U.PORTO

2º CICLO DE ESTUDOS
CIÊNCIAS JURÍDICO-POLÍTICAS

A categorização do risco no âmbito do Regulamento da Inteligência Artificial

Tomás Peixoto Ferreira

M

2025

Índice

Resumo

A Inteligência Artificial já se tornou parte integrante do quotidiano. Apesar dos amplos potenciais ganhos de produtividade e eficiência, esta tecnologia suscita inúmeros desafios éticos e jurídicos que não podem ser ignorados. Desta forma, a presente dissertação analisa de forma abrangente o Regulamento (UE) 2024/1689 sobre Inteligência Artificial, identificando e comparando as categorias de risco por ele previstas, nomeadamente: práticas de IA proibidas (risco inaceitável), sistemas de risco elevado, obrigações de transparência e modelos de finalidade geral com e sem risco sistémico. A investigação assenta num exame detalhado dos textos legais, das orientações da Comissão Europeia, confrontando-os com alguma literatura e a jurisprudência relevante. Destacam-se as linhas de orientação sobre práticas proibidas (e.g., *social scoring*, técnicas subliminares e a inferência de emoções no local de trabalho), a complexidade das obrigações impostas a sistemas de risco elevado (desde a qualidade dos dados à supervisão humana) e o regime específico de transparência (artigo 50.º). Conclui-se que, embora o Regulamento represente um esforço pioneiro e capaz de fomentar um "ecossistema de confiança" em torno da IA, existem algumas imperfeições que, ultimamente, podem ser ultrapassadas através dos mecanismos regulatórios previstos.

Abstract

Artificial Intelligence has already become an integral part of everyday life. Despite its significant potential to enhance productivity and efficiency, this technology raises numerous ethical and legal challenges that cannot be overlooked. This dissertation provides a comprehensive analysis of Regulation (EU) 2024/1689 on Artificial Intelligence, identifying and comparing the risk categories it establishes, namely: prohibited AI practices (unacceptable risk), high-risk systems, transparency obligations, and general-purpose AI models with and without systemic risk. The research is grounded in the examination of the legal texts and the European Commission's guidelines, considering some relevant academic literature and case law. Particular attention is given to the Commission's guidelines on prohibited practices (e.g., social scoring, subliminal techniques, and emotion inference in the workplace), the complexity of requirements imposed on high-risk systems (ranging from data quality to human oversight), and the specific transparency regime set out in Article 50. The dissertation concludes that, although the Regulation constitutes a pioneering effort capable of fostering a trustworthy AI ecosystem, certain imperfections remain, which may still be ultimately overcome through the available regulatory mechanisms.

1. Introdução

A Inteligência Artificial, nas suas várias faces, deixa de ser, em menos de uma década, um conceito distante e reservado aos confins da ficção científica. As evoluções no âmbito do poder computacional, *machine learning* e *big data*[1] permitiram o amanhecer de um "verdadeiro mundo novo" em que a Inteligência Artificial (doravante também identificada como IA), além de acessível, é cada vez mais capaz e prolífera. Os efeitos da ascensão da IA são, em primeira análise, extremamente entusiasmantes: a promessa, geralmente, de melhorar a qualidade de vida das pessoas e tornar os processos de tomada de decisão mais eficiente tem repercussões positivas em quase todos os setores da vida humana.

De forma superficial, é um fenómeno que pode ser visto meramente como progresso técnico-científico. No entanto, levanta questões complexas em diversos planos, como o político, a ética e, sobretudo, jurídico. A proliferação destes sistemas, de forma geral, gera, além de decisões autónomas, preocupações legítimas quanto ao seu efeito e relação com os direitos fundamentais, à equidade das suas decisões ou resultados e até mesmo à responsabilização das entidades que os desenvolvem, comercializam, implementam, importam e fabricam. A discussão em torno deste tema tem vindo a ganhar relevância, em particular na última década, dados os avanços já mencionados a nível computacional. A relevância da IA está, inegavelmente, no seu auge. Assim surge o esforço regulatório ainda em curso a nível europeu, liderado pelo Regulamento da Inteligência Artificial (Regulamento (UE) 2024/1689, também conhecido como Regulamento da Inteligência Artificial ou RIA). A Proposta do Regulamento do Parlamento Europeu e do Conselho, publicada em 2021, surgindo no seguimento do *Livro Branco sobre a inteligência artificial* da Comissão Europeia (publicado, por sua vez, em 2020) e da comunicação da Comissão intitulada *Inteligência Artificial para a Europa* em 2018, propunha um modelo essencialmente baseado em dois pilares, já previamente identificados e discutidos, a saber: a promoção da excelência e a nutrição de um verdadeiro ecossistema baseado na confiança.

Também se entendeu, como veremos, que uma abordagem com base no risco seria, em vez de uma aproximação regulatória totalmente uniforme que se poderia revelar demasiado rígida,

[1] OCDE – *Artificial Intelligence in Society*. Paris: OECD Publishing, 2019, p. 23. Disponível em: https://doi.org/10.1787/eedfee77-en. – as observações feitas pela OCDE em 2019 permanecem relevantes, na medida em que os seus pressupostos ainda são válidos. O *machine learning* e poder computacional continuam a evoluir nos dias de hoje, tal como a disponibilidade de dados. Desta forma, o alcance da IA continua a crescer dramaticamente.

o mais adequado. Isto na medida em que a mera tecnicidade não seria potencialmente eficaz, dada a opacidade e complexidade dos algoritmos, a dificuldade de prever e compreender os processos de tomada de decisão e a probabilidade real de causar danos graves. Ora, a regulação baseada no risco entenderá, por sua vez, a premissa de que nem todas as utilizações da IA são igualmente problemáticas e sujeitas a criar danos. Desta forma, uma intervenção normativa de geometria variável, com fortes simetrias com outros modelos europeus de regulação (*v.g.* a proteção de dados na UE)[2], e que se ajusta perante o grau de risco apresentado, foi o caminho escolhido, como será visto doravante.

1.1 Evolução da Inteligência Artificial como desafio regulatório.

Ora, é importante compreender que os potenciais benefícios da IA são incontornáveis e numerosos, desde o impacto no setor da saúde e mobilidade até ao setor agrícola. Nos transportes aproveitamos a diminuição da sinistralidade rodoviária, de custos associados aos transportes públicos e de congestionamento rodoviário. Na agricultura, extraímos os frutos da automação através da robótica aplicada aos processos de cultivo e colheita, tal como beneficiamos de novas formas de vigiar a fertilidade dos solos e prever a rentabilidade de uma cultura. Na banca, automatizam-se processos de suporte ao cliente, de conformidade regulatória, de deteção de fraude e utilizam-se algoritmos que determinam o perfil de crédito. O mundo digital torna-se também mais seguro através de monitorização ativa de ameaças cibernéticas, vigilância digital em tempo real e melhoria na segurança das redes. As mais-valias são incontestáveis e já foram amplamente discutidas[3]. A questão, todavia, que surge eventualmente em todas as discussões proveitosas sobre as vantagens da IA gira à volta dos seus possíveis riscos. Se os prós são significativos, os contras também acabam por ser, por sua vez, substanciais e não passíveis de ser ignorados.[4]

---

[2] Para uma análise de algumas abordagens regulatórias europeias e a sua evolução (RGPD, Regulamento dos Serviços Digitais e Regulamento da IA), cf. DE GREGORIO, Giovanni; DUNN, Pietro – *The European risk-based approaches: connecting constitutional dots in the digital age.* [em linha] In: Common Market Law Review, vol. 59, n. º 2 (2022), pp. 473–500. Disponível em: https://ssrn.com/abstract=4071437 ou http://dx.doi.org/10.2139/ssrn.4071437

[3] O previamente mencionado relatório da OCDE é, novamente, um bom exemplo disto. O capítulo 3 debruça-se mais aprofundadamente sobre as potenciais utilizações da IA e os seus benefícios, bem como alguns potenciais efeitos negativos pp. 49-73.

[4] O Conselho da Europa, por exemplo, já explorou algumas implicações dos algoritmos avançados no âmbito dos Direitos Fundamentais. Além disso, Leão, Anabela Costa explora alguns dos riscos e efeitos da nova era digital, incluindo a IA, no escopo dos Direitos Fundamentais, identificando alguns desafios constitucionais e, acima de tudo, os riscos fundamentais destas novas ferramentas, páginas 16 e ss. *COMMITTEE OF EXPERTS ON INTERNET INTERMEDIARIES (MSI-NET) – Algorithms and human rights: study on the human rights dimensions*

As decisões autónomas (e sistemas inteiros) podem ser facilmente corrompidas por dados enviesados, existindo também a possibilidade de assistirmos ao aumento da vulnerabilidade de sistemas críticos alvos a ataques cibernéticos, à própria manipulação de dados e informações, à desinformação e à própria perda de controlo de sistemas autónomos e à utilização de sistemas de IA secretos ou encobertos. Estes são apenas alguns dos riscos e preocupações já levantadas pelo Grupo de Peritos de Alto Nível sobre a Inteligência Artificial.[5] Tendo em conta que os sistemas de IA utilizam dados como *input,* é fácil compreender como o *output,* ou o resultado do sistema, dependerá e estará naturalmente suscetível a possíveis enviesamentos. Isto dependerá, evidentemente, das bases de dados utilizadas, que podem apresentar vieses, ainda que de forma não intencional. Estes casos podem estender-se desde o agravamento de preconceitos preexistentes a nível social, como a nível económico.[6] Dá-se o exemplo do caso de um grupo étnico ou religioso que, historicamente e com base nos dados estatísticos utilizados, tem uma taxa de aprovação menor em entrevistas de emprego. Esse grupo, como consequência, terá menos probabilidade de ser convidado para uma entrevista de emprego, na medida em que a decisão do algoritmo replica e amplifica aquilo que poderá ser resultado de um preconceito humano. Não só, é importante indicar que isto também pode resultar, simplesmente, da menor quantidade de dados disponíveis sobre o grupo em questão. Por outro lado, é também possível que tal resulte da qualidade dos dados dessa mesma categoria, que pode variar. O principal risco é, claramente, o do alcance e impacto destes sistemas, que é potencialmente muito mais vasto e difícil de detetar em comparação aos sistemas de tomada de decisão ditos tradicionais. Coloca-se imediatamente, desta forma, o problema de auditar algoritmos e sistemas de inteligência artificial. Os aumentos da complexidade dos algoritmos atuais tornam esta tarefa cada vez mais difícil. Quanto mais complexo o sistema, mais complexa a tarefa de o analisar e interpretar, logo pela quantidade dos dados utilizados e pelo próprio funcionamento de redes neurais. A falta de clareza e dificuldade em analisar os processos internos dos algoritmos acabam por se tornar um obstáculo

of automated data processing techniques and possible regulatory implications. [em linha] Estrasburgo*:* Conselho da Europa*,* 2018*.* Disponível em*:* https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5

[5] COMISSÃO EUROPEIA, Direção-Geral das Redes de Comunicação, Conteúdos e Tecnologias; GRUPO DE PERITOS DE ALTO NÍVEL SOBRE INTELIGÊNCIA ARTIFICIAL – Orientações éticas para uma IA de confiança [em linha]. Luxemburgo: Serviço das Publicações da União Europeia, 2019. Disponível em: https://data.europa.eu/doi/10.2759/2686

[6] A Agência dos Direitos Fundamentais da União Europeia evidencia que é possível, e não tão incomum, agravar a discriminação através de decisões tomadas com base em dados. Esta realidade é explorada no artigo: AGÊNCIA DOS DIREITOS FUNDAMENTAIS DA UNIÃO EUROPEIA, *Big Data: Discrimination in data-supported decision making* [em linha], Viena: FRA, 2018, disponível em: http://fra.europa.eu/en/publication/2018/big-data-discrimination

à própria identificação e correção de vieses. Remetendo ao exemplo anterior, é possível imaginar como, face a um manancial de dados, seja difícil na prática aferir se a menor probabilidade de convidar pessoas de um determinado grupo minoritário resulta de um preconceito humano, de uma simples menor representação/amostra ou de uma questão de qualidade de dados.

A este ponto, acrescentam-se os casos de manipulação deliberada de dados de forma a corromper intencionalmente processos automatizados de decisão ou sistemas inteiros. Como se torna evidente, o tratamento e a elevada qualidade dos dados como a "matéria-prima" da IA, em conjunto com os algoritmos e a capacidade computacional, são, *ab initio* uma preocupação e um ponto de contenção[7]. Estes são alguns dos desafios que motivam a construção de uma estrutura regulatória robusta, na medida em que, frente aos riscos, é essencial criar um ambiente de confiança e responsabilidade que permita, antes de mais, extrair o potencial valor dos sistemas de IA e, por outro lado, mitigar os riscos associados[8]. O objetivo basilar é garantir os valores preconizados pelo artigo 2.º do Tratado da União Europeia e todo o quadro normativo geral da União, desde a Carta dos Direitos Fundamentais ao direito secundário (v.g. Regulamento (UE) 2016/679 RGPD). Do ponto de vista da competência legislativa, a UE goza de competência partilhada através do domínio do mercado interno (Artigo 4.º, n. º2, al. a) TFUE, sendo os sistemas de IA empregues no mercado interno europeu), mas havendo particular relevância do artigo 114.º do TFUE, procurando prevenir a fragmentação do mercado interno e a redução da segurança jurídica resultante da eventual aplicação de diferentes regras nacionais no âmbito da Inteligência Artificial. Entendeu-se que seria necessário assegurar um nível de proteção robusto e, acima de tudo, coerente em toda a União, de forma a alcançar uma

[7] COMISSÃO EUROPEIA, Direção-Geral das Redes de Comunicação, Conteúdos e Tecnologias; GRUPO DE PERITOS DE ALTO NÍVEL SOBRE INTELIGÊNCIA ARTIFICIAL – Orientações éticas para uma IA de confiança [em linha]. Luxemburgo: Serviço das Publicações da União Europeia, 2019. Disponível em: https://data.europa.eu/doi/10.2759/2686 destaca a importância da qualidade e integridade dos dados nas suas orientações para uma IA de confiança, reconhecendo a possibilidade de erros e enganos, bem como a possibilidade da existência de enviesamentos socialmente construídos e até mesmo a introdução deliberada de dados maliciosos que alteram significativamente o comportamento de sistemas de IA que dependam da autoaprendizagem ou *machine learning*. Destaca-se ainda o apelo do Grupo e Especialistas à eliminação dos enviesamentos identificáveis numa fase inicial de recolha de dados, na medida do possível.

[8] Comunicação da Comissão ao Parlamento Europeu, ao Conselho Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões Inteligência artificial para a Europa, Bruxelas, 25.4.2018, COM (2018) 237 final. Disponível em: https://eur-lex.europa.eu/legalcontent/PT/TXT/HTML/?uri=CELEX:52018DC0237&from=EN, a Comissão reconheceu a importância da criação de um ambiente de confiança e responsabilidade assente num quadro ético e jurídico apropriado, baseado em sistemas de IA explicáveis, transparentes e de qualidade. Já nesta comunicação se assumia a importância da IA no mundo, havendo uma identificação clara desta tecnologia como a mais estratégica deste século e uma verdadeira catalisadora da mudança do mundo em que vivemos – pp. 2, 16 e 17.

IA de confiança e garantir o bom funcionamento do mercado interno. Noutro prisma, releva ainda o artigo 16.º do TFUE, tendo em conta que o Regulamento dispõe de regras próprias aplicáveis à proteção de pessoas singulares, particularmente no que diz respeito ao tratamento de dados pessoais. Veja-se o exemplo das restrições previstas para a utilização de sistemas de IA para a identificação biométrica à distância e avaliação de risco em relação de risco em relação a pessoas singulares (veja-se, a título de exemplo, no Regulamento, o artigo 5.º, n. º1, alínea g) e seguintes).[9]

Com efeito, o verdadeiro desafio regulatório revela-se ser justamente este: o de assegurar a extração dos benefícios desta nova tecnologia, garantindo sincronamente a mitigação dos riscos inerentes à sua utilização. Neste sentido, é possível dizer que estamos perante um verdadeiro desafio regulatório global. Assim, procuram-se equilibrar os dois eixos fundamentais da inovação e da proteção dos direitos. Entende-se a possibilidade de, sem estabelecer salvaguardas e mecanismos de controlo eficazes, vermos os benefícios da inovação tecnológica serem inteiramente ofuscados pelos riscos e consequências indesejadas da mesma. Da mesma forma, e por outro lado, reconhece-se a hipótese da intervenção regulatória, dependendo da intensidade, sufocar a inovação. Apurar a dose de regulação ideal é, desta forma, um ponto crucial da discussão.

1.2 A relevância da categorização do risco

Neste âmbito, importa abordar, como já foi feito supra, a regulação do risco e a sua relevância particular. O objetivo da regulação do risco poderá ser eliminar ou proibir o risco por completo, reduzi-lo a um patamar tolerável ou, como já foi mencionado, atingir um equilíbrio proporcional. Neste caso, a forma de flexibilizar e medir de forma proporcional a aplicação das obrigações é através da definição de patamares de risco, como uma pirâmide de risco[10]. Como já referido, o Regulamento da Inteligência Artificial acaba por adotar, em primeiro lugar, diferentes graus de responsabilidade, impondo obrigações específicas consoante

[9] O considerando 3 do Regulamento aborda justamente esta questão.

[10] Lilkov aborda o conceito de Pirâmide de Risco, identificando ainda algumas problemáticas na então Proposta da Comissão. Explica-se que, por exemplo, o topo da pirâmide corresponde ao "risco inaceitável", que correspondem a práticas que vão contra os valores da União e/ou violam direitos fundamentais. Atualmente, são as práticas de IA proibidas plasmadas no Artigo 5.º do Regulamento da Inteligência Artificial. A versão final, agora em vigor, do Regulamento endereça algumas das críticas expostas por Lilkov, nomeadamente a referência a "danos psicológicos" como conceito indeterminado, alterando o texto da norma, que agora inclui a referência a "danos significativos a essa ou a outra pessoa, ou a um grupo de pessoas". LILKOV, Dimitar, *Regulating artificial intelligence in the EU: A risky game*, in European View, vol. 20, n. º 2, 2021, pp. 166-174, pp. 168 ss., DOI: https://doi.org/10.1177/17816858211059248.

o nível de risco. Ora, esse nível de risco é definido numa perspetiva *top-down*[11], pelo Regulamento em si, que identifica diretamente várias categorias e âmbitos de risco[12]. Ultimamente, a regulação baseada no risco terá sempre como objetivo eliminar ou mitigar o risco em si.

A identificação e distinção de quatro níveis de risco é a peça fundamental do quadro regulatório. É desta forma que se pretende proteger a inovação e, mitigar os riscos. A premissa é, como já abordado, tornar proporcional a intensidade da intervenção regulatória à perigosidade potencial da aplicação do sistema de IA. Assim, surge o risco inaceitável, o risco elevado, o risco limitado e o risco mínimo. Naturalmente, as obrigações e normativos são gradualmente mais exigentes, desde a inexistência de regulação no caso dos sistemas em que se aceita o risco até à preterição total no caso dos sistemas de risco inaceitável. Isto permite, na prática, evitar onerar excessivamente certos operadores económicos (por exemplo, alguém que desenvolva e coloque no mercado um filtro de *spam* não estará sujeito a obrigações de transparência específicas).

Este enquadramento não é meramente técnico, assumindo uma relevância estrutural do diploma. Pressupõe que nem todas as aplicações de IA representam ameaças equivalentes e, assim, não devem ser reguladas da mesma forma. Daqui surge a ideia de que o risco, no Regulamento, não é apenas o alvo ou consequência a evitar, mas sim um verdadeiro conceito organizador e o prisma através do qual todo o diploma se estrutura. Surge desta centralidade a importância e relevância da categorização do risco, dada a regulamentação baseada no risco e a regulação do risco, como abordado supra.[13]

Infra, serão abordadas todas as categorias de risco existentes, acompanhadas de alguns exemplos e ocasionais considerações críticas sobre os regimes aplicáveis. Em geral, o objetivo será estabelecer uma estrutura interpretativa para cada uma das categorias, ainda que de forma breve.

---

[11] Sobre as distinções *top-down* e *bottom-up*, consultar: DE GREGORIO; DUNN, op. cit.

[12] ALMADA, Marco, *Will the EU AI Act Shape Global Regulation? Network Law Review*, 2025, disponível em: https://www.networklawreview.org/almada-ai-act/.

[13] Não só: entende-se também que o risco, em geral, tornou-se um conceito fundamental na legislação comunitária contemporânea, particularmente no âmbito da estratégia europeia para a era digital, ainda que, de forma muito diferente consoante as áreas abordadas (seja a proteção de dados, serviços digitais ou IA). Esta aparente transição da abordagem ao risco, verificada, por exemplo, entre o RGPD e o *AI Act* pode ser vista como um passo em direção ao "constitucionalismo europeu digital". Todavia, tanto o RGPD, Regulamento dos Serviços Digitais e Regulamento da IA podem ser vistos, em uníssono, como diplomas de teor constitucional, na medida em que adotam direitos fundamentais como força motriz. DE GREGORIO; DUNN, op. cit.

