DOI: 10.33242/rbdc.2024.04.007

IA, RISCOS E RESPONSABILIDADE – UMA REFLEXÃO EM TORNO DO REGULAMENTO IA E DO PROJETO DE LEI BRASILEIRO Nº 2338, DE 2023[1]

AI, RISKS AND RESPONSIBILITY – A REFLECTION ON THE AI ACT AND THE BRAZILIAN PROJECT OF LAW NO. 2338, OF 2023

Mafalda Miranda Barbosa

Universidade de Coimbra, Instituto Jurídico da Faculdade de Direito da Universidade de Coimbra/University of Coimbra Institute for Legal Research, Faculdade de Direito da Universidade de Coimbra. Professora Associada com Agregação. Orcid: https://orcid.org/0000-0003-0578-4249 *E-mail*: mcnmb@fd.uc.pt

Resumo: O regulamento IA recentemente aprovado, ao consagrar uma série de deveres que impendem sobre prestadores que coloquem no mercado ou coloquem em serviço sistemas de IA ou que coloquem no mercado modelos de IA de finalidade geral no território da União, sobre responsáveis pela implantação de sistemas de IA, sobre importadores e exportadores de IA, fabricantes de produtos que coloquem no mercado conjuntamente com o seu produto um sistema de IA e sob o seu nome ou marca e mandatários dos prestadores que não estejam estabelecidos na EU, pode ter implicações importantes em sede de responsabilidade civil. Nas páginas que se seguem, depois de uma breve análise do regulamento, procuraremos tecer algumas considerações acerca desse impacto, articulando o diploma com a proposta de diretiva em matéria de responsabilidade pela IA.

Palavras-chave: IA. Responsabilidade civil. Responsabilidade subjetiva e objetiva. *AI act*.

Abstract: The recently approved AI act, by establishing a series of duties that apply to providers who place AI systems on the market or put them into service, or who place general-purpose AI models on the market within the territory of the Union, to those responsible for the deployment of AI systems, to importers and exporters of AI, to manufacturers of products that place an AI system on the market together with their product under their name or brand, and to representatives of providers not established in the EU, may have significant implications in terms of civil liability. In the following pages, after a brief analysis of the regulation, we will attempt to make some considerations about this impact, linking the regulation with the proposed directive on AI liability.

[1] O texto que agora se publica foi escrito na sequência do convite para participar na conferência "Responsabilidade Civil e Inteligência Artificial: Debate na União Europeia", promovida pelo Grupo de Estudos Direito e Tecnologia (*TechLaw*) do Instituto de Estudos Avançados Polo Ribeirão Preto da USP, em colaboração com o Centro de Estudos Avançados de Direito e Inovação da USP, o Center for Artificial Intelligence e a iniciativa UAI (Understanding Artificial Intelligence), do IEA, no dia 23 de maio de 2024.

Keywords: AI. Tort law. Fault liability. Strict liability. AI Act.

1 Primeiras palavras

São imensos os desafios que a inteligência artificial nos coloca, não só do ponto de vista jurídico, mas também do ponto de vista filosófico e humano. Não pretendemos, nas páginas que se seguem, refletir sobre o fenómeno em termos mais genéricos,[2] mas tão-só pensar nas implicações que o surgimento de sistemas autónomos pode ter ao nível da responsabilidade civil. O tratamento do tema não é, para nós, novidade. Contudo, decorrido que foi algum tempo desde os nossos primeiros escritos na matéria, os diversos ordenamentos jurídicos foram dando passos firmes no sentido da definição do que poderá vir a ser a responsabilidade pela IA. Nessa medida, mais do que explicitar por que razão as características da IA podem tornar inoperantes os sistemas de responsabilidade clássicos e mais do que, em termos teóricos, ponderar diversas vias de solução para os problemas, procuraremos analisar crítico-reflexivamente as soluções que se desenham no horizonte europeu.

2 As particularidades da IA

Os sistemas de IA, cada vez mais sofisticados, acabam por condicionar, fruto das suas peculiaridades, as respostas que os ordenamentos jurídicos pensaram ao longo dos tempos para os problemas de danos causados por outro sujeito.

Entre as diversas características que podem arrastar problemas está, desde logo, a *conectividade*. Os sistemas de inteligência artificial funcionam em rede, o que faz com que, não só a conexão se possa perder, como com que os sistemas fiquem expostos a atos de pirataria.[3] Como bem se compreenderá esta peculiaridade

[2] Cf., para tanto, mas ainda assim de forma muito fragmentária, Mafalda Miranda Barbosa, *Inteligência artificial. Entre a utopia e a distopia. Alguns problemas jurídicos*, Gestlegal, 2021.

[3] Cf. COM (2020) 64 final, 6, dando dois exemplos particularmente interessantes. Um primeiro caso extrai-se de uma notificação feita pela Islândia no âmbito do Sistema de Troca Rápida de Informação da UE, relativa a um relógio inteligente para crianças, que poderia ser utilizado como um meio de acesso a tais crianças, graças a um sistema de localização. Um segundo caso referente a uma notificação apresentada pela Alemanha, no que respeita a um automóvel de passageiros, na medida em que o software do rádio de um veículo apresentava falhas de segurança que permitiriam o acesso não autorizado de terceiros aos

arrasta consigo problemas complexos do ponto de vista do estabelecimento de um nexo de causalidade, ou, para usarmos uma linguagem mais próxima da nossa visão sobre o requisito, de um nexo de imputação objetiva.

Por outro lado, os sistemas são caracterizados pela sua *autonomia*. As capacidades de autoaprendizagem da máquina e a possibilidade de decidir para além da programação inicial podem determinar alterações significativas no seu funcionamento, durante o período útil de vida, ao mesmo tempo que podem ser geradoras de danos, até porque, se os sistemas de inteligência artificial se mostram aptos a cumular e computar biliões de dados, a uma velocidade que não está ao alcance do ser humano, também é verdade que a inteligência não ultrapassa, a este propósito, uma ideia de *agency*¸ não se mostrando a máquina capaz de aceder a uma dimensão semântica, isto é, ao significado dos signos que mobiliza, de tal forma que se tornam propensos a gerar distorções. Mas, nem por isso se pode, *prima facie*, falar de um comportamento culposo do utilizador do sistema; e o juízo imputacional torna-se, também a este nível, complexo.

Acresce que a inteligência artificial se mostra particularmente dependente dos dados, que, se não forem exatos, condicionam o funcionamento do sistema e os resultados a que com base nele se pode chegar. A segurança do sistema fica, assim, dependente de circunstâncias exteriores que podem não ser totalmente controladas pelo programador ou pelo utilizador. Mais uma vez se percebe por que motivo pode ser tão difícil o estabelecimento de um nexo de imputação objetiva: o sistema é, na verdade, um ecossistema, não se conseguindo discernir, na maioria das situações, qual a verdadeira causa da deturpação que a aplicação de IA gerou ou do dano que fez emergir. Repare-se, aliás, que muitas vezes a corrupção dos dados pode ter tido origem num sistema de IA. As dificuldades agravam-se pelo facto de a nova realidade digital implicar que muitas aplicações possam ser descarregadas em produtos, com impacto não negligenciável, bem como uma cooperação estreita entre diversos agentes económicos e os utilizadores, que podem alterar o sistema. As dificuldades aumentam com a opacidade característica do funcionamento dos algoritmos.

Os autores costumam falar, a este propósito, de uma tripla opacidade:[4] opacidade corporativa, deliberadamente gerada como forma de resguardar os segredos

sistemas de controlo do veículo, podendo determinar a ocorrência de acidentes, se tais vulnerabilidades fossem exploradas por um terceiro. Os exemplos foram oferecidos no âmbito de uma ponderação relativa à eventual necessidade de reponderar os problemas relacionados com a segurança dos produtos e uma obrigação geral de segurança.

4 Cf. Jenna, Burrell, "How the machine thinks: understanding opacity in machine learning algorithms", 2015, http://ssrn.com/abstract=2660674, acesso em 24-6-2021; F. Pasquale, *The black box society*, Harvard University Press, 2015, 79 s.; Mariana Marques Rielli, "Críticas ao ideal de transparência como solução para a opacidade de sistemas algorítmicos", *Direito Digital e Inteligência Artificial: Diálogos entre Brasil*

de negócios das empresas que desenvolvem os algoritmos; opacidade cognitiva, resultante da incapacidade que as pessoas em geral (e o titular dos dados em especial) têm de entender o funcionamento do algoritmo e de perceber a linguagem que o mesmo utiliza;[5] e opacidade técnica, inerente ao recurso ao *deep learning*, inviabilizador da explicitação do percurso decisório do *software*, mesmo por parte dos seus programadores.[6]

3 Entre a responsabilidade objetiva e a responsabilidade subjetiva

Afastada que seja a hipótese de responsabilização direta dos sistemas de inteligência artificial, a implicar a subjetivação destes entes,[7] o debate ao nível europeu, no tocante à responsabilidade civil por danos causados por sistemas autónomos, oscila entre a defesa da consagração de uma hipótese de responsabilidade objetiva e a manutenção da regra da responsabilidade assente na culpa.

