3

Supervisão, classificação e certificação dos sistemas de IA na Proposta de Regulamento sobre Inteligência Artificial

(https://doi.org/10.47907/DireitoemMudanca/2023/3)

*José Ricardo Marcondes Ramos**

Resumo: Diante das oportunidades e dos riscos oriundos do uso crescente da Inteligência Artificial, a União Europeia vem desenvolvendo medidas e documentos legais para regular o desenvolvimento e uso deste tipo de tecnologia, tendo como objetivo incentivar suas vantagens e ao mesmo tempo tutelar os riscos inerentes. Como resultado desses esforços, foi apresentada a Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial e cria um quadro normativo básico relativo à governação, à supervisão e à responsabilidade. Para consolidar uma abordagem equilibrada que garanta uma gestão eficiente do risco sem prejudicar a inovação, a Proposta de Regulamento da União Europeia sobre inteligência artificial classifica os sistemas de IA em níveis de risco e correlaciona deveres de comportamento específicos e proporcionais a cada tipo. Além disso, também estabelece um conjunto de regras harmonizadas para o desenvolvimento, a colocação no mercado e a utilização de sistemas de IA na UE, e cria requisitos essenciais e obrigatórios para cada tipo de sistema, com foco na transparência e prestação de informações. Neste contexto, o objetivo deste estudo é analisar a Proposta de Regulamento sobre inteligência artificial, e visa examinar os critérios de classificação dos sistemas de IA, as

* Doutorando em Ciências Jurídico-Criminais da Faculdade de Direito da Universidade de Coimbra; investigador-colaborador do IJ.

relações regulatórias que serão instituídas pela legislação na Europa e os deveres de comportamento previstos para mitigar os riscos inerentes aos sistemas de IA.

Palavras-chave: Inteligência artificial, regulação, proposta de regulamento, certificação algorítmica.

1. Introdução

A Inteligência Artificial (IA) já é uma das tecnologias mais transformadoras do século XXI e vem sendo cada vez mais presente no nosso dia a dia, muitas vezes de forma imperceptível, seja indicando o melhor caminho no GPS, fazendo uma gestão otimizada da bateria de dispositivos eletrónicos, identificando emails como spam, ou mesmo recomendando conteúdos online. Nos últimos anos, o uso da inteligência artificial tem também trazido uma série de inovações bastante disruptivas a exemplo dos assistentes virtuais como a Alexa ou a Siri, dos sistemas de criação de imagens a partir de texto desenvolvidos por empresas como Dall-E 2, Crayion e Midjourney, e dos Grandes Modelos de Linguagem que baseiam o funcionamento de ChatBots como o Bing da Microsoft, o Bard do Google e o, ChatGPT da Open AI.

Por suas capacidades de melhorar previsões, otimizar as operações e a afetação de recursos de empresas e instituições e personalizar o fornecimento de serviços, o uso da inteligência artificial já tem mostrado que pode contribuir para diversos resultados benéficos para a sociedade e para a economia. Ocorre, porém, que as mesmas habilidades e capacidades técnicas que vêm sendo decisivos para diversos benefícios sociais e económicos têm também demonstrado diversos riscos que representam ameaças à saúde, à segurança e aos direitos fundamentais das pessoas.

Tendo como foco incentivar e maximizar as vantagens que a inteligência artificial pode trazer para a sociedade e, simultaneamente, tutelar os riscos inerentes que esta família de tecnologias pode trazer, a União Europeia vem desenvolvendo uma série de medidas e documentos legais para regular o desenvolvimento e o uso de sistemas de inteligência artificial em âmbito europeu. Levando em consideração a velocidade recente da evolução desta tecnologia e os possíveis desafios sociais daí decorrentes, a UE está empenhada em alcançar uma

abordagem equilibrada que permita garantir uma gestão eficiente do risco de sistemas de inteligência artificial sem prejudicar a inovação. Como fruto destes esforços, foi apresentada pela Comissão Europeia e pelo Parlamento Europeu a Proposta de Regulamento que estabelece regras harmonizadas em matéria de inteligência artificial.

O presente trabalho tem como objeto de estudo a *Proposta de Regulamento* sobre a inteligência artificial, tendo-se como objetivo analisar os critérios de classificação de sistemas de inteligência artificial, as relações regulatórias que serão criadas por esta legislação em âmbito europeu e, finalmente, os deveres de comportamento obrigatórios que os sistemas de inteligência artificial deverão adotar para tutelar os seus riscos inerentes.

2. Classificação de sistemas de Inteligência Artificial

A *Proposta de Regulamento* insere-se em um contexto amplo de regulação da Inteligência Artificial em âmbito Europeu, iniciada ainda em 2018 com a Comunicação da Comissão Europeia *Inteligência Artificial para a Europa (Estratégia IA)* e aprofundada gradualmente em diversos outros documentos como o *Plano coordenado para o desenvolvimento e utilização da inteligência artificial «Made in Europe»* (publicado inicialmente em 2019 e posteriormente revisado em 2021), as *Orientações Éticas para uma IA de Confiança* do grupo de Peritos de Alto Nível sobre a Inteligência Artificial (2019), a Comunicação *Aumentar a Confiança numa Inteligência Artificial Centrada no Ser Humano* (2019) e o *Livro Branco sobre a Inteligência Artificial: uma abordagem europeia virada para a excelência e a confiança* (2020), entre outros.

Como reflexo e consolidação dos princípios e objetivos delineados nos documentos anteriores, a *Proposta de Regulamento* aborda a necessidade de garantir um arcabouço ético e jurídico apropriado para o desenvolvimento da inteligência artificial[1] através da criação um

[1] Enquanto, por exemplo, a Comunicação *Estratégia IA*, de 2018, prevê a garantia de um quadro ético e jurídico apropriado baseado nos valores da União e em consonância com a Carta dos Direitos Fundamentais da União Europeia como um dos 3 pilares bases da abordagem europeia à inteligência artificial (os outros dois pilares são (i) reforçar a capacidade tecnológica e industrial da UE e a aceitação da IA em toda a economia e (ii) preparar a sociedade para as mudanças socioeconómicas decorrentes da IA); em suas *Orientações Éticas para uma IA de Confiança*, o grupo

quadro normativo básico relativo à governação, à supervisão e à responsabilidade em sistemas de inteligência artificial. Seguindo de perto a abordagem europeia centrada no binómio excelência e confiança e no duplo objetivo de promover da adoção da IA enquanto simultaneamente aborda os seus riscos inerentes, delineada no *Livro Branco* que cria as bases da regulação europeia sobre IA, a *Proposta de Regulamento* apresenta um quadro regulamentar horizontal, equilibrado e proporcionado baseado em quatro objetivos específicos, nomeadamente:

- Garantir que os sistemas de IA colocados no mercado da União e utilizados sejam seguros e respeitem a legislação em vigor em matéria de direitos fundamentais e valores da União
- Garantir a segurança jurídica para facilitar os investimentos e a inovação no domínio da IA
- Melhorar a governação e a aplicação efetiva da legislação em vigor em matéria de direitos fundamentais e dos requisitos de segurança aplicáveis aos sistemas de IA
- Facilitar o desenvolvimento de um mercado único para as aplicações de IA legítimas, seguras e de confiança e evitar a fragmentação do mercado.

Para alcançar estes objetivos de forma eficaz e gerir os riscos inerentes aos sistemas de inteligência artificial, a *Proposta de Regulamento* estabelece regras harmonizadas para o desenvolvimento, a colocação no mercado e a utilização de sistemas de IA na União Europeia, o que é feito através da criação de requisitos essenciais e obrigatórios para determinados tipos de sistemas de IA e além de deveres de comportamento relacionados à transparência e à prestação de informações por parte de fornecedores (art. 16), fabricantes (art. 24), importadores (art. 26) e distribuidores (art. 27) de produtos e serviços que utilizem algoritmos de inteligência artificial. Conforme previsto na *Proposta de*

---

de Peritos de Alto Nível sobre a Inteligência Artificial identifica a necessidade de ser Legal, cumprindo toda a legislação e regulamentação aplicáveis como uma das 3 componentes essenciais para uma IA de confiança (as outras duas componentes são (i) ser Ética, garantindo a observância de princípios e valores éticos, e (ii) ser Sólida, tanto do ponto de vista técnico como do ponto de vista social, uma vez que, mesmo com boas intenções, os sistemas de IA podem causar danos não intencionais). De outro lado, o *Livro Branco* ressalta a importância da existência de um quadro regulamentar claro e adaptado às características específicas da inteligência artificial, principalmente a sua capacidade de aprendizagem automática e a sua autonomia decisória.

Regulamento, toda esta estrutura será supervisionada por mecanismos nacionais e europeus de avaliação de conformidade e acompanhamento de sistemas de inteligência artificial.

Consolidando a abordagem europeia centrada no binómio excelência e confiança descrita no *Livro Branco*, a *Proposta de Regulamento* destaca a importância de o quadro regulamentar aplicado ao domínio da inteligência artificial ser eficaz na gestão dos riscos, mas sem ser demasiadamente prescritivo a ponto de inviabilizar a pesquisa e a inovação e também sem criar um encargo desproporcionado principalmente para pequenas e médias empresas[2]. Como forma de criar uma intervenção jurídica equilibrada e proporcional, a *Proposta de Regulamento* propõe um quadro jurídico que é simultaneamente sólido, centrado em uma *abordagem baseada no risco*, que classifica os sistemas de inteligência artificial a partir dos níveis de risco criados pelos sistemas e prevê uma intervenção jurídica às situações concretas em que existe um motivo de preocupação justificado presente ou razoavelmente antecipado num futuro próximo, e simultaneamente flexível, incluindo mecanismos que permitam a sua adaptação dinâmica à medida que a tecnologia evolui e surgem novas situações preocupantes.