2. A Categorização do Risco no Regulamento da Inteligência Artificial

2.1 Os sistemas de IA de risco mínimo, risco limitado, risco elevado e as práticas proibidas

Como foi exposto, o Regulamento da Inteligência Artificial institui uma abordagem regulatória baseada no risco para sistemas de IA, distinguindo quatro níveis principais de risco, seguindo uma lógica piramidal: riscos inaceitáveis (as práticas proibidas que constituem o topo da pirâmide), risco elevado (ao qual é dedicada grande parte do normativo regulatório), risco limitado ou de transparência (sujeitos a meras obrigações de transparência) e risco mínimo (sem obrigações específicas, abrange os sistemas de risco mínimo e risco inexistente). Não esquecendo os riscos específicos dos modelos de finalidade geral, também conhecidos como *GPAI*, abordados no capítulo V do diploma, que aqui serão analisados de forma autónoma. Cada categoria corresponde a um regime normativo distinto, com diferentes requisitos, obrigações e fiscalizações. Alguns dos impactos práticos desta categorização serão abordados mais à frente. Neste momento, importará compreender o enquadramento normativo de cada categoria, através do texto do diploma, considerando não só os artigos, mas também os anexos relevantes, fornecendo ainda alguns exemplos concretos.

2.2 Práticas de IA proibidas

O Artigo 5.º do Regulamento[14] elenca exaustivamente as práticas de IA proibidas, consideradas inaceitáveis por implicarem, essencialmente, riscos significativos ao bem-estar ou a direitos fundamentais. De forma geral, salvaguardam alguns dos direitos protegidos pela CDF, a saber: a dignidade do ser humano (artigo 1.º), a proteção de dados pessoais (artigo 8.º), o respeito pela vida privada e familiar (artigo 7.º), a igualdade perante a lei e a não discriminação (artigos 20.º e 21.º, respetivamente), os direitos das crianças (artigo 24.º), a liberdade de expressão e de informação (artigo 11.º) , a liberdade de reunião e de associação (artigo 12.º), a liberdade de pensamento, de consciência e de religião (artigo 10.º) e o direito à ação e a um tribunal imparcial, também como a presunção de inocência e direitos de defesa (artigos 47.º e

[14] JORNAL OFICIAL DA UNIÃO EUROPEIA, Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento da Inteligência Artificial), 2024, disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32024R1689.

48.º). Em cada alínea, exploraremos contextualmente a relevância de cada um destes direitos garantidos pela Carta, quando aplicável.

Mais uma vez, o alcance jurídico desta categoria é absoluto: tais sistemas não podem ser colocados no mercado, disponibilizados ou utilizados de forma a produzirem efeitos no território da União Europeia, sob pena de sanções. O regulamento estabelece, assim, uma proibição total, não apenas quanto ao uso, mas também quanto à comercialização ou operação desses sistemas, independentemente da sua origem geográfica[15]. Apenas se admite uso restrito de reconhecimento biométrico em tempo real em crimes específicos (caso (h), mediante avaliação de proporcionalidade e registo prévio pelas autoridades competentes (cf. Artigo 27.º e Artigo 49.º). Não há qualquer mecanismo de conformidade para reabilitar esses usos: é vedada a sua utilização.

2.3 Impactos práticos da categorização – as orientações da Comissão sobre práticas proibidas de inteligência artificial tal como definidas no Regulamento

Para abordar os impactos práticos da categorização, debruçar-nos-emos com especial atenção sobre o artigo 5.º. Esta opção justifica-se por diversas razões, em primeiro lugar dispõe, desde fevereiro de 2025, de orientações práticas sobre a sua aplicação.[16] Embora esta diretriz interpretativa não seja vinculativa, espera-se que influencie de forma substancial a interpretação do Regulamento da IA pelo Tribunal de Justiça da União Europeia, pelas autoridades de supervisão e tribunais nacionais. Como é evidente, o objetivo das orientações passa por assegurar, na medida do possível, uma aplicação consistente, efetiva e uniforme do Regulamento. Representam, contudo, a visão interpretativa da Comissão, o que se revela particularmente pertinente no âmbito desta investigação. Acresce, doutra forma, que ao contrário do que acontece com os requisitos técnicos aplicáveis aos sistemas de riscos elevados, também brevemente analisados infra, a estrutura do artigo 5.º exige um esforço interpretativo mais intenso do que poderá ser inicialmente percetível. Como veremos, as proibições previstas

[15] Vejam-se ainda os artigos 79.º e 81.º do Regulamento da IA, dentro de uma lógica de atribuição à Comissão Europeia do poder de decidir, em sede de procedimento de salvaguarda da União específico, se determinado sistema constitui uma prática proibida. Visando, desta maneira, assegurar uma aplicação harmonizada das proibições em todos os Estados-Membros e reforçar a segurança jurídica no território da União.

[16] Publicadas pela Comissão Europeia a 4 de fevereiro de 2025, serão a referência primária da análise exposta doravante sobre as práticas de IA proibidas - COMISSÃO EUROPEIA, *Approval of the content of the draft Communication from the Commission – Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act),* anexo à Comunicação C(2025) 884 final, Bruxelas, 4.2.2025, disponível em: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

revelam um grande potencial interpretativo, derivado da quantidade de condições, remissões internas e avulsas, bem como da utilização de conceitos indeterminados, de certo grau de abstração ou simplesmente de conceitos que carecem de uma contextualização sistémica. Tal complexidade traduz-se numa matéria que desperta sobejo interesse e, na ótica adotada, justifica uma análise particularmente densa, na justa medida do possível, mesmo quando comparada com as demais categorias do regulamento.

Começando na alínea a), são vedados à "colocação no mercado, colocação em serviço ou a utilização" de sistemas de IA que, primeiro, empreguem técnicas subliminares ou manipulativas capazes de distorcer o comportamento de pessoas de forma a causar prejuízo significativo ou de forma que seja razoavelmente suscetível de os causar. E ainda, na alínea b), que explorem vulnerabilidades de indivíduos (em razão de idade, deficiência, fragilidade socioeconómica) para influenciá-los ou de maneira a distorcer substancialmente o seu comportamento, mais uma vez, de forma nociva, i.e., de forma que cause danos ou seja razoavelmente suscetível de os causar. Aqui destaca-se o conceito de "danos significativos"[17], na medida em que ambas alíneas abordam sistemas que causam danos, ou que, de forma abstrata, se entenda que são suscetíveis de os causar. Aqui, as práticas manipuladoras já não se fazem depender da intenção manipuladora, aplicando-se tanto à causa, como ao fim ou efeito.[18]

Em suma, estas duas proibições iniciais debruçam-se essencialmente sobre a proteção de indivíduos e pessoas vulneráveis, particularmente nos casos de manipulação. Entende-se que o objetivo principal é a da proteção da própria dignidade humana (cf. Carta dos Direitos Fundamentais da União Europeia, artigo 1.º), que constitui a base de todos os direitos fundamentais, incluindo a autonomia pessoal como seu aspeto essencial. Estas alíneas afiguram-se como a expressão ideal de uma inteligência artificial centrada no ser humano,

[17] O considerando 29 menciona "em especial com repercussões negativas suficientemente importantes na saúde física, psicológica ou nos interesses financeiros".

[18] Tal como já havia sido apontado pela Presidência Eslovena em 2021, o texto do regulamento não deveria fazer depender da intenção manipuladora a proibição da alínea a) do Artigo 5.º. A intenção seria, naturalmente, difícil de provar, apesar de não constituir uma *probatio diabolica*, constituiria um entrave à operacionalização do preceito. Além disso, os sistemas de IA estão sujeitos a ser configurados pelo próprio utilizador, que poderá ser diferente em cada caso, ainda que o sistema seja o mesmo. Assim, as intenções podem ser variadíssimas, mesmo com o mesmo sistema. *COUNCIL OF THE EUROPEAN UNION*, *Presidency compromise text*, 14278/21, disponível em: https://data. consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf

transparente e segura, que não permite a instrumentalização e redução dos indivíduos a meios para alcançar fins.[19]

A título de exemplo, um sistema de IA poderá, de forma propositada, empregar técnicas manipulativas de teor sensorial, como a reprodução de áudio ou exibição de imagens que levem a alterações ao estado psicológico de uma pessoa, causando ansiedade, influenciando, ultimamente, o comportamento da mesma ao ponto de criar danos significativos. De forma concreta, estamos a falar de sistemas de IA que empreguem mensagens subliminares visuais, podendo mostrar imagens ou texto de forma muito breve, quase impercetível (ainda que tecnicamente visíveis), registadas apenas de forma subconsciente de forma a influenciar de alguma forma um indivíduo. O mesmo se aplica a mensagens subliminares de cariz auditivo, tal como a reprodução de mensagens de áudio mascaradas por outros sons, não percetíveis a nível consciente, mas com efeitos inconscientes. Outro exemplo será o de um sistema de IA que, através dos dados pessoais de um utilizador, reconhece as suas vulnerabilidades e cria mensagens altamente persuasivas que influenciam o seu comportamento ao ponto de causar, mais uma vez, danos significativos (ou que seja suscetível de os causar). É o caso de um brinquedo desenhado para manter ao máximo a atenção de uma criança através de um ciclo de reforço positivo e incentivo dado pelo cumprimento de desafios crescentemente perigosos, tal como subir a móveis e lidar com objetos afiados. No espectro oposto da idade, existem ainda sistemas de IA que poderão tirar proveito de situações de solidão e vulnerabilidade das populações mais velhas por via de fraude e burla, causando danos financeiros. Por exemplo, agente de IA que se faz passar por um técnico de suporte por via telefónica e convence ou força alguém, em situação de vulnerabilidade, a ter certas atitudes que, doutra forma não teriam e que causam um dano significativo (como o pagamento de uma importância, constituindo burla). Por fim, dá-se o exemplo de um sistema de IA que personifica um familiar de uma pessoa através da reprodução sintética da sua voz ou até mesmo imagem (com objetivo de burlar ou influenciar alguém, por exemplo). O que releva é o nexo causal entre o *social scoring* e o tratamento desfavorável da(s) pessoa(s) avaliada(s). O tratamento tem de ser consequência dessa classificação. Não só, basta que a classificação ou avaliação seja capaz de resultar num tratamento desfavorável para que a proibição se aplique (que seja plausível que isso aconteça).

[19] Mais sobre técnicas de manipulação em sistemas de IA: CARROLL, M. et al., *Characterising Manipulation from AI Systems, in Equity and Access in Algorithms, Mechanisms, and Optimization*, [em linha] 2023, Boston, DOI: https://doi.org/10.1145/3617694.3623226

A alínea c) debruça-se sobre sistemas que realizam avaliação/classificação contínua de pessoas com base em características pessoais ou comportamentais e que resulte em *social scoring* e num subsequente tratamento desfavorável injustificado. Compreende-se, novamente, que este tipo de classificação não depende do resultado, podendo ser proibida mesmo antes da materialização de qualquer tratamento desfavorável. Neste caso, surge a preocupação vincada pelo princípio de não discriminação, pela equidade no acesso às oportunidades, pela proteção de dados e pela proteção da vida familiar e privada. Note-se que os conceitos de avaliação e classificação são distintos. Ao passo que o conceito de avaliação sugere o juízo sobre uma pessoa ou grupo de pessoas, a mera classificação de indivíduos com base em características objetivas como a idade ou o sexo pode não levar necessariamente à sua avaliação. Assim, o alcance da classificação é mais amplo do que o da avaliação, na medida em que poderá incluir a mera categorização de indivíduos com base em critérios que não implicam um juízo ou análise concreta. Neste caso, proíbe-se o *social scoring* que acarreta tratamento desfavorável ou até mesmo tratamento preferencial a certos indivíduos ou grupo de pessoas (na medida em que isso implica o detrimento de outros indivíduos).

Por sua vez, a avaliação está relacionada com o conceito de definição de perfis. Ainda que não seja explicitamente mencionado pelo Regulamento da Inteligência Artificial[20], é relevante para a proibição da alínea c) nos casos em que este processo de avaliação é feito de maneira automatizada por um sistema de IA com recurso a dados pessoais (cf. Com o artigo 22.º do RGPD). Dar-se-á o exemplo de uma autoridade tributária que utiliza IA para selecionar indivíduos sujeitos a inspeção ou auditoria, considerando dados não relacionados com a sua situação tributária, como os seus hábitos sociais ou ligações à internet (em vez de utilizar exclusivamente dados como os rendimentos anuais, património, número de dependentes, etc.). Isto é, uma ferramenta de Inteligência Artificial que seleciona contribuintes para inspeção especial com base em dados não relacionados com a sua situação contributiva, inferindo características (ou recolhendo por outros meios) derivados de contextos sociais como a origem étnica, o comportamento nas redes sociais, o desempenho no trabalho, os locais frequentados, etc.[21] A questão aqui cinge-se ao tratamento desfavorável ou prejudicial com base em contextos

[20] Cf. COMISSÃO EUROPEIA, *Approval of the content of the draft Communication from the Commission – Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act),* anexo à Comunicação C(2025) 884 final, Bruxelas, 4.2.2025, pp. 52 e ss. Disponível em: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

[21] Um caso sobre definição de perfis na Polónia, por parte do Ministério do Trabalho Polaco, que usava um sistema posteriormente considerado inconstitucional para definir os perfis de pessoas desempregadas, sobre pretensa de

sociais não relacionados com os contextos nos quais os dados foram originalmente gerados ou recolhidos – i.e., um indivíduo vê-se sujeito a uma inspeção tributária devido a fatores e dados não relacionados com a sua situação fiscal (artigo 5.º, alínea c), ponto i).

Outro caso será o do tratamento prejudicial ou desfavorável que é injustificado ou desproporcionado face ao seu comportamento social ou à gravidade do mesmo. A propósito, o exemplo do caso em que uma agência pública utiliza um sistema de IA para proteger menores em situação de risco através de critérios como a situação de emprego dos pais e registos da saúde mental (relacionados com o bem-estar do menor), mas também com informações sobre o seu comportamento social. Com base no perfil resultante, o menor é considerado em risco e é retirado da tutela dos pais, incluindo em casos de menor gravidade, como o facto dos pais terem recebido uma contraordenação no trânsito, ou terem falhado ocasionalmente uma consulta no médico. Como se afigura evidente, este tratamento será considerado injustificado ou desproporcionado face ao comportamento social dos pais (artigo 5.º, alínea c), ponto ii).

Ora, haverá casos em que ambos os pontos i) e ii) são aplicáveis, bastando a utilização simultânea de contextos sociais não relacionados com os dados gerados ou recolhidos e um tratamento desproporcionado e injustificado base ao comportamento social apresentado. O mesmo se aplica a utilizadores públicos e privados. Por outro lado, também existem casos que ficam além do âmbito desta alínea, como o caso de empresas que apresentam um interesse legítimo e até obrigação de prevenir a fraude financeira (como as instituições bancárias) e avaliam os seus clientes com base em dados relevantes, (tal como o seu histórico de transações e outros fatores provenientes de fontes objetivamente relevantes para determinar o risco de fraude, podendo até incluir fatores como o local de residência) e cujo tratamento prejudicial seja justificado e proporcional como consequência de um comportamento fraudulento.

A alínea d), por sua vez, compreende a proibição de "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"[22]. Isto é, compreende os sistemas que preveem ou avaliam o risco de um indivíduo cometer uma infração penal com base na

combater o desemprego. Cf. NIKLAS, Jędrzej; SZTANDAR-SZTANDERSKA, Karolina; SZYMIELEWICZ, Katarzyna; BACZKO-DOMBI, A.; WALKOWIAK, A., *Profiling the unemployed in Poland: social and political implications of algorithmic decision making,* [em linha] *Fundacja Panoptykon*, 2015, disponível em: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf.

[22] Transcrição direta do disposto no Regulamento da Inteligência Artificial.

definição de perfis ou avaliação de traços de personalidade e características pessoais. Destaca-se ainda a segunda parte da norma, que aponta a exclusão dos sistemas de IA utilizados no apoio da avaliação humana do envolvimento de uma pessoa numa atividade criminosa concreta, que apresenta factos objetivos e verificáveis diretamente ligados à atividade criminosa. É importante destacar a função de "apoio da avaliação humana", na medida em que não estaremos a falar de uma avaliação autónoma feita pelo sistema, mas sim de uma função de apoio à avaliação feita pelas autoridades de segurança pública, por exemplo. Como veremos, o busílis desta questão é precisamente o da relação entre estas duas dimensões.

É importante clarificar que os "Sistemas de IA concebidos para serem utilizados por autoridades responsáveis pela aplicação da lei, ou em seu nome, ou por instituições, órgãos ou organismos da União em apoio das autoridades responsáveis pela aplicação da lei para avaliar o risco de uma pessoa singular cometer uma infração penal ou reincidir não exclusivamente com base na definição de perfis de pessoas singulares na aceção do artigo 3.º, ponto 4, da Diretiva (UE) 2016/680, ou para avaliar os traços e características da personalidade ou o comportamento criminal passado de pessoas singulares ou grupos"[23] são, por sua vez, ainda que excluídos do artigo 5.º, classificados expressamente como sistemas de IA de risco elevado pelo Artigo 6.º, n.º2 do Regulamento, através do Anexo III, nº6, alínea d). Desta forma, são-lhes aplicáveis as obrigações e requisitos previstos no Regulamento para sistemas de risco elevado.

É uma proibição com uma *ratio* bastante clara, pretendendo proteger alguns dos direitos fundamentais mencionados no início do capítulo, nomeadamente os artigos 48.º e 21.º da Carta dos Direitos Fundamentais da União Europeia, na presunção da inocência e na não discriminação (a fim de evitar um julgamento com base em comportamento previsto pela IA, com base num perfil definido com base em características como a nacionalidade, idade, local de residência, número de filhos, etc. sem suspeita razoável de envolvimento numa atividade criminosa sustentada em factos verificáveis e objetivos). Existe, claramente, a preocupação de garantir que os cidadãos são julgados com base na sua conduta concreta.[24]

Surgem depois mais duas condições cumulativas para o preenchimento da alínea d), nomeadamente a avaliação ou previsão do risco de uma pessoa singular cometer uma infração

[23] Transcrição direta do disposto no Regulamento da Inteligência Artificial.
[24] O Regulamento da Inteligência artificial expõe alguns destes pontos no considerando 42, página 12, que parece servir como base da *ratio legis* desta alínea.

penal e, por último, o fundamento exclusivo na definição de perfil e/ou avaliação dos seus traços e características de personalidade. No que respeita a este último critério, basta que a previsão ou avaliação se baseie numa das modalidades referidas para que a condição esteja preenchida, podendo, contudo, assentar simultaneamente em ambas.

A previsão ou predição de crimes não dispõe de uma definição consensual, mas os temas gerais respeitantes à Inteligência Artificial e policiamento já foram estudados[25]. A questão que se coloca surge, mais uma vez, em relação ao reforço de certos vieses que os sistemas de IA podem agravar exponencialmente, conduzindo a situações de discriminação ou de reforço da mesma. As variáveis em causa que não estejam incluídas no conjunto de dados analisado podem, por exemplo, fornecer informações importantes sobre os casos concretos que não são tidas em conta e que são, ultimamente, ignoradas. Destaca-se que a proibição não se aplica apenas a indivíduos, mas também a grupos de pessoas. Incluindo a definição de perfis de grupo, que podem ser aplicados por um sistema de IA a um indivíduo específico (um indivíduo que preencha as características do perfil de grupo), como por exemplo: alguém que se enquadra num perfil de grupo previamente definido como "terrorista".

A alínea d) do Artigo 5.º utiliza de forma explícita o conceito de "definição de perfis", que está diretamente ligado ao Artigo 4.º, n. º4 do Regulamento Geral da Proteção de Dados (EU) 2016/679 por remissão do Artigo 3.º, n. º52 do Regulamento da Inteligência Artificial[26]. Com efeito, o RGPD identifica a avaliação de certos aspetos pessoais de uma pessoa singular como um elemento fundamental da definição de perfis. Neste âmbito, a avaliação de certos aspetos pessoais de uma pessoa singular é feita de forma a prever o risco de cometer de um crime. O conceito de "avaliação dos seus traços e características de personalidade" parece ser próxima e até mesmo quase redundante quando comparado ao conceito de definição de perfis. No entanto, assemelha-se evidente que o Regulamento da Inteligência Artificial procurou utilizar esta redundância como uma espécie de mecanismo duplo de reforço conceptual. Se não for possível

[25] A Europol já publicou estudos que ilustram alguns dos desafios (e benefícios) da IA no policiamento, incluindo alguns casos concretos já em vigor. *EUROPOL*, *The Second Quantum Revolution – The impact of quantum computing and quantum technologies on law enforcement*, [em linha] *Europol Innovation Lab observatory report, Publications Office of the European Union*, Luxemburgo, 2023, disponível em: https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf

[26] O RGPD define definição de perfis como "qualquer forma de tratamento automatizado de dados pessoais que consista em utilizar esses dados pessoais para avaliar certos aspetos pessoais de uma pessoa singular, nomeadamente para analisar ou prever aspetos relacionados com o seu desempenho profissional, a sua situação económica, saúde, preferências pessoais, interesses, fiabilidade, comportamento, localização ou deslocações", Regulamento(UE) n.º 679/2016, de 27 de Abril, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE.

garantir que a prática constitui uma definição de perfil, será possível, provavelmente com maior facilidade, argumentar que a alternativa se aplica (i.e., a avaliação dos traços e características de personalidade, que poderá ser uma ligação mais fácil de sustentar). O considerando 42, já mencionado, providencia um elenco ilustrativo destas características e traços que podem ser utilizados na avaliação do risco.

