JUN. 26

NOTA INFORMATIVA

TECNOLOGIA, MEDIA E TELECOMUNICAÇÕES

Transparência dos conteúdos gerados por IA

Código de Conduta

I. Enquadramento

O AI Office publicou, em 10 de junho de 2026, o Código de Conduta sobre Transparência dos Conteúdos Gerados por IA, visando densificar o artigo 50.º do Regulamento (UE) 2024/1689 (Regulamento da IA) relativo às obrigações de transparência aplicáveis a sistemas de Inteligência Artificial (IA) generativa. O Código está atualmente a ser objeto de uma avaliação de adequação pela Comissão Europeia e pelo Comité para a IA, prevendo-se que seja complementado por orientações da Comissão sobre o âmbito das obrigações de transparência estabelecidas no artigo 50.º do Regulamento da IA, cuja consulta pública decorreu até ao início do mês de julho.

A evolução da IA generativa permite a criação de conteúdos com um grau de realismo que os torna difíceis de identificar, o que levanta preocupações quanto ao potencial de indução em erro do público, com impacto na opinião pública e na tomada de decisões informadas pelos cidadãos. Para tanto, o n.º 4 e 5 do artigo 50.º do Regulamento da IA obriga que se mencione "que os conteúdos foram artificialmente gerados ou manipulados".

O Código pode ser subscrito para apoiar o cumprimento do Regulamento da IA, prevendo-se medidas que auxiliam os i) *providers* ou fornecedores a marcarem e detetarem conteúdo criado ou modificado por IA e os ii) *deployers* ou utilizadores/responsáveis pela implantação a rotularem conteúdo criado ou modificado por IA.

A adesão a este Código não constitui, por si, prova de cumprimento das obrigações previstas no artigo 50.º do Regulamento da IA. Os fornecedores e *deployers* que decidam dar cumprimento a estas obrigações através de outros meios terão de demonstrar individualmente que essas medidas são adequadas, o que será avaliado pelas autoridades nacionais de fiscalização do mercado.

Pedro Lomba
Mafalda Sequeira Roldão
Equipa de Tecnologia, Media e Telecomunicações

A evolução da IA generativa permite a criação de conteúdos com um grau de realismo que os torna difíceis de identificar.

II. Medidas previstas no Código

1. Marcação e deteção de conteúdo criado ou manipulado por IA *(providers)*

À luz do n.º 2 do artigo 50.º do Regulamento da IA, os fornecedores ou *providers* (entidades que desenvolvem e colocam no mercado sistemas de IA generativa, que permitam gerar áudio, imagem, vídeo ou texto) deverão proceder à deteção e marcação do conteúdo criado ou modificado por IA através de meios suscetíveis de serem lidos por máquinas *(machine-readable)*. O Código incentiva as entidades a desenvolverem soluções de marcação ou deteção de IA, ainda que não sejam diretamente responsáveis pela marcação.

Os signatários do Código podem adotar vários tipos de marcação como:

- **Metadados assinados digitalmente e certificados temporalmente, de forma segura e resistente a adulterações.** Recomenda-se ainda a inclusão de outros metadados, desde que não contenham dados pessoais ou comerciais.

- **Marca de água *(watermark)* impercetível e de difícil separação do conteúdo.** Recomenda-se a implementação de *watermarking* no modelo.

- ***Fingerprinting*** (mais adequado para áudio e conteúdo visual) ***ou logging*** (mais adequado para texto), são apresentadas como medidas complementares, visto que, isoladamente, não são suficientes para respeitar os requisitos do artigo 50.º (eficácia, interoperabilidade, solidez e fiabilidade).

Utilizar um único tipo de marcação é considerado proporcional e suficiente i) sempre que esteja em causa um sistema de IA generativo incorporado em produtos físicos num ambiente tecnicamente controlado e fechado, de natureza predominantemente instrutiva, e onde existam medidas técnicas eficazes que impeçam a saída desses outputs do ambiente do produto, e ii) quando esteja em causa texto livre, visto não ser suscetível de transportar metadados.[1] Em todos os outros casos, propõe-se que seja implementada uma marcação multicamada, garantindo que os outputs dos sistemas incluem pelo menos dois tipos de marcação legível por máquina.