A responsabilidade objetiva partiria da ideia de que os sistemas em questão envolvem uma grande propensão para causar danos, podendo alguns atingir magnitude considerável e podendo resultar da lesão de bens jurídicos fundamentais, a que se associaria uma outra – a de que, com base na culpa, é praticamente impossível responsabilizar o programador ou o utilizador, tendo em conta que as lesões podem ser geradas a partir da atuação autonómica da máquina.

No horizonte europeu, a primeira proposta de regulamentação da responsabilidade civil pela IA orientava-se por estas considerações de dupla natureza.

Nos termos da Resolução do Parlamento Europeu de 20 de outubro de 2020 (Resolução do Parlamento Europeu 2020/2014 (INL), que vinha propor um Regulamento em matéria de responsabilidade civil pela IA e que cobria as hipóteses de dano à vida, à saúde, à integridade física de uma pessoa singular, à propriedade de uma pessoa singular ou coletiva ou de ocorrência de uma lesão imaterial significativa que cause um dano económico (*significant immaterial harm resulting in a verifiable economic loss*), previa-se que a responsabilidade devia ser assacada ao operador, de acordo com um esquema dúplice. Incluía-se, aí, quer o *frontend*, quer

e Europa (coord. Mafalda Miranda Barbosa/Filipe Braga Netto/Michael César Silva/José Luiz de Moura Faleiros Júnior), Editora Foco, 2021, 440.

5 Burrell, "How the machine thinks: understanding opacity in machine learning algorithms"; Rielli, "Críticas ao ideal de transparência como solução para a opacidade de sistemas algorítmicos", 443.

6 Burrell, "How the machine thinks: understanding opacity in machine learning algorithms"; Rielli, "Críticas ao ideal de transparência como solução para a opacidade de sistemas algorítmicos", 443.

7 Para uma aprofundada reflexão acerca da questão, cf. Mafalda Miranda Barbosa, *Inteligência artificial. Entre a utopia e a distopia*, 7 s.

o *backend operator*. O primeiro surgia definido como a pessoa singular ou coletiva que exercia um qualquer nível de controlo sobre um risco ligado ao funcionamento de um sistema de inteligência artificial e beneficiava com tal operação; o *backend operator*, por seu turno, era a pessoa singular ou coletiva que, de forma contínua, definia os recursos tecnológicos e providenciava o acesso aos dados e um serviço de suporte necessário, de tal modo que também exercia um nível de controlo sobre o risco ligado ao funcionamento do sistema de inteligência artificial.

De acordo com o artigo 4º/1, o operador seria objetivamente responsável por qualquer dano que fosse causado por uma atividade física ou virtual ou por qualquer processo que envolvesse inteligência artificial, desde que estivesse em causa um sistema de alto risco (*high-risk AI system*), não podendo exonerar-se pela invocação de que atuou diligentemente ou que o dano ou lesão teriam sido causados por uma atividade autónoma ou processo conduzido por um sistema de inteligência artificial. A exclusão da responsabilidade ocorria unicamente por via da invocação de força maior.

Os sistemas de alto risco a que se aludem eram definidos, nos termos do artigo 3º/c), em função da potencialidade de um sistema de inteligência artificial causar danos a uma ou mais pessoas de maneira aleatória e de forma que ultrapassasse o que era razoavelmente expectável, desde que constasse do anexo de regulamento. A potencialidade a que se aludia dependia da gravidade do possível dano ou lesão, do grau de autonomia do sistema de decisão, da probabilidade de materialização do risco e do contexto de utilização do sistema de inteligência artificial.

Em todas as outras situações que envolvessem a utilização da inteligência artificial e não se configurassem como sistemas de alto risco (*other AI-systems*), a responsabilidade do operador baseava-se na culpa. Podia, assim, nos termos do artigo 8.º/2, excluir-se a responsabilidade do operador se este provasse a ausência de culpa, designadamente se provasse que o sistema de inteligência artificial tinha sido ativado sem o seu conhecimento, apesar de terem sido adotadas todas as medidas razoáveis e necessárias para evitar tal ativação; que tinha sido observada a diligência devida na execução de determinados processos, designadamente na seleção do adequado sistema de inteligência artificial para o desempenho da função, no momento em que o sistema começou a operar, na monitorização das atividades, e na regular atualização do software. Do mesmo modo, excluir-se-ia a responsabilidade com base na força maior. Mas não se excluiria a responsabilidade com base na ideia de que a lesão havia sido causada por uma atividade autónoma ou processo levado a cabo pelo processo de inteligência artificial. A responsabilidade subjetiva a que assim éramos conduzidos surgia agravada e implicava uma presunção de culpa.

Tal agravamento era notório, também, pelo facto de o operador responder pelos danos causados pela interferência de um terceiro no sistema de inteligência artificial, pela modificação do seu modo de funcionamento ou dos seus feitos, desde que aquele terceiro não fosse identificado ou não tivesse possibilidade de pagar a indemnização. Tratava-se de uma hipótese de responsabilidade objetiva, por facto alheio, dependente de requisitos estritos, que obviamente não afastava a possibilidade de responsabilização, em geral, do operador nas hipóteses de lesão causada imediatamente por um terceiro, quando aquela interferência tivesse sido potenciada pela violação de deveres de cuidado por parte do referido operador.

A Resolução do Parlamento Europeu 2020/2014 (INL) constituiu um importante passo na definição de um quadro normativo em matéria de responsabilidade civil por danos causados pela inteligência artificial. Vários foram os aspetos aí consagrados que mereciam aplauso.

Em primeiro lugar, louva-se o abandono de uma ideia inicialmente avançada de responsabilização direta do ente dotado de inteligência artificial/algoritmo.

Do mesmo modo, era de saudar que a previsão de seguros obrigatórios surgisse paredes-meias com a definição de critérios de imputação baseados no risco. Repare-se, aliás, que a responsabilidade objetiva que se parecia definir no horizonte não se quedava numa pura responsabilidade pela causalidade – que, aliás, resulta muitas vezes problemática a este nível –, mas, excecionada a previsão excecionalíssima de uma responsabilidade-garantia, no quadro da responsabilidade por culpa, antes se configurava como uma responsabilidade assente no risco, que era ponderado e alocado em função das especificidades de cada sistema de inteligência artificial.

O risco a que se aludia era compreendido em termos não unívocos. Por um lado, lidava-se com o risco inerente a qualquer sistema de inteligência artificial, resultado da possibilidade de atuação autónoma, de interferência de terceiros, da necessária conectividade dos sistemas, com relevo não só para a fundamentação de base de todo o sistema, como para a resolução de alguns problemas relativos ao nexo de imputação, outrora tratados sob a égide da causalidade; por outro lado, lidava-se com o risco específico de determinadas atividades e com o perigo de causação de um dano de proporções significativas, a justificar a discriminação entre sistemas de inteligência artificial.

Questionáveis eram, porém, algumas formulações dogmáticas pouco rigorosas. Assim, em matéria de ilicitude, atuante quando esteja em causa uma hipótese de responsabilidade subjetiva, mas também em matéria de causalidade, relativamente à qual o diploma era (quase) omissivo, não obstante alguns dados relevantes que oferecia que podiam ser compreendidos à luz da transmutação de uma perspetiva estritamente causal – absolutamente improcedente (em geral e neste domínio) – numa perspetiva imputacional.

Prevendo-se uma responsabilidade objetiva, a prescindir da culpa, havia nódulos problemáticos que não podiam deixar de ser tidos em conta. Assim, a responsabilidade independente de culpa, estando normalmente associada a tetos de indemnização, poderia não permitir a compensação de todos os danos; por outro lado, não garantiria a realização de todas as outras finalidades da responsabilidade civil, na medida em que abdica do requisito culpa, ao qual não pode deixar de estar associado um valor axiológico relevante, apesar do sentido imputacional com que vinha pensada. Na verdade, temos vindo a defender que não é absolutamente indiferente a previsão de uma hipótese de responsabilidade objetiva ou subjetiva, em termos valorativos, atendendo ao fundamento último da juridicidade. Mas isso não significa que, em certos domínios, atentas as idiossincrasias da atividade concretamente em causa, não seja meritória a sua previsão.