Por outro lado, importa mencionar que o termo "com base exclusivamente" deve ser acautelado de forma a evitar a circunvenção da norma. Desta forma, os elementos que se associam à avaliação do risco devem ser substanciais para que a proibição não se aplique. Por exemplo, se houver *ab initio* uma suspeita razoável em relação a um indivíduo, motivada por uma avaliação humana feita com base em factos verificáveis e objetivos, deve justificar a exclusão da proibição, tal como a última frase da norma indica. Doutra forma, damos exemplos de casos em que a proibição se deve aplicar, v.g.: um sistema de IA que prevê o risco de terrorismo com base na idade, nacionalidade, estado civil e morada de um indivíduo. Este sistema considera indivíduos com estas características mais propensos a cometerem o crime de terrorismo no futuro com base exclusiva nas suas características pessoais. É um caso em que a proibição se deve, claramente, aplicar. Há ainda casos que se podem sobrepor com a tipificação da alínea anterior, respeitante ao "social scoring". Por exemplo: uma autoridade tributária usa um sistema de IA que cria um perfil individual para cada contribuinte com o propósito de identificar possíveis casos de crime de fraude fiscal. Para isso, utiliza dados como os traços pessoais dos contribuintes, como a nacionalidade, naturalidade, número de filhos e outras variáveis. Este sistema parece cair dentro da previsão da alínea d), contudo, poderá também sobrepor-se com o escopo de aplicação da alínea c), por constituir um tratamento desfavorável com base em dados não relacionados (assumindo que a Autoridade procede a uma investigação com base na predição do algoritmo). Haverá potencialmente, então, casos excecionais em que as duas alíneas se poderão sobrepor.

No âmbito da última frase da alínea d) do Artigo 5.º, importa compreender a exclusão de sistemas de IA utilizados no apoio à avaliação humana que já se baseia em factos objetivos e verificáveis diretamente ligados a uma atividade criminosa. Além desta exceção, contudo, prevê-se que o não cumprimento de um dos três requisitos cumulativos resulte no afastamento da proibição.[27] Isto não obstante o sistema de IA poder ser classificado, ainda assim, como

[27] De facto, o facto de os requisitos serem cumulativos, em conjunto com as exceções previstas, pode enfraquecer a proibição e torná-la numa norma a evitar e não a cumprir. Há quem aponte o risco da proibição se tornar, para efeitos práticos, num gesto simbólico em vez de um escudo substancial dos direitos fundamentais. As fendas da

sistema de risco elevado, como é o caso da exclusão da previsão prevista na letra da norma. No caso em que o sistema caia no âmbito de aplicação da exclusão da proibição prevista pela última parta da alínea d), importa sublinhar a importância da avaliação humana. Além da avaliação humana, será necessário cumprir as exigências dos artigos 14.º e 26.º do Regulamento da Inteligência Artificial, com destaque para o Artigo 14.º, que estabelece a necessidade de supervisão humana eficaz, com ênfase na prevenção e mitigação de riscos para a segurança e direitos fundamentais e com estrita aderência à finalidade prevista. Assim, entende-se que a ação e avaliação humana é um ponto fundamental tanto da própria exclusão como das exigências inerentes ao risco elevado. Não se poderá desconsiderar, em qualquer caso, o outro requisito de base em factos objetivos e verificáveis diretamente ligados a uma atividade criminosa.

Ainda importa mencionar que não são apenas as autoridades e forças de segurança que estão sujeitas a estes princípios. Para reforçar a eficácia da norma e garantir que a mesma não é contornada é importante considerar casos em que entidades privadas dentro do âmbito de aplicação da alínea d). A letra da norma acomoda isto, na medida em que está aberta a essa interpretação. Isto é, à interpretação de que o preceito deve ser aplicado no caso em que certa entidade privada atua no exercício de poderes públicos na prevenção e investigação de crimes. A dimensão gramatical permite claramente esta interpretação, pelo que não menciona explicitamente a aplicabilidade exclusiva a autoridades de segurança pública, por exemplo. A título de exemplo: uma força de segurança pública pode recorrer a uma empresa privada para analisar dados (dados que se baseiam nas características de personalidade dos sujeitos) com a finalidade de prever o risco de certos indivíduos cometerem crimes. Neste caso, a empresa parece estar no âmbito de aplicação da proibição do artigo 5.º.

Noutro caso, supõe-se um banco que atua sobre a obrigação de prevenção de branqueamento de capitais da UE[28] e avalia os seus clientes para prevenir este tipo de crimes. Se houver recurso a um sistema de IA, isso deve ser feito com base em dados objetivos e verificáveis que possam determinar o nível de risco de um cliente em relação ao branqueamento de capitais. Além disso,

proibição, por onde escapam alguns casos, poderão reduzir a maioria das utilizações a risco elevado, o que se poderá entender como um regime manifestamente insuficiente para esta situação, como indica: LEVANO, Jessie, *Predictive policing in the AI Act: meaningful ban or paper tiger?* [em linha]. *European Law Blog*, julho 2024. Disponível em: https://doi.org/10.21428/9885764c.6d0aa28c

[28] Regulamento (UE) 2024/1624 do Parlamento Europeu e do Conselho de 31 de maio de 2024 relativo à prevenção da utilização do sistema financeiro para efeitos de branqueamento de capitais ou de financiamento do terrorismo.

tem de existir obrigatoriamente uma avaliação humana para garantir que a proibição do artigo 5.º d) não se aplique.

Em geral, há certos sistemas que, por definição, se encontrarão fora do âmbito de aplicação da alínea d). Serve de exemplo os sistemas que se baseiam na identificação de locais de risco. Neste caso, a predição assenta na probabilidade de um certo tipo de crime ser cometido numa zona específica. Ora, se esta avaliação não se basear na avaliação de um indivíduo ou grupos, estará fora do alcance da proibição. Aqui trata-se da avaliação de risco num local e não do risco de uma pessoa específica.[29] Em todos os casos, é importante, todavia, ter em conta que o risco do local não deve ser transmissível e atribuível ao indivíduo. Se o risco de um local for posteriormente associado a um perfil específico de uma pessoa, é evidente que o sistema passará a ser baseado em pessoas, pelo que a proibição se voltará a aplicar. A este propósito, surge a questão de se o mesmo se aplicará a pessoas coletivas. Como resulta do texto da norma, as pessoas coletivas estarão excluídas desta proibição – o texto é claro em relação à avaliação de risco de pessoas singulares, não de pessoas coletivas. Por outro lado, levanta-se a questão do facto de que as pessoas coletivas, no normal decorrer da sua atividade, agirem através das pessoas individuais. Logo, não acabariam por incidir também sobre estas? Para todos os efeitos, podem existir casos de fronteira, como os de atuação de uma pessoa singular por intermédio de uma pessoa coletiva, como numa sociedade unipessoal, em que a proibição se possa aplicar, assumindo que estão preenchidos todos os requisitos normativos. Parece revelar, acima de tudo, o efeito material sobre as pessoas singulares, que deve ser evitado. No âmbito das pessoas coletivas, a exclusão da proibição parece estar pensada para casos como o da análise de dados objetivos sobre empresas, como dados sobre transações, declarações fiscais, dados aduaneiros e outros indicadores que permitam avaliar o risco de uma pessoa coletiva de cometer certos ilícitos. O exemplo mais prático é o de uma autoridade fiscal que utiliza estes dados para avaliar o risco de uma empresa cometer fraude fiscal ou algum crime aduaneiro. Quando a análise depende da avaliação do risco dos trabalhadores da empresa, com base nos seus traços de personalidade, características e perfil, a proibição parece ser aplicável novamente.

Surge ainda a questão, no entanto, de considerar se esta proibição fará sentido tendo em conta a exclusão do âmbito de aplicação do regulamento previsto para a segurança nacional no

[29] São casos reais que já parecem ser utilizados, como identificado na página 16 do documento da *EUROPOL*, *The Second Quantum Revolution – The impact of quantum computing and quantum technologies on law enforcement*, [em linha] *Europol Innovation Lab observatory report, Publications Office of the European Union*, Luxemburgo, 2023, disponível em: https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf

Artigo 2.º, n. º3. Podem surgir dúvidas derivadas da proximidade conceptual de "segurança nacional" e "segurança pública". Ora, a delimitação clara do conceito de "segurança nacional" no âmbito da exclusão do âmbito do Artigo 2.º é de sobeja importância para a aplicação correta do Regulamento no campo da segurança (em sentido lato). Carrega particular importância no caso de colocação no mercado e utilização em serviço de sistemas de IA por forças de segurança que atuam no foro criminal, tal como no caso abordado acima, não só neste caso do Artigo 5.º, n. º1, alínea d), mas também no caso da alínea h), que será tratado posteriormente (relativo à identificação biométrica remota em tempo real em espaços públicos por forças de segurança, exceto em casos de investigação de crimes graves). Isto porque as forças de segurança, como as forças de polícia, quando atuam na prevenção e investigação criminal, na aplicação da lei e, geralmente, com fins de segurança pública, estão sujeitos à aplicação do Regulamento da Inteligência Artificial, isto é, estão dentro do âmbito de aplicação do diploma. O Artigo 3.º, n. º46 define claramente o termo «aplicação da lei» no âmbito do regulamento, identificando-o como "as atividades realizadas por autoridades responsáveis pela aplicação da lei ou em nome destas 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". Resulta, assim, de forma inequívoca, deste conceito, a aplicabilidade do Regulamento a questões criminais e de segurança pública. Nesta discussão, o considerando 24 do Regulamento esclarece que todos os sistemas de IA que sejam utilizados para fins civis, humanitários, de aplicação da lei ou de segurança pública, ainda que de forma temporária, estão abrangidos pelo âmbito de aplicação do mesmo. Disto resulta o englobamento dos sistemas de finalidade dupla (ou *dual use systems*) no âmbito de aplicação do regulamento, na medida em que sejam também utilizados para um fim que não se enquadra na exclusão do Artigo 2.º.

Esta distinção torna-se mais clara se a interpretação do Tribunal de Justiça da União Europeia (doravante, TJUE) sobre o conceito de segurança nacional for tida em consideração e, num segundo momento, for confrontada com a definição de "aplicação da lei" segundo o Regulamento. Segundo o TJUE, a segurança nacional, que é da exclusiva responsabilidade de cada Estado-Membro segundo o artigo 4.º, n.º2 do TUE, corresponde "ao interesse primordial de proteger as funções essenciais do Estado e os interesses fundamentais da sociedade e inclui a prevenção e a repressão de atividades suscetíveis de desestabilizar gravemente as estruturas constitucionais, políticas, económicas ou sociais fundamentais de um país, em especial de ameaçar diretamente a sociedade, a população ou o Estado enquanto tal, como, nomeadamente,

as atividades terroristas."[30]. Torna-se evidente que o conceito de segurança nacional não abrange questões relacionadas à luta contra a criminalidade em geral, por exemplo. De forma geral, o Acórdão de 6 de outubro de 2020 é bastante claro na delimitação destas questões. Fica, todavia, desde já claro, o alcance da exclusão do âmbito de aplicação relativo à segurança nacional. Como já foi referido, é especialmente relevante considerar cuidadosamente a diferença entre segurança nacional e pública. Outra questão, que merece análise distinta, é o da exclusão do âmbito de aplicação das autoridades públicas de países terceiros e nos casos que caem no âmbito da cooperação internacional ou de acordos internacionais para efeitos de cooperação policial e judiciária com a União ou com um ou vários Estados-Membros, sob condição de esse país terceiro ou organização internacional apresentar salvaguardas adequadas em matéria de proteção de direitos e liberdades fundamentais das pessoas, previstos no Artigo 2.º, n.º4. Essa é uma questão distinta que será abordada mais à frente.

Passando à alínea e), que aborda os 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 (ou TVCF). Neste caso, a *ratio legis* também parece ser clara, estando em causa a privacidade e a proteção de dados dos indivíduos. A questão, contudo, vai além da proteção de eventuais violações dos direitos fundamentais como o direito à privacidade, posto que a proibição também se relaciona com o aumento do sentimento de vigilância em larga escala que estes tipos de sistemas causam[31]. Apesar de não ter sido inicialmente prevista, esta proibição foi incluída no Regulamento em consideração de desenvolvimentos no domínio da proteção de dados, incluindo decisões de autoridades nacionais de proteção de dados em relação a casos como o *Clearview AI*[32]. Do ponto de vista dos requisitos, verifica-se novamente o pressuposto tripartido transversal: o da colocação no mercado, a colocação em serviço ou a utilização de sistemas de IA. Esta tríade surge como requisito essencial nas alíneas do artigo 5.º e deve ser considerada como o primeiro requisito a analisar. Cumulativamente, a alínea e) estabelece mais três condições, nomeadamente a da

[30] TRIBUNAL DE JUSTIÇA DA UNIÃO EUROPEIA (Grande Secção), Acórdão de 6 de outubro de 2020, *La Quadrature du Net* e outros contra *Premier ministre* e outros, Processos apensos C-511/18, C-512/18 e C-520/18.EU:C:2020:791, parágrafo 135.

[31] Esta é uma questão referida explicitamente no considerando 43 do Regulamento da Inteligência Artificial, página 12.

[32] Sobre o caso *Clearview* e outras práticas de vigilância proibidas pelo regulamento, consultar: BARKANE, Irena; BUKA, Lolita – *Prohibited AI surveillance practices in the Artificial Intelligence Act: promises and pitfalls in protecting fundamental rights.* em BARKANE, Irena; BUKA, Lolita. *Critical perspectives on predictive policing.* Cheltenham*: Edward Elgar Publishing,* 2024. Cap. 6, p. 110–129. Disponível em: https://www.elgaronline.com/view/book/9781035323036.

criação ou expansão de base de dados de reconhecimento facial (entenda-se com o propósito de criar ou expandir), a recolha aleatória de imagens faciais, com proveniência da internet ou TVCF. Para clarificar, compreende-se que a recolha aleatória de imagens artificiais para preencher as bases de dados tenha de ser feita, naturalmente, com recurso a ferramentas ou sistemas de Inteligência Artificial, tal é o âmbito da proibição. Semelhante ao previamente visto, os requisitos enunciados são cumulativos e todos devem ser preenchidos para que a proibição seja aplicável. Mais uma vez, estamos perante uma proibição aplicável tanto a prestadores como responsáveis pela implementação.[33]

Em relação às bases de dados, as mesmas não devem ser entendidas em sentido estrito, ou seja, não devem ser englobadas na proibição de forma exclusiva as bases de dados especificamente e unicamente utilizadas para reconhecimento facial. Parece ser suficiente que a base de dados em questão seja suscetível de ser utilizada para reconhecimento facial. A recolha aleatória de imagens faciais, por outro lado, já poderá suscitar outras questões, como por exemplo a da recolha parametrizada ou direcionada. Se a recolha de imagens for parametrizada ou direcionada a um indivíduo em específico, ou até mesmo a um grupo pré-definido de pessoas, estamos perante uma recolha parametrizada e não uma recolha aleatória. Ao que tudo indica, este tipo de recolha está fora do âmbito da alínea e) e pode ser utilizada, por exemplo, com objetivo de encontrar um criminoso, uma pessoa desaparecida ou para identificar um grupo específico de vítimas (como o caso de vítimas de tráfico humano). Ora, a principal preocupação reside em evitar que esta exclusão do âmbito seja utilizada de forma indigna. Assim, é importante que a interpretação da alínea seja feita de forma a evitar a sua circunvenção. Uma recolha de imagens faciais pode, em primeira análise, aparentar ser parametrizada, mas ultimamente ter um efeito análogo ao da recolha aleatória; por exemplo, no caso em que se seguem parâmetros distintos em várias recolhas e que, no final, o resultado seja funcionalmente igual ao da recolha aleatória. Desta forma, é importante que se considere a finalidade da recolha ou a sua utilidade final. Em relação à proveniência das imagens (da Internet ou TVCF), também se poderá colocar a questão das imagens presentes nas redes sociais ou outras plataformas em que as próprias pessoas tenham disponibilizado o seu retrato. Ora, o mero facto de alguém ter disponibilizado o seu retrato numa plataforma online não indica o seu consentimento para que

[33] O conceito de "responsável pela implantação" é delineado pelo Artigo 3.º, ponto 4) do Regulamento, e consiste no seguinte: "uma pessoa singular ou coletiva, autoridade pública, agência ou outro organismo que utilize um sistema de IA sob a sua própria autoridade, salvo se o sistema de IA for utilizado no âmbito de uma atividade pessoal de caráter não profissional;". É um conceito que, doravante, deve ser entendido desta forma.

as imagens sejam utilizadas numa base de dados de reconhecimento facial.[34] Não parece existir, em nenhum caso, consentimento tácito em relação a esta questão.

Fora do escopo de aplicação da alínea e) do Artigo 5.º, n. º1, estarão, naturalmente, os dados biométricos que não sejam imagens faciais, como impressões digitais e amostras da voz. Se a recolha aleatória não envolver sistemas de IA, também será seguro assumir que não estamos perante este âmbito de aplicação, tal como no caso em que as bases de dados não são de reconhecimento facial (e sejam, por exemplo, utilizadas para treinar modelos de IA, não identificando pessoas). Por outro lado, há que ter em consideração a ação recíproca entre o Regulamento da Inteligência Artificial e outros diplomas comunitários, nomeadamente o Regulamento Geral de Proteção de Dados, que já abrangerá esta matéria.

Após a relativa simplicidade da alínea e), surge a alínea f), que aporta os sistemas de IA que inferem 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. É uma alínea que reconhece imediatamente exceções aplicáveis e lida com questões que justificam uma análise atenta, como veremos. Desde logo, importa destacar que os sistemas de reconhecimento de emoções que não se enquadram nesta proibição são classificados como sistemas de risco elevado, nos termos do ponto 1., alínea c) do Anexo III do Regulamento. Da mesma forma, também são aplicáveis as obrigações de transparência tipificadas pelo Artigo 50.º, n. º3.

Antes de mais, reconhecem-se preocupações em relação à base científica dos sistemas de inferência de emoções baseados em IA graças à complexidade própria da emoção humana e a ampla variedade de formas de exprimir emoções consoante a cultura e a situação em causa. Por outro lado, o reconhecimento de emoções já é objeto de estudo em várias áreas e a sua aplicação é discutida[35]. Identificam-se, todavia, potencialidades na área da saúde mental, por exemplo. Apesar de tudo, os fatores como a fiabilidade limitada, falta de especificidade e possibilidade limitada de generalização revelam riscos inaceitáveis na utilização destes sistemas. Ultimamente, a condução a resultados discriminatórios e a intrusão nos direitos, liberdades e

[34] Cf.: TRIBUNAL DE JUSTIÇA DA UNIÃO EUROPEIA (Quarta Secção), Acórdão de 4 de outubro de 2024, *Schrems contra Meta Platforms Ireland Ltd*., Processo C-446/21, ECLI:EU:C:2024:804. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62021CJ0446

[35] De forma geral, o assunto já foi pensado e abordado há anos, particularmente na área do Marketing. E, Nick; BRODERICK, Amanda J.; CHAMBERLAIN, Laura, *What is 'neuromarketing'? A discussion and agenda for future research, International Journal of Psychophysiology* [em linha], vol. 63, n.º 2, 2007, pp. 199-204, ISSN 0167-8760, DOI: https://doi.org/10.1016/j.ijpsycho.2006.03.007, disponível em: https://www.sciencedirect.com

garantias das pessoas em causa determinam a razão desta proibição, particularmente nos casos agravados pelo contexto laboral ou educativo, que são pautados pelas assimetrias nas relações de poder[36].

Concretamente, a proibição da alínea f) prevê, à semelhança do que vimos anteriormente, um rol cumulativo quadripartido de condições. Como temos visto, inclui a tríade de colocação no mercado, em serviço ou utilização de sistemas de IA, que dispensa densificação. O segundo critério, todavia, prende-se à existência de um sistema de IA que infira emoções de uma pessoa singular, sendo o terceiro ligado à área/local de trabalho e/ou instituições de ensino, considerando ainda, por último, não se verifica a exceção identificada na última parte da alínea, i.e., que não se destine a ser instalado ou introduzido no mercado por razões médicas ou de segurança.

