Orientações sobre práticas proibidas de IA

Transformative Legal Experts

PLMJ

Do AI Act às Orientações sobre práticas proibidas de IA

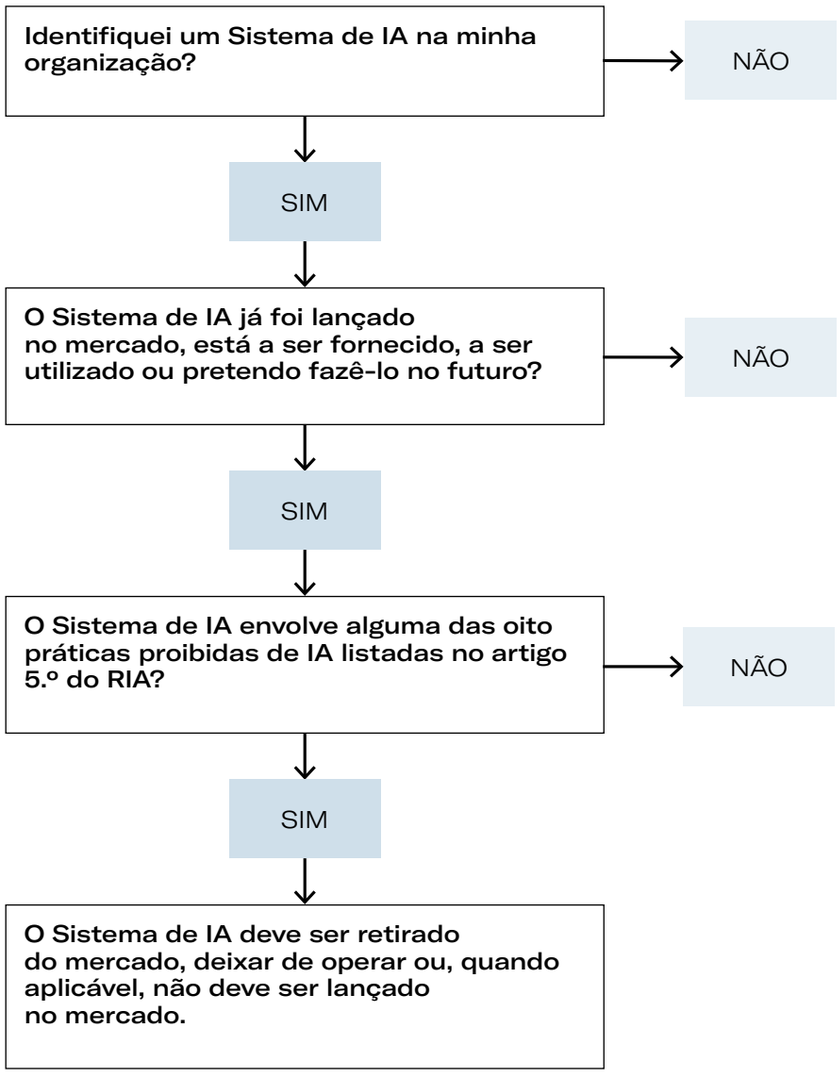
Após a entrada em vigor do Regulamento da Inteligência Artificial ('**RIA**'), as organizações já estão obrigadas ao cumprimento das regras sobre práticas proibidas de IA desde o dia 2 de fevereiro do presente ano.

A Comissão Europeia publicou agora orientações sobre práticas proibidas de IA[1] ('**Orientações**'). Vejamos algumas implicações práticas destas Orientações para as organizações.

Qual o papel das Orientações?

As Orientações da Comissão ajudam a concretizar e a identificar as práticas que são consideradas proibidas no domínio dos Sistemas de IA[2] por representarem riscos inaceitáveis[3] para a segurança e os direitos fundamentais dos cidadãos.

Mais concretamente, o RIA elenca, no seu artigo 5.º, oito práticas proibidas de IA. É sobre estas práticas que a Comissão Europeia emitiu, recentemente, as suas diretrizes. Resta às organizações, num trabalho conjunto entre equipas de IT e equipas jurídicas, conseguir identificar os Sistemas de IA[4] que são proibidos e que, por isso, devem deixar de operar, não ser lançados e/ou retirados do mercado.



1 Brussels, 4.2.2025, C(2025) 884 final, COMMUNICATION TO THE COMMISSION, Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act).

2 Os sistemas de IA podem ser confundidos com outras tecnologias como o software tradicional, algoritmos determinísticos ou mesmo assistentes virtuais ou chatbots pré-programados baseados em scripts.