Neste contexto, para além de excluir expressamente a sua aplicação aos sistemas de IA desenvolvidos ou utilizados exclusivamente para fins militares (art. 2.º, n. 3), a *Proposta de Regulamento* distingue os sistemas de inteligência artificial em quatro categorias diferentes de risco, correlacionando deveres de comportamento específicos e proporcionais a cada tipo, nomeadamente: *riscos inaceitáveis* (Título II), que são práticas proibidas em território europeu; *riscos limitados* (Título IV), para os quais existem deveres de informação e transparência para com consumidores; *riscos mínimos* (Título IX), para os quais não

---

[2] Conforme descrito no Livro Branco "por uma questão de princípio, o novo quadro regulamentar para a IA deve ser eficaz para atingir os seus objetivos, mas não excessivamente prescritivo, de forma a não criar um encargo desproporcionado, especialmente para as PME. Para atingir este equilíbrio, a Comissão considera que deve seguir uma abordagem baseada no risco". A *Proposta de Regulamento* também reflete esta busca pelo equilíbrio regulatório, estabelecendo em sua exposição de motivos que "a presente proposta apresenta uma abordagem regulamentar horizontal equilibrada e proporcionada ao domínio da inteligência artificial, que se limita aos requisitos mínimos necessários para dar resposta aos riscos e aos problemas associados à IA, sem restringir ou prejudicar indevidamente a evolução tecnológica ou aumentar desproporcionalmente o custo de colocação no mercado das soluções de IA".

existem deveres e obrigações específicos; e, finalmente, *riscos elevados* (Título III), foco principal da *Proposta de Regulamento* para os quais, por ameaçarem direitos fundamentais e a segurança, são previstas uma série de requisitos relacionados à qualidade dos dados, à documentação e à rastreabilidade, à transparência, à supervisão humana, à exatidão e à solidez.

3. Sistemas de inteligência artificial com riscos inaceitáveis

O primeiro tipo de risco identificado pela *Proposta de Regulamento* são os chamados *riscos inaceitáveis* (Título II, art. 5.º), que englobam práticas que a Comissão entende que devem ser proibidas por serem particularmente prejudiciais à população por violarem os valores da União Europeia como a dignidade humana, a liberdade, a igualdade, a democracia e o Estado de direito, bem como direitos fundamentais da União, incluindo o direito à não discriminação, à proteção de dados pessoais e à privacidade, e os direitos das crianças. Neste ponto, a *Proposta de Regulamento* tem como foco quatro tipos de prática, das quais duas relacionadas à manipulação da população com o potencial para causar danos físicos ou psicológicos à pessoa manipulada ou a terceiros, uma relacionada à avaliação e à classificação de pessoas para uso geral por parte das autoridades públicas e, por fim, uma prática relacionada à identificação biométrica à distância em «tempo real» em espaços acessíveis ao público para efeitos de manutenção da ordem pública.

Quanto ao uso de sistemas de inteligência artificial para práticas manipuladoras, exploratórias e de controlo social, por entender que algoritmos concebidos para distorcer o comportamento humano desrespeitam os valores da União, a *Proposta de Regulamento* proíbe a colocação no mercado ou em serviço e a utilização de sistemas de IA (i) baseados em técnicas subliminares que contornem a consciência das pessoas para distorcer substancialmente o seu comportamento (art. 5.º, n. 1, al. a)), bem como algoritmos (ii) que explorem vulnerabilidades de grupos específicos de pessoas associadas à sua idade ou deficiência física ou mental, a fim de distorcer substancialmente o seu comportamento (art. 5.º, n. 1, al. b)).

Em segundo lugar, a *Proposta de Regulamento* proíbe o uso, por parte de autoridades públicas, de sistemas de classificação social que avaliam ou classificam a credibilidade de pessoas singulares com base

no seu comportamento social em diversos contextos ou com base em características de personalidade ou pessoais, conhecidas ou previsíveis (art. 5.º, n. 1, al. c)). Esta proibição decorre do entendimento da Comissão de que este tipo de sistema pode criar resultados discriminatórios, levar a tratamentos prejudiciais, desfavoráveis, injustificados ou desproporcionados de pessoas singulares ou grupos sociais ou mesmo levar à exclusão de grupos inteiros de pessoas, principalmente quando a classificação social obtida por meio desses sistemas for aplicada em contextos sociais não relacionados com o contexto nos quais os dados foram originalmente gerados ou recolhidos (Considerando 17).

A última prática de inteligência artificial proibida pela *Proposta de Regulamento* relaciona-se ao uso de sistemas de identificação biométrica de pessoas singulares à distância que, na definição adotada no artigo 3.º, n. 36, pode ser compreendido como "um sistema de IA concebido para identificar pessoas singulares à distância por meio da comparação dos dados biométricos de uma pessoa com os dados biométricos contidos numa base de dados de referência, sem que o utilizador do sistema de IA saiba antecipadamente se a pessoa em causa estará presente e pode ser identificada". Diferentemente das proibições anteriores, a vedação trazida no artigo 5.º, n. 1, al. d) da *Proposta de Regulamento*, não está relacionada ao sistema de inteligência artificial em si mesmo, mas à forma como este tipo de sistema pode vir a ser utilizado, motivo pelo qual a proibição está relacionada à verificação simultânea de três elementares normativas principais: que a identificação biométrica à distância seja feita «em tempo real», realizada em espaços acessíveis ao público e com o propósito específico de realizar a manutenção da ordem pública.

Com relação ao primeiro elemento, a *Proposta de Regulamento* diferencia a identificação biométrica à distância «em tempo real» e «em diferido»[3] descrevendo a primeira forma como "a utilização «ao vivo» ou «quase ao vivo» de materiais, como vídeos, criados por uma câmara ou outro dispositivo com uma funcionalidade semelhante"

---

[3] Em contraste, o artigo 3.º, n. 37 define os sistemas de identificação biométrica à distância em diferido como "um sistema de identificação biométrica à distância que não seja um sistema de identificação biométrica à distância em «tempo real»". De outro lado, porém, o considerando 8 esclarece que "no caso dos sistemas «em diferido», os dados biométricos já foram recolhidos e a comparação e a identificação ocorrem apenas após um atraso significativo. Estes sistemas utilizam materiais, tais como imagens ou vídeos, criados por câmaras de televisão em circuito fechado ou dispositivos privados antes de o sistema ser utilizado relativamente às pessoas singulares em causa".

(Considerando n.º 8), ou como a recolha de dados biométricos, a comparação com uma base de dados de referência e a identificação de pessoas singulares sem atraso significativo, de forma imediata ou quase imediata, ou em todo o caso, sem um atraso significativo (art. 3.º, n. 36). Diante da importância desta elementar normativa, o texto legal ainda inclui a previsão expressa de que "[p]ara evitar que as regras sejam contornadas, tal inclui não apenas a identificação instantânea, mas também a identificação com ligeiro atraso".

A delimitação da proibição aos espaços acessíveis ao público – definido pelo artigo 3.º, n. 39) como "qualquer espaço físico aberto ao público, independentemente da eventual aplicação de condições de acesso específicas" – está relacionada não apenas à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais mas, principalmente, à gestão dos riscos inerentes à utilização de sistemas de identificação biométrica de pessoas singulares. Isto porque a Comissão Europeia entende que a utilização deste tipo de sistema "é considerada particularmente intrusiva para os direitos e as liberdades das pessoas em causa, visto que pode afetar a vida privada de uma grande parte da população, dar origem a uma sensação de vigilância constante e dissuadir indiretamente o exercício da liberdade de reunião e de outros direitos fundamentais" (Considerando 19). Além disso, a Comissão também ressalta que existem riscos acrescidos para os direitos e as liberdades das pessoas decorrentes do impacto imediato e das oportunidades limitadas para a realização de controlos adicionais ou correções da utilização desses sistemas que funcionam «em tempo real».

Finalmente, a última elementar normativa que interessa a esta proibição está relacionada ao contexto em que os sistemas de identificação biométrica à distância são aplicados, isto é, para efeitos de manutenção da ordem pública o que é compreendido como "atividades realizadas por autoridades policiais para efeitos de prevenção, investigação, deteção ou repressão de infrações penais ou execução de sanções penais, incluindo a proteção contra ameaças à segurança pública e a prevenção das mesmas" (art. 3.º, n. 41). Esta elementar é relevante uma vez que é precisamente no contexto da manutenção da ordem pública que a *Proposta de Regulamento* descreve três situações de exceção em que a utilização de sistemas de identificação biométrica de pessoas singulares à distância em tempo real é permitida, nomeadamente, a procura e investigação seletiva de potenciais vítimas específicas de crimes,

principalmente crianças desaparecidas; a prevenção de uma ameaça específica, substancial e iminente à vida ou à segurança física de pessoas singulares ou de um ataque terrorista; e, finalmente, a deteção, localização, identificação ou instauração de ações penais relativamente a infratores ou suspeitos de infrações penais a que se refere a Decisão-Quadro 2002/584/JAI do Conselho[4], desde que tal infração seja punível no Estado-Membro em causa com pena ou medida de segurança privativas de liberdade de duração máxima não inferior a três anos e tal como definidas pela legislação desse Estado-Membro.

Muito embora estas três exceções sejam permitidas, uma vez que a Comissão entende que são situações em que a utilização de sistemas de identificação biométrica de pessoas singulares à distância em tempo real é estritamente necessária por motivos de interesse público importante e cuja importância prevalece sobre os seus riscos inerentes, a *Proposta de Regulamento* condiciona esta utilização à autorização expressa e específica de uma autoridade judiciária competente ou de uma autoridade administrativa independente. Ainda assim, apesar desta exigência de autorização jurisdicional prévia, a *Proposta de Regulamento* também prevê a possibilidade da apresentação de um pedido durante o uso do sistema ou logo após, desde que haja uma situação de emergência devidamente justificada, "ou seja, quando a necessidade de utilizar os sistemas em causa torna efetiva e objetivamente impossível obter uma autorização antes de iniciar essa utilização" (Considerando

---

[4] As infrações em questão são as seguintes: participação numa organização criminosa; terrorismo; tráfico de seres humanos; exploração sexual de crianças e pedopornografia; tráfico ilícito de estupefacientes e de substâncias psicotrópicas; tráfico ilícito de armas, munições e explosivos; corrupção; fraude, incluindo a fraude lesiva dos interesses financeiros das Comunidades Europeias na aceção da convenção de 26 de julho de 1995, relativa à proteção dos interesses financeiros das Comunidades Europeias; branqueamento dos produtos do crime; falsificação de moeda, incluindo a contrafação do euro; cibercriminalidade; crimes contra o ambiente, incluindo o tráfico ilícito de espécies animais ameaçadas e de espécies e essências vegetais ameaçadas; auxílio à entrada e à permanência irregulares; homicídio voluntário, ofensas corporais graves; tráfico ilícito de órgãos e de tecidos humanos; rapto, sequestro e tomada de reféns; racismo e xenofobia; roubo organizado ou à mão armada; tráfico de bens culturais incluindo antiguidades e obras de arte; burla; extorsão de proteção e extorsão; contrafação e piratagem de produtos; falsificação de documentos administrativos e respetivo tráfico; falsificação de meios de pagamento; tráfico ilícito de substâncias hormonais e outros fatores de crescimento; tráfico ilícito de materiais nucleares e radioativos; tráfico de veículos roubados; violação; fogo-posto; crimes abrangidos pela jurisdição do Tribunal Penal Internacional; desvio de avião ou navio; sabotagem.