A primeira densificação dos conceitos deve ser apontada aos "sistemas de reconhecimento de emoções". Esta definição surge no Artigo 3.º, n.º39) do Regulamento, traduzindo-se num "sistema de IA concebido para identificar ou inferir emoções ou intenções de pessoas singulares com base nos seus dados biométricos". Ora, a alínea f) não utiliza a expressão "sistema de reconhecimento de emoções", mas cobre sim sistemas de IA que inferem emoções de pessoas singulares. A própria inferência abrange, em princípio, uma identificação *a priori*, pelo que aqui parece fazer sentido incluir os sistemas de reconhecimento de emoções, ou seja, incluir tanto a identificação, como a inferência de emoções e não só: de intenções também[37]. Não fará sentido interpretar de forma isolada a alínea f), pelo que se afigura relevante fazer a remissão tanto ao Artigo 50.º do Regulamento como ao Anexo III, ponto 1., alínea c) e limitar o seu alcance às inferências feitas com base nos dados biométricos das pessoas singulares. Esta visão holística justifica uma interpretação da alínea f) em harmonia com as restantes normas aplicáveis a sistemas de reconhecimento de emoções, prevenindo incongruências e promovendo uma interpretação verdadeiramente sistemática.

Ainda do ponto de vista interpretativo, é importante considerar novamente os considerandos 12 e 18 do Regulamento para efeitos da aplicação da proibição em análise. Primeiro, é

[36] O considerando 44 do Regulamento aborda a razão de ser desta proibição. Novamente, os riscos de lesão dos direitos das pessoas em causa são determinantes, nomeadamente o direito à privacidade, dignidade humana e liberdade de consciência. As relações assimétricas como as laborais e educativas colocam trabalhadores e estudantes em posições de vulnerabilidade, o que justifica o perímetro. Doutra forma, entende-se que no perímetro da saúde os benefícios superem significativamente os riscos, pelo que se prevê a exceção.

[37] Cf. considerando 18 do Regulamento e a sua menção aos dados biométricos, exclusão de mera deteção de expressões, gestos ou movimentos rapidamente visíveis, etc.

aconselhável entender o conceito de inferência em sintonia com a observação do considerando 12, que define a capacidade de fazer inferências como uma característica principal dos sistemas de IA. Esta capacidade prende-se com a aptidão dos sistemas de irem além do tratamento básico de dados e alcançarem processos de aprendizagem, raciocínio ou modelação com fim de obter resultados como previsões, conteúdos, recomendações ou até mesmo decisões autónomas (que ultimamente influenciem ambientes físicos ou virtuais, ainda que apenas potencialmente) através dos dados. Como já foi referido, é diferente da mera identificação, que se assemelha mais a um processamento de dados, que permite comparar e reconhecer emoções com base em emoções fixas previamente programadas. Sucede-se em contraposição com a inferência, que deriva de abordagens de aprendizagem automática e de base lógica do sistema para obtenção de modelos e algoritmos que viabilizam, a partir de dados, a obtenção de resultados de forma autónoma e dinâmica. Disto resulta, naturalmente, que a inferência não será feita exclusivamente com base nos dados biométricos de uma pessoa natural, pelo que haverá inclusão destes processos de autoaprendizagem. Com efeito, os dados inferidos são nova informação resultante da análise probabilística de dados com objetivo de encontrar, de forma autónoma, correlação e padrões em conjuntos de dados.

Por sua vez, o considerando 18 auxilia-nos na circunscrição do que pode ser considerado uma "emoção" para efeito do regulamento. Refere-se explicitamente a "felicidade, a tristeza, a raiva, a surpresa, a repugnância, o embaraço, o entusiasmo, a vergonha, o desprezo, a satisfação e o divertimento", com exclusão explícita de estados físicos como a dor e a fadiga. Desde já, conclui-se que a leitura desta lista nunca deve ser feita de forma exaustiva, na medida em que o elenco apresentado parece ser meramente exemplificativo. A extensão de emoção deve ser tão abrangente quanto o necessário para, mais uma vez, evitar a circunvenção da norma (através de, por exemplo, alegação de que um sistema reconhece atitudes e não emoções). Por outro lado, os sistemas de deteção do cansaço no âmbito da aviação comercial com objetivo de prevenção de acidentes, por exemplo, não se incluem neste âmbito, tal como a mera deteção de expressões aparentes como os gestos e movimentos rapidamente visíveis, a não ser que por esse meio se identifiquem ou infiram emoções. A deteção de um estado depressivo já parece, por sua vez, estar incluído no âmbito da deteção de emoções. Em suma, identificar um sorriso não é considerado reconhecimento de emoções. Doutra forma, chegar à conclusão de que alguém está feliz, otimista ou orgulhoso com base no seu sorriso, postura corporal e movimentos já é considerado reconhecimento de emoções. Inferir, contudo, que um motorista está cansado com base nos seus dados biométricos já não é considerado reconhecimento de emoções.

Suscita-se a questão de saber se modelos de reconhecimento de emoções com base em dados de texto[38] são abarcados pela proibição. Desde logo, entende-se que não, a não ser que sejam utilizados dados biométricos em conjunto com dados de texto para inferir emoções. Releva deixar claro que os dados biométricos devem ser entendidos de forma ampla, como por exemplo: se um sistema de IA utilizar dados de texto em combinação com a forma específica de escrever no teclado, a postura corporal e expressão facial de um indivíduo para identificar/inferir a emoção, então será proibido. A forma de andar, de escrever no teclado, os dados dactiloscópicos, o odor, os batimentos cardíacos, eletrocardiogramas, e outros sinais biométricos que podem parecer "invisíveis" também se incluem no conceito de dados biométricos. A combinação de vários dados pode aumentar a robustez da identificação ou inferência, mas aumenta também a problemática a nível da proteção de dados pessoais. Resumidamente, é importante ter em consideração, na leitura desta proibição, o ponto 39) e 34) do artigo 3.º do regulamento, que definem sistemas de reconhecimento de emoções e dados biométricos, respetivamente.

Note-se ainda a relação entre "dados biométricos" na conceção do Regulamento da Inteligência Artificial e na do Regulamento Geral sobre a Proteção de Dados. Para todos os efeitos, o entendimento do Regulamento da IA afigura-se como sendo mais abrangente, na medida em que não reconhece a exigência de permitir ou confirmar a identificação única de uma pessoa singular, que está presente no RGPD. Assim, o Regulamento da IA engloba, particularmente no âmbito da alínea f), qualquer dado físico ou comportamental que permita a inferência de emoções, mesmo que não vise ou permite identificar unicamente o indivíduo. Por outro lado, nos casos em que o processamento desses dados envolva informação pessoal que o permita esta identificação única, as obrigações do RGPD parecem subsistir[39].

Como já foi referido, esta proibição aplica-se ao contexto do trabalho e da educação, tal como é esclarecido pelo considerando 44. A noção de local de trabalho, por exemplo, deve ser também interpretada de forma atual e ampla. Isto traduzir-se-á na inclusão de locais tradicionalmente associados ao trabalho como escritórios, fábricas, armazéns, lojas até locais

[38] Segue o exemplo de um modelo de linguagem de grande escala que identifica emoções exclusivamente com base em dados de texto: HUGGING FACE. emotion-english-distilroberta-base. [em linha] Disponível em: https://huggingface.co/j-hartmann/emotion-english-distilroberta-base. Acedido em 22 jun. 2025.

[39] Sobre a definição de "dados biométricos" e as restrições relativas ao tratamento de categoriais especiais de dados, consultar: REGULAMENTO (UE) 2016/679 DO PARLAMENTO EUROPEU E DO CONSELHO, de 27 de abril de 2016, relativo à proteção de pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados (Regulamento Geral sobre a Proteção de Dados). Jornal Oficial da União Europeia, L 119, 4 maio 2016, p. 1–88, arts. 4.º, n.º 14; 9.º, n.º 1 e 2.

como a habitação própria do trabalhador (ex. dos trabalhadores em regime remoto). De facto, entender-se-á local de trabalho como qualquer sítio onde o trabalho é prestado pelo trabalhador, variando com base na natureza e tipo de trabalho. Além disso, esta noção também se aplicará ao trabalho por conta própria e aos candidatos num processo de recrutamento. A interpretação de local de trabalho deve seguir uma margem ampla, tal como se tem vindo a verificar noutros conceitos ligados a proibições. Posto isto, entende-se que a utilização de reconhecimento de voz para identificar as emoções dos trabalhadores durante uma reunião é proibida, tal como a utilização da webcam para reconhecer quando um trabalhador está entediado durante o período de trabalho remoto ou durante uma reunião em linha.

Noutra perspetiva, a noção de instituições de ensino também deve ser entendida num sentido lato. Ou seja, deve incluir tanto instituições públicas como privadas, sem limite ou exclusão da proibição com base na idade, tipo de ensino ou formação (seja em linha, presencial ou híbrida), formação contínua, superior ou básica, profissional, industrial, etc[40]. Aqui também se englobam os candidatos no processo de admissão ou candidatura. Por exemplo, uma aplicação de IA que utilize reconhecimento de emoções e que seja escolhida voluntariamente por um utilizador para aprender uma língua fora de qualquer contexto educativo formal não é proibida pela alínea f). Por outro lado, se uma instituição de ensino integrar esse mesmo sistema e exigir a sua utilização pelos alunos como parte do programa curricular, o seu uso fica vedado. A abrangência da noção de instituições de ensino não inclui aplicações escolhidas voluntariamente por alguém fora de uma instituição. Noutros casos, ainda a título de exemplo, é permitido utilizar sistemas que detetem conversas paralelas entre alunos (seja fisicamente, por via de telemóveis ou até mesmo por outras vias de comunicação em linha, nos casos do ensino remoto), sendo que não existe inferência de emoções neste caso. Por outro lado, esse mesmo sistema será proibido se detetar qualquer tipo de emoção, como excitação, ansiedade, interesse ou envolvimento emocional na matéria. Curiosamente, a Comissão reconhece que estes sistemas podem ser

[40] Certos usos de sistemas de inteligência artificial no contexto da formação profissional têm sido apontados como geradores de interferências significativas com um vasto leque de direitos fundamentais. Este risco foi sublinhado, por exemplo, na avaliação de impacto que acompanhou a proposta da Comissão Europeia relativa ao Regulamento da IA, onde se destacaram casos concretos de aplicações de IA em instituições de ensino e formação profissional. Também a investigação promovida pelo Parlamento Europeu enfatiza os desafios que estas tecnologias levantam no setor educativo. Consultar: COMISSÃO EUROPEIA. Documento de trabalho dos serviços da Comissão. Avaliação de impacto. Anexos. SWD(2021) 84 final, Parte 2/2, p. 43. e TUOMI, I. *The use of Artificial Intelligence (AI) in education,* [em linha], Parlamento Europeu, 2020, pp. 9–10, disponível em: https://op.europa.eu/en/publication-detail/-/publication/92273e25-f7c3-11ea-991b-01aa75ed71a1/language-en

utilizados para fins exclusivamente educativos, como no caso de treino de atores (na medida em que os resultados avaliativos não sejam influenciados pelo sistema).

Cabe assinalar novamente, contudo, que a alínea f) prevê explicitamente uma exceção aplicável aos sistemas de reconhecimento de emoções no local de trabalho e instituições de ensino por razões médicas ou de segurança, nomeadamente para utilização terapêutica (considerando 44). Entende-se que a interpretação desta exceção deve ser mais restrita, tal como deve ser o caso em todas as exceções. A Comissão defende, inclusive, que a exceção não se aplica à deteção de depressão ou esgotamento no local de trabalho e instituições de ensino, no caso da saúde. No caso da segurança, entender-se-á que a exceção é procedente quando está em causa a proteção da vida e da saúde e não a proteção da propriedade contra roubo ou fraude, por exemplo. Em geral, deve atender a finalidades estritamente conexas à segurança e saúde (e não às necessidades do empregador ou da instituição de ensino) e limitadas ao necessário, i.e., deve obedecer inteiramente ao princípio da proporcionalidade. Por exemplo, se se entender que o mesmo fim seria cumprido por outra via menos intrusiva, então a exceção não se aplica, na medida em que o princípio da proporcionalidade não foi observado. Remetendo novamente ao considerando 18, percebe-se que vários sistemas ligados à segurança já estariam fora deste âmbito por definição, como o caso referido dos sistemas de deteção de fadiga em motoristas e pilotos profissionais na aviação e logística para a prevenção de acidentes.

Relembrando ainda também que os sistemas de reconhecimento de emoções abrangidos pela exceção da proibição podem ainda ser qualificados como sistemas de alto risco, nos termos do Artigo 6.º, n.º 2, e do Anexo III, ponto 1, alínea c), do Regulamento, devendo cumprir os requisitos aplicáveis aos sistemas de alto elevado previstos no Capítulo III, Secção 2, do mesmo, bem como a obrigação de transparência estabelecida no Artigo 50.º, n.º 3. Isto inclui os sistemas de reconhecimento de emoções que não estão no âmbito de aplicação da alínea f), tal como aqueles que são utilizados fora do local de trabalho e instituições de ensino (que, ainda assim, podem estar proibidos por outros diplomas comunitários, como o RGPD ou até mesmo por outras disposições do Artigo 5.º, como a alínea a) e b)).

É ainda importante sublinhar, dado que a alínea f) afeta também direitos dos trabalhadores, que o disposto no Artigo 2.º, alínea 11) do Regulamento é aplicável, sendo esta uma norma que reconhece explicitamente uma cláusula de norma mais favorável, i.e., os Estados Membros podem introduzir e manter disposições legislativas, regulamentares ou administrativas mais favoráveis para os trabalhadores em termos de proteção dos seus direitos no que diz respeito à

utilização de sistemas de IA por empregadores e de incentivarem ou permitirem a aplicação de convenções coletivas mais favoráveis para os trabalhadores. Esta disposição aplica-se diretamente à alínea f), pelo que os Estados Membro poderão adotar normas que garantam um maior nível de proteção dos trabalhadores no âmbito dos sistemas de reconhecimento de emoções. Dá-se o exemplo de sistemas de reconhecimento de emoções que podem ser utilizados em contextos de marketing digital ou em linha, para interagir com clientes, que em princípio não estão abrangidos pela proibição da alínea f). Sistemas que analisam o tom de voz, mensagens escritas ou padrões de escrita no teclado para fins de marketing personalizado, incluindo publicidade em ambientes inteligentes (como painéis interativos), não são proibidos pela alínea f). Contudo, estas mesmas práticas podem estar abrangidas pelas proibições de técnicas manifestamente manipuladoras ou enganadoras ou que explorem vulnerabilidades previstas nas alíneas a) e b) do Artigo 5.º, caso tal se aplique[41].

A alínea g), por sua vez, tipifica a proibição de sistemas de categorização biométrica que classifiquem individualmente pessoas singulares baseando-se nos seus dados biométricos como forma a deduzir ou inferir características sensíveis (particularmente a raça, filiação sindical, vida e orientação sexual, convicção religiosa, filosófica e política, etc.). Esta proibição não abrangerá a filtragens, rotulagens e categorizações de dados biométricos que sejam legalmente obtidos em conformidade com o direito da União ou dos Estados-Membros, particularmente nos casos da aplicação da lei.

O fundamento desta proibição não será muito diferente do que tem sido observado. A principal questão cinge-se ao leque de informações pessoais que podem ser extraídas, inferidas ou deduzidas com base nos dados biométricos de uma pessoa individual. Particularmente, existe a preocupação em relação às informações sensíveis que podem ser obtidas por esta via, sem consentimento nem conhecimento das partes envolvidas. É, num primeiro momento, imediatamente evidente o risco que se afigura em relação à proteção da privacidade das pessoas singulares. Numa segunda análise, é fácil compreender o perigo significativo para outros direitos fundamentais, tal como a igualdade, a não-discriminação e a própria dignidade humana. Mais uma vez, a proibição é motivada pelo risco significativo aos direitos fundamentais e ao

[41] Nunca esquecendo, mais uma vez, a potencial aplicabilidade das restantes disposições comunitárias relacionadas à proteção de dados e proteção dos consumidores, na medida em que a maioria dos visado não padecem de vulnerabilidades. Cf.: PARLAMENTO EUROPEU; CONSELHO DA UNIÃO EUROPEIA. Diretiva (UE) 2019/2161, de 27 de novembro de 2019, que altera as Diretivas 93/13/CEE, 98/6/CE, 2005/29/CE e 2011/83/UE no que respeita à melhor aplicação e modernização das regras da União em matéria de defesa do consumidor. Jornal Oficial da União Europeia, L 328, 18 dez. 2019

bem-estar. É uma proibição que não depende, ao contrário da anterior, de um local, sendo de âmbito mais abrangente. De qualquer modo, expressa-se a preocupação em relação à proteção dos dados biométricos de forma a evitar, por exemplo, que uma realidade de discriminação algorítmica totalmente automatizada e sistematizada se materialize.

As condições para a aplicação da proibição da alínea g) são, à semelhança do que se tem verificado, cumulativas. Parecem ser, desta vez, cinco condições distintas que se devem verificar para que o preceito se aplique, nomeadamente, como é transversal a todas as proibições, a colocação no mercado para o fim específico, em serviço ou a utilização de um sistema de IA. Mais uma vez, basta que a prática constitua um ato de colocação do mercado para que o pressuposto esteja preenchido. Aqui, como em todos os casos, a utilização do sistema é o facto mais gravoso, pelo que esse ato poderá constituir a materialização do dano, ao passo que as anteriores representarão, em princípio, mero perigo ou iminência de dano. Para o efeito, isto tem que ver com a garantia de responsabilidade ao longo de toda a cadeia, procurando evitar a concentração de responsabilidade num ponto único. Neste caso, a segunda condição prende-se com o tipo de sistema, que terá de ser necessariamente um sistema de categorização biométrica. Para cumprir a terceira condição, esse sistema terá de classificar/categorizar individualmente pessoas singulares. Em quarto, o sistema terá de se basear, na sua classificação, em dados biométricos das pessoas singulares em questão. Por último, o fim prosseguido terá de ser o da inferência ou dedução da raça, opinião política, filiação sindical, convicção religiosa ou filosófica, vida sexual ou orientação sexual da pessoa singular. Só através do preenchimento das cinco condições referidas é que a proibição se aplicará, salvaguardando a exceção explícita na última parte da norma, ou seja, os casos em que as rotulagens e filtragens de conjuntos de dados biométricos são legalmente adquiridos ou são utilizados no domínio da aplicação da lei.

Para compreender a questão da categorização biométrica, é importante perceber o processo de categorização biométrica no âmbito destes sistemas. A categorização ou classificação de uma pessoa individual por este tipo de sistemas não tem que ver com a simples identificação do indivíduo ou com a mera verificação da sua identidade. Como o termo indica, estamos perante sistemas que atribuem categorias a indivíduos num processo de rotulagem. Consiste no processo de associar os dados biométricos de alguém correspondem a um grupo (ou rótulo). É uma prática recorrente na publicidade digital, onde os utilizadores são alvo de anúncios personalizados de acordo com algumas categorias como a idade, género, estado civil e área

geográfica[42]. Doutra forma, a categorização biométrica também pode ser usada, como é feito regularmente, para fins estatísticos, como no caso dos censos demográficos. No âmbito do regulamento, um sistema de categorização biométrica é definido como "um sistema de IA destinado a afetar pessoas singulares a categorias específicas com base nos seus dados biométricos, a menos que seja acessório a outro serviço comercial e estritamente necessário por razões técnicas objetivas"[43]. Como já foi abordado, o conceito de "dados biométricos" corresponde a dados pessoais como imagens faciais e dados dactiloscópicos, tal como definido pelo ponto 34) do Artigo 3.º do Regulamento. Excluir-se-ão, então, acessórios, indumentária e atividade concreta nas redes sociais, por exemplo. A problemática está nos casos em que a indumentária possa servir a inferência da raça, ainda que com margem de erro. E como é evidente, a questão mais pertinente, como já referido, é justamente a categorização ou a criação de categorias com base em determinadas características sensíveis como a raça. Como é claro, levantam-se sérias preocupações em relação aos direitos fundamentais, particularmente no campo da não-discriminação. Além destes dados, contudo, a categorização também se poderá basear, além dos dados de imagens faciais e afins, em aspetos comportamentais (que não a mera atividade nas redes sociais, como dito) como a forma de andar ou de escrever em teclado, como também já foi mencionado.

Doutra forma, atentando ainda no ponto 40 do artigo 4.º do regulamento, torna-se claro o alcance material dos sistemas de categorização biométrica (e entenda-se, a categorização biométrica como conceito). Com efeito, se o sistema for acessório a outro serviço comercial e estritamente necessário por razões técnicas objetivas, fica fora do alcance material do conceito. Ora, entende-se que estas razões técnicas tenham de ser claramente demonstráveis. Na análise da norma, é de sobeja relevância considerar e articular todos os pressupostos de aplicabilidade de forma sistémica. Seguindo a mesma lógica, afigura-se ainda essencial considerar o disposto no considerando 16 do regulamento, que torna ainda mais claro o conceito. De acordo com o considerando, corroboramos o facto da categorização biométrica se prender com a atribuição de pessoas singulares a categorias específicas com os dados biométricos, como previamente abordado. Ainda mais, reforçamos o entendimento de que, para que uma prática esteja fora

[42] Já no âmbito da preparação do RGPD se levantava a questão da categorização biométrica e o seu papel em anúncios, por exemplo. GRUPO DE TRABALHO DO ARTIGO 29.º PARA A PROTECÇÃO DE DADOS. *Opinion 02/2012 on facial recognition in online and mobile services*. Bruxelas: Comissão Europeia, 2012. (WP 193). Disponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf, p. 6.