3 O RIA define quatro níveis de risco (risco mínimo; risco específico em matéria de transparência; risco elevado; risco inaceitável). Os sistemas de IA estão sujeitos a regras diferentes consoante o grau de risco identificado caso-a-caso. Quanto às práticas de IA de risco inaceitável, a regra é simples: a proibição. Ou seja, os sistemas de IA que comportam um nível de risco inaceitável para a segurança dos cidadãos e para os direitos fundamentais estão proibidos de serem lançados no mercado, estarem em serviço ou utilização.

4 Definição do que é um Sistema de IA: cfr. artigo 3.º, n.º1 do RIA e as Orientações sobre a definição de Sistema de Inteligência Artificial.

Quem deve estar atento às Orientações?

As Orientações são especialmente úteis para as organizações que desenvolvem, fornecem e/ou utilizam sistemas de IA, ainda que tenham sido lançados no mercado antes da data do início da aplicação das regras do RIA sobre práticas de proibidas de IA.

Além disso, embora não-vinculativas, serão, naturalmente, consideradas pelos tribunais e pelas autoridades de fiscalização.

Como identificar práticas proibidas de IA?

Após a identificação de um sistema de IA, as Organizações deverão perceber se o sistema de IA envolve práticas enquadradas na lista das oito práticas proibidas (cfr. artigo 5.º do RIA).

Para que as organizações estejam capazes de reconhecer práticas proibidas de IA e atuar adequadamente no cumprimento da legislação em vigor, a Comissão mencionou alguns exemplos, seja em função das técnicas que são utilizadas no seu desenvolvimento, seja em função dos efeitos e/ou danos que estas práticas possam causar.

A Comissão menciona alguns exemplos concretos de práticas proibidas de IA, seja pelas técnicas que são utilizadas no seu desenvolvimento, seja pelos efeitos e/ou danos que estas possam causar.

▲ VOLTAR AO INÍCIO

De forma simples, destacamos alguns:

LISTA DE PROIBIÇÕES (CFR. ARTIGO 5.º DO RIA):	PALAVRAS, CONCEITOS E COMPONENTES – CHAVE DAS PROIBIÇÕES:	SÃO PROIBIDOS SISTEMAS DE IA QUE UTILIZEM AS SEGUINTES PRÁTICAS (CFR. EXEMPLOS DAS ORIENTAÇÕES):	SALVO NOS SEGUINTES CASOS:
Técnicas de manipulação e técnicas enganosas [5.º, n.º 1, al. a) do RIA]	○ Técnicas subliminares ○ Técnicas propositadamente manipuladoras ○ Técnicas enganosas ○ Distorção substancial do comportamento	○ Utilização de técnicas subliminares (ex. mensagens subliminares visuais ou auditivas, estímulos subvisuais ou subaudíveis, desvio da atenção, manipulação da perceção do tempo causando a impaciência e dependência do utilizador, brain spyware). ○ Utilização de técnicas propositadamente manipuladoras (ex. sistema de IA que utiliza som ou imagens de fundo provocando alterações de humor, aumentando a ansiedade e o sofrimento do utilizador). ○ Utilização de técnicas enganosas como a apresentação de informações falsas e/ou enganosas (ex. chatbot de IA que se faz passar por um humano causando burlas e danos significativos).	○ Sistema que apresenta um comportamento manipulador de forma meramente incidental e desde que tenham sido tomadas medidas preventivas e atenuantes adequadas no caso de ser razoavelmente provável a ocorrência de danos (ex. sistema de IA generativa que alucina[5] e, por isso, apresenta informações enganosas). ○ Sistema com recurso a técnicas de persuasão lícitas cujos objetivos de funcionamento são transparentes e respeitam a autonomia das pessoas (ex. um sistema que analisa emoções para melhorar as interações com os clientes e prestar-lhes apoio) (fora do âmbito da proibição).
Exploração prejudicial de vulnerabilidades [5.º, n.º 1, al. b) do RIA]	○ Exploração ○ Vulnerabilidades ○ Crianças ○ Idosos ○ Pessoas com deficiência ○ Situação socio-económica ○ Distorção substancial do comportamento ○ Razoavelmente suscetível de causar danos significativos	○ Exploração de vulnerabilidades de um individuo ou grupo de indivíduos com o objetivo ou o efeito de distorcer o comportamento desse individuo ou grupo, com a suscetibilidade de gerar danos significativos (ex. chatbot de IA que visa grupos desfavorecidos incitando-os a cometer atos de violência; ou um sistema de IA que explora vulnerabilidades cognitivas de pessoas idosas direcionando especialmente para estas pessoas tratamentos médicos mais dispendiosos).	○ Sistema que embora utilize a persuasão não utiliza a manipulação (ex. sistemas de IA que apoiam crianças na aprendizagem escolar e nos jogos) (fora do âmbito da proibição).

