O REGULAMENTO DA INTELIGÊNCIA ARTIFICIAL: ANÁLISE INTRODUTÓRIA

Pelo Prof. Doutor Nuno Sousa e Silva(*)

Sumário:

1. Introdução

A rápida evolução tecnológica das últimas décadas — gerando um vasto acervo de informação digitalizada e acessível (possibilitada pela Internet) e os avanços em termos de *hardware* e de *software* —, permitiu que certas técnicas matemáticas (ditas de aprendizagem automática), até então incipientes, se tornassem operativas e mesmo revolucionárias. Isso está na base dos vertiginosos desenvolvimentos em matéria de Inteligência Artificial registados nos últimos anos.

(*) Advogado e Prof. Auxiliar da Universidade Católica Portuguesa (Porto). E: <nsilva@ucp.pt>. W: <www.nss.pt>. Este texto beneficiou da troca de ideias e sugestões de Rafael Dias Almeida e Pedro Sousa e Silva, que se registam e agradecem. Este artigo coincide parcialmente com um texto em inglês "The Artificial Intelligence Act: Critical Overview" aceite para publicação na JIPITEC.

Contudo, apesar das inúmeras vantagens que esta evolução proporciona(1), o tom catastrofista tem ganhado proeminência(2).

Desde a segunda década deste século XXI que a segurança em Inteligência Artificial (adiante "**IA**") se tem afirmado como ramo interdisciplinar de estudo, indo para além de considerações éticas no desenvolvimento de sistemas(3). Existem discussões relativas à transparência e possibilidade de explicação de decisões tomadas por sistemas de IA(4), ao potencial discriminatório ou gerador de injustiça na utilização destes sistemas(5), e a questões relativas ao controlo e alinhamento do sistema com valores humanos(6). Sublinha-se a necessidade premente de garantir a robustez e qualidade técnica destes sistemas(7). Assinalam-se as práticas

---

(1) Entre muitas outras, refira-se a aceleração do desenvolvimento de medicamentos [J. Jumper, *et al.*, 'Highly accurate protein structure prediction with AlphaFold' *Nature* 596 (2021), pp. 583-589] e vacinas [A Sharma, *et al.*, *Artificial Intelligence-Based Data-Driven Strategy to Accelerate Research, Development, and Clinical Trials of COVID Vaccine.* BioMed research international (2022)], do combate às alterações climáticas (J. Cowls, *et al.*, 'The AI gambit: leveraging artificial intelligence to combat climate change — opportunities, challenges, and recommendations' *in* AI & Society 38 (2023), pp. 283-307) e da criação de novos materiais [Phil De Luna (ed.), *Accelerated Materials Discovery: How to Use Artificial Intelligence to Speed Up Development* (De Gruyter 2022)].

(2) Entre as obras mais influentes nesta linha contam-se Nick Bostrom, S*uperintelligence: Paths, Dangers, Strategies* (OUP 2014) e, anterior, Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Viking 2005). Com uma visão mais equilibrada, cf. Henry A Kissinger/Eric Schmidt/Daniel Huttenlocher, *The Age of AI: And Our Human Future* (Little, Brown and Company 2021). Um dos principais arautos do apocalipse é Eliezer Yudkowsky.

(3) R.V. Yampolskiy, 'Artificial intelligence safety engineering: Why machine ethics is a wrong approach' *in* AAVV, *Philosophy and Theory of Artificial Intelligence* (Springer 2013) pp. 389-396.

(4) Trata-se daquilo que é designado por XAI (*explainable AI*). Sobre o tema numa perspetiva mais ampla, cf. Frank Pasquale, T*he Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2016). Discutindo a existência de um direito à explicação ao abrigo do art. 22.º RGPD, cf. o debate entre Sandra Wachter/Brent Mittelstadt/Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' International Data Privacy Law, Vol. 7(2) (2017), pp. 76-99 e Gianclaudio Malgieri/Giovanni Comande, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' International Data Privacy Law, 2017, Vol. 7(4), pp. 243-265. O consenso parece ir no sentido de não existir um direito à explicação detalhada da decisão, mas apenas do enunciado dos seus critérios e parâmetros básicos [AAVV, *General Data Protection: Article-by-article commentary* (Hart C. H. Beck 2023), p. 541]. Sobre o debate, cf. Diogo Morgado Rebelo, *Inteligência Artificial e Scoring no Crédito ao Consumo* (Almedina 2023), p. 326, ss.