Pareceu-nos ser esse o caso ao nível da responsabilidade pelos danos causados pela IA. Sem que, contudo, essa pareça ser a opção mais recente ao nível europeu.

Em 28 de setembro de 2022, a Comissão adotou uma Proposta de Diretiva do Parlamento Europeu e do Conselho relativa à adaptação das regras da responsabilidade civil extracontratual à inteligência artificial (Diretiva Responsabilidade da IA), na sequência da Resolução 2020/2014 (INL) do Parlamento Europeu.

A Comissão entendeu que seria importante ter em conta as diferenças entre as tradições jurídicas nacionais e o facto de os tipos de produtos e serviços equipados com sistemas de IA suscetíveis de afetar o público em geral e pôr em risco direitos jurídicos importantes, como o direito à vida, à saúde e de propriedade, e, por conseguinte, suscetíveis de estar sujeitos a um regime de responsabilidade objetiva, ainda não estarem amplamente disponíveis no mercado, razão pela qual, em detrimento daquela que tinha sido a ideia central na proposta de resolução do Parlamento Europeu, optou preferencialmente por um modelo de responsabilidade assente na culpa.

Nessa medida, a diretiva é aplicável a ações de indemnização de direito civil extracontratual por danos causados por um sistema de IA, sempre que tais ações sejam intentadas ao abrigo de regimes de responsabilidade culposa, o que significa, como a própria Comissão esclarece, que as medidas previstas podem inserir-se nos sistemas de responsabilidade civil existentes sem com eles entrarem em conflito.

Mais do que isso, sendo esta uma diretiva de harmonização mínima, por um lado, e sendo o seu âmbito de aplicação consideravelmente circunscrito, nada impede que cada ordenamento jurídico dos diversos Estados-membros venha prever outras regras no que respeita à prova dos requisitos da responsabilidade ou mesmo no que respeita à tipificação de hipóteses de responsabilidade pelo risco.

Aliás, a responsabilidade que assim se desenha surge paredes-meias com a alteração da disciplina da responsabilidade do produtor ao nível europeu.

A responsabilidade subjetiva, situando-se, em termos de opção de fundo, num patamar axiológico mais elevado, por responder mais completamente às exigências ditadas pela ideia do direito enquanto direito, não deixa de colocar, como sublinhado *supra*, dificuldades a este nível como consequência necessária das características próprias da IA.

Nessa medida, torna-se fundamental a consagração de especiais deveres que vinculem os diversos sujeitos envolvidos na utilização de sistemas autónomos. Foi essa a opção do legislador europeu. Ao aprovar o Regulamento IA, consagrou amplas esferas de responsabilidade (no sentido da *role responsibility*) que, a serem postas em causa, facilitarão o juízo imputacional imprescindível para a afirmação da *liability*.

Ao nível comunitário, a grande preocupação parece ser a de melhorar o funcionamento do mercado, aumentando para isso a confiança das pessoas na utilização de sistemas de IA. A lógica não é tanto repressiva, mas preventiva, o que não quer dizer que não se facilite, em termos dogmáticos, a articulação e prova dos diversos requisitos delituais. Este é, aliás, como veremos, o ponto chave da proposta de Diretiva. Tal facilitação é, porém, potenciada pelos níveis de prevenção a que somos conduzidos.

Ou seja, a responsabilidade civil não pode, portanto, senão ser pensada a este nível em articulação com as regras ditadas pelo novo regulamento.

4 Regulamento IA

4.1 O âmbito de aplicação

O regulamento IA aplica-se a todos os sistemas de inteligência artificial, embora, como veremos, se estabeleçam discriminações entre eles, em função do risco que comportam. Por sistema de IA entende-se um sistema baseado em máquinas concebido para funcionar com níveis de autonomia variáveis e que pode apresentar capacidade de adaptação após a implantação e que com base nos dados de entrada que recebe infere de forma a gerar resultados.

Em termos subjetivos, aplica-se a:

– prestadores que coloquem no mercado ou coloquem em serviço sistemas de IA ou que coloquem no mercado modelos de IA de finalidade geral no território da União, independentemente de estarem fisicamente presentes ou estabelecidos na União ou num país terceiro, sendo definido

o prestador como a pessoa que "desenvolve ou manda desenvolver um sistema de IA ou um modelo de IA de finalidade geral e o coloque no mercado ou coloque o sistema de IA em serviço sob o seu próprio nome ou a sua própria marca, a título oneroso ou gratuito";

- responsáveis pela implantação de sistemas de IA que tenham o seu local de estabelecimento na UE ou num país terceiro, se o resultado produzido pelo sistema for utilizado na União, entendendo-se por responsável pela implantação aquele que "utiliza o sistema de IA sob a sua autoridade, salvo se o sistema de IA for usado no âmbito de uma atividade pessoal de caráter não profissional";
- importadores e exportadores de IA;
- fabricantes de produtos que coloquem no mercado conjuntamente com o seu produto um sistema de IA e sob o seu nome ou marca;
- mandatários dos prestadores que não estejam estabelecidos na UE;
- pessoas afetadas localizadas na UE.

O operador do sistema é compreendido em termos muito latos, de modo a abarcar o prestador, o fabricante de produtos, o responsável pela implantação, o mandatário, o importador e o distribuidor.

Por seu turno, a colocação no mercado traduz-se na primeira disponibilização de um sistema de IA ou de um modelo de IA de finalidade geral no mercado da União; entendendo-se por disponibilização no mercado o fornecimento de um sistema de IA ou de um modelo de IA de finalidade geral para distribuição ou utilização no mercado da União; e por colocação em serviço o fornecimento, diretamente ao responsável pela implantação ou para utilização própria, de um sistema de IA para a primeira utilização na União.

Prevêem-se algumas exclusões em termos objetivos. Assim, o regulamento não se aplica:

- quando os sistemas de IA forem usados no âmbito de uma atividade puramente pessoal de caráter não profissional;
- a sistemas de IA que estejam em fase de testagem, embora a testagem em condições reais não seja abrangida pela exclusão;
- a sistemas de IA colocados em serviço exclusivamente para fins de investigação e desenvolvimento científico;
- a sistemas de IA lançados ao abrigo de licenças gratuitas e de código aberto, a menos que sejam colocados no mercado ou em serviço como sistemas de IA de risco elevado ou que sejam sistemas de IA abrangidos pelo âmbito de aplicação dos artigos 5º ou 50º.

4.2 Os diversos níveis de risco

A disciplina estabelecida pelo Regulamento estrutura-se em função de diversos níveis de risco, resultado da combinação da probabilidade de ocorrência de danos com a gravidade desses danos.

Desde logo, há determinados sistemas que são considerados de *risco inaceitável*, sendo absolutamente proibidos:

- a colocação no mercado, a colocação em serviço ou a utilização de um sistema de IA que empregue técnicas subliminares que contornem a consciência de uma pessoa, ou técnicas manifestamente manipuladoras ou enganadoras, com o objetivo ou o efeito de distorcer substancialmente o comportamento de uma pessoa ou de um grupo de pessoas prejudicando de forma considerável a sua capacidade de tomar uma decisão informada e levando, assim, a que tomem uma decisão que, caso contrário, não tomariam, de uma forma que cause ou seja razoavelmente suscetível de causar danos significativos a essa ou a outra pessoa, ou a um grupo de pessoas;
- colocação no mercado, a colocação em serviço ou a utilização de um sistema de IA que explore vulnerabilidades de uma pessoa singular ou de um grupo específico de pessoas devidas à sua idade, incapacidade ou situação socioeconómica específica, com o objetivo ou o efeito de distorcer substancialmente o comportamento dessa pessoa ou de uma pessoa pertencente a esse grupo de uma forma que cause ou seja razoavelmente suscetível de causar danos significativos a essa ou a outra pessoa;
- a colocação no mercado, a colocação em serviço ou a utilização de sistemas de IA para avaliação ou classificação de pessoas singulares ou grupos de pessoas durante um certo período com base no seu comportamento social ou em características de personalidade ou pessoais, conhecidas, inferidas ou previsíveis, em que a classificação social conduza a um tratamento prejudicial ou desfavorável de certas pessoas singulares ou grupos de pessoas em contextos sociais não relacionados com os contextos nos quais os dados foram originalmente gerados ou recolhidos ou a um tratamento prejudicial ou desfavorável de certas pessoas singulares ou grupos de pessoas que seja injustificado ou desproporcionado face ao seu comportamento social ou à gravidade do mesmo;
- a colocação no mercado, a colocação em serviço para esta finalidade específica ou a utilização de um sistema de IA para a realização de avaliações de risco de pessoas singulares a fim de avaliar ou prever o risco de uma pessoa singular cometer uma infração penal, com base exclusivamente na definição de perfis de uma pessoa singular ou na avaliação dos seus traços e características de personalidade, não se aplicando

esta proibição aos sistemas de IA utilizados para apoiar a avaliação humana do envolvimento de uma pessoa numa atividade criminosa, que já se baseia em factos objetivos e verificáveis diretamente ligados a uma atividade criminosa;