21). Ainda assim, segundo previsão do texto legal, o uso dos sistemas de identificação biométrica em situações de emergência deve limitar-se ao mínimo absolutamente necessário e a autoridade policial deve obter uma autorização o quanto antes, apresentando as razões para não ter efetuado o pedido mais cedo.

4. Sistemas de inteligência artificial de risco limitado e de risco mínimo

Entre os sistemas de inteligência artificial cuja utilização é permitida em âmbito Europeu, a *Proposta de Regulamento* identifica em seu artigo 52.º (Título IV) três tipos de sistemas de IA que geram *riscos limitados* para os quais são previstas regras de transparência harmonizadas que criam deveres de prestação de informações para seus usuários, de forma a garantir que as pessoas possam tomar decisões informadas ou distanciar-se de determinadas situações – o que inclui a obrigação de que essas informações e notificações devem ser fornecidas em formatos acessíveis a pessoas com deficiência. Em específico, estas obrigações de transparência são aplicáveis a (i) sistemas autónomos que interagem com pessoas singulares (art. 52.º, n. 1), (ii) sistemas de reconhecimento de emoções ou de categorização biométrica (art. 52.º, n. 2), e (iii) sistemas de geração ou manipulação de conteúdo (art. 52.º, n. 3).

Inicialmente, por entender que sistemas autónomos que interagem com pessoas singulares e sistemas de criação e manipulação de conteúdo "podem representar riscos específicos de usurpação de identidade ou fraude" (Considerando 70), a *Proposta de Regulamento* cria obrigações de transparência específicas. Assim, em primeiro lugar, fornecedores de sistemas de IA destinados a interagir com seres humanos têm o dever de garantir que este tipo de algoritmo seja concebido e desenvolvido de maneira a que as pessoas sejam informadas de que estão a interagir com um sistema de IA e não com outro ser humano, salvo quando as circunstâncias e o contexto de utilização do algoritmo revelem esta interação automatizada de forma óbvia[5] (art. 52.º, n. 1).

---

[5] Cumpre salientar que a *Proposta de Regulamento* prevê uma segunda exceção estabelecendo que "esta obrigação não se aplica a sistemas de IA legalmente autorizados para detetar, prevenir, investigar e reprimir infrações penais, salvo se esses sistemas estiverem disponíveis ao público para denunciar uma infração penal".

Em segundo lugar, utilizadores que recorram a sistemas de inteligência artificial para gerar ou manipular conteúdos de imagem, áudio ou vídeo que sejam consideravelmente semelhantes a conteúdos autênticos – ou seja, que sejam consideravelmente semelhantes a pessoas, locais ou acontecimentos reais e que, falsamente, pareçam ser autênticos a outrem, as chamadas «falsificações profundas» – têm o dever de informar aos usuários que o conteúdo em questão foi gerado ou manipulado artificialmente (art. 52.º, n. 3). O texto legislativo prevê como exceção a estas obrigações de transparência a utilização destes sistemas para fins legítimos, descrevendo duas situações específicas: a sua utilização no contexto da manutenção da ordem pública e em casos de liberdade de expressão.

Finalmente, os deveres de transparência e informação são também aplicáveis a utilizadores de sistemas de inteligência artificial de reconhecimento de emoções e de categorização biométrica de pessoas singulares (art. 52.º, n. 2). Assim, ainda como forma de garantir que as pessoas possam tomar decisões informadas ou mesmo distanciar-se de determinadas situações, existe uma obrigação de informação aos usuários quando as suas emoções ou características são reconhecidas por meios automatizados ou quando estão sujeitas à avaliação e associação com categorias (sociais) com base em dados biométricos. Aqui, a única exceção em que esta obrigação não se aplica relaciona-se à utilização de sistemas de IA usados para categorização biométrica que sejam legalmente autorizados para detetar, prevenir e investigar infrações penais.

Relativamente aos deveres de transparência dos sistemas de inteligência artificial de risco limitado, é interessante notar que enquanto os deveres de comportamento relacionados aos sistemas de reconhecimento de emoções e categorização biométrica e aos sistemas de manipulação de conteúdo aplicam-se aos seus *utilizadores* – que na aceção do artigo 3.º, n. 4, compreende "uma pessoa singular ou coletiva, autoridade pública, agência ou outro organismo que utilize, sob a sua autoridade, um sistema de IA, salvo se o sistema de IA for utilizado no âmbito de uma atividade pessoal de caráter não profissional" – as obrigações de informação relacionadas aos sistemas automáticos de interação (ou *chatbots*) aplicam-se aos *fornecedores*, compreendidos como "uma pessoa singular ou coletiva, autoridade pública, agência ou outro organismo que desenvolva um sistema de IA ou que tenha um sistema de IA desenvolvido com vista à sua colocação no mercado ou

colocação em serviço sob o seu próprio nome ou marca, a titulo oneroso ou gratuito" (art. 3.º, n. 2).

Relativamente aos sistemas de inteligência artificial de risco mínimo, a *Proposta de Regulamento* não menciona este tipo de sistema diretamente, motivo pelo qual eles são identificados de forma subsidiária por não estarem em nenhuma das classificações anteriores. Muito embora para este tipo de sistema não existam obrigações legais aplicáveis ou deveres de comportamento específicos, por força do Título IX da *Proposta de Regulamento* a Comissão Europeia encoraja e facilita a adoção de códigos de conduta voluntários por todos os sistemas de inteligência artificial utilizados na União Europeia, independentemente dos deveres de comportamento específicos correlacionados a cada tipo de risco.

5. Sistemas de inteligência artificial de risco elevado

Para além das obrigações de transparência e informação relacionadas a sistemas de risco limitado, a *Proposta de Regulamento* cria também um conjunto de requisitos horizontais obrigatórios que garantam uma IA de confiança, procedimentos de avaliação de conformidade antes da sua colocação no mercado e obrigações previsíveis, proporcionadas e claras para garantir a segurança e o respeito da legislação em vigor. Ainda como reflexo da abordagem baseada no risco e seguindo o objetivo de criar uma intervenção jurídica equilibrada e proporcional, a *Proposta de Regulamento* descreve estes deveres de comportamento específicos como "estritamente necessários para atenuar os riscos" colocados pela inteligência artificial nos domínios da saúde, segurança e direitos fundamentais, restringindo esta intervenção às situações concretas em que existe um motivo de preocupação justificado presente ou razoavelmente antecipado num futuro próximo, naquilo que identifica como *sistemas de inteligência artificial de risco elevado* (Capítulo III)[6].

---

[6] Conforme descrito no Considerando 28: "a dimensão dos impactos adversos causados pelo sistema de IA nos direitos fundamentais protegidos pela Carta é particularmente importante quando se classifica um sistema de IA como sendo de risco elevado. Esses direitos incluem o direito à dignidade do ser humano, o respeito da vida privada e familiar, a proteção de dados pessoais, a liberdade de expressão e de informação, a liberdade de reunião e de associação, a não discriminação, a defesa dos consumidores, os direitos dos trabalhadores, os direitos das pessoas com deficiência, o

Reconhecendo que o condicionamento da introdução de bens e serviços no mercado europeu ao cumprimento de deveres específicos implica restrições à liberdade de empresa e à liberdade das artes e das ciências, em seu artigo 6.º (Título III, Capítulo 1) a *Proposta de Regulamento* delimita duas regras claras para a classificação de um sistema de inteligência artificial como sendo de risco elevado, especificamente: (i) caso o sistema de IA destine-se a ser utilizado como componente de segurança de um produto, ou seja ele mesmo um produto que já é objeto de um procedimento de avaliação da conformidade por força da legislação da União Europeia; e (ii) caso seja um sistema de IA autónomo que, em função da sua finalidade prevista, represente um risco elevado de danos para a saúde e a segurança ou de prejuízo para os direitos fundamentais das pessoas, tendo em conta a gravidade dos possíveis danos e a probabilidade dessa ocorrência, e se forem utilizados num conjunto de domínios especificamente predefinidos no regulamento.

Primeiramente, com relação aos sistemas de IA que constituem componentes de segurança ou são eles mesmos produtos já objeto de avaliação e certificação, a aplicação de deveres de comportamento e a obrigatoriedade de certificação para a segurança dos consumidores têm como foco garantir a integração da regulação da inteligência artificial à legislação de segurança setorial em vigor, assegurando a coerência legislativa, evitando duplicações e minimizando os encargos adicionais. Assim sendo, tendo em consideração a lista de produtos já sujeitos à certificação por força da legislação europeia setorial, descrita no Anexo II da *Proposta de Regulamento*, tem-se que são classificados como sistemas de IA de risco elevado os algoritmos utilizados como sistemas de segurança dos seguintes produtos: máquinas, brinquedos, ascensores, aparelhos e sistemas de proteção destinados a ser utilizados em atmosferas potencialmente explosivas, equipamentos de rádio, equipamentos sob pressão, equipamentos de embarcações de recreio, instalações por cabo, aparelhos a gás, dispositivos médicos e dispositivos médicos para diagnóstico *in vitro*.