(5) Entre muitos, Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Brown 2016); Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (NYU Press 2018); Meredith Broussard, *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech* (MIT Press 2023).

(6) Vejam-se as obras de Brian Christian, *The Alignment Problem* (Atlantic Books 2020) e Stuart Russel, *Human Compatible: AI and the Problem of Control* (Penguin 2019).

(7) Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Penguin 2017) propõe uma análise assente em quatro vetores: verificação, validação, controlo e segurança.

extrativas, tanto de dados (alguns protegidos por direitos de propriedade intelectual), como de minerais e alerta-se para o consumo de energia associado à IA([8]).

Nos últimos anos os juristas e o poder político começaram a olhar para o tema, procurando adotar regras adequadas para fazer face aos múltiplos desafios por ele colocados. De um ponto de vista jurídico, os temas são muitos e têm impacto designadamente em matéria de Direitos Fundamentais (liberdade, trabalho e emprego, privacidade, igualdade e não discriminação, participação democrática, acesso à justiça, liberdade de expressão e informação, organização política, proteção do meio ambiente), de responsabilidade civil e penal, de proteção de dados pessoais, da privacidade e dos direitos de personalidade, da Propriedade Intelectual, do Direito da Concorrência, Direito do Ambiente, Direito Penal, Direito Fiscal e do Direito Administrativo([9]).

([8]) Kate Crawford *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021).

([9]) As monografias sobre Direito e Inteligência Artificial têm-se multiplicado. Numa primeira fase o estudo (e a abordagem do Parlamento Europeu) focou-se na robótica cabendo destacar entre as obras mais gerais Ugo Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts* (Springer 2013); Alain Bensoussan/Jérémy Bensoussan, *Droit des Robots* (Larcier 2015) e Ryan Calo/Michael Froomkin/Ian Kerr (eds.), *Robot Law* (EE 2016). Aliás, a tendência de focar a análise na robótica estendia-se para além do Direito como evidencia a obra geral de Patrick Lin/Keith Abney/George A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press 2011). Nestas obras abordavam-se essencialmente temas de personalidade, crime, contratos e ilícitos civis (responsabilidade). Estes assuntos estavam também ligados aos aspetos relacionados com a condução autónoma, que tem vindo a ser objeto de monografias, inclusivé em Portugal [Sofia Alcaide, *A Responsabilidade Civil por Danos Causados por Veículos Autónomos* (Almedina 2021)]. Outros como Moisés Barrio Andrés (eds.), *Derecho de los Robots* (Wolters Kluwer 2018) iam já mais longe tratando também de temas de Direito do Trabalho, Direito Financeiro e Tributário, Direito da Saúde e do seu impacto nas profissões jurídicas. Ainda sobre a perspetiva de Direito e Robótica, mas centrando já a análise na Inteligência Artificial pode ver-se Jacob Turner, *Robot Rules* (Palgrave 2019) e Ryan Abbott, *The Reasonable Robot* (Cambridge University Press 2020). Obedecendo à tendência mais geral, também a nível regulatório os juristas passaram a preferir a análise centrada na Inteligência Artificial. Vai já nesse sentido o pequeno opúsculo de Henrique Sousa Antunes, *Direito e Inteligência Artificial* publicado pela Universidade Católica Editora em 2020. Levantamentos monográficos de cariz geral incluem Matt Hervey/Matthew Lavy (eds.), *The Law of Artificial Intelligence* (Sweet & Maxwell 2020); Woodrow Barfield/Ugo Pagallo, *Advanced Introduction to Law and Artificial Intelligence* (Edward Elgar 2020); Woodrow Barfield/Ugo Pagallo (eds.), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar 2020); Jan De Bruyne/Cedric Vanleenhove (eds.), *Artificial Intelligence and the Law* (Intersentia 2021); Hoeren/Pinelli, *Künstliche Intelligenz — Ethik und Recht* (C. H. Beck 2022); Mafalda Miranda Barbosa, *Inteligência Artificial* (Gestlegal 2021) e Charles Kerrigan, *Artificial Intelligence: Law and Regulation* (Edward Elgar 2022). Pela sua abrangência merece ainda destaque a obra Ebers/Heinze/Krügel/Steinrötter, *Künstliche Intelligenz und Robotik* (C.H. Beck 2020) que, em mais de mil páginas, faz também análises sectoriais. Esta análise tem vindo a ser feita essencialmente de uma perspetiva dogmática, mas existem igualmente estudos empíricos, teorias críticas e análise económica do