- a colocação no mercado, a colocação em serviço para esta finalidade específica ou a utilização de sistemas de IA que criam ou expandem bases de dados de reconhecimento facial através da recolha aleatória de imagens faciais a partir da Internet ou de imagens de televisão em circuito fechado;
- a colocação no mercado, a colocação em serviço para esta finalidade específica ou a utilização de sistemas de IA para inferir emoções de uma pessoa singular no local de trabalho e nas instituições de ensino, exceto nos casos em que o sistema de IA se destine a ser instalado ou introduzido no mercado por razões médicas ou de segurança;
- a colocação no mercado, a colocação em serviço para este fim específico, ou a utilização de sistemas de categorização biométrica que classifiquem individualmente as pessoas singulares com base nos seus dados biométricos para deduzir ou inferir a sua raça, opiniões políticas, filiação sindical, convicções religiosas ou filosóficas, vida sexual ou orientação sexual;
- a utilização de sistemas de identificação biométrica à distância em «tempo real» em espaços acessíveis ao público para efeitos de aplicação da lei, a menos e na medida em que essa utilização seja estritamente necessária para a busca seletiva de vítimas específicas, de rapto, tráfico de seres humanos ou exploração sexual de seres humanos, bem como a busca por pessoas desaparecidas; a prevenção de uma ameaça específica, substancial e iminente à vida ou à segurança física de pessoas singulares ou de uma ameaça real e atual ou real e previsível de um ataque terrorista; a localização ou identificação de uma pessoa suspeita de ter cometido uma infração penal, para efeitos da realização de uma investigação criminal, ou instauração de ação penal ou execução de uma sanção penal por alguma das infrações referidas no anexo II e puníveis no Estado-Membro em causa com pena ou medida de segurança privativa de liberdade de duração máxima não inferior a quatro anos.

Por seu turno, os *sistemas de IA de risco elevado* os sistemas destinados a ser usados como um componente de um produto ou os sistemas que sejam produtos e que estejam previstos no anexo I; os produtos cujo componente de segurança seja um sistema de IA ou os sistemas que sejam sujeitos a uma avaliação de conformidade por terceiros com vista à sua colocação em serviço, nos termos dos atos enumerados no anexo I; os sistemas constantes do anexo III, desde que

cumpram as especificações previstas no regulamento. Este elenco não é fixo, podendo ser alargado ou diminuído, segundo os critérios do artigo 7º.

Assim, um sistema de IA a que se refere o Anexo III não pode ser considerado de risco elevado se não representar um risco significativo de danos para a saúde, a segurança ou os direitos fundamentais das pessoas singulares, nomeadamente se não influenciarem de forma significativa o resultado da tomada de decisões. Mas, os sistemas de IA a que se refere o anexo III devem ser sempre considerados de risco elevado nos casos em que executarem a definição de perfis de pessoas singulares.

Prevê-se, ainda, que a qualquer momento a comissão possa atualizar a listagem do anexo III. Para tanto, é necessário que se preencham determinados requisitos: os sistemas de IA destinem-se a ser utilizados em qualquer um dos domínios enumerados no anexo III; e os sistemas de IA representem um risco de danos para a saúde e a segurança ou de repercussões negativas nos direitos fundamentais, e esse risco seja equivalente ou superior ao risco de danos ou repercussões negativas representado pelos sistemas de IA de risco elevado já referidos no anexo III. Entre os diversos critérios a ter em conta conta-se:

- a finalidade prevista do sistema de IA;
- o grau de utilização efetiva ou a probabilidade de utilização de um sistema de IA;
- a natureza e a quantidade dos dados tratados e utilizados pelo sistema de IA e, em particular, o facto de serem tratadas categorias especiais de dados pessoais;
- a medida em que o sistema de IA atua de forma autónoma e a possibilidade de um ser humano anular decisões ou recomendações que possam causar danos;
- a medida em que a utilização de um sistema de IA já tenha causado danos para a saúde e a segurança, tenha tido repercussões negativas nos direitos fundamentais ou tenha suscitado preocupações significativas quanto à probabilidade de esses danos ou essas repercussões negativas ocorrerem;
- a potencial dimensão desses danos ou dessas repercussões negativas, nomeadamente em termos de intensidade e de capacidade para afetar várias pessoas, ou para afetar de forma desproporcionada um determinado grupo de pessoas;
- a medida em que as pessoas que sofreram potenciais danos ou repercussões negativas dependem dos resultados produzidos por um sistema de IA, em especial se, por razões práticas ou jurídicas, não lhes for razoavelmente possível autoexcluir-se desse resultado;

– a medida em que existe um desequilíbrio em termos de poder ou em que as pessoas que sofreram potenciais danos ou repercussões negativas se encontram numa posição vulnerável em relação ao responsável pela implantação de um sistema de IA, em particular por motivos relacionados com o estatuto, a autoridade, o conhecimento, as circunstâncias económicas ou sociais, ou a idade;
– a medida em que os resultados produzidos com o envolvimento de um sistema de IA são facilmente corrigíveis ou reversíveis, tendo em conta as soluções técnicas disponíveis para os corrigir ou reverter, sendo que os resultados com uma repercussão negativa na saúde, na segurança ou nos direitos fundamentais não podem ser considerados como facilmente corrigíveis ou reversíveis;
– a magnitude e a probabilidade dos benefícios da implantação do sistema de IA para as pessoas, os grupos ou a sociedade em geral, incluindo possíveis melhorias na segurança dos produtos;
– a medida em que o direito da União em vigor prevê medidas de reparação eficazes em relação aos riscos representados por um sistema de IA, com exclusão de pedidos de indemnização; medidas eficazes para prevenir ou minimizar substancialmente esses riscos.

Do mesmo modo, a Comissão pode, a qualquer momento, deixar de qualificar um sistema como sendo de risco elevado, sempre que o sistema de IA de risco elevado em causa deixe de representar um risco significativo para os direitos fundamentais, a saúde ou a segurança, e a supressão não diminua o nível geral de proteção da saúde, da segurança e dos direitos fundamentais ao abrigo do direito da União.

Para além do risco elevado, prevê-se a existência de sistemas de risco moderado e limitado.

São, ademais, tratados de forma específica os sistemas de inteligência artificial de finalidade geral, isto é, aqueles que têm capacidade para servir para diversas finalidades, tanto para utilização direta, como para integração noutros sistemas de IA. Quanto a estes há que estabelecer uma linha divisória entre os que importam risco sistémico e os que não envolvem. Os primeiros são os que apresentam capacidades de alto impacto, avaliadas com base em ferramentas e metodologias técnicas apropriadas, incluindo indicadores e referências, ou que, com base em uma decisão da Comissão, *ex officio* ou após um alerta qualificado pelo painel científico, sejam vistos como modelos de IA que tenham capacidades ou impacto equivalentes àqueles. Esta linha divisória será fundamental para se determinarem os deveres que vinculam os prestadores destes modelos.

4.3 O risco elevado e as obrigações associadas

Aos sistemas de risco elevado está associado um conjunto mais exigente de deveres.

a) Requisitos de conceção e desenvolvimento

Tratando-se de um sistema de risco elevado, é deve ser criado, implantado, documentado e mantido um sistema de gestão de riscos, nos termos do artigo 9º, no âmbito do qual se têm de ter em conta apenas os riscos que possam ser razoavelmente atenuados ou eliminados aquando do desenvolvimento ou da conceção do sistema de IA de risco elevado ou por meio da prestação de informações técnicas adequadas.

Além disso, os sistemas de IA de risco elevado que utilizem técnicas que envolvam o treino de modelos com dados devem ser desenvolvidos com base em conjuntos de dados de treino, validação e teste que cumpram os critérios de qualidade previstos no artigo 10º. Mais concretamente, devem estar sujeitos a práticas de governação e gestão de dados adequadas à finalidade prevista do sistema de IA. E os conjuntos de dados de treino, validação e teste devem ser pertinentes, suficientemente representativos e, tanto quanto possível, isentos de erros e completos, tendo em conta a finalidade prevista. Devem, ainda, ter as propriedades estatísticas adequadas, nomeadamente, quando aplicável, no tocante às pessoas ou grupos de pessoas em relação às quais se destina a utilização do sistema de IA de risco elevado.