Na medida em que a *Proposta de Regulamento* esclarece que a classificação deste tipo de sistema como de risco elevado tem como objetivo "prevenir e atenuar devidamente os riscos de segurança que possam ser

---

direito à ação e a um tribunal imparcial, a presunção de inocência e o direito de defesa e o direito a uma boa administração".

criados por um produto devido aos seus componentes digitais, incluindo sistemas de IA" (Considerando 28), como consequência da integração da *Proposta de Regulamento* ao arcabouço legislativo europeu de proteção aos consumidores, tem-se que a verificação do cumprimento dos requisitos aplicáveis aos sistemas de IA será realizado no âmbito do chamado *novo quadro legislativo* (NQL), com a avaliação e a certificação dos requisitos obrigatórios descritos na *Proposta de Regulamento* em conjunto com os mecanismos de conformidade e execução *ex ante* e *ex post* aplicáveis aos produtos acima identificados[7].

Relativamente ao segundo tipo de sistemas de inteligência artificial classificado como de risco elevado, a *Proposta de Regulamento* esclarece que esta classificação é apropriada se, em função da finalidade prevista, os algoritmos "representarem um risco elevado de danos para a saúde e a segurança ou de prejuízo para os direitos fundamentais das pessoas, tendo em conta a gravidade dos possíveis danos e a probabilidade dessa ocorrência, e se forem utilizados num conjunto de domínios especificamente predefinidos no regulamento" (Considerando 32). Desta forma, em seu Anexo III, a *Proposta de Regulamento* traz uma lista exaustiva com um número limitado de sistemas de IA cujos riscos já se materializaram ou são suscetíveis de se materializar num futuro próximo, incluídos em um dos seguintes domínios[8]: identificação biométrica e caracterização de pessoas singulares à distância em tempo real ou em diferido, nas exceções autorizadas; gestão e funcionamento de infraestruturas críticas, como trânsito rodoviário e redes de abastecimento de água, gás, aquecimento e eletricidade; educação e formação profissional; acesso a serviços privados e públicos essenciais; manutenção

---

[7] Segundo descreve a Comissão "a principal diferença é que os mecanismos de *ex ante* e *ex post* assegurarão o cumprimento não só dos requisitos estabelecidos pela legislação setorial, mas também dos requisitos estabelecidos pelo presente regulamento".

[8] Por força do artigo 7.º da *Proposta de Regulamento,* a Comissão Europeia mantém poderes para adotar Atos Delegados (art. 73.º) para atualizar a lista do Anexo III, aditando os sistemas de IA considerados de risco elevado, desde que preencham cumulativamente 2 requisitos: que os sistemas de IA destinem-se a ser utilizados em qualquer um dos domínios enumerados no anexo III, pontos 1 a 8 e que representem um risco de danos para a saúde e a segurança ou um risco de impacto adverso nos direitos fundamentais que, em termos de gravidade e probabilidade de ocorrência, é equivalente ou superior ao risco de danos ou impacto adverso representado pelos sistemas de IA de risco elevado já referidos no anexo III.

da ordem pública; gestão de migração, asilo e controlo de fronteiras; e administração da justiça e processos democráticos.

Com relação aos sistemas de IA concebidos para a identificação biométrica à distância de pessoas singulares, a Comissão esclarece que este tipo de sistema deve ser considerado de risco elevado na medida em que pode conduzir a resultados enviesados e ter efeitos discriminatórios, particularmente no que diz respeito à idade, à etnia, ao sexo ou a deficiências das pessoas. Assim, tendo em consideração seus riscos inerentes, os sistemas de identificação biométrica, tanto em sua forma «em tempo real» quanto «em diferido», devem estar sujeitos a requisitos específicos relativos às capacidades de registo e à supervisão humana (Considerando 33). Aliás, o cuidado da Comissão com este tipo de sistema é tal que os sistemas de identificação biométrica de pessoas é o único tipo de sistema de inteligência artificial de risco elevado cuja avaliação de conformidade não poderá ser feita pelo próprio fornecedor, sendo obrigatoriamente necessária a participação de um organismo notificado (Considerando 64).

De outro lado, a utilização de sistemas de IA na gestão e funcionamento de infraestruturas críticas como componentes de segurança no controlo do tráfego rodoviário e das redes de abastecimento de água, gás, aquecimento e eletricidade, é tida como de risco elevado uma vez que a Comissão entende que uma falha ou anomalia nestes sistemas pode pôr em risco a vida e a saúde das pessoas em larga escala e provocar perturbações substanciais das atividades sociais e económicas normais (Considerando 34).

Outro domínio no qual sistemas de inteligência artificial são considerados de risco elevado é o da educação ou formação profissional, designadamente quando utilizados para determinar o acesso ou a afetação de pessoas a instituições de ensino e de formação profissional ou para avaliar testes que as pessoas realizam no âmbito da sua educação ou como pré-condição para a mesma. Aqui, a Comissão destaca no Considerando 35 que o uso de sistemas de inteligência artificial possui um risco inerente uma vez que, caso sejam indevidamente concebidos e utilizados, estes sistemas podem violar o direito à educação e à formação, bem como o direito a não ser alvo de discriminação e de perpetuação de padrões históricos de discriminação, o que possui importância central para a sociedade na medida em que este tipo de violação influencia diretamente o percurso académico e profissional

das pessoas e, por consequência, a sua capacidade de garantir a própria subsistência.

Refletindo e complementando a tutela anterior da capacidade de subsistência dos trabalhadores da União Europeia, também são considerados de risco elevado os sistemas de inteligência artificial aplicados ao domínio do emprego, da gestão de trabalhadores e de acesso ao emprego por conta própria. Na medida em que a utilização de sistemas de IA como assistente decisório pode levar recrutadores, empregadores e gestores de recursos humanos a decisões erradas ou enviesadas que perpetuam padrões históricos de discriminação (por exemplo, contra as mulheres, certos grupos etários, pessoas com deficiência ou pessoas de uma determinada origem racial ou étnica ou orientação sexual), a *Proposta de Regulamento* procura gerir o uso da inteligência artificial em dois âmbitos principais de relações trabalhistas: a seleção e recrutamento de novos trabalhadores, designadamente para divulgação de vagas, aplicações de triagem ou filtragem de currículos, e avaliação de candidatos; e na avaliação de desempenho durante a vigência de relações de trabalho, na tomada de decisões sobre promoções ou despedimentos, sobre a repartição de tarefas e, por fim, no controlo e avaliação do desempenho e do comportamento dos trabalhadores (Considerando 36).

Também sob a perspetiva de evitar a ocorrência de discriminação de pessoas ou grupos, evitar a criação de impactos discriminatórios e impedir a perpetuação de padrões históricos de discriminação em razão da origem étnica ou racial, deficiência, idade ou orientação sexual, ou criar novas formas de impactos discriminatórios, são classificados como sendo de risco elevado os sistemas de inteligência artificial utilizados para gerir o acesso a determinados serviços e prestações essenciais, tanto de cariz privado quanto público, e o usufruto dos mesmos. Segundo destaca a Comissão (Considerando 37), a classificação destes sistemas como sendo de risco elevado decorre do impacto potencial que sistemas discriminatórios podem trazer para as pessoas e da importância dos serviços públicos e privados para que as pessoas participem plenamente na sociedade ou melhorem o seu nível de vida.

No âmbito dos serviços privados, o foco da *Proposta de Regulamento* é o acesso da população ao setor financeiro e à possibilidade das pessoas de terem acesso a recursos financeiros ou a serviços essenciais, como o alojamento, a eletricidade e os serviços de telecomunicações, motivo pelo qual são considerados de risco elevado os sistemas de IA

concebidos para avaliar a classificação de crédito ou a capacidade de endividamento de pessoas singulares[9]. De outro lado, relativamente à utilização de sistemas de IA para gerir o acesso aos serviços públicos, ainda que reconheça a importância de evitar a criação de obstáculos ao desenvolvimento e à utilização de abordagens inovadoras na administração pública, a *Proposta de Regulamento* destaca a importância dos cuidados no desenvolvimento deste tipo de algoritmo, na medida em que as pessoas singulares que se candidatam ou que recebem prestações e serviços de assistência pública dependem dos mesmos e estão numa posição vulnerável face às autoridades competentes motivo pelo qual os sistemas de IA podem ter um impacto significativo na subsistência das pessoas e podem infringir os seus direitos fundamentais, como o direito à proteção social, à não discriminação, à dignidade do ser humano ou à ação.

Por este motivo, são considerados como de risco elevado não apenas os algoritmos concebidos para serem utilizados no envio ou no estabelecimento de prioridades no envio de serviços de resposta a emergências, incluindo bombeiros e assistência medica, mas também os sistemas de IA concebidos para serem utilizados por autoridades públicas ou em nome de autoridades públicas para avaliar a elegibilidade de pessoas singulares quanto a prestações e serviços públicos de assistência, bem como para conceder, reduzir, revogar ou recuperar tais prestações e serviços.

Ainda no âmbito da administração pública, outro tipo de sistema considerado de risco elevado são os algoritmos de inteligência artificial utilizados em ações das autoridades policiais para a manutenção da ordem pública, contexto social particularmente sensível por ser caracterizado por um grau substancial de desequilíbrio de poder e que pode conduzir à vigilância, detenção ou privação da liberdade de uma pessoa singular, bem como ter outros impactos adversos nos direitos fundamentais garantidos pela Carta. Neste ponto, a exatidão, a fiabilidade e a transparência dos sistemas de inteligência artificial possuem

---

[9] A única exceção aplicável são sistemas de avaliação de crédito desenvolvidos e utilizados por fornecedores de pequena dimensão: "Tendo em conta a dimensão bastante limitada do impacto e as alternativas disponíveis no mercado, é apropriado isentar os sistemas de IA utilizados para efeitos de avaliação da capacidade de endividamento e de classificação de crédito que sejam colocados em serviço por fornecedores de pequena dimensão para utilização própria" (Considerando 37).

importância central para evitar impactos adversos na sociedade, reter a confiança do público e assegurar a responsabilidade e vias de recurso eficazes, particularmente porque, primeiro, se não for treinado com dados de alta qualidade, não cumprir os requisitos adequados em termos de exatidão ou solidez ou não tiver sido devidamente concebido e testado antes de ser colocado no mercado ou em serviço, o sistema de IA pode destacar pessoas de uma forma discriminatória ou incorreta e injusta; e, segundo, o exercício de importantes direitos fundamentais processuais, como o direito à ação e a um tribunal imparcial, a presunção de inocência e o direito de defesa, pode ser prejudicado[10] (Considerando 28).