Embora as iniciativas regulatórias tenham lugar um pouco por todo o mundo, deve ser dado especial destaque ao papel da União Europeia(10). Efetivamente, o Parlamento Europeu começou por adotar, a 16 de Fevereiro de 2017, uma resolução com recomendações à Comissão Europeia sobre regras de Direito Civil sobre robótica(11). Nessa resolução reconhecem-se os perigos e oportunidades da robótica e da inteligência artificial e são feitas várias sugestões para a respectiva regulação instando-se a Comissão a apresentar uma proposta legislativa sobre questões jurídicas relacionadas com o desenvolvimento e a utilização da robótica e da Inteligência Artificial. Em anexo a esse documento foram ainda apresentadas recomendações relativas ao conteúdo de uma tal proposta — incluindo a definição de robot, a criação de um sistema de registo gerido por uma agência europeia, regras de responsabilidade civil, seguros e fundos de garantia e o estabelecimento de regras de interoperabilidade — e uma "Carta da Robótica", um código de conduta voluntário dirigido a investigadores e *designers* em robótica. Esta resolução de 2017 acelerou a discussão sobre os temas jurídicos ligados à inteligência artificial e robótica(12).

No ano seguinte a Comissão apresentou duas comunicações "Inteligência artificial para a Europa"(13) e "Plano coordenado para a Inteligência Artificial"(14), tendo-se sucedido as resoluções, estudos e relatórios, mere-

---

Direito [v.g. Georgios Zekos, *Economics and Law of Artificial Intelligence* (Springer 2021)]. Em Portugal entre as publicações recentes devem destacar-se Diogo Morgado Rebelo, *ob. cit.*; Anabela Miranda Rodrigues/Susana Aires de Sousa (eds.), *I Congresso Inteligência Artificial e Direito* (Almedina 2023) e Luís Manuel Pica, *A Inteligência Artificial no Direito Tributário — Fundamentos e Limites Constitucionais* (Almedina 2023).

(10) Para lá da UE, a 17 de Maio de 2024 foi aprovada Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law no âmbito do Conselho da Europa (organização internacional autónoma que transcende a União Europeia) (neste texto "**Convenção CoE**"). No mesmo mês de 2024 as guidelines da OCDE (Recommendation on Artificial Intelligence) de 2019 foram revista (C/MIN(2024)16/FINAL). Nos EUA existe legislação sectorial, iniciativas (e.g. USC 15 Chpater 19 — National Intelligence Iniative), legislação estadual e ordens executivas, mas ainda não foi aprovada nenhuma lei federal de caráter geral. Uma parte dos países, como Austrália, Japão, Israel, Singapura ou Índia, tem seguido abordagens de *soft law*, complementadas por intervenções sectoriais. Têm surgido algumas propostas de legislação p. ex. no Brasil e no Canadá. Em Julho de 2023 o Peru adotou a Ley n.º 31814 com vista à promoção do uso de IA. Para um acompanhamento dos desenvolvimentos legislativos e regulatórios nesta matéria, cf. <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker>.

(11) [2015/2103(INL)]. A resolução, aprovada com 396 votos a favor (123 contra e 85 abstenções), teve como relatora a luxemburguesa Mady Delvaux.

(12) Sobre o estado do tema nessa data veja-se o texto, escrito na sequência desta resolução, Nuno Sousa e Silva, 'Direito e Robótica: Uma primeira *aproximação*' Revista da Ordem dos Advogados [2017], pp. 485-551.

(13) COM(2018)237 final, de 25 de abril de 2018.

(14) COM(2018)795 final, de 7 de dezembro de 2018.

cendo especial destaque o "Livro Branco sobre a Inteligência Artificial" apresentado pela Comissão em Fevereiro de 2020([15]).

A 20 de outubro de 2020 o Parlamento Europeu aprovou uma resolução com recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial([16]). Esse documento continha o texto de um projeto de regulamento sobre a responsabilidade pela operação de sistemas de IA. No considerando 9 dessa proposta pode ler-se:

> *A Diretiva 85/374/CEE do Conselho («Diretiva relativa à responsabilidade decorrente dos produtos») demonstra ser há mais de 30 anos um meio eficaz para obter indemnização pelos danos causados por um produto defeituoso. Portanto, também deverá ser utilizada no que respeita a ações de responsabilidade civil de uma parte que sofra prejuízos ou danos contra o produtor de um sistema de IA defeituoso*.