Deve, igualmente, ser elaborada documentação técnica, antes da colocação no mercado ou da colocação em serviço do sistema, devendo a mesma ser mantida atualizada. Esta documentação deve ser elaborada de maneira que demonstre que o sistema de IA de risco elevado cumpre os requisitos estabelecidos na presente secção e deve facultar às autoridades nacionais competentes e aos organismos notificados, de forma clara e completa, as informações necessárias para aferir a conformidade do sistema de IA com esses requisitos, consoante prescreve o artigo 11º.

O sistema deve, também, ser concebido de modo a permitir o registo automático de eventos durante a sua vida útil, nos termos do artigo 12º, e bem assim de forma a assegurar que o seu funcionamento seja suficientemente transparente para permitir aos responsáveis pela implementação interpretar os resultados e utilizá-los de forma adequada. O artigo 13º impõe, ainda, que sejam acompanhados de instruções de utilização concisas, completas, corretas, claras, pertinentes, acessíveis, compreensíveis pelos responsáveis pela implantação.

Nos termos do artigo 14º, devem ser concebidos e desenvolvidos de modo a poderem ser eficazmente supervisionados por pessoas singulares durante o

período em que estão em utilização; e, de acordo com o artigo 15º, de maneira que alcancem um nível apropriado de exatidão, solidez e cibersegurança.

b) Obrigações dos prestadores de serviços

Os prestadores de sistemas de IA de risco elevado devem cumprir uma série de deveres, designadamente: assegurar que os seus sistemas de IA de risco elevado cumpram os requisitos previstos no regulamento; indicar no sistema de IA de risco elevado ou, se tal não for possível, na embalagem ou na documentação que o acompanha, consoante o caso, o seu nome, o nome comercial registado ou a marca registada e o endereço no qual podem ser contactados; dispor de um sistema de gestão da qualidade que cumpra o disposto no artigo 17º; conservar a documentação nos termos do artigo 18º; quando tal esteja sob o seu controlo, manter os registos gerados automaticamente pelos sistemas de IA de risco elevado que disponibilizam, conforme previsto no artigo 19º; assegurar que o sistema de IA de risco elevado seja sujeito ao procedimento de avaliação da conformidade aplicável, tal como previsto no artigo 43º, antes da colocação no mercado ou da colocação em serviço; elaborar uma declaração UE de conformidade, nos termos do artigo 47º; apor a marcação CE no sistema de IA de risco elevado ou, se tal não for possível, na embalagem ou na documentação que o acompanha, para indicar a conformidade com o regulamento; respeitar as obrigações de registo a que se refere o artigo 49º; adotar as medidas corretivas necessárias e prestar as informações, tal como estabelecido no artigo 20º; mediante pedido fundamentado de uma autoridade nacional competente, demonstrar a conformidade do sistema de IA de risco elevado com os requisitos estabelecidos pelo regulamento.

Os prestadores de serviço estabelecidos em países terceiros devem, ainda, antes de disponibilizarem os seus sistemas de IA de risco elevado no mercado, através de mandato escrito, designar um mandatário estabelecido na União, que fica vinculado pelas obrigações constantes no artigo 22º.

São também impostas obrigações aos importadores (artigo 23º) e aos distribuidores (artigo 24º).

Repare-se que qualquer distribuidor, importador, responsável pela implantação ou outro terceiro é considerado um prestador de um sistema de IA de risco elevado para efeitos do presente regulamento e fica sujeito às obrigações dos prestadores estabelecidas no artigo 16º, se colocar o seu nome ou marca num sistema de IA de risco elevado já colocado no mercado ou colocado em serviço, sem prejuízo de disposições contratuais que estipulem uma atribuição diferente das obrigações; se introduzir uma modificação substancial num sistema de IA de risco elevado que já tenha sido colocado no mercado ou colocado em serviço, de forma que o mesmo continue a ser um sistema de IA de risco elevado nos termos

do artigo 6º; se modificar a finalidade prevista de um sistema de IA, incluindo um sistema de IA de finalidade geral, que não tenha sido classificado como sendo de risco elevado e que já tenha sido colocado no mercado ou colocado em serviço, de forma que o sistema de IA em causa se torne um sistema de IA de risco elevado. O prestador que inicialmente colocou no mercado ou colocou em serviço o sistema de IA deixa de ser considerado um prestador desse sistema de IA específico, ficando, no entanto, obrigado a cooperar com o segundo, disponibilizando as informações necessárias e facultando o acesso técnico e a assistência razoavelmente esperados e necessários para o cumprimento das obrigações a que este passa a estar vinculado, exceto se tiver declarado que o seu sistema não podia ser alterado.

c) Obrigações dos responsáveis pela implantação

Sobre os responsáveis pela implantação recaem igualmente inúmeros deveres:

- dever de adotar medidas técnicas e organizativas adequadas para garantir que utilizam esses sistemas de acordo com as instruções de utilização que os acompanham;
- dever de atribuir a supervisão humana a pessoas singulares que possuam as competências, a formação e a autoridade necessárias, bem como o apoio necessário;
- nas hipóteses em que exerça controlo sobre os dados de entrada, dever de assegurar que os dados de entrada sejam pertinentes e suficientemente representativos tendo em vista a finalidade prevista do sistema de IA de risco elevado;
- dever de controlar o funcionamento do sistema de IA de risco elevado com base nas instruções de utilização;
- dever de manter os registos gerados automaticamente pelo sistema de IA de risco elevado, desde que esses registos estejam sob o seu controlo, por um período adequado à finalidade prevista do sistema de IA de risco elevado;
- dever de realizar uma avaliação de impacto sobre a proteção de dados;
- tratando-se de sistemas de risco elevado previstos no anexo III, que tomam decisões ou ajudam a tomar decisões relacionadas com pessoas singulares, dever de informar as pessoas singulares de que estão sujeitas à utilização do sistema de IA;
- dever de cooperar com as autoridades competentes em todas as medidas que essas autoridades tomarem em relação a um sistema de IA de risco elevado.

4.4 Obrigações de transparência em relação a determinados sistemas

Nos termos do artigo 50º, quando os sistemas de IA se destinarem a interagir diretamente com pessoas singulares, os prestadores de serviço devem concebe-los e desenvolvê-los de maneira a que as pessoas singulares em causa sejam informadas de que estão a interagir com um sistema de IA, salvo se tal for óbvio do ponto de vista do sujeito razoavelmente informado, atento e advertido, tendo em conta as circunstâncias e o contexto de utilização.

Por outro lado, os prestadores de sistemas de IA, incluindo sistemas de IA de finalidade geral, que gerem conteúdos sintéticos de áudio, imagem, vídeo ou texto, devem assegurar que os resultados do sistema de IA sejam marcados num formato legível por máquina e detetáveis como tendo sido artificialmente gerados ou manipulados.

Por seu turno, os responsáveis pela implantação de um sistema de reconhecimento de emoções ou de um sistema de categorização biométrica devem informar as pessoas expostas a esse sistema do seu funcionamento e tratar os dados pessoais em conformidade com o RGPD; devem, quando o sistema de IA que gere ou manipule conteúdos de imagem, áudio ou vídeo que constituam uma falsificação profunda, revelar que os conteúdos foram artificialmente gerados ou manipulados; e devem, quando o sistema de IA que gere ou manipule texto publicado com o objetivo de informar o público sobre questões de interesse público, revelar que o texto foi artificialmente gerado ou manipulado.

Repare-se que estes deveres de transparência não visam combater uma opacidade corporativa, tanto mais que se reconhecem como dignos de tutela os segredos de negócios das empresas que desenvolvem os algoritmos; nem a opacidade técnica que, sendo inerente ao deep learning, não pode ser apagada sem mais; mas uma opacidade cognitiva, dotando os sujeitos expostos de informações necessárias que lhes permitam lidar com os resultados dos sistemas autónomos, capacitando-os para diversas particularidades. É essa também a lógica subjacente aos deveres que, ao nível da conceção e desenvolvimento dos sistemas, de acordo com o artigo 13º, implicam que a colocação no mercado seja acompanhada de instruções de utilização concisas, completas, corretas, claras, pertinentes, acessíveis, compreensíveis pelos responsáveis pela implantação.

4.5 Os sistemas de IA de finalidade geral

Como vimos anteriormente, os sistemas de IA que tem finalidade geral, isto é, que têm capacidade para servir para diversas finalidades, tanto para utilização

direta, como para integração noutros sistemas de IA, podem envolver ou não um risco sistémico.