Assim, tendo em conta a natureza das atividades em causa e os riscos associados às mesmas, serão considerados de risco elevado os sistemas de inteligência artificial concebidos para serem utilizados pelas autoridades policiais (i) em avaliações individuais de riscos (tanto para determinar o risco de uma pessoa singular cometer infrações ou voltar a cometer infrações ou quanto para avaliar o risco para potenciais vítimas de infrações penais), (ii) em instrumentos utilizados para detetar o estado emocional de uma pessoa singular (como polígrafos ou instrumentos semelhantes), (iii) para detetar «falsificações profundas», (iv) para avaliar a fiabilidade dos elementos de prova em processos penais, (v) para prever a ocorrência ou a recorrência de uma infração penal real ou potencial com base na definição de perfis de pessoas singulares ou para avaliar os traços de personalidade e as características ou o comportamento criminal passado de pessoas singulares ou grupos, e (vi) para a definição de perfis no decurso da deteção, investigação ou repressão de infrações penais, bem como para o estudo analítico de crimes relativos a pessoas singulares.

O penúltimo domínio de sistema de inteligência artificial aplica-se igualmente à administração pública, especificamente relacionado à utilização na gestão da migração, do asilo e do controlo de fronteiras[11].

---

[10] Segundo prevê a *Proposta de Regulamento,* "os sistemas de IA especificamente concebidos para serem utilizados em processos administrativos por autoridades fiscais e aduaneiras não devem ser considerados sistemas de IA de risco elevado utilizados por autoridades policiais para efeitos de prevenção, deteção, investigação e repressão de infrações penais" (Considerando 38).

[11] Para além de classificar este tipo de sistema de IA como de risco elevado, como forma de assegurar a integração da certificação deste tipo de algoritmo à legislação

Segundo descreve a Comissão (Considerando 39), por tratar-se de um âmbito que afeta pessoas que, via de regra, encontram-se numa posição particularmente vulnerável e que dependem do resultado das ações das autoridades públicas competentes, a exatidão, a natureza não discriminatória e a transparência dos sistemas de IA utilizados nesses contextos são particularmente importantes para garantir o respeito dos direitos fundamentais das pessoas em causa, nomeadamente os seus direitos à livre circulação, à não discriminação, à proteção da vida privada e dos dados pessoais, à proteção internacional e a uma boa administração.

Com isso, são considerados de risco elevado os sistemas de inteligência artificial utilizados neste domínio para detetar o estado emocional de uma pessoa singular (como polígrafos ou instrumentos similares), para avaliar determinados riscos colocados pelas pessoas singulares que entram no território de um Estado-Membro ou pedem um visto ou asilo; para verificar a autenticidade dos documentos apresentados pelas pessoas singulares; para auxiliar as autoridades públicas competentes na análise dos pedidos de asilo, de visto e de autorização de residência e das queixas relacionadas, com o objetivo de estabelecer a elegibilidade das pessoas singulares que requerem determinado estatuto.

Finalmente, o último tipo de sistema de inteligência artificial considerado de risco elevado são aqueles aplicáveis à administração da justiça e aos processos democráticos, ou seja, algoritmos concebidos para auxiliar as autoridades judiciárias na investigação e na interpretação de factos e do direito e na aplicação da lei a um conjunto específico de factos. O foco central da *Proposta de Regulamento* está associado ao potencial de impacto negativo significativo que enviesamentos, erros e opacidade de sistemas de IA podem trazer à democracia, ao Estado de direito e às liberdades individuais, bem como ao direito à ação e a um tribunal imparcial. Tendo em conta os riscos específicos que o texto legislativo busca tutelar, sistemas de IA concebidos para atividades administrativas puramente auxiliares que não afetam a administração efetiva da justiça em casos individuais (a exemplo da anonimização ou a pseudonimização de decisões judiciais, documentos ou dados,

---

setorial em vigor, a *Proposta de Regulamento* ainda determina que os sistemas de IA aplicáveis no domínio da gestão da migração, do asilo e do controlo das fronteiras devem cumprir os requisitos processuais estabelecidos na Diretiva 2013/32/UE do Parlamento Europeu e do Conselho, no Regulamento (CE) n.º 810/2009 do Parlamento Europeu e do Conselho e noutra legislação aplicável.

comunicações entre pessoal, tarefas administrativas ou afetação de recursos) não estão abrangidas por esta classificação.

6. Supervisão e certificação de sistemas de inteligência artificial de risco elevado

Tendo em conta os riscos potenciais à sociedade criados pelos sistemas de inteligência artificial de risco elevado e como forma de tutelar eventuais danos para a saúde, a segurança ou para os direitos fundamentais das pessoas, a *Proposta de Regulamento* cria a obrigatoriedade de que este tipo de sistema cumpra um conjunto de requisitos obrigatórios horizontais para uma IA de confiança, que serão subsequentemente operacionalizados por via de normas técnicas harmonizadas. Como forma de integrar a supervisão dos sistemas de inteligência artificial ao arcabouço legislativo da União Europeia de proteção aos consumidores, de segurança dos produtos e de garantia da livre circulação de produtos e serviços no mercado europeu, a *Proposta de Regulamento* segue o modelo já utilizado no âmbito do *novo quadro legislativo* (NQL), instituído para a avaliação e certificação de determinados produtos, e cria dois tipos de relações regulatórias relacionados, primeiro, à testagem, prestação de informações e documentação *antes* da colocação de um sistema de IA no mercado e, segundo, de controlo, manutenção de registos e prestação de informações sobre incidentes graves ou anomalias no *pós-comercialização*, durante todo o ciclo de vida do algoritmo.

Primeiramente, como forma de assegurar um nível elevado de fiabilidade dos sistemas de IA de risco elevado perante os consumidores europeus, antes de serem colocados no mercado ou em serviço estes sistemas deverão passar por um procedimento de *avaliação de conformidade* (art. 19), que vai verificar o cumprimento de todos os requisitos obrigatórios determinados pelo Capítulo 2, Título III, da *Proposta de Regulamento* (Considerando 62). Conforme aponta a Comissão Europeia, estas "obrigações relativas à testagem *ex ante*, à gestão de riscos e à supervisão humana também facilitarão o respeito de outros direitos fundamentais, graças à minimização do risco de decisões assistidas por IA erradas ou enviesadas em domínios críticos como a educação e a formação, o emprego, serviços essenciais, a manutenção da ordem pública e o sistema judicial".

Aqui, surge um traço característico do sistema europeu: na busca de criar uma intervenção jurídica equilibrada e proporcional e como forma de minimizar os encargos impostos aos desenvolvedores, especialmente no caso de pequenas e médias empresas, esta avaliação de conformidade será uma autoavaliação realizada pelo fornecedor sob a sua própria responsabilidade (Considerando 64), que vai consolidar o cumprimento de todos os requisitos obrigatórios em uma documentação técnica (art. 11). Como já se mencionou, a *Proposta de Regulamento* prevê uma única exceção aplicável aos sistemas de IA concebidos para serem utilizados para a identificação biométrica à distância de pessoas singulares, caso em que os sistemas deverão sempre ser certificados pelos chamados organismos notificados que verificarão o cumprimento de todos os requisitos técnicos obrigatórios (Considerando 65).

Conforme explicitado no artigo 43.º da *Proposta de Regulamento* e detalhado em seu Anexo VI, via de regra, a avaliação de conformidade será feita com base no controlo interno, sendo realizada pelo fornecedor uma autoavaliação acerca do cumprimento de todos os requisitos obrigatórios e das normas técnicas aplicáveis, posteriormente consolidada em uma *documentação técnica* que ficará à disposição da autoridade de supervisão (os chamados organismos notificados) pelo prazo de 10 anos (art. 50.º). Nesta documentação técnica (art. 11.º e Anexo IV), deverão ser disponibilizadas aos organismos notificados todas as informações necessárias para aferir a conformidade do sistema de IA com esses requisitos, incluindo: uma descrição geral do sistema[12]; uma descrição pormenorizada dos elementos do sistema de IA e do respetivo processo de desenvolvimento[13]; informações pormenorizadas sobre

---

[12] Conforme descrito no Anexo IV, n. 1, neste ponto devem constar: a) A finalidade prevista, a(s) pessoa(s) responsáveis pelo seu desenvolvimento, a data e a versão do sistema; b) De que forma o sistema de IA interage ou pode ser utilizado para interagir com *hardware* ou *software* que não faça parte do próprio sistema de IA, se for caso disso; c) As versões do *software* ou *firmware* instalado e quaisquer requisitos relacionados com a atualização das versões; d) A descrição de todas as formas sob as quais o sistema de IA é colocado no mercado ou colocado em serviço; e) A descrição do *hardware* no qual se pretende executar o sistema de IA; f) Se o sistema de IA for um componente de produtos, fotografias ou ilustrações que revelem as características externas, a marcação e a disposição interna desses produtos; e g) Instruções de utilização para o utilizador e, se for caso disso, instruções de instalação.