De facto, o regime dessa Diretiva, transposta em Portugal pelo DL 383/89, de 6 de Novembro([17]), enfrenta algumas dificuldades de aplicação relativamente a *software* e, mais genericamente, a conteúdo digital ou bens com conteúdo digital (incluindo agentes de inteligência artificial e robots autónomos). Além disso, colocam-se algumas restrições e obstáculos práticos à obtenção de indemnizações([18]). Por isso, no dia 28 de setembro de 2022, a Comissão Europeia apresentou duas propostas: uma revisão da Diretiva relativa à responsabilidade decorrente dos produtos defeituosos, que visa substituir a Diretiva 85/374/CE([19]) e uma nova Diretiva relativa à adaptação das regras de responsabilidade civil extracontratual à inteligência artificial([20]).

No entanto, a principal abordagem regulatória deste fenómeno consiste no Regulamento da Inteligência Artificial, conhecido ***AI Act***([21]). Este regulamento nasceu de uma proposta apresentada pela Comissão Europeia

---

([15]) Com o subtítulo "— Uma abordagem europeia virada para a excelência e a confiança" (COM(2020)65 final). Sobre o tema *vide* Graça Enes, 'A estratégia europeia para a Inteligência Artificial' *in* Maria Raquel Guimarães/Rute Teixeira Pedro, *Direito e Inteligência Artificial* (Almedina 2023), pp. 37-91.

([16]) 2020/2014(INL).

([17]) A Diretiva 85/374 foi alterada pela Diretiva 1999/34/CEE de 10 de Maio, por sua vez transposta pelo DL 131/2001 de 24 de Abril. Sobre o regime a obra fundamental continua a ser J. Calvão da Silva, *Responsabilidade Civil do Produtor* (Almedina 1990).

([18]) Henrique Sousa Antunes, 'Responsabilidade Civil do Produtor: os danos ressarcíveis na era digital' Revista de Direito da Responsabilidade, Ano 1, 2019, pp. 1476-1485.

([19]) COM(2022)495 final.

([20]) COM(2022)496 final.

([21]) Neste texto referido como "**Regulamento**" e ao qual, salvo indicação ou contexto contrário, pertencerão as normas citadas sem outra indicação. A base legislativa utilizada é dupla: os arts. 16.º (relativo à proteção de dados) e 114.º (relativo ao mercado interno), ambos do TFUE.

em Abril de 2021(22). A proposta foi objeto de intensas negociações (incluindo uma maratona de 36 horas entre representantes da Comissão Europeia, Parlamento Europeu e Conselho), profundas alterações e um *corrigendum* (de 19 de Abril de 2024), tendo sido aprovada a 13 de Junho de 2024 e publicada a 12 de Julho sob o número 2024/1689(23).

Este artigo procura fazer uma apresentação sucinta dos principais aspetos deste Regulamento.

2. Estrutura, objetivos e abordagem

O Regulamento é um bom exemplo da tendência que já foi chamada de "brutalidade regulatória"(24). Como se tornará claro nas próximas páginas, a legislação é particularmente complexa, envolvendo 68 definições, 113 artigos, 13 anexos e 180 considerandos. As sanções previstas são extremamente severas (podendo chegar a 7% da receita global do infrator ou 35 milhões de euros), o âmbito de aplicação territorial é particularmente amplo e a supervisão é feita ao nível nacional e europeu, estabelecendo-se uma nova arquitetura regulatória, que inclui, a nível da União, o Serviço para a IA (*EU AI Office*), o Comité Europeu para a IA (*EU AI Board*), um fórum consultivo e um painel científico de peritos independentes (arts. 64.º e ss.) e, a nível nacional, pelo menos uma autoridade nacional notificadora e uma autoridade nacional de fiscalização de mercado (art. 70.º).