Em geral, os prestadores de serviços de sistemas de IA de finalidade geral devem elaborar e manter atualizada a documentação técnica do modelo, incluindo o seu processo de treino e de testagem e os resultados da sua avaliação; elaborar, manter atualizadas e disponibilizar informações e documentação aos prestadores de sistemas de IA que pretendam integrar o modelo de IA de finalidade geral nos seus sistemas de IA. Além disso, sem prejuízo da necessidade de respeitar e proteger os direitos de propriedade intelectual e as informações comerciais de caráter confidencial ou segredos comerciais, devem permitir que os prestadores de sistemas de IA tenham uma boa compreensão das capacidades e limitações do modelo de IA de finalidade geral e cumpram as suas obrigações. E devem elaborar e disponibilizar ao público um resumo suficientemente pormenorizado sobre os conteúdos utilizados para o treino do modelo de IA.

Estas obrigações ficam excluídas se os sistemas de IA forem disponibilizados ao abrigo de uma licença gratuita e aberta que permita o acesso, a utilização, a modificação e a distribuição do modelo, e cujos parâmetros, incluindo as ponderações, as informações sobre a arquitetura do modelo e as informações sobre a utilização do modelo, sejam disponibilizados ao público, exceto se o sistema envolver risco sistémico.

Neste caso, tratando-se de um sistema que envolva risco sistémico, os prestadores de modelos de IA de finalidade geral devem, para além daquelas obrigações, cumprir as que se encontram previstas no artigo 53º, designadamente devem realizar a avaliação do modelo em conformidade com protocolos e instrumentos normalizados que reflitam o estado da arte, incluindo a realização e documentação de testagens antagónicas do modelo, com vista a identificar e atenuar os riscos sistémicos; avaliar e atenuar eventuais riscos sistémicos a nível da União, incluindo as respetivas fontes, que possam resultar do desenvolvimento, da colocação no mercado ou da utilização de modelos de IA de finalidade geral com risco sistémico; acompanhar, documentar e comunicar, sem demora injustificada, ao Serviço para a IA e, se for caso disso, às autoridades nacionais competentes, as informações pertinentes sobre incidentes graves e eventuais medidas corretivas para os resolver; assegurar um nível adequado de proteção em termos de cibersegurança para o modelo de IA de finalidade geral com risco sistémico e a infraestrutura física do modelo.

4.6 A delimitação de uma esfera de responsabilidade: consequências dogmáticas

O regulamento IA permite-nos delimitar, por referência aos diversos sujeitos envolvidos no período de vida do sistema, uma esfera de responsabilidade no

sentido da *role responsibility*, conformada a partir dos deveres que sobre cada um deles isolada ou conjuntamente impende. Estando em causa deveres de conceção, deveres de cuidado, deveres de controlo, deveres de informação e deveres de transparência, aventa-se a possibilidade de, sempre que não tenham como destinatários as autoridades públicas, estarmos diante de deveres no tráfego. Ao mesmo tempo, e no que à ilicitude diz respeito, procura-se saber se algumas das normas contidas no Regulamento IA podem ser qualificadas como disposições legais de proteção de interesses alheios.

Estaríamos, assim, diante de uma daquelas hipóteses em que o fundamento normativo daqueles deveres seria encontrado numa concreta norma legal, ligando-se insofismavelmente à segunda modalidade de ilicitude, embora com possível conexão com a primeira, sempre que se desvelasse simultaneamente a lesão de um direito dotado de eficácia *erga omnes*.

Nessa medida, percebe-se que, em termos de articulação dogmática dos diversos pressupostos de imposição, é possível, a partir da violação normativa, presumir a culpa, nos termos gerais da responsabilidade civil e do que a doutrina tem vindo a defender a esse nível. Mais inquietante é tentar perceber em que medida as presunções de conformidade consagradas no Regulamento podem ou não ser ilididas.

Do mesmo modo, podemos considerar que, a partir do momento em que se viola um dos deveres previstos no Regulamento, a primitiva esfera de responsabilidade convola-se numa outra esfera de responsabilidade, no sentido da *liability*, podendo ser aí reconduzidas todas as lesões que poderiam ter sido evitadas com o cumprimento do dever. Em causa não está um juízo de probabilidade, mas um juízo normativo que implica que se analise o âmbito de proteção de cada um dos deveres impostos.

5 A Diretiva Responsabilidade da IA

Nos termos da Diretiva Responsabilidade da IA, cuja proposta foi adotada pela Comissão em setembro de 2022, presume-se a causalidade verificados que sejam determinados requisitos.

Nos termos do artigo 4º/1 da referida proposta, os tribunais nacionais presumem o nexo de causalidade entre o facto culposo do demandado e o resultado produzido pelo sistema de IA ou a incapacidade do sistema de IA de produzir um resultado, se estiverem preenchidas todas as seguintes condições:

– o demandante demonstrou ou o tribunal presumiu a existência de culpa do demandado, ou de uma pessoa por cujo comportamento o demandado é responsável, consistindo tal no incumprimento de um dever de

diligência previsto no direito da União ou no direito nacional diretamente destinado a proteger contra o dano ocorrido;
– pode-se considerar que é razoavelmente provável, com base nas circunstâncias do caso, que o facto culposo influenciou o resultado produzido pelo sistema de IA ou a incapacidade do sistema de IA de produzir um resultado;
– o demandante demonstrou que o resultado produzido pelo sistema de IA ou a incapacidade do sistema de IA de produzir um resultado deu origem ao dano.

Os termos da presunção, ilidível, concitam-nos as maiores dúvidas. Desde logo, não se percebe o que se presume na hipótese de o lesado demonstrar que o resultado produzido pelo sistema de IA ou a incapacidade do sistema de IA de produzir um resultado deu origem ao dano. Por outro lado, confunde-se a análise do âmbito de proteção do dever incumprido, a permitir uma presunção baseada na imputação, com uma ideia de probabilidade que nos aponta ainda para uma visão causalista e fisicista e com uma ideia de dificuldade probatória.

Esta conclusão é confirmada pelas restantes regras em matéria de presunção de causalidade.[8]

Assim, no caso de uma ação de indemnização intentada contra um fornecedor de um sistema de IA de risco elevado sujeito aos requisitos do Regulamento Inteligência Artificial ou uma pessoa sujeita às obrigações do fornecedor nos termos do mesmo Regulamento, a primeira condição só é cumprida se o autor da ação tiver demonstrado que o fornecedor ou, se for caso disso, a pessoa sujeita às obrigações do fornecedor não cumpriu algum dos seguintes requisitos estabelecidos nos referidos capítulos, tendo em conta as medidas tomadas e os resultados do sistema de gestão de riscos:
– o sistema de IA é um sistema que utiliza técnicas que envolvem o treino de modelos com dados que não foram desenvolvidos com base em conjuntos de dados de treino, validação e teste que cumprem os critérios de qualidade;
– o sistema de IA não foi concebido e desenvolvido de maneira que cumpra os requisitos de transparência;
– o sistema de IA não foi concebido e desenvolvido de maneira que permita uma supervisão eficaz por pessoas singulares durante o período de utilização do sistema de IA;

[8] Note-se que, em termos terminológicos, não existe sintonia entre a proposta de diretiva e o Regulamento IA. Tal deve-se ao facto de este ter conhecido diversas versões até à aprovação do texto final. Esta falta de sintonia não significa que não possa haver continuidade nas soluções pensadas a este propósito. A adaptação exige-se no futuro.

– o sistema de IA não foi concebido e desenvolvido de maneira que alcance, tendo em conta a finalidade prevista, um nível apropriado de exatidão, solidez e cibersegurança;
– ou as medidas corretivas necessárias não foram imediatamente tomadas para assegurar a conformidade do sistema de IA com as obrigações estabelecidas no Regulamento Inteligência Artificial ou para proceder à retirada ou recolha do sistema.

Por seu turno, no caso de uma ação de indemnização intentada contra um utilizador de um sistema de IA de risco elevado sujeito aos requisitos estabelecidos no título III, capítulos 2 e 3, do Regulamento Inteligência Artificial, a primeira condição é cumprida se o demandante provar que o utilizador não cumpriu as suas obrigações de utilizar ou controlar o sistema de IA em conformidade com as instruções de utilização que o acompanham ou que expôs o sistema de IA a dados de entrada sob o seu controlo que não são pertinentes tendo em conta a finalidade prevista do sistema.

Além disso, no caso de uma ação de indemnização relativa a um sistema de IA de risco elevado, o tribunal nacional não pode aplicar a presunção prevista no nº 1 se o demandado demonstrar que estão razoavelmente acessíveis ao demandante elementos de prova e conhecimentos especializados suficientes para provar o nexo de causalidade; e, no caso de uma ação de indemnização relativa a um sistema de IA que não seja um sistema IA de risco elevado, a presunção estabelecida só é aplicável se o tribunal nacional considerar que é excessivamente difícil para o demandante provar o nexo de causalidade.