[13] Segundo pontua o Anexo IV, n. 2, neste ponto devem constar: a) Os métodos utilizados e os passos dados com vista ao desenvolvimento do sistema de IA, incluindo, se for caso disso, o recurso a sistemas ou ferramentas previamente treinados

o acompanhamento, o funcionamento e o controlo do sistema de IA[14]; uma descrição pormenorizada do sistema de gestão de riscos; a descrição de todas as alterações introduzidas no sistema ao longo do seu ciclo

---

fornecidos por terceiros e de que forma estes foram utilizados, integrados ou modificados pelo fornecedor; b) As especificações de conceção do sistema, designadamente a lógica geral do sistema de IA e dos algoritmos; as principais opções de conceção, nomeadamente a lógica subjacente e os pressupostos utilizados, também no respeitante às pessoas ou grupos de pessoas em relação às quais se pretende que o sistema seja utilizado; as principais opções de classificação; o que se pretende otimizar com o sistema e a importância dos diferentes parâmetros; as decisões acerca de eventuais cedências em relação às soluções técnicas adotadas para cumprir os requisitos definidos no título III, capitulo 2; c) A descrição da arquitetura do sistema, explicando de que forma os componentes de *software* se incorporam ou enriquecem mutuamente e como se integram no processamento global; os recursos computacionais utilizados para desenvolver, treinar, testar e validar o sistema de IA; d) Se for caso disso, os requisitos de dados em termos de folhas de dados que descrevam as metodologias e técnicas de treino e os conjuntos de dados de treino utilizados, incluindo informações sobre a proveniência desses conjuntos de dados, o seu âmbito e as suas principais características; de que forma os dados foram obtidos e selecionados; procedimentos de rotulagem (por exemplo, para aprendizagem supervisionada), metodologias de limpeza de dados (por exemplo, deteção de valores atípicos); e) Análise das medidas de supervisão humana necessárias em conformidade com o artigo 14.o, incluindo uma análise das soluções técnicas necessárias para facilitar a interpretação dos resultados dos sistemas de IA pelos utilizadores, em conformidade com o artigo 13.o, n.o 3, alínea d); f) Se for caso disso, uma descrição pormenorizada das alterações predeterminadas do sistema de IA e do seu desempenho, juntamente com todas as informações pertinentes relacionadas com as soluções técnicas adotadas para assegurar a conformidade contínua do sistema de IA com os requisitos aplicáveis estabelecidos no título III, capítulo 2; g) Os procedimentos de validação e teste aplicados, incluindo informações sobre os dados de validação e teste utilizados e as principais características desses dados; as métricas utilizadas para aferir a exatidão, a solidez, a cibersegurança e a conformidade com outros requisitos aplicáveis estabelecidos no título III, capítulo 2, bem como potenciais impactos discriminatórios; registos dos testes e todos os relatórios de teste datados e assinados pelas pessoas responsáveis, incluindo no respeitante às alterações predeterminadas referidas na alínea f).

[14] Especialmente no que diz respeito às suas capacidades e limitações de desempenho, incluindo os níveis de exatidão no tocante a pessoas ou grupos de pessoas específicos em relação às quais se pretende que o sistema seja utilizado e o nível geral esperado de exatidão em relação à finalidade prevista; os resultados não pretendidos mas previsíveis e as fontes de riscos para a saúde e a segurança, os direitos fundamentais e a proteção contra a discriminação atendendo à finalidade prevista do sistema de IA; as medidas de supervisão humana necessárias em conformidade com o artigo 14.o, incluindo as soluções técnicas adotadas para facilitar a interpretação dos resultados dos sistemas de IA pelos utilizadores; especificações relativas aos dados de entrada, consoante apropriado (Anexo IV, n. 3).

de vida; uma lista das normas técnicas aplicadas total ou parcialmente; e uma descrição pormenorizada do sistema existente para avaliar o desempenho do algoritmo na fase de pós-comercialização.

De outro lado, porém, quando não for possível ao fornecedor aplicar total ou parcialmente normas harmonizadas desenvolvidas por organismos de normalização técnica ou quando estas normas não tiverem sido desenvolvidas, a avaliação de conformidade deverá ser baseada na análise do sistema de gestão de qualidade e da documentação técnica, caso em que, seguindo o procedimento detalhado no Anexo VII da *Proposta de Regulamento*, o processo de avaliação do cumprimento dos requisitos obrigatórios será realizado pelo organismo notificado que, ao final, vai emitir uma decisão fundamentada com as conclusões da sua avaliação acerca do sistema de inteligência artificial.

Importa mencionar que, para realizar esta avaliação, o organismo notificado deverá dispor de total acesso aos conjuntos de dados de treino e teste utilizados pelo fornecedor, incluindo através de interfaces de programação de aplicações ou outros meios e ferramentas adequadas que possibilitem o acesso remoto (Anexo VII, 4.3). Ao final da avaliação, o organismo notificado pode aprovar a documentação técnica emitindo um certificado UE de avaliação com validade máxima de 5 anos (prorrogável por iguais períodos mediante reavaliação, art. 44, n. 2), ou recusar a emissão do certificado UE informando o requerente do facto, fundamentando pormenorizadamente as razões da sua recusa. Em específico, caso a recusa se dê pelo não cumprimento dos requisitos relativos aos dados utilizados para treinar o sistema de IA, será necessário voltar a treinar o sistema de IA antes da apresentação do pedido de nova avaliação da conformidade (Anexo VII, n. 4.7).

Ainda como forma de aumentar a transparência e a supervisão públicas, de reforçar a supervisão pós-comercialização por parte das autoridades competentes e de garantir a confiança dos consumidores em produtos e serviços que utilizem algoritmos de inteligência artificial, após realizada a avaliação de conformidade necessária, os fornecedores deverão registar os sistemas de IA de risco elevado em uma base de dados unificada pública gerida pela Comissão Europeia na forma do artigo 60.º (Título VII). Finalmente, como forma de garantir a livre circulação dentro do mercado interno europeu, os sistemas de inteligência artificial de risco elevado devem apresentar, de modo visível, legível e indelével, a marcação CE para indicar o cumprimento de todos

os requisitos legais previstos na *Proposta de Regulamento* (Considerando 67 e artigo 49.º).

Tendo em consideração o caráter dinâmico e orgânico dos sistemas de inteligência artificial, que continuam a aprender depois de terem sido colocados no mercado ou em serviço, a tutela dos riscos inerentes aos sistemas de risco elevado não acaba em sua certificação técnica *ex ante*, motivo pelo qual, em seu Título VIII, a *Proposta de Regulamento* traz uma série de deveres de acompanhamento pós-comercialização na forma de obrigações de controlo e de comunicação aplicáveis aos fornecedores durante todo o ciclo de vida do algoritmo. Assim, em primeiro lugar, sob a perspetiva de tutelar eventuais riscos decorrentes de sistemas de IA de risco elevado, a *Proposta de Regulamento* obriga os fornecedores a informar as autoridades nacionais sobre incidentes graves ou anomalias que constituam violações do direito nacional e da União em matéria de direitos fundamentais assim que tomarem conhecimento das mesmas[15], bem como sobre eventuais recolhas ou retiradas de sistemas de IA do mercado (art. 62.º).

De outro lado, como forma de aproveitar a experiência adquirida na utilização de sistemas de IA de risco elevado, primeiro, para melhorar os algoritmos bem como os seus processos de conceção e desenvolvimento e, segundo, para assegurar uma resolução mais eficaz e atempada dos eventuais riscos observados após a colocação no mercado, a *Proposta de Regulamento* cria a obrigação para fornecedores de desenvolver e documentar um plano de acompanhamento pós-comercialização que seja proporcionado à natureza das tecnologias de IA e aos riscos do sistema (art. 61.º). Conforme detalham os artigos 3.º, n. 25 e 61.º do texto legislativo, o sistema de acompanhamento pós-comercialização deve recolher, documentar e analisar de forma ativa e sistemática os dados pertinentes fornecidos pelos utilizadores ou recolhidos por meio de outras fontes sobre o desempenho dos sistemas de IA de risco elevado ao longo da sua vida útil, bem como permitir ao fornecedor avaliar a contínua conformidade dos sistemas de IA com os requisitos estabelecidos no título III, capítulo 2.

---

[15] Conforme detalhado no artigo 62, n. 1, 2º parágrafo, "Essa notificação deve ser efetuada imediatamente após o fornecedor ter determinado uma relação causal entre o sistema de IA e o incidente ou anomalia ou a probabilidade razoável dessa relação e, em qualquer caso, o mais tardar 15 dias após o fornecedor ter conhecimento do incidente grave ou da anomalia".

7. Abordagem de caixa de areia e os ambientes de testagem da regulamentação

Neste contexto amplo da supervisão e certificação de sistemas de inteligência artificial de risco elevado, a Comissão Europeia destaca que, de um lado, as obrigações relativas à testagem *ex ante*, à gestão de riscos e à supervisão humana facilitarão o respeito de outros direitos fundamentais, graças à minimização do risco de decisões assistidas por IA erradas ou enviesadas em domínios críticos como a educação e a formação, o emprego, serviços essenciais, a manutenção da ordem pública e o sistema judicial. Complementarmente, os controlos e deveres de comportamento *ex post* inseridos no sistema de acompanhamento pós-comercialização funcionarão, para os consumidores europeus, como garantia de acesso a vias eficazes de recursos diante de eventuais violações a direitos fundamentais.

Ainda que assim seja, mesmo argumentando que "uma avaliação da conformidade *ex ante* abrangente por meio de controlos internos, aliada a uma forte execução *ex post*, poderá constituir uma solução eficaz e razoável para esses sistemas", a Comissão reconhece que o modelo de supervisão delineado na *Proposta de Regulamento* pode não ser o ideal para a tutela dos riscos conhecidos e desconhecidos oriundos da inteligência artificial, especialmente "dada a fase inicial da intervenção regulamentar e o facto de o setor da inteligência artificial ser bastante inovador e de só agora estarem a ser reunidos conhecimentos especializados para as auditorias".

Assim sendo, reconhecendo que a inteligência artificial é uma família de tecnologias em rápida evolução que exige novas formas de supervisão regulamentar e um espaço seguro para a experimentação, garantindo ao mesmo tempo uma inovação responsável e a integração de salvaguardas e medidas de atenuação dos riscos adequadas, a Comissão deixa uma margem de mudança e evolução às relações regulatórias através da adoção da chamada *abordagem de caixa de areia* (*sandbox approach*)[16] com a

---

[16] Esta abordagem de caixa de areia reflete o binómio excelência *vs* confiança estabelecido ainda no Livro Branco e pode ser encontrada também na revisão de 2021 do Plano Coordenado Para a Inteligência Artificial em que a Comissão esclarece que "Na sua essência, estes ambientes proporcionam uma instalação de experimentação para a regulamentação pública e permitem uma avaliação mais rápida do impacto da intervenção pública".

criação de *ambientes de testagem da regulamentação* que facilitem o desenvolvimento e o teste de sistemas de IA inovadores sob uma supervisão regulamentar rigorosa, antes que estes sistemas sejam colocados no mercado ou em serviço (Considerando 71).