É ainda recomendado que os signatários:

- Manter e não alterar marcações de conteúdos criados por outros sistemas de IA quando utilizados como input e subsequentemente transformados em output pelos seus sistemas de IA.

- Promover a proibição da remoção intencional ou da adulteração das marcações de metadados por *deployers* através de políticas de utilização aceitável, termos e condições ou outra documentação que acompanhe o sistema de IA.

- Fornecer, facultativamente, informação adicional sobre a origem dos conteúdos ao longo dos fluxos de trabalho, sempre que tal seja tecnicamente viável, nomeadamente:

  i) O nome do sistema de IA;

  ii) A denominação da entidade prestadora;

1 O Código recomenda a aplicação de watermarking nos textos livres com mais de 200 caracteres, sugerindo também que o acesso às ferramentas de deteção correspondentes seja restrito a utilizadores especialistas verificados.

iii) O *timestamp* da criação ou manipulação do conteúdo;

iv) O identificador e a versão do modelo subjacente;

v) O registo do tipo de operação, por exemplo, remoção de objetos. Havendo várias operações, recomenda-se que sejam identificadas num único tipo de marcação de metadados.

- Disponibilizar uma funcionalidade que permita aplicar diretamente uma etiqueta visível no momento da geração do output, facilitando o cumprimento dos deployers da sua obrigação de identificar *deepfakes*[2] e texto criados ou manipulados por IA (artigo 50.º, n.º 4, do Regulamento da IA).

- Promover a literacia no domínio da IA dos trabalhadores com funções relevantes para assegurar o cumprimento do n.º 2 e 5 do artigo 50.º e do artigo 4.º do Regulamento da IA.[3]

- Manter documentado um processo de conformidade que descreva, a um nível geral, a forma como aplicaram as diferentes medidas para assegurar o cumprimento do n.º 2 e 5 do artigo 50.º do Regulamento da IA (esta medida será aplicada de forma proporcional, tendo em conta a dimensão e os recursos do signatário, em especial no caso de PME e empresas em fase de arranque).

- Disponibilizar uma solução, em regra gratuita, que permita verificar se o conteúdo é criado ou manipulado por IA, independentemente do tipo de marcação, comunicando essa deteção de forma clara, compreensível e acessível ao público-alvo.

  i) O Código prevê a possibilidade dos resultados de deteção serem exportados com assinatura digital, incluindo *hash*, identificador e *timestamp*, que haja acesso restrito a especialistas para mecanismos menos fiáveis (ex.: deteção de *watermark* em texto livre) e mecanismos de deteção forense para conteúdos sem marcação (ex.: marcações removidas).

O Código prevê a possibilidade dos resultados de deteção serem exportados com assinatura digital.

As soluções técnicas utilizadas para deteção de conteúdos criados ou manipulados por IA têm de ser:

- **Eficazes** - As soluções serão consideradas eficazes quando as pessoas singulares consigam aceder e compreender o significado dos resultados de deteção. Não existe uma métrica quantitativa específica, sendo exigida uma avaliação centrada no utilizador.

- **Fiáveis** - A fiabilidade corresponde à capacidade de identificar corretamente a origem do conteúdo gerado ou manipulado por IA. Devem ser utilizadas métricas adequadas (por exemplo, taxa de erro de deteção) e demonstradas baixas taxas de erro em amostras variadas, abrangendo conteúdos próprios e de terceiros.

2 Nos termos do ponto 60 do Regulamento da IA - «Falsificações profundas», conteúdos de imagem, áudio ou vídeo gerados ou manipulados por IA, que sejam semelhantes a pessoas, objetos, locais, entidades ou acontecimentos reais, e que possam levar uma pessoa a crer, erroneamente, que são autênticos ou verdadeiros.