O enfoque da solução é probatória e orienta-se pela proteção do lesado. Repare-se, aliás, que se prevê a possibilidade de se contestar uma ação de responsabilidade quando baseada numa presunção de causalidade, que não ultrapassa o plano do ser, o que nos mostra a fragilidade da ponderação baseada na probabilidade.

Os desafios colocados pela inteligência artificial mostram-nos à saciedade que o esquema causal alicerçado na ligação causa-efeito não é prestável e tornam urgentes a assunção de uma perspetiva imputacional que temos vindo a defender em termos gerais.

O problema da causalidade pode (e deve, por motivos metodológicos, dogmáticos e axiológicos) ser solucionado de acordo com a perspetiva binária e ético-axiológica, com assento personalista.[9]

[9] Para uma análise dos argumentos que justificam, em geral, esta perspetiva cf. Mafalda Miranda Barbosa, *Do nexo de causalidade ao nexo de imputação*, Princípia, 2013

Sendo a esfera de risco é delineada *a priori* pelo legislador e encabeçada a montante pelo sujeito, independentemente de qualquer atuação concreta, ou seja, nas hipóteses de configuração da responsabilidade pela IA como uma responsabilidade objetiva, colocar-se-ão problemas no que respeita à comprovação da interferência do algoritmo na história de surgimento da lesão.[10] Mas os problemas são fácticos e não normativos, resultando da opacidade tecnológica e corporativa a que já aludimos, e não se poderão resolver por apelo a qualquer *but-for test* ou *NESS-test*.

Posteriormente, considerando-se que a programação e/ou a utilização de um algoritmo determina a assunção de uma esfera de risco/responsabilidade, haveremos de concluir que se reconduzem ao seu núcleo todas as lesões que possam ser causadas pelo algoritmo. O juízo de recondução a que aludimos não se baseia numa ideia de probabilidade, mas de possibilidade, sendo densificado por uma ideia de concretização do risco que levou à tipificação da hipótese responsabilizatória. No fundo, o que se procura saber é se a lesão se pode compreender como atualização do risco que levou o legislador a tipificar a hipótese de responsabilidade objetiva.

Tratando-se de uma hipótese assente na culpa, como parece ser a opção do legislador, a esfera de risco edifica-se a partir da preterição de deveres no tráfego que permitam concluir no sentido do aumento do risco. Àquela pertencerão todas as lesões que pudessem ser evitadas com o cumprimento do dever, o que implica a análise do âmbito de relevância do mesmo. Ulteriores presunções são estabelecidas ao nível da proposta de diretiva da responsabilidade civil da IA.

Não se detetando qualquer um dos fatores aptos a excluir *a priori* o nexo de ilicitude, impõe-se, subsequentemente, um confronto de esferas de risco/responsabilidade. Avulta com particular importância a contemplação quer da esfera de risco/responsabilidade do lesado, quer a contemplação da esfera de risco/responsabilidade de um terceiro. O primeiro pode ter contribuído, com o seu comportamento, para o surgimento ou o agravamento do dano, ou pode, fruto das suas particulares idiossincrasias – e dependendo do domínio em que a automação seja aplicada – sofrer um dano especialmente gravoso, devendo, portanto, recorrer-se aos critérios de imputação objetiva para lidar com estas hipóteses. O segundo (dito terceiro) pode concorrentemente com o primeiro lesante assumir uma esfera de risco/responsabilidade, restando saber se este obliterou determinados deveres que serviriam exatamente para evitar o comportamento do terceiro. É, assim, possível condenar solidariamente mais do que um sujeito ao pagamento de uma indemnização.

[10] Apesar de a proposta de Diretiva não o prever, tratando-se de uma diretiva de harmonização mínima, é possível aos legisladores nacionais criarem a referida previsão.

O artigo 10º da Proposta de Regulamento em matéria de responsabilidade civil pela IA dispunha a este propósito que "se a lesão ou o dano forem causados tanto por uma atividade física ou virtual ou um processo levado a cabo por um sistema de inteligência artificial, como pela ação da pessoa afetada ou de um terceiro pela qual ela responda, a responsabilidade por ser limitada na sua extensão", admitindo-se, inclusivamente, a sua exclusão em caso de exclusivo contributo do lesado para o surgimento do dano, o que se mostra em sintonia com a proposta por nós apresentada. Embora não se vá, no quadro legislativo europeu, tão longe quanto desejável continuando-se preso a uma ideia causal estrita,[11] a solução não pode deixar de ser aplaudida, por abrir, em termos interpretativos, as portas a um esquema imputacional como aquele que estamos a delinear. Note-se, porém, que ao nível da Diretiva Responsabilidade Civil IA, a solução não chega a ser pensada.

A par da incerteza que matiza os sistemas e que pode ser solucionada por este esta, colocam-se, como referido anteriormente, importantes problemas no que respeita à causalidade múltipla.

Podemos, desde logo, confrontar-nos com hipóteses de causalidade alternativa incerta, isto é, de confluência de múltiplas possíveis causas, não se conseguindo determinar qual delas é a efetiva causa da lesão. Ora, contra aquele que é o entendimento tradicional na matéria, estamos em crer ser possível defender a solução da solidariedade de responsáveis, tanto quanto, do ponto de vista imputacional que nos orienta, partimos da edificação de esferas de risco/responsabilidade. Naquela hipótese em que é incerto se a lesão resulta da programação inicial, das atualizações posteriores ou dos dados que foram recolhidos pelo sistema, parece-nos não se poder senão defender a responsabilização do sujeito nos termos anteriormente definidos, podendo excluir-se a responsabilidade de um dos putativos lesantes se este provar que a lesão não resultou da materialização do

[11] Na verdade, o grupo de peritos que esteve na base dos estudos conducentes a esta proposta europeia continua a afirmar que que "where the damage is of a kind that safety rules were meant to avoid, failure to comply with such safety rules, including rules on cybersecurity, should lead to a reversal of the burden of proving causation". No mais, embora afirmem que o ónus da prova da causalidade, em geral, impede sobre a vítima, o seu cumprimento pode ser aligeirado em algumas circunstâncias: "the likelihood that the technology at least contributed to the harm; the likelihood that the harm was caused either by the technology or by some other cause within the same sphere; the risk of a known defect within the technology, even though its actual causal impact is not self-evident; the degree of ex-post traceability and intelligibility of processes within the technology that may have contributed to the cause (informational asymmetry); the degree of ex-post accessibility and comprehensibility of data collected and generated by the technology; the kind and degree of harm potentially and actually caused". Cf. Expert Group on Liability and New Technologies, *Liability for Artificial Intelligence and other emerging digital technologies,* 59. Ora, a perspetiva por nós encabeçada permite que os problemas atinentes ao ónus da prova sejam compreendidos de outro modo. Veja-se, para uma melhor compreensão, Mafalda Miranda Barbosa, *Do nexo de causalidade ao nexo de imputação. Contributo para a compreensão da natureza binária e personalista do requisito causal ao nível da responsabilidade civil extracontratual*, Princípia, 2013.

risco ou que não resultou da preterição de quaisquer deveres. Faz-se, assim, recair sobre quem controla o sistema o peso da opacidade tecnológica, com a consequente redução dos contornos da opacidade corporativa.

Em segundo lugar, podem configurar-se hipóteses de causalidade cumulativa necessária. Estamos a falar de todos aqueles casos em que, por exemplo, a vulnerabilidade do sistema só pode ser aproveitada por um terceiro em virtude de um comportamento não cauteloso de um operador intermediário do sistema. É neste contexto que, podendo falar-se de *technological units*, haveremos de ser particularmente cautelosos para determinar quando é que é possível edificar uma esfera de risco/responsabilidade que seja partilhada por diversos sujeitos interdependentes.

Do exposto, podemos concluir que, sem embargo da necessidade de se introduzirem modificações no quadro normativo para fazer face aos desafios colocados pela inteligência artificial, a questão da causalidade não pode ser solucionada com base numa mera previsão de uma hipótese ressarcitória, sequer com apelo a uma presunção que se alicerce, de forma ilidível, num grau de probabilidade bastante. Ao invés, a solução do problema deve, neste domínio particular, acompanhar a mutação de perspetiva que, em termos gerais, advogamos, devendo alicerçar-se numa visão imputacional do requisito.

6 Confronto com o projeto brasileiro na matéria

Louva-se no modelo europeu a compreensão da responsabilidade à luz de uma esfera de deveres que são impostos. Contudo, há dois aspetos que nos parecem particularmente frágeis. Em primeiro lugar, a dependência da Comissão relativamente à qualificação de um sistema como sendo de risco elevado. Por outro lado, a inexistência de uma hipótese de responsabilidade objetiva que deixa na sombra situações como aquelas em que a lesão ocorre por força da corrupção de dados provocada pelo próprio algoritmo. Toda a disciplina é pensada por referência a um quadro em que é possível controlar o risco associado ao uso de sistemas autónomos, quando o que resulta da prática – e resultará ainda mais pela rápida evolução no setor – é a incontrolabilidade de muitas formas de atuação autonómica do algoritmo.