Para isso, a Comissão incentiva os Estados-Membros a criar ambientes controlados para testar tecnologias inovadoras durante um período limitado com base num plano de testagem acordado com as autoridades competentes que reflita quatro objetivos principais, nomeadamente: (i) fomentar a inovação no domínio da IA, mediante a criação de um ambiente controlado de experimentação e teste na fase de desenvolvimento e pré-comercialização que assegure o cumprimento da legislação aplicável; (ii) reforçar a segurança jurídica para os inovadores; (iii) melhorar a supervisão e a compreensão, por parte das autoridades competentes, das oportunidades, dos riscos emergentes e dos impactos da utilização da inteligência artificial; e (iv) acelerar o acesso aos mercados, nomeadamente por via da eliminação dos entraves para as pequenas e médias empresas (PME) e as empresas em fase de arranque (Considerando 72).

8. Deveres de comportamento e elementos obrigatórios em sistemas de inteligência artificial de risco elevado

Como já se mencionou, partindo do entendimento de que os sistemas de inteligência artificial de risco elevado possuem um potencial de risco relevante para a saúde, a segurança e os direitos fundamentais dos consumidores europeus, a *Proposta de Regulamento* estabelece regras harmonizadas para o desenvolvimento, a colocação no mercado e a utilização de sistemas de IA na União Europeia, consolidando em seu Título III, capítulo 2, uma série de requisitos de segurança aplicáveis a este tipo de sistema de inteligência artificial. Em coerência com as recomendações e princípios internacionais[17] e como resultado de um trabalho preparatório de dois anos desenvolvido pelo Grupo de Peritos de Alto Nível sobre a Inteligência Artificial a *Proposta de Regulamento*

---

[17] Entre os documentos internacionais relevantes, cumpre citar os relatórios *The Age of Digital Interdependence*, publicado pela Organização das Nações Unidas, e *Artificial Intelligence in Society*, publicado pela Organização para a Cooperação e Desenvolvimento Económico (OCDE).

estabelece os requisitos legais aplicáveis aos sistemas de IA de risco elevado relativamente aos dados (isto é, à qualidade dos conjuntos de dados e à governação de dados), à documentação e à manutenção de registos, à transparência e à prestação de informações aos utilizadores, à supervisão humana, à solidez, à exatidão e à segurança.

Neste contexto, como forma de atenuar eficazmente os riscos e proteger os consumidores europeus, os fornecedores de sistemas de inteligência artificial de risco elevado devem criar e manter um sistema de gestão de qualidade[18] (art. 17.º) sólido e que documente, de forma sistemática e ordenada, políticas, procedimentos e instruções escritas para o cumprimento da legislação e dos requisitos técnicos aplicáveis aos algoritmos, em torno de seis elementos centrais, nomeadamente, procedimentos de gestão e governação de dados (art. 10), um sistema permanente de gestão de risco (art. 9), procedimentos de manutenção de registos (art. 12), garantia da transparência e da prestação de

---

[18] Conforme detalha o artigo 17.º da Proposta de Regulamento, o sistema de gestão da qualidade deve ser proporcionado à dimensão da organização do fornecedor e deve incluir, no mínimo, uma estratégia clara o cumprimento da regulamentação, incluindo procedimentos de avaliação da conformidade e de gestão de modificações do sistema de IA de risco elevado; descrição das técnicas, procedimentos e ações sistemáticas que serão utilizadas para o desenvolvimento, o controlo da qualidade e a garantia da qualidade sistema e para a conceção, o controlo da conceção e a verificação da conceção do sistema a terceiros; procedimentos de exame, teste e validação a realizar antes, durante e após o desenvolvimento sistema e a frequência com a qual serão realizados; a descrição dos testes e procedimentos que devem ser feitos em ambientes controlados que facilitem o desenvolvimento, a testagem e a validação dos sistemas, sob supervisão e orientação das autoridades competentes (art. 53); as especificações técnicas do sistema, incluindo normas técnicas a aplicar, meios a usar para assegurar que o sistema cumpre os requisitos, se as normas harmonizadas em causa não forem aplicadas na íntegra; sistemas e procedimentos de gestão de dados, incluindo sua recolha, análise, rotulagem, armazenamento, filtragem, prospecção, agregação, conservação e quaisquer outras operações relativa aos dados que realizadas antes e para efeitos da colocação no mercado ou colocação em serviço; o sistema de gestão de riscos (artigo 9.º); estabelecimento, aplicação e manutenção de um sistema de acompanhamento pós-comercialização (art. 61.º); procedimentos de comunicação de incidentes graves e de anomalias (art. 62.º); a gestão da comunicação com autoridades nacionais competentes, incluindo as autoridades setoriais, organismos notificados, outros operadores, clientes ou outras partes interessadas; sistemas e procedimentos de manutenção de registos de toda a documentação e informação importante; gestão de recursos, incluindo medidas relacionadas com a segurança do aprovisionamento; um quadro que defina as responsabilidades do pessoal com funções de gestão e do restante pessoal no atinente a todos os aspetos elencados no presente número.

informações (art. 13), mecanismos de supervisão humana (art. 14) e garantia de exatidão, solidez e cibersegurança (art. 15).

Diante da importância da disponibilidade de dados de elevada qualidade para o desempenho de vários sistemas de IA com vista a assegurar que o sistema de IA de risco elevado funcione como pretendido e de modo seguro e não se torne a fonte de uma discriminação proibida pelo direito da União, sobretudo quando são utilizadas técnicas que envolvem o treino de modelos, o primeiro elemento relevante para o sistema de gestão de qualidade pode ser identificado nos procedimentos de governação de dados (Considerando 44 e art. 10). Em casos de modelos de inteligência artificial treinados com dados, os fornecedores devem garantir não apenas a fiabilidade e a solidez dos procedimentos de recolha, pré-processamento e processamento dos dados[19], mas também devem garantir a qualidade dos conjuntos de dados de treino, validação e teste, assegurando que são pertinentes, representativos, isentos de erros e completos; que têm as propriedades estatísticas adequadas no tocante às pessoas ou grupos de pessoas em que o sistema será utilizado; e que têm em conta as características, os elementos idiossincráticos do enquadramento geográfico, comportamental e funcional em que o algoritmo deve ser utilizado.

Um segundo elemento essencial aos sistemas de gestão de qualidade de sistemas de inteligência artificial de risco elevado são os procedimentos de manutenção de registos, de que trata o artigo 12.º da *Proposta de Regulamento*. Considerada essencial para verificar o cumprimento dos requisitos estabelecidos na *Proposta de Regulamento* e para a identificação e resolução e atempada de eventuais riscos identificados após a colocação no mercado, a manutenção de registos deve incluir informações relevantes como as características gerais, as capacidades e as limitações do sistema, os algoritmos, os dados e os processos de

---

[19] Conforme descreve o artigo 10.º, os fornecedores devem garantir que os conjuntos de dados de treino, validação e teste estão sujeitos a práticas adequadas de governação e gestão de dados relacionados às escolhas de conceção tomadas; aos procedimentos de recolha de dados; ao pré-processamento dos dados, isto é, à preparação e o tratamento de dados, tais como anotação, rotulagem, limpeza, enriquecimento e agregação dos dados; à formulação dos pressupostos aplicáveis, nomeadamente informações que os dados devem medir e representar; à avaliação prévia da disponibilidade, quantidade e adequação dos conjuntos de dados que são necessários; aos exames para detetar eventuais enviesamentos; à identificação de eventuais lacunas ou deficiências de dados e de possíveis soluções para as mesmas.

treino, teste e validação utilizados, bem como documentação sobre o sistema de gestão de riscos aplicado (Considerando 46).

Neste contexto, será obrigatório o desenvolvimento de um sistema de registo automático de eventos enquanto o algoritmo estiver em funcionamento, registos esses que deverão ser mantidos por um período adequado em função da finalidade prevista do sistema e das obrigações legais aplicáveis. Ademais, deverá ser assegurado um nível de rastreabilidade do funcionamento adequado à finalidade prevista do sistema, além de ser assegurado que as capacidades de registo permitirão o controlo do funcionamento do sistema contra riscos para a saúde e segurança ou para direitos fundamentais, contra modificações substanciais e para facilitar o acompanhamento pós-comercialização do algoritmo.

Outro elemento essencial para a gestão dos sistemas de inteligência artificial de risco elevado é o sistema permanente de gestão de riscos (art. 9.º), que tem como objetivo fazer com que eventuais riscos residuais associados a cada perigo e eventuais riscos globais associados ao algoritmo possam ser considerados aceitáveis, quando do uso normal do sistema ou quando houver um uso indevido razoavelmente previsível. Tendo este objetivo em mente, o sistema permanente de gestão de riscos deverá consistir num processo iterativo contínuo, executado ao longo de todo o ciclo de vida de um sistema de IA de risco elevado e regularmente submetido a atualizações sistemáticas, dividido em quatro etapas: (i) identificação e análise dos riscos conhecidos e previsíveis associados a cada sistema de IA de risco elevado; (ii) estimativa e avaliação de riscos que podem surgir quando o sistema de IA de risco elevado é usado em conformidade com a sua finalidade prevista e em condições de utilização indevida razoavelmente previsíveis; (iii) avaliação de outros riscos que possam surgir, baseada na análise dos dados recolhidos a partir do sistema de acompanhamento pós-comercialização; e (iv) adoção de medidas de gestão de riscos adequadas levando consideração, primeiro, os efeitos e eventuais interações resultantes da aplicação combinada dos requisitos obrigatórios e, segundo, o estado da técnica geralmente reconhecido em normas harmonizadas ou especificações comuns pertinentes.