3 Esta medida não prejudica a responsabilidade dos *deployers*, que mantêm a obrigação de divulgar de forma clara e percetível os conteúdos em causa, nos termos do n.º 4 e 5 do artigo 50.º do Regulamento da IA.

- **Sólidas** - As soluções devem manter níveis de desempenho sob diferentes condições (por exemplo, *screenshots*, redimensionamento, rotação, tradução, desfoque de rostos, remoção, cópia ou modificação de marcações ou tentativas de mascarar a origem do conteúdo). A solidez será avaliada com as mesmas métricas da fiabilidade.

- **Interoperáveis** - As soluções devem funcionar de forma integrada entre sistemas, fornecedores e contextos, permitindo a deteção independentemente da técnica utilizada. Na ausência de padrões consolidados, adotar-se-á uma implementação faseada, prevendo-se o desenvolvimento de soluções de interoperabilidade para *watermarking* **até 2 de fevereiro de 2027**.

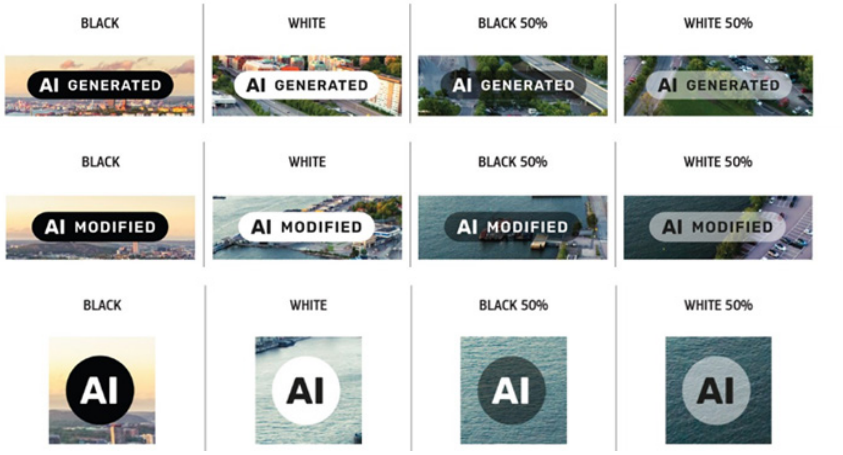
Os signatários poderão utilizar técnicas alternativas de marcação e deteção desde que, antes da colocação no mercado ou disponibilidade do sistema de IA generativa, demonstrem conformidade junto das autoridades competentes, com base em métodos e *benchmarks* de avaliação reconhecidos. Até à consolidação de métodos e *benchmarks* reconhecidos (nomeadamente aprovados pelo AI Office), os signatários devem testar e reportar o desempenho das suas soluções com base em *benchmarks* internos e melhores práticas do setor. Podem ainda envolver peritos independentes nos testes, designadamente para avaliação da solidez face a ataques *(red teaming)*, ou recorrer a ambientes de testagem da regulamentação da IA *(sandboxes regulatórias)* ao abrigo do artigo 57.º do Regulamento da IA.

2. Rotulagem de *deep fakes* e textos criados ou manipulados por IA *(deployers)*

Segundo o n.º 4 e 5 do artigo 50.º do Regulamento da IA, os *deployers* de sistemas de IA generativa têm a obrigação de proceder à rotulagem de conteúdo criado ou manipulado com IA. O Código introduz orientações sobre i) conteúdos de imagem, áudio ou vídeo que constituam ***deepfakes*** e ii) **texto com o propósito de informar o público sobre matérias de interesse público** criados ou manipulados através de IA, **sem que tenha havido processo de análise humana ou controlo editorial**.

O Código incentiva o respeito das seguintes medidas:

- Divulgar de forma consistente e eficaz a origem artificial de *deepfakes* ou texto, através de iconografia da UE ou rótulo equivalente que cumpra as especificações de design.