Talvez por isso seja importante refletir sobre o projeto brasileiro em matéria de IA e responsabilidade civil por danos causados por sistemas autónomos. Em causa o Projeto Lei nº 2338, de 2023, que dispõe sobre o uso da IA.

Depois de definir uma série de princípios (boa-fé; crescimento inclusivo, desenvolvimento sustentável e bem-estar; autodeterminação e liberdade de decisão e de escolha; participação humana no ciclo da inteligência artificial e supervisão

humana efetiva; não discriminação; justiça, equidade e inclusão; transparência, explicabilidade, inteligibilidade e auditabilidade; confiabilidade e robustez dos sistemas de inteligência artificial e segurança da informação; devido processo legal, contestabilidade e contraditório; rastreabilidade das decisões durante o ciclo de vida de sistemas de inteligência artificial como meio de prestação de contas e atribuição de responsabilidades a uma pessoa natural ou jurídica; prestação de contas, responsabilização e reparação integral de danos; prevenção, precaução e mitigação de riscos sistêmicos derivados de usos intencionais ou não intencionais e de efeitos não previstos de sistemas de inteligência artificial; e não maleficência e proporcionalidade entre os métodos empregados e as finalidades determinadas e legítimas dos sistemas de inteligência artificial), o projeto consagra diversos direitos das pessoas afetadas por sistemas de IA: informação, explicação sobre a decisão; contestação de decisões ou previsões de sistemas de IA; participação humana em decisões de sistemas de IA; não discriminação; privacidade.

Além disso, tal como o regulamento europeu parte de diversos níveis de risco, categorizando-os, de acordo com uma avaliação preliminar realizada pelo fornecedor, antes da sua colocação no mercado.

Tratando-se de sistemas de risco excessivo (isto é, que empreguem técnicas subliminares que tenham por objetivo ou por efeito induzir a pessoa natural a se comportar de forma prejudicial ou perigosa à sua saúde ou segurança ou contra os fundamentos desta Lei; que explorem quaisquer vulnerabilidades de grupos específicos de pessoas naturais, tais como as associadas a sua idade ou deficiência física ou mental, de modo a induzi-las a se comportar de forma prejudicial a sua saúde ou segurança; ou usados pelo poder público, para avaliar, classificar as pessoas naturais, com base no seu comportamento social ou em atributos da sua personalidade, por meio de pontuação universal, para o acesso a bens e serviços e políticas públicas, de forma ilegítima ou desproporcional), fica vedada a sua implementação e o seu uso.

Os sistemas de alto risco, por seu turno, são aqueles que são usados para as seguintes finalidades: aplicação como dispositivos de segurança na gestão e no funcionamento de infraestruturas críticas, tais como controle de trânsito e redes de abastecimento de água e de eletricidade; educação e formação profissional, incluindo sistemas de determinação de acesso a instituições de ensino ou de formação profissional ou para avaliação e monitoramento de estudantes; recrutamento, triagem, filtragem, avaliação de candidatos, tomada de decisões sobre promoções ou cessações de relações contratuais de trabalho, repartição de tarefas e controle e avaliação do desempenho e do comportamento das pessoas afetadas por tais aplicações de inteligência artificial nas áreas de emprego, gestão de trabalhadores e acesso ao emprego por conta própria; avaliação de critérios de

acesso, elegibilidade, concessão, revisão, redução ou revogação de serviços privados e públicos que sejam considerados essenciais, incluindo sistemas utilizados para avaliar a elegibilidade de pessoas naturais quanto a prestações de serviços públicos de assistência e de seguridade; avaliação da capacidade de endividamento das pessoas naturais ou estabelecimento de sua classificação de crédito; envio ou estabelecimento de prioridades para serviços de resposta a emergências, incluindo bombeiros e assistência médica; administração da justiça, incluindo sistemas que auxiliem autoridades judiciárias na investigação dos fatos e na aplicação da lei; veículos autónomos, quando seu uso puder gerar riscos à integridade física de pessoas; diagnósticos e procedimentos médicos; sistemas biométricos de identificação; investigação criminal e segurança pública, em especial para avaliações individuais de riscos pelas autoridades competentes, a fim de determinar o risco de uma pessoa cometer infrações ou de reincidir, ou o risco para potenciais vítimas de infrações penais ou para avaliar os traços de personalidade e as características ou o comportamento criminal passado de pessoas singulares ou grupos; estudo analítico de crimes relativos a pessoas naturais, permitindo às autoridades policiais pesquisar grandes conjuntos de dados complexos, relacionados ou não relacionados, disponíveis em diferentes fontes de dados ou em diferentes formatos de dados, no intuito de identificar padrões desconhecidos ou descobrir relações escondidas nos dados; investigação por autoridades administrativas para avaliar a credibilidade dos elementos de prova no decurso da investigação ou repressão de infrações, para prever a ocorrência ou a recorrência de uma infração real ou potencial com base na definição de perfis de pessoas singulares; ou gestão da migração e controle de fronteiras.

A autoridade competente na matéria pode atualizar a lista dos sistemas qualificados como implicando risco excessivo ou alto risco, identificando novas hipóteses com base nos critérios firmados no artigo 18º.

Aos agentes de IA (fornecedores e operadores) são, subsequentemente, impostos diversos deveres. Designadamente, devem estabelecer estruturas de governança e processos internos aptos a garantir a segurança dos sistemas e a salvaguarda dos direitos das pessoas afetadas. São, igualmente, impostas medidas de transparência, medidas de gestão de dados, medidas de segurança, medidas de governança, que conhecem um figurino agravado para os sistemas de alto risco.

O paralelo com o regulamento europeu é evidente. Com técnicas legislativas muito diversas, parte-se dos conhecimentos atuais em matéria de inteligência artificial e procura-se proteger quer a pessoa, titular de direitos fundamentais, quer o próprio modelo civilizacional, sem se criarem obstáculos desmedidos ao desenvolvimento científico e tecnológico. A IA é, nas duas latitudes, pensada com

base na centralidade da pessoa e da sua confiança. A grande diferença passa pelo pendor mais regulamentar, tão tipicamente conhecido das instâncias comunitárias europeias, e pela necessária fundamentação da legislação europeia numa ideia de construção de um mercado único ágil e livre.

Há, porém, um ponto em que o afastamento é claro. Nos termos do artigo 27º do Projeto Lei brasileiro, "o fornecedor ou operador de sistema de inteligência artificial que cause dano patrimonial, moral, individual ou coletivo é obrigado repará-lo integralmente, independentemente do grau de autonomia do sistema. Quando se tratar de sistema de inteligência artificial de alto risco ou de risco excessivo, o fornecedor ou operador respondem objetivamente pelos danos causados, na medida de sua participação no dano. Quando não se tratar de sistema de inteligência artificial de alto risco, a culpa do agente causador do dano será presumida, aplicando-se a inversão do ônus da prova em favor da vítima".

Os agentes de inteligência artificial não serão, contudo, responsabilizados quando comprovarem que não colocaram em circulação, empregaram ou tiraram proveito do sistema de inteligência artificial; ou quando comprovarem que o dano é decorrente de fato exclusivo da vítima ou de terceiro, assim como de caso fortuito externo.

Louva-se no quadro brasileiro a previsão de uma hipótese de responsabilidade objetiva. Mas, fica-se sem perceber por que motivo a responsabilização do agente de IA fica limitada à sua participação no dano. Não só pode não haver efetiva participação no dano – no sentido causalista do termo –, como, em moldes imputacionais, o grande problema pode ser não se conseguir discernir o grau de contribuição de cada interveniente no ciclo de vida do sistema. Significa isto que, mesmo dando-se um passo em frente, parece denotar-se aqui uma certa prisão a quadros dogmáticos que devem ser superados (em geral e, muito em particular, no tocante à IA).

Informação bibliográfica deste texto, conforme a NBR 6023:2018 da Associação Brasileira de Normas Técnicas (ABNT):

BARBOSA, Mafalda Miranda. IA, riscos e responsabilidade – uma reflexão em torno do Regulamento IA e do Projeto Lei brasileiro nº 2338, de 2023. *Revista Brasileira de Direito Civil – RBDCivil*, Belo Horizonte, v. 33, n. 4, p. 163-189, out./dez. 2024. DOI: 10.33242/rbdc.2024.04.007.

Recebido em: 18.07.2024
Aprovado em: 20.10.2024