5 Termo utilizado para descrever uma falha técnica num sistema de IA generativa.

LISTA DE PROIBIÇÕES (CFR. ARTIGO 5.º DO RIA):	PALAVRAS, CONCEITOS E COMPONENTES – CHAVE DAS PROIBIÇÕES:	SÃO PROIBIDOS SISTEMAS DE IA QUE UTILIZEM AS SEGUINTES PRÁTICAS (CFR. EXEMPLOS DAS ORIENTAÇÕES):	SALVO NOS SEGUINTES CASOS:
Avaliação ou classificação social (scoring) [5.º, n.º 1, al. c) do RIA]	○ Comportamento social ○ Características pessoais ou de personalidade ○ Definição de perfis ○ Tratamento prejudicial ou desfavorável ○ Tratamento injustificado ou desproporcionado	○ Desenvolvimento e utilização de um sistema para avaliar e classificar indivíduos com base em comportamentos e/ou características pessoais ao longo do tempo podendo conduzir a resultados desfavoráveis ou prejudiciais (ex. sistema de IA pelas finanças que recorre aos dados das declarações fiscais para selecionar indivíduos específicos para inspeção; ou um banco que utiliza um sistema de IA para determinar a capacidade de crédito e decidir se determinado individuo deve ou não obter um crédito à habitação com base em características pessoais não relacionadas). ○ Desenvolvimento e utilização de um sistema para avaliar e classificar indivíduos com base em comportamentos e/ou características pessoais ao longo do tempo podendo conduzir a resultados injustificados ou desproporcionais (ex. uma autoridade pública que utiliza um sistema de IA para controlar fraude no processo de atribuição de bolsas de estudo considerando o estatuto familiar, o nível de estudos dos pais, como fator distintivo do nível de fraude).	○ Sistema de classificação social de tratamento justificado e proporcional (ex. sistema de IA, utilizado por uma empresa com interesse legítimo para o efeito, de deteção de fraudes financeiras se a avaliação se basear em dados relevantes como o comportamento transacional) (fora do âmbito da proibição).
Avaliação e previsão do risco de infração penal [5.º, n.º 1, al. d) do RIA]	○ Avaliação do risco ○ Previsão da criminalidade ○ Definição de perfis ○ Traços e características de personalidade ○ Avaliação humana ○ Factos objetivos e verificáveis relacionados com a atividade criminosa	○ Avaliação de risco para prever a probabilidade de uma pessoa cometer uma infração penal (ex. algoritmo que a partir da nacionalidade e etnia é capaz de prever a criminalidade). ○ Avaliação baseada exclusivamente na definição de perfis e/ou na avaliação de traços de personalidade, sem integrar dados objetivos e verificáveis relativos a atividade criminal (ex. ferramentas que preveem crimes a partir de traços de personalidade como a idade ou es(tado civil).	○ Sistema de IA utilizado para apoiar a avaliação do envolvimento de uma pessoa na atividade criminosa que se baseia em factos objetivos e verificáveis diretamente ligados à atividade criminosa (ex. utilização de um sistema de IA para a definição de perfis e categorização de comportamentos perigosos razoavelmente suspeitos numa multidão, que indiquem que alguém se está a preparar e é suscetível de cometer um crime) (fora do âmbito da proibição).
Recolha aleatória de imagens faciais a partir da Internet / CCTV para desenvolver bases de dados de reconhecimento facial [5.º, n.º 1, al. e) do RIA]	○ Reconhecimento facial ○ Recolha não direcionada ○ Bases de dados ○ Imagens faciais ○ Internet ○ CCTV	○ Recolha de imagens faciais de forma indiscriminada a partir da Internet ou de imagens de CCTV, permitindo a recolha de tanta informação quanto possível, sem foco num determinado grupo ou indivíduo (ex. um software de reconhecimento facial treinado a partir de imagens das redes sociais). ○ Criação ou expansão de bases de dados de reconhecimento facial capaz de corresponder um rosto humano de uma imagem/vídeo a um rosto da base de dados. (ex. um software que analisa imagens de segurança para criar bases de dados sem consentimento dos indivíduos).	○ Sistemas com recurso a recolha não direcionada de dados biométricos que não sejam a partir de imagens faciais (ex. amostras de voz) (fora do âmbito da proibição).