- Posicionar o acrónimo "AI" de modo visível e incorporado diretamente por tempo suficiente para ser notado no conteúdo, preferencialmente com informação sobre se foi gerado ou modificado com IA.

- Utilizar um *disclaimer* auditivo no início do conteúdo e repetição periódica ao longo do conteúdo, se a divulgação visual não for possível.

- Assegurar a acessibilidade a todos, nomeadamente a pessoas com necessidades especiais, em conformidade com o Direito da União Europeia, designadamente a Diretiva (UE) 2019/882 (transposta pelo Decreto-Lei n.º 82/2022, de 6 de dezembro, relativo aos requisitos de acessibilidade de produtos e serviços) e a Diretiva (UE) 2016/2102 (transposta pelo Decreto-Lei n.º 83/2018, de 19 de outubro, relativo aos requisitos de acessibilidade dos sítios web e de aplicações móveis de organismos públicos), através de *disclaimers* auditivos, soluções tácteis ou alternativas para conteúdos visuais, iconografia de alto contraste e suscetíveis de deteção por tecnologias de assistência.

- Contribuir para a criação de uma *task force* para desenvolver a iconografia da UE, incluindo:

  i) melhoria do design e usabilidade;

  ii) criação de soluções interativas;

  iii) harmonização de práticas de divulgação;

  iv) partilha de boas práticas setoriais.

- Implementar processos internos adequados, incluindo um processo de conformidade interna com documentação pronta para ser partilhada com as autoridades competentes, exemplos de implementação e divulgação pública de soluções adotadas.

- Criação de mecanismos de reporte de erros de rotulagem e célere correção de incumprimentos.

- Formar trabalhadores com funções relevantes sobre esta obrigação, implementação de designs e especificidades de posicionamento.

- Implementar políticas internas que assegurem revisão humana do conteúdo e identificação pública da entidade responsável editorialmente. Os signatários que sejam prestadores de serviços de media, na aceção do n.º 2 do artigo 2.º do Regulamento (UE) 2024/1083 e que já estejam sujeitos a obrigações editoriais podem recorrer à exceção do n.º 4, 2.º parágrafo, do artigo 50.º, aplicando os seus procedimentos de revisão, controlo editorial e padrões profissionais já existentes.

Implementar políticas internas que assegurem revisão humana do conteúdo e identificação pública da entidade responsável editorialmente.

III. Conclusão e *next steps*

O Código de Conduta representa um passo significativo na operacionalização do Regulamento da IA, antecipando-se que venha a ser utilizado como documento orientador para a uniformização da fiscalização de mercado pelas autoridades competentes em toda a União. O Código promove uma aplicação coerente, prática e proporcionada das obrigações de transparência do Regulamento da IA, mas não substitui o Regulamento da IA nem as orientações da Comissão sobre o artigo 50.º. Proporciona, sim, um quadro prático reconhecido à escala da UE para os signatários demonstrarem o cumprimento dessas obrigações.

Apesar de as medidas do Código serem de carácter voluntário, as obrigações subjacentes (artigo 50.º do Regulamento da IA) são aplicáveis a partir de 2 de agosto de 2026.

Atendendo ao acima exposto, recomenda-se que *providers* procedam à:

- Adoção de soluções técnicas (metadata, *watermarking*);
- Revisão políticas de utilização aceitável/ termos e condições para prevenir remoção de marcas tanto internamente, como por utilizadores;
- Preparação de evidência de compliance *(audit trail)*.

Os *deployers* deverão:

- Rotular conteúdo com iconografia UE aquando da publicação de *deepfakes* ou textos com o propósito de informar sobre matérias de interesse público criados ou alterados por sistemas IA;
- Adaptar iconografia às regras de acessibilidade em vigor;
- Assegurar que as políticas internas editoriais, se existentes, estão conformes com estas medidas;
- Implementar mecanismos de reporte de erros de rotulagem e de alteração célere em casos de incumprimento;
- Preparar evidência de compliance *(audit trail)*. ■