the DSA, further underscoring the legal significance of proper marking and disclosure. We recommend the Code to clarify the applicability and scope of such overlapping provisions.