▲ VOLTAR AO INÍCIO

LISTA DE PROIBIÇÕES (CFR. ARTIGO 5.º DO RIA):	PALAVRAS, CONCEITOS E COMPONENTES – CHAVE DAS PROIBIÇÕES:	SÃO PROIBIDOS SISTEMAS DE IA QUE UTILIZEM AS SEGUINTES PRÁTICAS (CFR. EXEMPLOS DAS ORIENTAÇÕES):	SALVO NOS SEGUINTES CASOS:
Reconhecimento de emoções [5.º, n.º 1, al. f) do RIA]	○ Inferência de emoções ou intenções ○ Dados biométricos ○ Local de trabalho ○ Instituições de Ensino ○ Razões médicas ○ Razões de segurança	○ Deteção ou inferência de emoções e intenções em ambientes laborais ou instituições de ensino, baseados em dados biométricos. (ex. sistema de IA que monitoriza expressões e micro-expressões faciais para avaliar o nível de envolvimento e interesse dos alunos em sala de aula).	○ Se um sistema é capaz de detetar ou inferir emoções e intenções no local de trabalho ou instituições de ensino, baseado em dados biométricos por razões médicas ou de segurança em casos de necessidade explícita nos termos da legislação laboral (ex. sistema de IA utilizado para medir os níveis de stress dos trabalhadores quando o stress representa um perigo para os trabalhadores de máquinas perigosas).
Categorização biométrica [5.º, n.º 1, al. g) do RIA]	○ Categorização ○ Dados biométricos ○ Classificação ○ Deduzir ou inferir ○ Características sensíveis	○ Utilização de dados biométricos para inferir características sensíveis com vista a atribuir rótulos que possam resultar num tratamento discriminatório. (ex. sistema de IA que classifica indivíduos com base no tom de pele associando estes perfis a estatísticas de criminalidade). ○ Categorização individual de uma pessoa com base nos seus dados biométricos. (ex. sistema de IA que tem como objetivo deduzir a etnia de um indivíduo a partir da sua a orientação religiosa, voz, tatuagens ou características faciais).	○ Se rotulam ou fazem uma filtragem de dados biométricos que foram recolhidos em conformidade com a legislação aplicável ou efetuam categorização por razões técnicas objetivas para identificar e prevenir riscos específicos (ex. sistema de IA que categoriza pacientes a partir de imagens de acordo com a cor de pele para diagnosticar problemas oncológicos).
Identificação biométrica remota ("RBI") em tempo real [5.º, n.º 1, al. h) do RIA]	○ Identificação ○ Sistema de RBI ○ Dados biométricos ○ Em tempo real ○ Remoto ○ Espaços acessíveis ao público ○ Para efeitos de aplicação da lei	○ Identificação biométrica remota em tempo real (“RBI”) em espaços acessíveis ao público para fins de aplicação da lei (ex. utilização de sistemas de RBI em tempo real pela polícia para identificar um ladrão de uma loja e comparar as suas imagens faciais com outros os dados criminais da polícia judiciária).	○ Sistemas de RBI para fins específicos como a procura de pessoas desaparecidas, ou prevenção de ataques terroristas (ex. monitorização genética para prevenção de ameaças iminentes).

Por que devem as organizações conhecer estas Orientações?

Muito embora, repete-se, estas Orientações não sejam vinculativas, devem ser consideradas. Enquanto instrumento de soft law, trata-se de um importante apoio interpretativo para entidades públicas e privadas, como para as autoridades de fiscalização e para os tribunais na identificação daquelas que são consideradas práticas proibidas de IA.

As organizações deverão auditar e, se necessário, rever os sistemas de IA colocados no mercado, em serviço, ou que de momento, estejam a ser utilizados, bem como avaliar aqueles sistemas que estão a ser desenvolvidos e/ou implantados ou que, de futuro, pretendam desenvolver.

Relembremos que incumprimento das proibições previstas no artigo 5.º e agora concretizadas pelas Orientações da Comissão constitui uma infração grave ficando os responsáveis sujeitos à coima mais elevada – isto é, até 35.000.000,00EUR ou, se o infrator for uma empresa, até 7% do seu volume de negócios atual total a nível mundial relativo ao exercício financeiro do ano anterior, consoante o valor mais elevado.

Recorde-se que as organizações deverão fazer uma verificação anual da lista de práticas proibidas, que poderá ser alterada anualmente[6].

Enquanto instrumento de soft law, trata-se de um importante apoio interpretativo para entidades públicas e privadas, como para as autoridades de fiscalização e para os tribunais na identificação daquelas que são consideradas práticas proibidas de IA.

6 Artigo 112 do RIA.

▲ VOLTAR AO INÍCIO

Sobre a PLMJ

→ Quem somos

"Strengths include geographical presence, immediate response, deep knowledge and experience in market concepts and themes."

CLIENT REFERENCE FROM
THE LEGAL 500

Sobre a área de Tecnologia, Media e Telecomunicações

→ O que fazemos

KEY CONTACTS

Pedro Lomba

Sócio e coordenador da área de Tecnologia, Media e Telecomunicações.

pedro.lomba@plmj.pt

Benedita Cunha Pinto

Associada na área de Tecnologia, Media e Telecomunicações

benedita.cunhapinto@plmj.pt

Fevereiro 2025