Reconhecendo que os sistemas de inteligência artificial podem ser caracterizados por uma opacidade que os pode tornar incompreensíveis ou demasiado complexos para as pessoas singulares, a *Proposta de Regulamento* torna também obrigatória a garantia da transparência e a

prestação de informações para os utilizadores e consumidores europeus de forma a garantir que sejam capazes de interpretar o resultado do sistema e utilizá-lo de forma adequada (Considerando 47). Com isso, por força do artigo 13 do texto legislativo, os sistemas de IA de risco elevado devem ser acompanhados de documentação pertinente e instruções de utilização, em formato digital ou outro adequado, além de incluir informações concisas, completas, claras e compreensíveis relativas a possíveis riscos para os direitos fundamentais e de discriminação, se for caso disso[20].

Outra forma importante de tutela do risco inerente aos sistemas de IA previsto exigido pela Comissão Europeia são os mecanismos de supervisão humana (art. 14.º), em decorrência dos quais os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de maneira que permita a sua supervisão por pessoas singulares de forma a prevenir ou minimizar os riscos para a saúde, a segurança ou os direitos fundamentais que possam surgir quando um sistema de IA de risco elevado é usado em conformidade com a sua finalidade prevista ou em condições de utilização indevida razoavelmente previsí-

---

[20] Conforme detalha o n. 3 do artigo 13.º, as informações obrigatórias devem especificar:

a) A identidade e os dados de contacto do fornecedor e, se for caso disso, do seu mandatário; b) As características, capacidades e limitações de desempenho do sistema de IA de risco elevado, incluindo: i) a finalidade prevista do sistema, ii) o nível de exatidão, solidez e cibersegurança a que se refere o artigo 15.o relativamente ao qual o sistema de IA de risco elevado foi testado e validado e que pode ser esperado, bem como quaisquer circunstâncias conhecidas e previsíveis que possam ter um impacto nesse nível esperado de exatidão, solidez e cibersegurança, iii) qualquer circunstância conhecida ou previsível, relacionada com a utilização do sistema de IA de risco elevado de acordo com a sua finalidade prevista ou em condições de utilização indevida razoavelmente previsíveis, que possa causar riscos para a saúde e a segurança ou os direitos fundamentais, iv) o desempenho do sistema no tocante às pessoas ou grupos de pessoas em que o sistema se destina a ser utilizado, v) quando oportuno, especificações para os dados de entrada, ou quaisquer outras informações importantes em termos dos conjuntos de dados de treino, validação e teste usados, tendo em conta a finalidade prevista do sistema de IA; c) As alterações do sistema de IA de risco elevado e do seu desempenho que foram predeterminadas pelo fornecedor aquando da avaliação da conformidade inicial, se for caso disso; d) As medidas de supervisão humana a que se refere o artigo 14.o, incluindo as soluções técnicas adotadas para facilitar a interpretação dos resultados dos sistemas de IA pelos utilizadores; e) A vida útil esperada do sistema de IA de risco elevado e quaisquer medidas de manutenção e assistência necessárias para assegurar o correto funcionamento desse sistema de IA, incluindo no tocante a atualizações do software.

veis (Considerando 48). Neste ponto, a *Proposta de Regulamento* torna obrigatória a integração aos sistemas de IA de restrições operacionais que não possam ser neutralizadas pelo próprio sistema, compelindo os sistemas a responder ao operador humano, além de tornar obrigatório que as pessoas singulares a quem foi atribuída a supervisão humana tenham as competências, a formação e a autoridade necessárias para desempenhar essa função[21].

Por fim, o último elemento essencial previsto pela *Proposta de Regulamento* para a tutela dos riscos inerentes aos sistemas de inteligência artificial de risco elevado é a garantia da exatidão, solidez e cibersegurança do algoritmo. Conforme explica a Comissão Europeia nos Considerandos 49 a 50, é importante garantir não apenas que os sistemas de IA tenham um desempenho coerente ao longo de todo o seu ciclo de vida mas, igualmente, que estes sistemas sejam resistentes aos riscos associados às suas limitações inerentes – como erros, falhas, incoerências ou situações inesperadas – e às ações maliciosas suscetíveis de pôr em causa a segurança dos sistema de IA ou mesmo dar origem a comportamentos prejudiciais indesejáveis. Da mesma forma, é essencial garantir que os sistemas de IA sejam resistentes às ações de terceiros mal-intencionados que venham a tentar explorar as vulnerabilidades dos sistemas com o objetivo de lhes alterar a utilização, o comportamento e o desempenho ou por em causa as propriedades de segurança, explorando vulnerabilidades dos ativos digitais do sistema de IA ou

---

[21] Em específico, a *Proposta de Regulamento* torna obrigatória a adoção de medidas, pelo fornecedor, para garantir que as pessoas responsáveis pela supervisão humana dos sistemas de IA de risco elevado (i) Compreendam completamente as capacidades e limitações do sistema de IA de risco elevado e sejam capazes de controlar devidamente o seu funcionamento, de modo que os sinais de anomalias, disfuncionalidades e desempenho inesperado possam ser detetados e resolvidos o mais rapidamente possível; (ii) estejam conscientes da possível tendência para confiar automaticamente ou confiar excessivamente no resultado produzido pelo sistema de IA de risco elevado («enviesamento da automatização»), em especial relativamente aos sistemas de IA de risco elevado usados para fornecer informações ou recomendações com vista à tomada de decisões por pessoas singulares; (iii) sejam capazes de interpretar corretamente o resultado do sistema de IA de risco elevado, tendo em conta, nomeadamente, as características do sistema e as ferramentas e os métodos de interpretação disponíveis; (iv) em qualquer situação, sejam capazes de decidir não usar o sistema de IA de risco elevado ou ignorar, anular ou reverter o resultado do sistema de IA de risco elevado; (v) sejam capazes de intervir no funcionamento do sistema ou interromper o sistema por meio de um botão de «paragem» ou procedimento similar.

da infraestrutura de tecnologias da informação e comunicação (TIC) subjacente.

Por este motivo, os fornecedores devem garantir um nível apropriado de exatidão dos sistemas, informando aos utilizadores as métricas utilizadas para esta avaliação, além de garantir a solidez técnica e a cibersegurança dos algoritmos de acordo com o estado da técnica geralmente reconhecido, evitando a ocorrência de problemas de segurança possam afetar negativamente os direitos fundamentais, por exemplo, devido a decisões erradas ou a resultados errados ou enviesados gerados pelo sistema de IA. Para alcançar estes objetivos, o artigo 15.º da *Proposta de Regulamento* prevê que, tendo em conta a sua finalidade prevista, os sistemas de IA de risco elevado devem ser concebidos com um nível adequado de exatidão, solidez e cibersegurança, sendo resistentes a erros, falhas ou incoerências que possam ocorrer no sistema ou no ambiente em que aquele opera, em especial devido à interação com pessoas singulares ou outros sistemas; e sendo também resistentes a tentativas de terceiros não autorizados de alterar a sua utilização ou desempenho explorando as vulnerabilidades do sistema.

9. Conclusão

A *Proposta de Regulamento* sobre inteligência artificial vem consolidar os princípios e objetivos delineados em documentos anteriores sobre IA e, neste contexto, cria um quadro normativo básico relativo à governação, à supervisão e à responsabilidade em sistemas de inteligência artificial, estabelecendo regras harmonizadas para o desenvolvimento, a colocação no mercado e a utilização de sistemas de IA na União Europeia e criando requisitos essenciais e obrigatórios para determinados tipos de sistemas de IA e de deveres de comportamento relacionados à transparência e à prestação de informações.

Como reflexo da busca pelo equilíbrio regulatório e em consolidação da abordagem europeia centrada no binómio excelência e confiança descrita no *Livro Branco*, a *Proposta de Regulamento* busca criar uma intervenção jurídica equilibrada e proporcional. Para isso, propõe um quadro jurídico sólido centrado em uma *abordagem baseada no risco* que classifica os sistemas de inteligência artificial a partir dos níveis de risco criados pelos sistemas e correlaciona deveres de comportamento específicos e proporcionais a cada tipo, especificamente: *riscos*

inaceitáveis (Título II), que são práticas proibidas em território europeu; *riscos limitados* (Título IV), para os quais existem deveres de informação e transparência para com consumidores; *riscos mínimos* (Título IX), para os quais não existem deveres e obrigações específicos; e, *riscos elevados* (Título III), para os quais são previstas uma série de requisitos relacionados à qualidade dos dados, à documentação e à rastreabilidade, à transparência, à supervisão humana, à exatidão e à solidez.

Para avaliar o cumprimento dos requisitos obrigatórios e certificar os sistemas de inteligência artificial de risco elevado, a *Proposta de Regulamento* segue o modelo já utilizado no âmbito do *novo quadro legislativo*, criando relações regulatórias *ex ante*, relacionados à testagem, prestação de informações e documentação *antes* da colocação de um sistema de IA no mercado; e *ex post,* de controlo, manutenção de registos e prestação de informações sobre incidentes graves ou anomalias no *pós-comercialização*, durante todo o ciclo de vida do algoritmo.

Referências bibliográficas

Ebers, Martin, «Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act», in Larry A. DiMatteo / Michel Cannarsa / Cristina Poncibò, ed., *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, Cambridge University Press 2022.

Godinho, Inês Fernandes / Flores, Cláudio R. / Marques, Nuno Castro, «Consultation on The White Paper on Artificial Intelligence – A European Approach», *ILP Law Review / Revista de Direito da ULP* 14/1 (2021) 157-167, doi:10.46294.

Lilkov, Dimitar, «Regulating artificial intelligence in the EU: A risky game», *European View* 20/2 (2021) 166–174.

Ramos, José Ricardo Marcondes, «Relatório sobre a atual regulação normativa europeia e portuguesa em matéria de Inteligência Artificial», *Revista Portuguesa de Ciência Criminal* 31 (2022) 633-646.

Veale, Michael / Borgesius, Frederik Zuiderveen, «Demystifying the Draft EU Artificial Intelligence Act», *Computer Law Review International* 22/4 (2021) 97–112.