[43] Artigo 4º, nº40) do Regulamento da Inteligência Artificial.

deste âmbito, deve ser meramente acessória e ligada a outro serviço comercial de forma intrínseca. Significa, portanto, que nunca poderá ser utilizada sem o serviço comercial principal, devendo ser verdadeiramente inoperante sem o mesmo. Será importante provar esta inoperância de forma técnica e objetiva. A preocupação aqui é, mais uma vez, garantir na medida do possível que a exceção em causa não é usada para evitar a aplicação da norma. Posto isto, não basta a mera aparência de dependência em relação a uma atividade comercial principal, pelo que importa a existência objetiva de um nexo de operacionalidade entre o sistema de categorização e a atividade principal. Na prática, alguns exemplos não levantarão grandes dúvidas, como os casos apresentados pelo próprio considerando 16. Os filtros de categorização facial ou corporal utilizados por mercados em linha deverão, na maioria dos casos, constituir um elemento acessório estritamente necessário do ponto de vista objetivo e inteiramente dependente do serviço principal que é o da venda de produtos. Se o sistema apenas permitir a possibilidade de uma pessoa se pré-visualizar a usar o produto, poderá estar dentro da exceção. Noutros casos, como o das redes sociais, em que existe categorização através de filtros que permitem aos utilizadores acrescentarem ou alterarem imagens também poderemos estar perante um caso exclusivamente acessório enquadrado na exceção, considerando ainda que estes filtros não possam ser utilizados sem o serviço principal (que são as redes sociais e a partilha de conteúdos digitais em linha). Noutra ótica, de forma contrastante, existirão casos em que será mais difícil provar ou defender a aplicação da exceção. Por exemplo, no caso em que se categorizam utilizadores de uma rede social de acordo com a sua posição política presumida (ou inferida) através de dados biométricos (fotografias introduzidas pelo próprio utilizador na rede), com fim de exibir mensagens políticas direcionadas ou segmentadas. Ainda que o sistema seja meramente acessório da publicidade política efetuada, levantam-se sérias dúvidas em relação à sua necessidade estrita do ponto de vista técnico-objetivo, pelo que este tipo de utilização estará, na consideração adotada, vedado. Noutro exemplo, a categorização de pessoas numa rede social de acordo com a sua orientação sexual presumida da mesma forma, com objetivo de segmentação publicitária também se afigura inaceitável. Não se assemelhará a um caso de estrita necessidade deste elemento acessório, pelo que a exceção não se aplicará e o sistema será considerado de categorização biométrica. Na realidade, os casos poderão levantar várias dúvidas, pelo que parece importante considerar cuidadosamente o interesse de prevenção de circunvenção da norma e o de proteção de direitos fundamentais, particularmente a não-discriminação.

Reiterando, note-se que a aplicação da proibição depende da categorização a título individual. Se a categorização se limitar a agrupar coletividades (sem identificar cada elemento), a proibição não será aplicável. Ora, os casos de contagem demográfica que incluem idade, género e etnia, ou a identificação de indivíduos que têm uma característica singular (uma tatuagem no braço esquerdo, por exemplo) são casos de categorização individual. Outro exemplo será a da dedução ou inferência de características sensíveis, como a inferência ou dedução da raça de um indivíduo através da voz ou a dedução da orientação religiosa por via do rosto e tatuagens em combinação com acessórios como um crucifixo (aqui o que releva é a combinação de dados biométricos). Doutra maneira, um sistema que analise o ADN de vítimas de crimes tendo em vista a sua origem estará fora do âmbito da alínea g).

Importa relembrar que a proibição da alínea g) do Artigo 5.º, n. º1, não se aplica a sistemas de IA que rotulem e filtrem conjuntos de dados biométricos legalmente adquiridos, por exemplo no domínio da aplicação da lei[44]. Mas mesmo nestes casos é importante ter em conta que tal categorização biométrica só poderá ser feita de forma que garanta a justa representação de todos os grupos demográficos e não seja utilizada para representar de forma desproporcional um grupo específico. Mais uma vez, surge a preocupação em relação à qualidade dos dados em si. Os dados devem estar livres de vieses, sejam eles históricos ou resultantes de outras circunstâncias sistémicas subjacentes à sua forma de recolha. Nestes casos, poderá ser necessário, inclusive, recorrer a um processo de rotulagem dados sensíveis para garantir a qualidade elevada dos dados, especificamente na prevenção da discriminação. É algo que estará isento, em princípio, da proibição sob análise, na medida em que não visa a inferência de raça, opiniões políticas e outras. É, doutra forma, uma operação que até poderá estar prevista pelo próprio regulamento no âmbito das exigências relativas à governação de dados e sistemas de gestão de qualidade[45]. Nesta linha de raciocínio, dão-se exemplos de rotulagem ou categorização permissível, como: a rotulagem de dados biométricos com fim de evitar a discriminação de membros de uma etnia em processos de recrutamento devido a vieses nos

[44] O considerando 30, já mencionado, esclarece estes casos. Se os conjuntos de dados de dados biométricos categorizados forem adquiridos em conformidade com o direito da União ou o próprio direito nacional em matéria de dados biométricos, como é o caso da triagem de imagens em função da cor do cabelo ou dos olhos e são utilizadas para a aplicação da lei.

[45] Artigos 10º e 17º do regulamento, respetivamente. Destaca-se o processo de deteção e correção de enviesamentos do artigo 10º, nº5 e respetivas alíneas. Além disso, refere-se o artigo 10º, nº2, alínea f) em relação ao exame para deteção de eventuais enviesamentos suscetíveis de afetar a saúde e segurança, bem como ter repercussões negativas nos direitos fundamentais.

conjuntos de dados utilizados para treinar os sistemas de IA[46] e na categorização de pacientes através da sua cor de pele ou de olhos com objetivo de auxiliar um diagnóstico médico.

Em semelhança às demais alíneas, esta proibição deve ter cm conta as disposições de outros diplomas de direito da União, nomeadamente o RGPD (no seu artigo 9.º, n. º1, por exemplo) e outros diplomas que tratem de matérias relativas a dados pessoais, como a Diretiva de Proteção de Dados na Aplicação da Lei[47]. Os sistemas que não estão proibidos pela alínea g) do artigo 5.º do regulamento e utilizam categorização biométrica através de atributos sensíveis ou características protegidas pelo já mencionado artigo 9.º, n. º1 do RGPD, são classificados como sistemas de risco elevado[48]. Em suma, a alínea g) parece reforçar a proteção de dados pessoais na União e fortificar a rede regulatória relativa à proteção de dados. Como já foi referido, esta proibição parece estar em sintonia com os demais diplomas comunitários em relação à matéria em causa, devendo ser interpretada também de acordo com os mesmos, seguindo uma lógica de interpretação sistémica de todo o ordenamento jurídico europeu e até nacional, na medida do possível. A este propósito também surge o n. º8 do artigo 5.º, que salvaguarda justamente as proibições aplicáveis por outra legislação da União. Neste sentido, o Regulamento da IA parece atuar como regime geral, pelo que as proibições específicas não são afetadas.

Por último, a alínea h) do artigo 5.º evidencia-se como a mais densa proibição prevista na matéria do risco inaceitável. Para efeitos de concisão, procurar-se-á limitar a análise deste preceito a uma forma relativamente breve, abordando os traços gerais e questões essenciais[49]. A proibição em causa cinge-se com a utilização de sistemas de identificação biométrica à distância em tempo real, em espaços acessíveis ao público para efeitos de aplicação da lei, a

[46] É justamente o caso que já foi referido previamente, abordado em maior detalhe na já citada obra: AGÊNCIA DOS DIREITOS FUNDAMENTAIS DA UNIÃO EUROPEIA, *Big Data: Discrimination in data-supported decision making*. [Em linha]. Viena: FRA, 2018, disponível em: http://fra.europa.eu/en/publication/2018/big-data-discrimination

[47] UNIÃO EUROPEIA, Diretiva (UE) 2016/680 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais pelas autoridades competentes para efeitos de prevenção, investigação, deteção ou repressão de infrações penais ou execução de sanções penais, e à livre circulação desses dados, e que revoga a Decisão-Quadro 2008/977/JAI do Conselho. Jornal Oficial da União Europeia, L 119, 4.5.2016, p. 89–131. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32016L0680.

[48] De acordo com o Anexo III do Regulamento da Inteligência Artificial, com atenção ao disposto no considerando 54 do mesmo e tendo em conta a exceção prevista no nº1 da alínea a) do Anexo em causa.

[49] Para uma análise completa, consultem-se diretamente as linhas orientadoras da Comissão Europeia sobre as práticas de IA proibidas, que têm servido de base fundamental da análise aqui exposta: COMISSÃO EUROPEIA, *Annex to the Communication to the Commission – Approval of the content of the draft Communication from the Commission: Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, documento C(2025) 884 final, 4 fev. 2025, pp. 95-134, disponível em: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

menos e na medida em que essa utilização seja estritamente necessária para os fins explicitamente e exaustivamente previstos nos pontos (i) a (iii) da própria alínea. Estes três fins elencam as circunstâncias sob as quais estes tipos de sistemas podem ser utilizados, salvaguardando a autorização de legislação nacional e as condições dos seguintes números do Artigo 5.º (n.ºs 2 a 7, que preveem condições e salvaguardas específicas a ser observadas). Desde logo, o n. º5 do Artigo 5.º prevê que os Estados-Membros são livres de decidir autorizar total ou parcialmente, através de legislação nacional, (ou não autorizar de todo) em quais das três situações a utilização de identificação biométrica à distância em tempo real é permitida no seu território (de acordo com todos os termos do regulamento). A existência de legislação nacional nesta matéria é um pré-requisito essencial para a utilização deste tipo de sistemas. Se não houver legislação nacional que autorize e regule este tipo de sistemas, nos termos do Artigo 5.º, as forças de segurança e entidades competentes não poderão utilizar tais sistemas. Note-se, ainda, que a proibição se aplica não à colocação no mercado e colocação em serviço, mas sim a utilização do sistema (ao contrário de todas as proibições anteriores). Isto significa que a colocação no mercado e colocação em serviço destes sistemas, tal como outros tipos de sistemas de identificação biométrica à distância, não são proibidos, mas sim sujeitos às normas relativas a sistemas de risco elevado, de acordo com ao já referido artigo 6.º, n. º2 em articulação com a alínea a) do Anexo III do Regulamento da Inteligência Artificial. Isto também se assemelha relevante para os casos em que os Estados Membros autorizam a utilização de sistemas de identificação biométrica à distância em tempo real em espaços acessíveis ao público para efeitos da aplicação da lei, na medida em que nesses casos autorizados as regras para sistemas de risco alivado também serão aplicáveis[50].

O considerando 32 do Regulamento reconhece claramente os possíveis perigos dos sistemas de identificação biométrica à distância em tempo real, particularmente nos espaços acessíveis ao público, na medida em que se entende a natureza particularmente intrusiva dos mesmos. Em causa está a vida privada das pessoas singulares (de uma grande fatia da população), o sentimento de vigilância constante e o constrangimento do exercício da liberdade de reunião, expressão, religião e outros direitos fundamentais, como a não discriminação. Como se tem estabelecido como transversal, o enviesamento de resultados e efeitos discriminatórios práticos

[50] Nos casos em que a identificação não seja em tempo real, inclui o disposto no Artigo 26.º, n.º10 do Regulamento da IA, que salvaguarda o disposto na Diretiva (UE) 2016/680 e estabelece, grosso modo, que para a utilização de um sistema de IA de alto risco para identificação biométrica remota de um suspeito ou condenado, deve obter-se, salvo para identificação inicial baseada em factos objetivos, uma autorização vinculativa de autoridade judicial ou administrativa em até 48 h, limitando cada uso ao estritamente necessário à investigação da infração em causa

assumem particular relevância no que diz respeito à idade, raça, etnia, sexo ou deficiência. Além disso, de forma extremamente relevante, coloca-se em causa a questão dos sistemas deste tipo, que por atuarem em tempo real e terem impacto imediato, limitarão a possibilidade de realização eficaz em tempo útil de controlos adicionais, salvaguardas e correções à sua utilização. São sistemas que, dada essa característica, implicam riscos acrescidos para os direitos e liberdades das pessoas impactadas no decorrer das atividades de aplicação da lei[51]. Desta forma, de seguida, o considerando 33 reconhece e densifica a *ratio* por detrás das exceções aplicáveis, já referidas. Com efeito, entende-se que em certas situações elencadas exaustivamente, definidas de modo restrito, em que a utilização se afigura estritamente necessária por motivos de interesse público importante e cuja importância prevalece sobre os riscos para os direitos fundamentais, estes sistemas poderão ser utilizados para efeitos de aplicação da lei. Como já vimos, aplicam-se as exigências do Artigo 5.º, n.ºs 2 a 7.

Surge a questão de definir "espaço acessível ao público. Ora, veja-se que o conceito de espaço acessível ao público é definido pelo Regulamento no artigo 3.º, ponto 44, sendo: "…qualquer espaço físico, público ou privado, acessível a um número indeterminado de pessoas singulares, independentemente da eventual aplicação de condições de acesso específicas e independentemente das eventuais restrições de capacidade;". O considerando 19 clarifica ainda mais o conceito, providenciando algumas noções essenciais para a compreensão do conceito. Começando pela contrapartida, um espaço não é considerado acessível ao público se o seu acesso for limitado a pessoas singulares específicas e definidas (por ex., os escritórios de uma empresa que dependem de credenciais para entrar), seja por questão ligadas ao direito da União ou nacional, com questões de segurança pública ou pela própria vontade da pessoa que exerce autoridade sobre o espaço e o manifesta claramente. O facto de o acesso ser possível fisicamente, por si só, não implica que o espaço seja acessível ao público, por exemplo: um terreno pode não ter vedação, mas podem existir indicações ou circunstâncias claras que

[51] A propósito, destaca-se a crítica tecida por *Alice Giannini* e *Sarah Tas* à forma da proibição do Regulamento dos sistemas de identificação biométrica no regulamento, referindo particularmente o entendimento de que a identificação biométrica à distância em diferido poderá, discutivelmente, ter um impacto muito mais vasto e nefasto do que os sistemas em tempo real. O facto de os conjuntos de dados serem arquivados em grandes quantidades, dando aso a um tratamento mais refinado, pode repercutir-se nos efeitos a jusante. A ideia de que uma pessoa singular possa ser identificada meses depois de ter estado num certo espaço público (como um evento cultural), poderá ter tanto ou mais impacto do que a sua identificação em tempo real. Entende-se que o medo causado por esta possibilidade poderá ter o mesmo efeito limitador da liberdade de reunião, por exemplo. GIANNINI, Alice; TAS, Sarah, *AI Act and the Prohibition of Real-Time Biometric Identification: Much ado about nothing?* [em linha]. *VerfBlog*, 10 dez. 2024. Disponível em: https://verfassungsblog.de/ai-act-and-the-prohibition-of-real-time-biometric-identification/. DOI: 10.59704/a15aa6e58151c853

indicam o contrário, como sinais de proibição e restrição de acesso, etc. Em concreto, as prisões e zonas de controlo fronteiriço não são acessíveis ao público. Também existe a possibilidade de certos espaços serem constituídos por alguns espaços não acessíveis ao público e outros acessíveis ao público, como é o caso de um corredor de um edifício residencial privado necessário para aceder a um gabinete médico ou a um aeroporto. Os espaços em linha também não são abrangidos, sendo que não são espaços físicos. Em geral, entende-se que a determinação de um espaço como acessível ao público dependerá de uma análise casuística, tendo em causa as especificidades concretas. Por outro lado, um espaço deverá ser classificado como acessível ao público mesmo que haja eventuais restrições de capacidade ou de segurança, e até haja condições de acesso predeterminadas, desde que sejam potencialmente preenchidas por um número indeterminado de pessoas, como a compra de um bilhete ou título de transporte, inscrição prévia, ou uma determinada idade, como o local de um concerto ou um evento especialmente destinado à população acima dos 65 anos.

De forma concreta, contudo, quais são os requisitos para a aplicação da proibição da alínea h)? São cinco condições cumulativas que têm de ser preenchidas para que a proibição se aplique, a saber: primeiro, o sistema de IA tem de ser de identificação biométrica à distância, segundo, a conduta em causa tem de consistir na utilização desse mesmo sistema, terceiro, o sistema funciona em tempo real, quarto, em espaços acessíveis ao público e, por último, para efeitos de aplicação da lei. O termo "utilização", apesar de não ser explicitamente definido pelo regulamento, deve ser entendido num sentido lato, não particularmente restritivo. Em termos concetuais, um sistema de identificação biométrica à distância é definido no âmbito do regulamento, pelo ponto 41) do artigo 3.º, como um sistema de IA concebido para identificar pessoas singulares, sem a sua participação ativa, normalmente à distância, por meio da comparação dos dados biométricos de uma pessoa com os dados biométricos contidos numa base de dados de referência.

Em termos práticos, contudo, o que é que poderá ser apontado em relação à proibição da alínea h)? Por exemplo, a verificação ou autenticação biométrica estão fora do âmbito de aplicação da proibição. Isto significa que a autenticação biométrica de viajantes num aeroporto[52], comparando imagens faciais com a imagem presente no passaporte, não será proibida pela alínea h). O considerando 17 expõe e clarifica algumas exclusões de âmbito do

[52] Recordando que alguns espaços podem ter uma "função dupla", isto é, podem ser geralmente considerados acessíveis ao público, como os aeroportos nas suas áreas comuns (átrio, por exemplo), mas terem áreas dedicadas ao controlo de fronteiras e alfândegas que estão especialmente excluídos do âmbito desta proibição.

conceito de sistema de identificação biométricas à distância, como esta, ao passo que se entende que estes sistemas terão um impacto ligeiro nos direitos fundamentais das pessoas singulares quando comparados com sistemas de identificação biométrica à distância com alcance superior a nível de número de pessoas. Imaginemos, por outro lado, um caso de um atentado terrorista num parque público. As forças de segurança recebem informação de que um indivíduo se encontra num parque, armado com uma faca, a gritar ameaças e a perseguir transeuntes. Percebe-se que, pelas ameaças que grita, o ataque será de cariz terrorista. Neste caso, a identificação biométrica à distância em tempo real será admissível, assumindo que o estado-membro autorizou a utilização destes sistemas nos termos do Artigo 5.º, n. º1, alínea h), ponto ii) e as condições dos n.ºs 2 a 7 também foram observadas. Na prática, as forças de segurança poderão utilizar estes sistemas para identificar e localizar em tempo real a pessoa em questão, com objetivo de prevenir o ataque.

2.4 Sistemas de risco elevado

O Artigo 6.º identifica sistemas de IA que são automaticamente categorizados como de risco elevado, tal como quando são componentes de segurança de produtos regulados por legislação da UE [Artigo 6.º, n. º1, alínea a), b) e Anexo I], estando sujeitos a avaliação de conformidade obrigatória por terceiros. Por outras palavras, se o sistema de IA integrado é parte de um produto sujeito a certificação (ex.: equipamento médico, equipamento de transporte, dispositivo industrial), assume automaticamente a classificação de risco elevado. O Anexo I relaciona a extensa lista de normas de produto relevantes (por exemplo, diretivas sobre máquinas 2006/42/CE, brinquedos 2009/48/CE, reboques e veículos 2018/858/UE, dispositivos médicos 2017/745/UE, etc.). Assim, por exemplo, um sistema de IA que controle os sistemas de condução de um automóvel (como os travões e a direção) ou um software de diagnóstico robótico num equipamento médico, estando em produto sujeito a certificação, será de risco elevado.[53] Importa ainda destacar o considerando 51, que clarifica que a classificação de um sistema de IA como sendo de risco elevado não implicará obrigatoriamente que o produto em causa (que tem o sistema de IA como componente de segurança) ou o próprio sistema de IA

[53] Para clarificar: o conceito de "componente de segurança" está previsto no artigo 3º, ponto 14 do próprio Regulamento, sendo definido como "um componente de um produto ou sistema de IA que cumpre uma função de segurança nesse produto ou sistema de IA, ou cuja falha ou anomalia põe em risco a saúde e a segurança de pessoas ou bens;". É uma definição relativamente ampla, pelo que será sensato, em qualquer caso, proceder a uma interpretação em sintonia com a legislação europeia relevante, nomeadamente as previamente citadas diretivas sobre segurança de produto. Não implica, contudo, que esta interpretação em harmonia com os preceitos comunitários especiais não possa causar eventuais conflitos.

enquanto produto, seja de risco elevado. Por exemplo, o regulamento relativo aos dispositivos médicos prevê uma avaliação de conformidade por terceiros para produtos de risco médio e elevado, i.e., o sistema de IA pode ser classificado como sendo de risco elevado, mas isso não significa que o produto subjacente também o seja, por exemplo.