Este instrumento legislativo é composto por 13 capítulos(25):

---

(22) COM(2021)206 final. Para uma descrição dos antecedentes e principais traços da evolução das propostas até 2023 cf. Marta Boura, 'Inteligência Artificial. Quadro jurídico e reflexões sobre a Proposta de Regulamento de Inteligência Artificial' Revista Electrónica de Direito, Vol. 32(3) (2023) pp. 100-123; Nikos Th. Nikolinakos, *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies-The AI Act* (Springer 2023) e Carmen Muñoz García, *Regulación de la inteligencia artificial en Europa* (Tirant lo Blanch 2023), p. 13, ss.

(23) 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).

(24) V. Papakonstantinou/Paul De Hert, 'The Regulation of Digital Technologies in the EU: The law-making phenomena of "act-ification", "GDPR mimesis" and "EU law brutality"' Technology and Regulation [2022], pp. 48-60.

(25) A descrição que segue é simplifcada, ou seja, os títulos dos capítulos não são exatamente os que enuncio.

1) disposições gerais;
2) práticas proibidas;
3) sistemas de risco elevado;
4) obrigações de transparência para certo tipo de sistemas;
5) modelos de finalidade geral;
6) medidas de apoio à inovação;
7) governação;
8) base de dados relativo a sistema de risco elevado;
9) partilha de informações e fiscalização do mercado;
10) códigos de conduta e orientações;
11) delegação de poderes e procedimentos do Comité;
12) sanções; e
13) disposições finais.

A grande divisão do Regulamento é feita com base numa classificação de risco dos sistemas de IA([26]). Esta classificação tem em conta as utilizações ou aplicações dos sistemas de IA. Está assim em causa saber para que é o que sistema foi concebido, a chamada "finalidade prevista", definida no art. 3.º/12 como "*utilização a que o prestador destina o sistema de IA, incluindo o contexto específico e as condições de utilização, conforme especificado nas informações facultadas pelo prestador nas instruções de utilização, nos materiais e declarações promocionais ou de venda, bem como na documentação técnica*". Nesse sentido, a mesma algoritmia e *software* aplicados no sistema A pode levar a uma classificação com um nível de risco diferente do sistema B([27]). Por outro lado, à partida o presta-

([26]) Fala-se por isso numa *risk based approach*. "Risco" é definido no art. 3.º/2 do Regulamento como "*a combinação da probabilidade de ocorrência de danos com a gravidade desses danos*". Sobre a abordagem regulatória baseada no risco *vide* Giovanni De Gregorio/Pietro Dunn, 'The European risk-based approaches: Connecting constitutional dots in the digital age' Common Market Law Review, Vol. 59(2) (2022), pp. 473-500. Criticando a noção de risco no contexto do regulamento, cf. Marco Almada/Nicolas Petit, 'The EU AI act: a medley of product safety and fundamental rights?' RSC Working Paper 2023/59, pp. 19-20.

([27]) Frequentemente existirá uma dificuldade em determinar a utilização em causa — se o sistema tem várias aplicações possíveis e o Regulamento se aplica a toda a cadeia de distribuição do sistema poderá esse sistema ter níveis de risco diferentes ao longo da cadeia? A resposta deve ser afirmativa. Como se dá nota, o que importa para a classificação é a utilização prevista. Quando o sistema foi concebido para uma dada utilização, de risco reduzido está a ser efetivamente utilizado para uma apli-

dor poderá afastar a aplicação de certas normas ou mesmo do Regulamento como um todo, se for cuidadoso e explícito nas instruções e nos materiais que disponibilize([28]).

Existem essencialmente dois níveis de risco: risco intolerável (que leva à proibição de determinadas práticas ou utilizações de sistemas de IA — art. 5.º)([29]) e risco elevado([30]). A generalidade das regras são destinadas aos sistemas de risco elevado. O art. 5.º apresenta, como veremos, dificuldades de interpretação e delimitação. Nesse sentido, é essencial recorrer ao art. 6.º, que define os sistemas de risco elevado, para perceber qual o âmbito de aplicação das práticas proibidas. Se o Regulamento considera que uma determinada utilização do sistema de IA é de risco elevado, então esta não poderá ser incluída nas práticas proibidas.

O Regulamento disciplina também os chamados modelos de IA de finalidade geral, isto é, "*um modelo de IA* (…) *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 siste-*

---

cação de risco elevado, o art. 25.º prevê que essa alteração de finalidade possa alterar a qualificação do sujeito que a fez, passando este de "responsável pela implementação" (o utilizador) a "prestador" (o principal responsável por garantir o respeito pelo Regulamento). Além disso, o Regulamento trata dos modelos gerais (art. 51.º e ss.), que podem ser usados para muitas finalidades.