O Artigo 6.º, n.º2 faz referência direta ao anexo III, remetendo para este a identificação dos oito domínios nos quais os sistemas de IA serão considerados de risco elevado, como por exemplo: na área dos dados biométricos (incluindo identificação biométrica), das infraestruturas críticas, da educação e formação profissional (em casos específicos, como a deteção de práticas proibidas por parte de estudantes durante testes e a determinação do acesso ou admissão de pessoas singulares nos estabelecimentos de ensino), do emprego (em casos como na colocação de anúncios de emprego direcionados, filtragem de candidaturas, etc.), do acesso a serviços públicos e privados essenciais, da aplicação da lei (nos termos permitidos nos termos do direito comunitário) e da migração e controlo de fronteiras e por fim, a administração da justiça e processos democráticos. Esta remissão é um mecanismo essencial na flexibilização do regime dos sistemas de risco elevado, na medida em que permite à Comissão Europeia, por via de atos delegados, atualizar o Anexo. Este processo de alteração ou aditamento de novas condições está previsto de forma simplificada nos artigos 6.º, n. º6, 7 e 9 e artigo 7.º, em articulação com os termos do artigo 97.º, que por sua vez prevê o exercício desta delegação de poderes, conferido à Comissão. Esta possibilidade será indubitavelmente útil e permitirá a eventual expansão do elenco, seja por eventuais novas descobertas no decorrer da aplicação prática do regulamento ou como resposta a críticas que se têm apontado ao mesmo.[54] Destaca-se ainda o artigo 6.º, n.º5, que estabelece a disponibilização de orientações por parte da Comissão que especifiquem a aplicação do presente artigo, à semelhança das linhas orientadoras, já publicadas, relativas às práticas proibidas de IA proibidas, que foram extensamente analisadas e expostas na presente dissertação.[55]

[54] Sandra Wachter reconhece o Anexo III como um "bom começo", mas não deixa de tecer críticas em relação à omissão de algumas áreas de importância, tal como a da utilização de IA na multimédia, ciência e academia, mercados financeiros, seguros, etc. WACHTER, Sandra, *Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond* [em linha]. *Yale Journal of Law & Technology*, vol. 26, n. º 3, 2024. pp. 682 e ss. Disponível em: https://doi.org/10.2139/ssrn.4924553

[55] O artigo 6.º, n.º5 também estabelece a disponibilização de "uma lista exaustiva de exemplos práticos de casos de utilização de sistemas de IA de risco elevado e de risco não elevado.". Entender-se-á aqui, de facto, que a expressão "exaustiva" terá sido um erro de tradução, pelo que a lista não deverá ser literalmente exaustiva, mas abrangente (o tanto quanto possível), como sublinha SILVA, Nuno Sousa, O Regulamento da Inteligência Artificial: análise introdutória, [em linha] Revista da Ordem dos Advogados, 2024, p. 350. Disponível em: https://portal.oa.pt/media/147835/nuno-sousa-e-silva.pdf

Em contrapartida, reconhecem-se exceções à regra, além das exceções específicas do Anexo III, n. º1, alínea a) e n. º5, alínea b), respeitantes a sistemas de verificação de identidade e classificação de crédito utilizados no combate à fraude financeira, respetivamente. O Artigo 6.º, n. º3 prevê a derrogação do n. º2, permitindo o afastamento da classificação de risco elevado para os sistemas identificados pelo Anexo III de forma geral. Se, por sua vez, o sistema em causa "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.". É algo que, na prática, considerar-se-á aplicável caso se cumpram qualquer umas das seguintes circunstâncias, elencadas no próprio artigo, passando a citar:

a) "O sistema de IA destina-se a desempenhar uma tarefa processual restrita;
b) O sistema de IA destina-se a melhorar o resultado de uma atividade humana previamente concluída;
c) O sistema de IA destina-se a detetar padrões de tomada de decisões ou desvios em relação a padrões de tomada de decisões anteriores e não se destina a substituir nem influenciar uma avaliação humana previamente concluída, sem que se proceda a uma verificação adequada por um ser humano; ou
d) O sistema de IA destina-se a executar uma tarefa preparatória no contexto de uma avaliação pertinente para efeitos dos casos de utilização enumerados no anexo III."

Resulta, assim, que um sistema de IA que seja utilizado numa das áreas identificadas pelo Anexo III possa não ser considerado de risco elevado, logo que preencha qualquer uma das alíneas mencionadas.[56]Além disso, também é importante reter que, não obstante esta exceção, os sistemas que executem a definição de perfis de pessoas singulares serão sempre considerados de risco elevado, nos termos deste regime. De qualquer forma, um prestador que entenda que um sistema de IA não é de risco elevado deve seguir o procedimento indicado no artigo 6.º, n. º4, do qual a derrogação em questão depende. Em suma, a avaliação do sistema deve ser documentada antes da sua colocação no mercado ou em serviço, havendo ainda a obrigação de

[56] O considerando 53 é particularmente relevante neste âmbito, na medida em que nos fornece alguns exemplos de sistemas de IA que, pela sua ausência de influência significativa na substância e no resultado da tomada de decisões, poderão não ser considerados de risco elevado nestes termos. É o exemplo dos sistemas de IA que classificam meramente documentos recebidos, que detetam duplicações entre documentos e aplicações, apenas se destinam a melhorar a linguagem utilizada em documentos já redigidos (rem relação ao seu tom, seja ele mais académico, informal, profissional, etc.), verificam posteriormente desvios padrão na atribuição de notas de um professor (assinalando potenciais anomalias) ou fornecem soluções inteligentes para indexação, pesquisa, organização e processamento de documentos e até mesmo traduzam documentos iniciais.

registo do mesmo de acordo com o postulado no artigo 49.º, n. º2 (tendo também em conta o artigo 71.º, que tipifica a base de dados da UE relativa a estes sistemas de risco elevado). Na prática, cumpridos estes requisitos, os prestadores poderão colocar os sistemas no mercado, pelo que não necessitam de aguardar por aprovação ou validação da avaliação de risco. O que poderá acontecer, por outro lado, é a fiscalização por parte de uma autoridade de fiscalização do mercado segundo o previsto no artigo 80.º. A atuação da autoridade competente poderá resultar na proibição ou restrição da disponibilização ou colocação em serviço do sistema de IA em causa, no respetivo mercado nacional, segundo o artigo 79.º, n. º5, por exemplo (por remissão do disposto no artigo 80.º, n. º6). Parece, contudo, que na larga maioria dos casos as medidas corretivas exigíveis pela autoridade de fiscalização do mercado serão suficientes, além das possíveis coimas (artigo 99.º, por remissão do nº4 do artigo 80.º).

O resultado do regime em questão é difícil de antecipar, mas é possível reconhecer algumas críticas. Esta "avaliação de risco pré-mercado"[57] poderá revelar-se um entrave desnecessário à aplicabilidade do regime dos sistemas de IA de risco elevado, que de outra forma poderia ser simplificado através de uma classificação mais rígida. Uma aproximação baseada na perspetiva dos direitos fundamentais talvez justificasse um regime mais rígido, alicerçado numa classificação mediante a utilização pretendida, excluindo a acomodação de exceções tão complexas. Com efeito, as exceções parecem amplas, suscitando questões em relação à efetividade do regime. Por outro lado, todas as partes serão oneradas: quer os prestadores, que terão de efetuar a avaliação de risco e documentá-la, quer as autoridades competentes, que terão de os fiscalizar. Em suma, o regime dos sistemas de IA de risco elevado não é isento de crítica, pelo que o seu teor aparentemente compromissório poderá acabar por resultar em algumas ineficiências. Por outro lado, há vantagens inerentes a esta aproximação, nomeadamente a acomodação de dois polos diferentes: o da inovação e o da regulação.

Ora, as obrigações e requisitos aplicáveis aos sistemas de IA de risco elevado, em geral, são extensas (começando no artigo 9.º e acabando no 27.º, excluindo eventuais remissões e demais disposições aplicáveis, naturalmente). Desde logo, a criação e implementação de um sistema de gestão de riscos, segundo o artigo 9.º, que deve, além de ser um processo iterativo contínuo, documentado e mantido eficazmente, ter em consideração os potenciais efeitos e interações

[57] Segundo WACHTER, Sandra, *Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond* [em linha]. *Yale Journal of Law & Technology*, vol. 26, n. º 3, 2024. pp. 684–686. Disponível em: https://doi.org/10.2139/ssrn.4924553. A autora identifica algumas críticas relevantes ao modelo previsto pelo Artigo 6.º, aqui expostas.

resultantes da aplicação dos demais requisitos da secção 2 do Regulamento (segundo o Artigo 9.º, n. º4).[58] O Artigo 10.º, por sua vez, impõe alguns requisitos em relação à qualidade dos dados utilizados no treino de modelos. A finalidade é garantir alguns critérios de qualidade identificados nos n. º2 a 5 do próprio artigo, como a pertinência, representatividade, isenção e adequação dos dados, além da correção, deteção e mitigação de enviesamentos sempre que possível. É notável, neste âmbito, a criação de uma nova base de ilicitude[59] para o processamento de dados sensíveis com objetivo de asseverar, sempre na medida do estritamente necessário, a correção de enviesamentos. Isto não obsta, evidentemente, às obrigações previstas pelo RGPD. Na prática, contudo, formulam-se questões sobre a eficácia deste normativo, na medida em que parecem faltar algumas orientações concretas que indiquem, de forma clara, possíveis métodos e exemplos de mitigação de vieses, de prevenção, etc.[60]

Entre os demais requisitos, aplicáveis na sua maioria aos prestadores dos sistemas de IA, destaca-se o artigo 13.º e 14.º, que são elementos centrais da visão do Grupo de Peritos de Alto Nível sobre a IA[61]. Destaca-se, por sua vez, o artigo 14.º, que consagra, de certa forma, a supervisão humana, optando deixar os sistemas de IA de risco elevado sobre uma supervisão efetiva levada a cabo por pessoas. É uma exigência que, tal como indica a o n.º 1 do artigo, se aplica à própria génese dos sistemas, que devem ser concebidos e desenvolvidos, *ab initio*, com a supervisão humana em conta. Contudo, não aparenta ser uma medida estritamente técnica, mas também axiológica em certa medida. Releva a proporcionalidade e a efetividade da supervisão, pelo que não há reconhecimento concreto de um modelo a ser adotado, seja *human-*

[58] O considerando 65 oferece alguma clarificação sobre o sistema de gestão de riscos que deve ser implementado. Antes de mais, como é presumível, o sistema em questão deve ser executado ao longo de todo o ciclo de vida de um sistema de IA de risco elevado. Não só, é importante destacar que, além do processo de atualização regular, qualquer decisão e medida significativa tomada ao abrigo do regulamento deve ser justificada e documentada. Grosso modo, o sistema de gestão de risco, bem como as opções tomadas, deverá ser documentado em grande detalhe.

[59] Como identifica SILVA, Nuno Sousa, O Regulamento da Inteligência Artificial: análise introdutória, [em linha] Revista da Ordem dos Advogados, 2024, p. 352 e ss. Disponível em: https://portal.oa.pt/media/147835/nuno-sousa-e-silva.pdf

[60] Novamente, é uma crítica identificada por WACHTER, Sandra, *Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond* [em linha]. *Yale Journal of Law & Technology*, vol. 26, n. º 3, 2024. pp. 687–689. Disponível em: https://doi.org/10.2139/ssrn.4924553

[61] Isto no sentido em que o Grupo (e não só, na medida em que a discussão é anterior ao documento citado) defende elementos como a transparência e a ação e supervisão humana através de abordagens como *human-in-the-loop* e *human-in-control*, que se espelha na versão final do Regulamento. Mais do que isso, é uma evidência clara do objetivo geral de garantir uma IA de confiança. De modo geral, compreender o trabalho do Grupo é fundamental para entender o Regulamento. COMISSÃO EUROPEIA, Direção-Geral das Redes de Comunicação, Conteúdos e Tecnologias; GRUPO DE PERITOS DE ALTO NÍVEL SOBRE INTELIGÊNCIA ARTIFICIAL – Orientações éticas para uma IA de confiança [em linha]. Luxemburgo: Serviço das Publicações da União Europeia, 2019. Disponível em: https://data.europa.eu/doi/10.2759/2686

in-the-loop, *human-in-control*, entre outros.[62] Com efeito, a interpretação das exigências deste artigo não se deve traduzir na mera presença formal de um supervisor humano, correndo o risco de, na prática, o preceito se reduzir a um mero ritual formal que responsabiliza o utilizador final. Já o artigo 13.º dedica-se a garantir a transparência e explicabilidade dos sistemas de IA de risco elevado, particularmente em relação aos potenciais problemas que poderiam surgir graças à assimetria de informação entre os prestadores dos sistemas e os responsáveis pela implementação. Neste caso, o número 3 reconhece alguns elementos substanciais que devem ser incluídos nas instruções de utilização (algo concreto, isentando, em certa parte, o artigo de alguma crítica). É uma medida concreta que procura dar resposta às preocupações relacionadas com a opacidade e complexidade própria de alguns sistemas de IA. Em suma, garante que os responsáveis pela implantação conhecem as características, capacidades e limitações do desempenho do sistema em questão para que, em qualquer caso, exista uma base de apoio para que a tomada de decisões dos responsáveis seja informada, viabilizando a prevenção e utilização consciente.[63]

Modo geral, o esforço regulatório dos artigos 8.º a 15.º encontra-se na esfera dos prestadores, que são responsáveis pela implementação dos requisitos mencionados segundo o artigo 16.º. Além disso, garantem a criação de um sistema de gestão de qualidade (segundo ao artigo 17.º), mantêm a documentação nos 10 anos subsequentes à data de entrada no mercado, ou colocação em serviço do sistema de IA de risco elevado (artigo 18.º), tal como mantêm os registos gerados automaticamente (artigo 19.º). Os artigos 20.º e 21.º já incidem sobre um dever de colaboração com as autoridades competentes e de tomada de medidas corretivas relevantes (com remissão para o artigo 73.º e o dever de comunicação de incidentes graves). Além disso, reconhece-se a necessidade, no artigo 72.º, de dispor de um sistema de acompanhamento pós-comercialização, que também abrange o dever de informação do artigo 73.º.[64] Do ponto de vista

[62] Sobre a supervisão humana e o Artigo 14º do Regulamento, consultar: FINK, Melanie, *Human Oversight under Article 14 of the EU AI Act,* [em linha], em: MALGIERI, Gianclaudio; GONZÁLEZ FUSTER, Gloria; MANTELERO, Alessandro; ZANFIR-FORTUNA, Gabriela (eds.), *AI Act Commentary: A Thematic Analysis*. Oxford: Hart-Bloomsbury, 2026 (Forthcoming). Disponível em: SSRN: https://ssrn.com/abstract=5147196 or http://dx.doi.org/10.2139/ssrn.5147196

[63] Para uma perspetiva prática sobre a aplicação dos requisitos relativos aos sistemas de IA de risco elevado e sobre o artigo 13º, cf. SCHNURR, Daniel, *Effective implementation of requirements for high-risk AI systems under the AI Act* [em linha]. Oxford: *Centre for the Governance of AI*, 2024. Disponível em: https://ssrn.com/abstract=5147196.

[64] E é definido no artigo 3.º, n.º49 do Regulamento como "qualquer incidente ou anomalia num sistema de IA que, direta ou indiretamente, tenha alguma das seguintes consequências: a) morte de uma pessoa ou danos graves para a saúde de uma pessoa b) uma perturbação grave e irreversível da gestão ou do funcionamento de uma infraestrutura crítica, c) infração das obrigações decorrentes do direito da união destinadas a proteger os direitos fundamentais, d) danos graves a bens ou ao ambiente".

do processo e da burocracia que deve ser levada a cabo, as obrigações dos prestadores abrangem o seguimento do previsto no artigo 43.º, relativo ao procedimento de avaliação de conformidade, incluindo uma declaração de conformidade (artigo 47.º) e a aposição da marca CE (artigo 48.º). Note-se que o procedimento de avaliação de conformidade do artigo 43.º, que pode variar conforme a existência de normas harmonizadas ou não (e da aplicação do prestador das mesmas), poderá ser derrogado em casos específicos de acordo com o artigo 46.º, mediante pedido devidamente justificado. É uma derrogação que só deve ser aplicada em casos de situação de urgência e motivos verdadeiramente excecionais, justificados casuisticamente, podendo ser por motivos de segurança pública, ameaça específica e eminente para a vida ou segurança física de pessoas singulares, etc. Serão casos que, na prática, deverão ser residuais. Estas exigências serão, efetivamente, de cariz eminentemente burocrático e processual. Revela destacar, ainda dentro da mesma lógica, a obrigação do registo do sistema de risco elevado previsto no artigo 49.º[65], para efeitos da base de dados da UE relativa a estes sistemas prevista no artigo 71.º.[66]

Já na esfera dos responsáveis pela implantação contam com algumas obrigações previstas no artigo 26.º. De forma geral, as obrigações não suscitam grandes dúvidas: passam pela utilização dos sistemas de acordo com as instruções de utilização que os acompanham, a atribuição da supervisão humana a indivíduos que possuam a competência e formação devidas, exercer controlo efetivo sobre os dados de entrada (assegurando a sua pertinência e representatividade), a colaboração com as autoridades, manutenção de registos do sistema, etc. São, de certa forma, deveres de cuidado que os responsáveis pela implantação devem ter no exercício da sua função, que são, de certo ponto de vista, decorrentes da própria natureza do risco inerente a estes sistemas. Em relação às obrigações imputadas aos responsáveis pela implantação, todavia, destaca-se o artigo 27.º, que estabelece a obrigação, sob certos requisitos, de executar uma avaliação do impacto que as utilizações de certos sistemas de risco elevado possam ter nos direitos fundamentais. Isto significará que, quanto maior for o número de utilizadores (i.e. responsáveis pela implantação), maior será o número de avaliações de impacto nos direitos fundamentais. A questão é que o preceito em questão só se aplica a responsáveis

[65] Sobre a matéria de regulação de sistemas de risco elevado, ver também, novamente: SILVA, Nuno Sousa – O Regulamento da Inteligência Artificial: análise introdutória, Revista da Ordem dos Advogados, 2024, p. 352 e ss. Disponível em: https://portal.oa.pt/media/147835/nuno-sousa-e-silva.pdf

[66] O considerando 132 aborda a criação desta base de dados, esclarecendo que o objetivo será facilitar o trabalho da Comissão e dos Estados-Membros no domínio da Inteligência Artificial, além de aumentar a transparência para o público. É de notar que os responsáveis pela implantação que não sejam autoridades, agências ou organismos públicos, ainda que não estejam obrigados a fazê-lo, gozam do pleno direito de o fazer voluntariamente.

pela implantação que sejam organismos de direito público, ou entidades privadas que prestam serviços públicos e responsáveis pela implantação de sistemas de IA de risco elevado a que se refere o anexo III, ponto 5, alíneas b) e c), incluindo aqui os privados (entenda-se, primeiro, sistemas de avalização da capacidade de solvabilidade de pessoas singulares e classificação de crédito e, segundo, sistemas relacionados à fixação de preços no caso de seguros de vida e de saúde). O resultado consistirá, invariavelmente, num menor número de responsáveis sujeitos a esta obrigação. [67] Outro ponto a destacar em relação a esta obrigação relaciona-se com a relação ou interoperabilidade desta avaliação de impacto com a avaliação de impacto sobre a proteção de dados prevista pelo Regulamento Geral da Proteção de Dados no seu artigo 35.º. Ora, o próprio Regulamento da IA antevê a possível sobreposição, garantindo, na sua letra, a harmonização entre as duas avaliações, no n.º 4 do artigo 27.º. A norma reconhece a possibilidade de algumas obrigações previstas já terem sido cumpridas através da avaliação de impacto sobre a proteção de dados, indicando claramente que, nesse caso, a avaliação de impacto sobre os direitos fundamentais de um sistema de risco elevado deverá complementar a primeira.

Em suma, o regime dos sistemas de risco elevado é, sem sombra de dúvida, o mais extenso do Regulamento. Com efeito, constitui o corpo normativo ou conjunto principal de obrigações previstas, que não deixam de apresentar um certo grau de complexidade. Esta complexidade é, discutivelmente, necessária à flexibilização do sistema. Por um lado, reconhece-se a necessidade de rigidez em relação a sistemas utilizados como componente de segurança em produtos sujeitos a certificação, por outro, entende-se que os sistemas do elenco de áreas do Anexo III devem usufruir de alguma flexibilidade (certamente para viabilizar o acolhimento de evoluções tecnológicas e criar um caminho mais livra à inovação). Do ponto de vista do extenso regime de requisitos e obrigações, percebe-se, na ótica adotada, que os sistemas desta natureza carecem de *safeguards* que assegurem a sua fiabilidade. Não restam dúvidas de que o equilíbrio entre a proteção dos direitos e a inovação é de difícil alcance. Contudo, restará saber, através da prática (que viabilizará análises definitivas, de impacto e nos dotará com a bênção da retrospetiva) se o regime granjeará o sucesso que almeja, já que o mesmo depende de uma

[67] É também um ponto abordado por: WACHTER, Sandra, *Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond* [em linha], *Yale Journal of Law & Technology*, vol. 26, n. º 3, 2024. pp. 687–689. Disponível em: https://doi.org/10.2139/ssrn.4924553

conjugação muito precisa de rigor técnico, agilidade regulatória, proporcionalidade e atuação eficiente das autoridades competentes.

2.5 Obrigações de transparência

Existem ainda obrigações de transparência aplicáveis a determinados sistemas de IA (previstas unicamente no artigo 50.º, que por si só constitui a globalidade do Capítulo IV do Regulamento), que sujeitam tanto prestadores quanto responsáveis pela implantação, ao passo que os dois primeiros números do artigo dizem respeito aos prestadores, enquanto os números 3 e 4 visam os responsáveis pela implantação. Desde logo, importa deixar claro que estas obrigações se aplicam independentemente da classificação de um sistema como sendo de risco elevado ou não, sem prejuízo dos respetivos requisitos e obrigações aplicáveis. O artigo 50.º, n. º6 não deixa margem para dúvidas quando estabelece que o disposto nos seus números 1 a 4 não afetam os requisitos estabelecidos no capítulo dedicado aos sistemas de IA de risco elevado, salvaguardando ainda obrigações de transparência avulsas, que podem estar previstas no direito da União ou até mesmo no direito nacional. Ainda assim, parece admissível reconhecer o artigo 50.º como precursor de uma terceira categoria de risco que, não afetando as demais exigências previstas noutros regimes de risco, está dotada de alguma autonomia, tendo exceções e obrigações próprias. Em suma, o presente regime aplica-se independentemente da classificação de risco do sistema de IA, *mutatis mutandis*. Por exemplo, se um sistema estiver excluído do âmbito de aplicação do Regulamento por força do Artigo 2.º, nº6 (relativo aos sistemas ou modelos de IA desenvolvidos e colocados em serviço exclusivamente para fins de investigação e desenvolvimento científicos), então as obrigações não se aplicarão. Desta forma, não será inteiramente correto indicar que as obrigações de transparência se aplicam a todos os sistemas de IA, na medida em que existem sempre pressupostos e exceções a ter em conta, naturalmente. Por outro lado, a observação do artigo 50.º, atentando ao considerando 137, não pode ser interpretada como indicadora de licitude na utilização de um sistema de IA. Como é evidente, se um sistema de IA está dentro de aplicação do artigo 5.º, sendo proibido, o cumprimento das obrigações de transparência não constituirão uma presunção ou efeito de licitude. Já no caso dos sistemas de risco elevado, como foi mencionado, estas obrigações terão de ser observadas de forma cumulativa. I.e., deverão ser cumpridos tanto os requisitos subjacentes à classificação de sistema de IA de risco elevado e os requisitos de transparência do presente artigo. Até mesmo dentro do regime de risco elevado podemos ver algumas cláusulas de salvaguarda do artigo 50.º: veja-se o artigo 26.º, nº.11, que indica que os responsáveis pela implantação de sistemas

de IA de risco elevado devem informar as pessoas singulares de que estão sujeitas à utilização do sistema de IA de risco elevado. Disto resulta, em termos práticos, que nas situações em que ambos os artigos se aplicam, haverá uma dupla obrigação a ser cumprida. O efeito mais relevante, todavia, parece ser o da garantia do dever de informação quando o artigo 50.º não se aplica em virtude de uma exceção (na medida em que o artigo 26.º continua a ter de ser cumprido).[68]

Mas quais são as exceções previstas no âmbito do artigo 50.º, e antes de mais, que sistemas de IA estão em causa? O n. º1 debruça-se sobre "sistemas de IA destinados a interagir diretamente com pessoas singulares", abrangendo principalmente sistemas conversacionais, e.g. *chatbots*, assistentes virtuais inteligentes (ou assistentes de voz/*voice assistants*) e outros sistemas do mesmo tipo. Nos casos aplicáveis, os sistemas deverão ser concebidos e desenvolvidos de forma que as pessoas singulares que interajam com os mesmos sejam informadas de que estão a interagir com um sistema de IA, sendo esta a exigência fundamental. A interação direta com pessoas singulares excluirá, em princípio, sistemas que interagem com pessoas de forma indireta, como um sistema de IA que recomenda conteúdos online (ou sugere um catálogo de filmes com base no histórico de visualização, por exemplo). Neste caso, a interação parece ser indireta, mas este exemplo é um caso relevante, na medida em que é discutível.[69] As exceções identificadas pela própria norma são duas: primeiro, o requisito não se aplica aos casos em que a natureza artificial da interação seja óbvia; segundo, não se aplica aos "sistemas de IA legalmente autorizados para detetar, prevenir, investigar ou reprimir ações penais, sob reserva de garantias adequadas dos direitos e liberdades de terceiros, salvo se esses sistemas estiverem disponíveis ao público para denunciar uma infração penal". Importa aqui destacar, particularmente em relação à primeira exceção, de que o critério utilizado deve ser o de uma pessoa "razoavelmente informada, atenta e advertida, tendo em conta as circunstâncias e o contexto de utilização". Na prática, os prestadores deverão, idealmente, antever o público-alvo do seu sistema e ter em conta, especialmente, as circunstância e contexto de utilização. Isto na medida em que, como é evidente, diversos sistemas de IA poderão ter circunstâncias e

[68] Para uma análise exaustiva do regime do artigo 50.º, pela qual, em parte, este capítulo se guia, consultar: GILS, Thomas, *A Detailed Analysis of Article 50 of the EU's Artificial Intelligence Act* [em linha]. In: PEHLIVAN, C. N.; FORGÓ, Nikolaus; VALCKE, Peggy (eds.) – *The EU Artificial Intelligence (AI) Act: A Commentary*. Alphen aan den Rijn: *Kluwer Law International* (forthcoming). Disponível em: https://doi.org/10.2139/ssrn.4865427.

[69] O facto de um sistema de recomendação algorítmico, que seja identificado como sistema de IA, interagir com pessoas naturais é inegável, a questão é saber se esta interação é direta ou indireta. A posição adotada admite que existe um certo grau de separação e um impacto limitado deste tipo de interação, pelo que será mediata, como uma espécie de curadoria digital.

contextos de utilização amplamente distintos. Um sistema de IA utilizado num contexto específico, como numa empresa de informática, terá um contexto de utilização muito diferente de um sistema de IA de acesso geral (que seja acessível, por exemplo, à generalidade do público, incluindo crianças e idosos).[70] Com efeito, um sistema de IA utilizado numa empresa de informática especializada em análise de dados para uma função processual restrita, constituirá um contexto que justifica a isenção da obrigação. Desde logo, pelo contexto e circunstância própria da utilização do sistema, sendo os utilizadores presumivelmente proficientes na área e dispondo de formação contínua. Nestes casos, assume-se que a interação com um sistema automatizado será obviamente reconhecida pelo utilizador. O caso oposto será, evidentemente, o de um sistema de acesso livre, que terá uma base de utilizadores indeterminada e extremamente heterogénea.

Já o nº2, que recai ainda sobre os prestadores de sistemas de IA, aborda os sistemas de IA generativa, ou seja, "que geram conteúdos sintéticos de áudio, imagem, vídeo ou texto". Neste caso, a principal obrigação será de marcar o conteúdo de forma legível por máquina e detetável facilmente como tendo sido artificialmente gerado ou manipulado. Isto significa que o próprio prestador deve assegurar, do ponto de vista técnico, soluções eficazes, interoperáveis, sólidas e fiáveis na medida do possível, para assegurar esta exigência de transparência. Do ponto de vista prático, os conteúdos deverão ser identificados com um "selo" ou "marca de água" que identifiquem o seu cariz sintético. As exceções previstas passam, mais uma vez, pelos casos de utilização autorizada por lei no âmbito da prevenção e investigação penais ou, mais especificamente, nos casos em que o próprio sistema desempenhe uma mera função de apoio à edição normalizada ou não altere de forma substancial os dados de entrada (não tendo um impacto significativo nos mesmos, incluindo na sua semântica). Em geral, estas exceções serão difíceis de aplicar na prática, pelo que provar o preenchimento de algumas destas exceções

[70] Estas são questões pertinentes. O autor citado refere que os prestadores devem levar a cabo um exercício em que hipoteticamente consideram o tipo de pessoa que utilizará o sistema, tendo em conta um caso específico de utilização esperada. É essa pessoa que, segundo o regulamento, se presume razoavelmente informada, atenta e advertida. A questão é que isto exige uma sistematização ou previsão do público-alvo pretendido, que pode variar de forma substancial na realidade. Na prática, o público real poderá ser muito mais abrangente do que antecipado, ou poderá simplesmente ser manifestamente mais heterogéneo, tornando o exercício consideravelmente mais espinhoso. Por outro lado, a literacia digital poderá ser muito diversa mediante a região e estatuto socioeconómico dos utilizadores. Em suma, o exercício pode ser tão complexo que, para evitar o risco de não conformidade, a melhor alternativa para os prestadores será mesmo observar a exigência de transparência. Sobre esta questão, em alternativa, consultar novamente: GILS, Thomas, *A Detailed Analysis of Article 50 of the EU's Artificial Intelligence Act* [em linha]. Em: PEHLIVAN, C. N.; FORGÓ, Nikolaus; VALCKE, Peggy (eds.) – *The EU Artificial Intelligence (AI) Act: A Commentary*. Alphen aan den Rijn: Kluwer Law International (forthcoming). Disponível em: https://doi.org/10.2139/ssrn.4865427

aparenta ser melindroso (por exemplo, provar que o conteúdo, face ao seu contexto de utilização, é obviamente artificial ou provar que o sistema não vai além das funções de apoio à edição normalizada).[71]

Por último, já abrangendo os responsáveis pela implantação, surge no n. º3 uma obrigação de informar as pessoas expostas a sistemas de reconhecimento de emoções ou de categorização biométrica do seu funcionamento. Isto é, as pessoas devem ser informadas sobre a utilização deste tipo de sistemas. Como vimos, estes sistemas podem ser proibidos ou classificados como sendo de risco elevado, pelo que haverá sobreposição de normativos, incluindo as exigências relativas a sistemas de risco elevado, além da justaposição evidente que existe em relação à proteção de dados, que o próprio n. º3 identifica. Mais uma vez, estão fora do âmbito de aplicação os sistemas legalmente autorizados para detetar, prevenir ou investigar infrações penais. De outra forma, considerando o n. º4, "os responsáveis pela implantação de um sistema de IA que gere ou manipule conteúdos de imagem, áudio ou vídeo que constituam uma falsificação profunda, devem revelar que os conteúdos foram artificialmente gerados ou manipulados". Estas falsificações profundas são conhecidas pela expressão anglófona *deepfakes*.[72] Além destes casos, também se prevê o mesmo dever de informação no caso de notícias geradas ou manipuladas por um sistema IA. Isto é, "texto publicado com o objetivo de informar o público sobre questões de interesse público". Neste caso, aplica-se novamente a exceção relacionada à utilização autorizada por lei na investigação de infrações penais. Por outro lado, opera outra exceção que se destaca, sendo que, nos casos em que os conteúdos que são gerados por IA tiverem passado por um processo de análise humana ou por um controlo editorial, havendo uma pessoa singular ou coletiva (um órgão de comunicação social, v.g.) que seja responsável editorial pela publicação dos mesmos, a exceção aplica-se.[73]

[71] Nestes casos, na maioria das vezes, parece ser mais sensato assumir as obrigações da transparência, em alternativa à tentativa de preenchimento de exceções que são, por vezes, aparentemente difíceis de aplicar.

[72] O considerando 134 esclarece alguns pontos sobre a norma, clarificando, em certa medida, o alcance da sua interpretação, afirmando que o normativo em causa não deverá ser interpretado de forma a limitar totalmente o direito à liberdade de expressão, liberdade das artes e das ciências e até mesmo da liberdade de fruição artística e cultural. Aliás, identifica-se especialmente a proteção dos conteúdos que fazem parte de uma obra de cariz manifestamente criativo, artístico, satírico e ficcional (sempre com reservas de garantias adequadas de direitos de terceiros). Não se deverá, através desta obrigação de transparência, formar um entrave intransponível à fruição e exibição de obras artísticas. O efeito prático, contudo, ainda está por ser testemunhado.

[73] Isto não se traduz na obrigatoriedade da existência de um autor humano, pelo que bastará efetivamente a execução de um controlo humano para preencher a exceção. Ora, parece evidente que a maioria dos órgãos de comunicação social terão, quase presumidamente, uma derrogação garantida, logo que exista controlo editorial das peças publicadas. Ao contrário de várias exceções previamente vistas, esta parece ser uma exceção de fácil aplicabilidade. Do ponto de vista prático, não parece ser exigível que o conteúdo seja analisado por um jornalista,

Concluindo, reconhece-se uma terceira categoria de obrigações, dotada de autonomia, com aplicabilidade específica, mas que se encontrará frequentemente sobreposta com demais requisitos. Mais uma vez, o regime das exceções aplicáveis poderão limitar o alcance real da regulação, apesar de ser manifestamente difícil avaliar o potencial impacto das mesmas. Na ótica adotada, será benéfico, tanto a prestadores como responsáveis pela implantação, observar as exigências de transparência previstas, na medida em que parecem ser proporcionais ao risco subjacente.

2.6 Sistemas de risco mínimo

As restantes aplicações de IA caem na categoria de risco mínimo (ou nenhum risco). O Regulamento deixa claro que não impõe novas regras específicas para esses casos. Isso inclui a vasta maioria dos sistemas atuais, como filtros *anti-spam*, aplicações de jogos, sistemas de recomendação comuns, entre outros (entenda-se, quando não reconhecem emoções ou recaem sobre os mesmos outras obrigações específicas, por ex.). Tais sistemas não estão sujeitos a nenhuma obrigação adicional além do já previsto em outras normas gerais (ex.: proteção de dados). O nível de supervisão legal para risco mínimo é quase nulo, na medida em que o máximo que o texto sugere são incentivos voluntários, como a elaboração de códigos de conduta (Artigo 56.º), que poderão ajudar na adoção de boas práticas. Mas em termos de alcance jurídico, os sistemas de risco mínimo não necessitam estritamente de cumprir nenhum requisito ou registo específico, nem sofrem proibições particulares. Por outras palavras, a sua utilização é geralmente livre, ficando apenas sob o crivo de normas gerais aplicáveis no setor, já aplicadas previamente (direitos dos consumidores, proteção de dados etc.).

Desta forma, a categorização do Regulamento da IA garante que apenas os sistemas que apresentam riscos relevantes às pessoas sejam regulados de maneira onerosa (risco elevado, por exemplo) ou diretamente proibidos, enquanto os sistemas menos perigosos ficam sujeitos a exigências proporcionais (meras exigências de transparência) ou nenhuma obrigação adicional. Essa estrutura normativa constitui o núcleo legal da Secção 2, servindo de base para as regras de conformidade e fiscalização. A categoria dos modelos de finalidade geral (*GPAI*), que dispõem de um regime especial, serão abordados de forma autónoma de seguida.

pelo que bastará um processo de análise humana geral e uma pessoa singular ou coletiva que seja responsável editorial pela publicação (na prática, um órgão de comunicação social, por exemplo).

2.7 Os modelos de IA de finalidade geral

O Capítulo V do regulamento dedica-se inteiramente aos modelos de IA de finalidade geral, estabelecendo um regime próprio, separado dos graus de risco que foram abordados acima. Efetivamente, este regime estabelece dois graus de obrigações distintas conforme o risco. Isto é, estabelece um leque de obrigações para os modelos classificados como de finalidade geral (tipificado no artigo 53.º) e outro leque de obrigações, mais onerosas, previstas para os modelos classificados como de finalidade geral com risco sistémico. Neste último, as obrigações são previstas no artigo 55.º, ao passo que as regras de classificação de modelos de finalidade geral como modelos de IA de finalidade geral com risco sistémico estão previstas no artigo 51.º.

Colocam-se, todavia, variadas questões em relação a este regime, pelo que é importante estabelecer, de modo geral, uma estrutura interpretativa em relação ao Capítulo V. À semelhança das linhas orientadoras publicadas pela Comissão em relação às práticas de IA proibidas, serão publicadas também linhas orientadoras[74] no sentido de clarificar o escopo das obrigações previstas para prestadores de modelos de IA de finalidade geral. À data desta investigação, contudo, estas linhas orientadoras, esperadas para maio ou junho de 2025, ainda não foram publicadas. Todavia, a abril de 2025 foi publicado um *working document*[75] pelo *AI Office* que, apesar de não ser um documento final como o citado para as práticas proibidas, deverá contar como um documento importante, ainda que precoce, para a análise e compreensão do regime previsto no regulamento para os modelos de IA de finalidade geral.

Na ausência da versão final da Comissão, procuraremos então compreender e clarificar os conceitos-chave relativos aos modelos de IA de finalidade geral no regulamento, particularmente as definições de modelo de IA de finalidade geral, o entendimento de prestadores de modelos de IA de finalidade geral, bem como a identificação e presunção do risco sistémico. Além disso, relevará a densificação das obrigações previstas. Desde já, caberá também mencionar o Código de práticas previsto pelo artigo 56.º. Em concreto, será da

[74] Não confundir com o Código de Conduta para a Inteligência Artificial de Finalidade Geral publicado a 10 de julho de 2025. As orientações da Comissão complementarão o Código de Conduta, esclarecendo conceitos fundamentais relacionados com os modelos de IA de finalidade geral. Cf. COMISSÃO EUROPEIA, *Contents Code for Generative and General-Purpose AI Models* [em linha], Digital Strategy, 2024, disponível em: https://digital-strategy.ec.europa.eu/pt/policies/contents-code-gpai

[75] À semelhança da análise feita em relação às práticas de IA proibidas, o que será doravante exposto basear-se-á primordialmente no documento aqui citado: COMISSÃO EUROPEIA, *Targeted consultation in preparation of the Commission guidelines to clarify the scope of the obligations of providers of general-purpose AI models in the AI Act* [em linha], Bruxelas: Comissão Europeia, 22 de abril de 2025, disponível em: https://ec.europa.eu/newsroom/dae/redirection/document/114768

competência do Código detalhar, 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 em conformidade com o regulamento, por exemplo. Apesar de não serem vinculativas, os *providers* de modelos poderão escolher aderir ao Código para demonstrar compromisso com a conformidade regulatória antes da adoção de *standards*/padrões, esperados em 2027. Isto inclui tanto as obrigações do artigo 53.º como as do artigo 55.º, de modelos de IA de finalidade geral e de modelos de IA de finalidade geral com risco sistémico, respetivamente. O cumprimento deste Código é particularmente relevante para os sujeitos da norma, na medida em que o seu cumprimento acarreta uma presunção de conformidade. Por sua vez, em semelhança às linhas orientadoras relativas às práticas de IA proibidas, o exposto no *working document* não é vinculativo e dependerá, em larga medida, da interpretação jurisprudencial do TJUE suscitadas em sede de reenvio prejudicial, por exemplo.

Em termos concretos, o que é que se entende como modelo de IA de finalidade geral no âmbito do regulamento? O conceito em causa está previsto no Artigo 3.º, ponto 63 do Regulamento da IA [76]. Do ponto de vista prático, o âmago da questão parece depender de saber se o modelo apresenta uma generalidade significativa e é capaz de executar de forma competente uma vasta gama de tarefas distintas ou não. Modo geral, um modelo de IA de finalidade geral terá de, na sua essência, apresentar esta capacidade. Entende-se que, caso o modelo não se afigure capaz de executar de forma competente uma gama de tarefas distintas, não poderá ser considerado de finalidade geral por definição. O considerando 98 menciona que modelos com, pelo menos, "mil milhões de parâmetros e treinados com uma grande quantidade de dados utilizando a autossupervisão em escala para apresentar uma generalidade significativa e executar com competência uma vasta de gama de tarefas distintas" deverão ser considerados de finalidade geral. Esta formulação, contudo, deixa algumas dúvidas, nomeadamente a de saber o que é considerado uma "grande quantidade de dados", tal como a definição do artigo 3.º. Com efeito, reconhece-se a dificuldade de definir e regular os modelos de finalidade geral, seja pela sua opacidade e complexidade ou pelo colossal conjunto de dados que representam. Curiosamente, o *AI Office* adota uma aproximação preliminar que consiste na presunção de que

[76] "«Modelo de IA de finalidade geral», um modelo de IA, inclusive se for treinado com uma grande quantidade de dados utilizando a autossupervisão em escala, que apresenta uma generalidade significativa e é capaz de executar de forma competente uma vasta gama de tarefas distintas, independentemente da forma como o modelo é colocado no mercado, e que pode ser integrado numa variedade de sistemas ou aplicações a jusante, exceto os modelos de IA que são utilizados para atividades de investigação, desenvolvimento ou criação de protótipos antes de serem colocados no mercado;"

um modelo que consegue gerar texto e/ou imagens é um modelo de IA de finalidade geral, se o poder computacional utilizado no seu treino (training compute, a nível de hardware), seja maior do que 10^{22} FLOP.[77] Este entendimento é preliminar, sendo que se entende que esta métrica não será a mais apropriada para determinar a generalidade e capacidade de um modelo[78], contudo, compreende-se que, para efeitos de segurança jurídica, poderá não haver melhor alternativa atualmente. É mais uma evidência da dificuldade de regular a IA e, em particular, os *GPAI*, que são um desenvolvimento recente.

Ora, como já foi referido, os artigos 53.º e 54.º estabelecem as obrigações dos modelos de IA de finalidade geral, ao passo que o artigo 55.º impõe deveres adicionais aos modelos de finalidade geral com risco sistémico. Conhecendo a definição de modelos de IA de finalidade geral, definidos no artigo 3.º, importará agora saber o que é considerado um risco sistémico, definido no ponto 65 do mesmo artigo como: "um risco específico das capacidades de elevado impacto dos modelos de IA de finalidade geral que têm um impacto significativo no mercado da União devido ao seu alcance ou devido a efeitos negativos reais ou razoavelmente previsíveis na saúde pública, na segurança, na segurança pública, nos direitos fundamentais ou na sociedade no seu conjunto, que se pode propagar em escala ao longo da cadeia de valor;". Por sua vez, o exercício de classificar um modelo de IA de finalidade geral com risco sistémico passará pelos preceitos do artigo 51.º, que deixa claro que existe risco sistémico se o modelo tiver, segundo o n.º1., alínea a), "capacidades de elevado impacto avaliadas com base em ferramentas e metodologias técnicas adequadas, incluindo indicadores e parâmetros de referência", ou, segundo a alínea b), tiver "capacidades ou um impacto equivalentes às estabelecidas na alínea a), tendo em conta os critérios estabelecidos no anexo XIII, com base numa decisão da Comissão, *ex officio* ou na sequência de um alerta qualificado do painel científico.". As "capacidades de elevado impacto" são também definidas no artigo 3.º, ponto 64, como "capacidades ou um impacto equivalentes às estabelecidas na alínea a), tendo em conta os critérios estabelecidos no anexo XIII, com base numa decisão da Comissão, *ex officio* ou na sequência de um alerta qualificado do painel científico.". De forma simples, há uma remissão evidente desta matéria para uma concretização posterior através de atos delegados,

[77] Veja-se a página 4 do documento: COMISSÃO EUROPEIA, *Targeted consultation in preparation of the Commission guidelines to clarify the scope of the obligations of providers of general-purpose AI models in the AI Act* [em linha], Bruxelas: Comissão Europeia, 22 de abril de 2025, disponível em: https://ec.europa.eu/newsroom/dae/redirection/document/114768

[78] Segundo WACHTER, Sandra, *Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond* [em linha]. *Yale Journal of Law & Technology*, vol. 26, n. º 3, 2024. p. 698. Disponível em: https://doi.org/10.2139/ssrn.4924553

por parte da Comissão e para critérios puramente técnico-científicos (artigo 51.º/3 estabelece que a Comissão adota atos delegados nos termos do artigo 97.º, podendo alterar os limiares do artigo 51.º e até mesmo complementar os parâmetros de referência de acordo com os desenvolvimentos futuros desta tecnologia, procurando refletir, a todo o momento, o estado da arte, flexibilizando a classificação).

Já o número 2 do artigo 51.º, doutro modo, define um limiar técnico concreto que, ao ser ultrapassado, aciona a presunção de que o modelo detém capacidades de elevado impacto. Note-se que esta presunção é ilidível. O limiar em causa é o da quantidade acumulada de cálculo utilizado no treino do modelo que, medido em operações de vírgula flutuante por segundo, seja superior a 10^{25}.[79] Este nível mínimo de FLOPS é criticado por aparentemente apenas cobrir alguns modelos atuais, como o *GPT-4* da *OpenAI*, o *Gemini* da *Google* e o *Llama 3.1* da *Meta*. Isto significa que uma fatia significativa de modelos populares e acessíveis à data de hoje estariam excluídos da presunção, tal como o modelo *GPT-3.5*, que é disponibilizado de forma gratuita aos utilizadores do *ChatGPT*. Entende-se, por outro lado, que no futuro, com os avanços tecnológicos, este limiar sofra do problema inverso, ao ser demasiado baixo para as realidades computacionais vindouras. De qualquer das formas, as críticas apontadas parecem pertinentes.[80]

Curiosamente, em relação à classificação de um modelo de IA de finalidade geral como tendo risco sistémico, o artigo 52.º, n.º2 prevê que o prestador pode, no prazo de duas semanas, "apresentar, com a sua notificação, argumentos suficientemente fundamentados para demonstrar que, excecionalmente, embora preencha esse requisito, o modelo de IA de finalidade geral não apresenta, devido às suas características específicas, riscos sistémicos e, por conseguinte, não deverá ser classificado como um modelo de IA de finalidade geral com

[79] O artigo 3.º define no ponto 67 as operações de vírgula flutuante como "qualquer operação matemática ou atribuição que envolva números em vírgula flutuante, que são um subconjunto dos números reais normalmente representados em computadores por um número inteiro de precisão fixa escalado por um expoente inteiro de uma base fixa;". É uma definição estritamente técnica, que se traduz, simplificadamente, num valor utilizado para medir o desempenho ou poder computacional do *hardware* utilizado para treinar um certo modelo de IA. Note-se que o artigo 51.º refere operações de vírgula flutuante por segundo.

[80] Particularmente relevante para este parágrafo é o entendimento e olhar crítico de WACHTER, Sandra, *Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond* [em linha]. *Yale Journal of Law & Technology*, vol. 26, n. º 3, 2024. pp. 715 e ss. Disponível em: https://doi.org/10.2139/ssrn.4924553. A autora sugere a redução do limiar computacional para incluir modelos como o *GPT-3.5*, que têm um impacto semelhante aos modelos maiores. Além disso, defende a inclusão de um critério de número de utilizadores finais, sendo que isso alinharia o entendimento do Regulamento da IA com o Regulamento dos Serviços Digitais e a noção de plataformas digitais de muito grande dimensão, por exemplo. Entende-se, todavia, que a flexibilidade do Artigo 51º e o recurso aos atos delegados permitirá à Comissão alterar estas figuras.

risco sistémico.", estabelecendo um procedimento de ilidibilidade em certos casos.[81] Um ponto relevante a destacar, por outro lado, é o do artigo 54.º, que expressa a necessidade dos prestadores estabelecidos em países terceiros, mediante mandato escrito, designarem um mandatário estabelecido na União. Esse mandatário terá, por sua vez, de observar os deveres impostos por este regime.

Na essência, contudo, quais são os deveres previstos para os modelos de IA de finalidade geral, i.e., para os prestadores destes sistemas, incluindo os de risco sistémico? Identificam-se, essencialmente, os seguintes deveres no artigo 53.º:

a) manter documentação técnica adequada e atualizada [n.º 1/b) e Anexo XI];
b) facilitar integração e interoperabilidade com o seu sistema [n.º 1/b), Anexo XII];
c) aplicar uma política de respeito pelo direito de autor [n.º 1/c)], em especial garantindo que o sistema respeita a reserva de direitos prevista no art. 4.º da Diretiva 2019/790 no contexto de *text and data mining*; e
d) disponibilizar ao público um resumo sobre os conteúdos utilizados para o treino do modelo [n. º1/d), que prevê a elaboração de um modelo desse resumo pelo AI Office]. Além disso, está previsto um dever geral de colaboração com as autoridades (art. 53.º/3).

No que concerne aos modelos com risco sistémico, aplicam-se os deveres supra mais os previstos no n. º1 do artigo 55.º, a saber:

a) realizar testes e avaliações do modelo com vista a identificar e atenuar os riscos sistémicos;
b) avaliar e atenuar eventuais riscos;
c) acompanhar, documentar e comunicar informações pertinentes sobre incidentes graves e eventuais medidas corretivas para os resolver; e
d) assegurar um nível adequado de proteção em termos de cibersegurança.

[81] O considerando 110 do Regulamento identifica vários exemplos de riscos sistémicos, que serão, em princípio, dificilmente ilidíveis, tais como: "quaisquer efeitos negativos reais ou razoavelmente previsíveis em relação a acidentes graves, perturbações de setores críticos e consequências graves para a saúde e a segurança públicas; quaisquer efeitos negativos, reais ou razoavelmente previsíveis, em processos democráticos e na segurança pública e económica; a divulgação de conteúdos ilegais, falsos ou discriminatórios.".

À semelhança do que acontece em relação aos sistemas de risco elevado, o cumprimento das normas harmonizadas/*standards* ou Código de Práticas implicará uma presunção de conformidade de acordo com o artigo 40.º, n. º1.

De modo geral, há críticas que podem ser apontadas ao regime dos modelos de IA de finalidade geral, algumas já mencionadas, tal como a crítica ao limiar computacional em FLOPS utilizado. A questão central, que merece ser destacada, parece ser mesmo essa: será que o modelo díptico se justifica quando alguns riscos sistémicos, não identificados, poderão estar presentes em modelos que estão abaixo do limiar, escapando à presunção? Existe o risco, mais uma vez, de modelos que estão abaixo do limiar não serem identificados como sendo de risco sistémico, ainda que, na prática, deles possam resultar efeitos negativos reais que não foram antecipados. Desde logo, o modelo pode apresentar problemas de explicabilidade, que por si só dificulta a sua avaliação, levando a que na prática, no mercado, revele tendência a alucinar resultados, providenciar desinformação, não respeitar a proteção de dados etc. Além disso, é possível que a existência do limiar motive, em alguns casos, os *providers* a reduzir ou otimizar os modelos para que os mesmos permaneçam abaixo do limiar, ainda que o seu risco seja invariável (e seja significativo independentemente da sua posição em relação ao limite).[82]

3. Conclusão

Ao longo a presente dissertação, procedeu-se à análise dos regimes previstos no Regulamento da Inteligência Artificial da União Europeia. A abordagem adotada consistiu num percurso abrangente pelas diversas categorias de risco previstas, desde as práticas de IA proibidas, consideradas de risco inaceitável, até aos modelos de finalidade geral. A análise mais extensa foi dedicada às práticas de risco inaceitável, com base nas orientações publicadas pela Comissão Europeia, que foram concebidas justamente para garantir a aplicação coerente, eficaz e uniforme deste regime. Estas diretrizes abordam especificamente práticas como o *social scoring*, a identificação biométrica à distância em tempo real e as avaliações de risco de pessoas

[82] Cf. WACHTER, Sandra – Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond [em linha]. Yale Journal of Law & Technology, vol. 26, n. º 3, 2024. pp. 694 e ss. Disponível em: https://doi.org/10.2139/ssrn.4924553 – a autora ainda refere que a utilização de FLOPS como métrica de risco ou perigo é incomportável, oferecendo algumas alternativas referidas supra. Por outro lado, reconhece-se que a utilização de operações de vírgula flutuante por segundo como métrica poderá ser útil na avaliação de impacto ambiental do treino (e operação) de modelos de IA de finalidade geral. Por outro lado, traz-se à atenção o facto de no futuro, em virtude de otimizações dos modelos, seja possível que modelos que estão abaixo do limiar possam ter um impacto ou capacidade igual ou superior a modelos que são atualmente de risco sistémico. O acompanhamento do desenvolvimento técnico-científico destes modelos representará um desafio interessante para a Comissão Europeia e o AI Office.

singulares a fim de avaliar ou prever o risco de uma pessoa singular cometer uma infração penal. Desta forma, através das linhas orientadoras da Comissão, foi possível rever alguns exemplos práticos e expor interpretações que poderão, na medida do possível, ser relevantes às partes interessadas que queiram compreender os requisitos aplicáveis. Por outro lado, levantam-se algumas questões e pontos de contenção. O meso se aplica à análise dos sistemas de IA de risco elevado, que apresentam um leque de obrigações e exceções por vezes meandroso. As obrigações aplicáveis a estes sistemas, desde a qualidade dos dados, supervisão humana, transparência, registo e gestão de risco, são extensas e precisas (aplicando-se ora a prestadores, ora a responsáveis pela implantação). Além disso, refletiu-se sobre as obrigações de transparência previstas no artigo 50.º, apontando a sua relevância e suscitando algumas questões. Por fim, procedeu-se à breve análise do regime específico para modelos de IA de finalidade geral com e sem risco sistémico.

Não obstante a riqueza normativa exposta, permanece a tensão entre a precisão técnica e a aplicabilidade prática, por vezes onerosa, dos preceitos do regulamento. A ambição de criar um ecossistema de confiança em torno da IA é louvável e a apreciação global do esforço é positiva. Restará no futuro, através de uma abordagem empírica, tirar conclusões sobre a eficácia real das obrigações previstas. A articulação eficaz e reforçada entre atores públicos e privados será impreterível para assegurar a aplicação eficiente dos regimes aplicáveis. Mais do que isso, reconhece-se a importância do trabalho contínuo que terá de ser levado a cabo pela Comissão, na adoção de atos delegados, quando aplicáveis, e publicação de demais orientações práticas de suporte. Entende-se, de forma clara, que o Regulamento da IA dependerá de um processo contínuo, apenas possível através da flexibilização presente no seu texto, de atualização e acompanhamento das evoluções técnico-científicas do futuro. Não só, a atuação das autoridades competentes designadas por cada Estado-Membro também será fulcral para a execução adequada das normas regulatórias previstas.

Ultimamente, contudo, verifica-se uma tendência positiva na regulação da IA, mais relevante que nunca, através do diálogo entre todas as partes interessadas, com uma preocupação vincada de alcançar soluções inovadoras que atinjam um compromisso adequado entre inovação e direitos. Mais do que isso, é importante considerar a evolução permanente da tecnologia, pelo que o Regulamento aparentar estar capacitado, através dos mecanismos de flexibilização, de ferramentas eficazes para dar resposta aos problemas emergentes, garantido um nível de resiliência face aos desafios de amanhã.

Bibliografia

AGÊNCIA DOS DIREITOS FUNDAMENTAIS DA UNIÃO EUROPEIA, *Big Data: Discrimination in data-supported decision making*. Viena: FRA, 2018. Disponível em: http://fra.europa.eu/en/publication/2018/big-data-discrimination

ALMADA, MARCO, *Will the EU AI Act shape global regulation?* Network Law Review, 2025. Disponível em: https://www.networklawreview.org/almada-ai-act/

BARKANE, IRENA; BUKA, LOLITA, *Prohibited AI surveillance practices in the Artificial Intelligence Act: promises and pitfalls in protecting fundamental rights*. In: BARKANE, IRENA; BUKA, LOLITA (eds.), *Critical perspectives on predictive policing*. Cheltenham: Edward Elgar Publishing, 2024, cap. 6, p. 110–129. Disponível em: https://www.elgaronline.com/view/book/9781035323036

CARROLL, M.; et al., *Characterising manipulation from AI systems*. In: *Equity and Access in Algorithms, Mechanisms, and Optimization*. Boston: ACM Press, 2023. Disponível em: https://doi.org/10.1145/3617694.3623226

COMISSÃO EUROPEIA, *Artificial Intelligence for Europe*. COM(2018) 237 final. Bruxelas: Comissão Europeia, 2018. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52018DC0237&from=EN

COMISSÃO EUROPEIA, *Approval of the content of the draft Communication from the Commission — Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*. Comunicação C(2025) 884 final. Bruxelas: Comissão Europeia, 4 fev. 2025. Disponível em: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

COMISSÃO EUROPEIA, *Targeted consultation in preparation of the Commission guidelines to clarify the scope of the obligations of providers of general-purpose AI models in the AI Act*. Documento de trabalho. Bruxelas: Comissão Europeia, 22 abr. 2025. Disponível em: https://ec.europa.eu/newsroom/dae/redirection/document/114768

COMISSÃO EUROPEIA; GRUPO DE PERITOS DE ALTO NÍVEL SOBRE INTELIGÊNCIA ARTIFICIAL, *Orientações éticas para uma IA de confiança*. Luxemburgo: Serviço das Publicações da União Europeia, 2019. Disponível em: https://data.europa.eu/doi/10.2759/2686

COUNCIL OF THE EUROPEAN UNION, *Presidency compromise text*. 14278/21. Bruxelas: Council of the European Union, set. 2021. Disponível em: https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf

DE GREGORIO, GIOVANNI; DUNN, PIETRO, *The European risk-based approaches: connecting constitutional dots in the digital age*. Common Market Law Review, vol. 59, n.º 2, 2022, p. 473–500. Disponível em: https://ssrn.com/abstract=4071437

EUROPOL, *The Second Quantum Revolution — The impact of quantum computing and quantum technologies on law enforcement*. Europol Innovation Lab observatory report. Luxemburgo: Publications Office of the European Union, 2023. Disponível em: https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf

FINK, MELANIE, *Human Oversight under Article 14 of the EU AI Act*. In: MALGIERI, GIANCLAUDIO; GONZALEZ FUSTER, GLORIA; MANTELERO, ALESSANDRO; ZANFIR-FORTUNA, GABRIELA (eds.), *AI Act Commentary: A Thematic Analysis*. Oxford: Hart-Bloomsbury, 2026 (forthcoming). Disponível em: https://ssrn.com/abstract=5147196

GIANNINI, ALICE; TAS, SARAH, *AI Act and the prohibition of real-time biometric identification: much ado about nothing?* VerfBlog, 10 dez. 2024. Disponível em: https://verfassungsblog.de/ai-act-and-the-prohibition-of-real-time-biometric-identification/ . DOI: 10.59704/a15aa6e58151c853

GILS, THOMAS, *A detailed analysis of Article 50 of the EU's Artificial Intelligence Act*. In: PEHLIVAN, C. N.; FORGÓ, NIKOLAUS; VALCKE, PEGGY (eds.), *The EU Artificial Intelligence (AI) Act: A Commentary*. Alphen aan den Rijn: Kluwer Law International, forthcoming. Disponível em: https://doi.org/10.2139/ssrn.4865427

GRUPO DE TRABALHO DO ARTIGO 29.º PARA A PROTEÇÃO DE DADOS, *Opinion 02/2012 on facial recognition in online and mobile services*. Bruxelas: Comissão Europeia, 2012. (WP 193). Disponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf

HUGGING FACE, *emotion-english-distilroberta-base*. Disponível em: https://huggingface.co/j-hartmann/emotion-english-distilroberta-base

LILKOV, DIMITAR, *Regulating artificial intelligence in the EU: a risky game*. European View, vol. 20, n.º 2, 2021, p. 166–174. Disponível em: https://doi.org/10.1177/17816858211059248

NIKLAS, JĘDRZEJ; SZTANDAR-SZTANDERSKA, KAROLINA; SZYMIELEWICZ, KATARZYNA; BACZKO-DOMBI, ANITA; WALKOWIAK, ANTONIO, *Profiling the unemployed in Poland: social and political implications of algorithmic decision making*. Fundacja Panoptykon, 2015. Disponível em: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf

OCDE, *Artificial Intelligence in Society*. Paris: OECD Publishing, 2019. Disponível em: https://doi.org/10.1787/eedfee77-en

PARLAMENTO EUROPEU; CONSELHO DA UNIÃO EUROPEIA, *Diretiva (UE) 2019/2161, de 27 de novembro de 2019, que altera as Diretivas 93/13/CEE, 98/6/CE, 2005/29/CE e 2011/83/UE no que respeita à melhor aplicação e modernização das regras da União em matéria de defesa do consumidor*. Jornal Oficial da União Europeia, L 328, 18 dez. 2019

REGULAMENTO (UE) 2016/679, *Regulamento Geral sobre a Proteção de Dados (RGPD)*. Jornal Oficial da União Europeia, L 119, 4 maio 2016

REGULAMENTO (UE) 2024/1624, *Relativo à prevenção da utilização do sistema financeiro para efeitos de branqueamento de capitais ou financiamento do terrorismo*. Jornal Oficial da União Europeia, L 135, 7 jun. 2024

REGULAMENTO (UE) 2024/1689, *Regulamento da Inteligência Artificial*. Jornal Oficial da União Europeia, L 168, 14 jun. 2024. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32024R1689

SCHNURR, DANIEL, *Effective implementation of requirements for high-risk AI systems under the AI Act*. Oxford: Centre for the Governance of AI, 2024. Disponível em: https://ssrn.com/abstract=5147196

SILVA, NUNO SOUSA, *O Regulamento da Inteligência Artificial: análise introdutória*. Revista da Ordem dos Advogados, 2024. Disponível em: https://portal.oa.pt/media/147835/nuno-sousa-e-silva.pdf

TUOMI, I., *The use of Artificial Intelligence (AI) in education*. Parlamento Europeu, 2020. Disponível em: https://op.europa.eu/en/publication-detail/-/publication/92273e25-f7c3-11ea-991b-01aa75ed71a1/language-en

TRIBUNAL DE JUSTIÇA DA UNIÃO EUROPEIA, *Acórdão de 4 de outubro de 2024, Schrems – Meta Platforms Ireland Ltd.* Processo C-446/21. ECLI:EU:C:2024:804

WACHTER, SANDRA, *Limitations and loopholes in the EU AI Act and AI Liability Directives: what this means for the European Union, the United States, and beyond*. Yale Journal of Law & Technology, vol. 26, n.º 3, 2024. Disponível em: https://doi.org/10.2139/ssrn.4924553

FACULDADE DE **DIREITO**