([28]) Art. 8.º. Mesmo assim, o Regulamento obriga o produtor de um sistema de risco elevado a ter um sistema de gestão de riscos, onde se inclui [art. 9.º/2/*b*)] a estimativa e avaliação dos riscos que podem surgir de uma "utilização indevida razoavelmente previsível", definida como "*utilização de um sistema de IA de uma forma não conforme com a sua finalidade prevista, mas que pode resultar de um comportamento humano razoavelmente previsível ou de uma interação razoavelmente previsível com outros sistemas, incluindo outros sistemas de IA*" (art. 3.º/13).

([29]) Pode perguntar-se se esta abordagem faz sentido. Se a mesma aplicação ou prática ocorresse sem recurso a sistemas de IA seria lícita? Se a resposta for negativa, então a associação aos sistemas de IA será irrelevante. Na verdade, o que está em causa o art. 5.º é a regulação de condutas.

([30]) O art. 50.º não diz respeito a "baixo risco" ou "risco limitado", aplica-se à luz da utilização em causa, independentemente da classificação de risco do sistema. Assinala-se frequentemente que existem sistemas de IA, como videojogos e filtros de *spam*, que não estão abrangidas pelo Regulamente, constituiriam uma outra categoria de "ausência de risco". Ora, tudo na vida tem os seus riscos… Parece-me preferível assinalar apenas que esses sistemas não são abrangidos pelo Regulamento. Marco Almada/Nicolas Petit, *ob. cit.*, pp. 8-9 falam em três regimes: risco intolerável (art. 5.º), alto risco (coberto pelo Regulamento) e outros sistemas de IA (que não são abrangidos pelo Regulamento, mas que poderão estar sujeitos i.a. ao Regulamento 2023/988 relativo à segurança geral dos produtos). No sentido, do qual discordo, de existirem quatro níveis de risco cf. Maria João Vaz, *et al.*, 'Análise crítica da Proposta de Regulamento sobre Inteligência Artificial: considerações sobre os sistemas de identificação biométrica em especial' (2023). JusGov Research Paper N.º 2023-06 *in* <https://ssrn.com/abs tract=4420154>, p. 5. Falando em três níveis, cf. Emilija Leinarte, 'The Classification of High-Risk AI Systems Under the EU Artificial Intelligence Act' Journal of AI Law and Regulation, Vol. 1(3) (2024), pp. 262-264.

mas 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" (art. 3.º/63), em especial aqueles que apresentem risco sistémico (arts. 51.º e ss.).

A abordagem seguida no Regulamento alinha-se com a legislação sobre segurança e liberdade de circulação de mercadorias, nomeadamente o Regulamento (UE) 2023/988 de 10 de maio de 2023 relativo à segurança geral dos produtos([31]), e os instrumentos normativos sectoriais sobre brinquedos([32]), cosméticos([33]), ou dispositivos médicos([34]). Em todos estes regimes está em causa a regulação com vista à garantia de segurança no mercado comum([35]). Esta opção trata os sistemas de inteligência artificial como mercadorias, sujeitando os sistemas de risco elevado a uma marca de conformidade (*CE* — abreviatura de *conformité européenne*) através da qual se confirma que houve uma verificação e que o sistema de IA (de risco elevado) está de acordo com a legislação europeia aplicável (art. 48.º)([36]). A forma mais simples de evitar ambiguidades e dificuldades interpretativas passará por seguir os padrões e normas técnicas aprovadas ao abrigo

---

([31]) Em 2008 a UE adotou o chamado "Novo Quadro Legislativo" (New Legislative Framework) atualizando as regras gerais para garantia da segurança e conformidade de produto, acompanhado de regras especiais para determinadas categorias (nesta data 26 categorias, incluindo elevadores, material de construção, explosivos, rádio, fertilisantes, baterias, máquinas e *drones*). O Regulamento faz agora parte dessa categoria de legislação.

([32]) Directiva 2009/48/CE do Parlamento Europeu e do Conselho de 18 de Junho de 2009 relativa à segurança dos brinquedos

([33]) Regulamento (CE) n.º 1223/2009 do Parlamento Europeu e do Conselho, de 30 de Novembro de 2009, relativo aos produtos cosméticos.

([34]) Regulamento (UE) n.º 2017/745 do Parlamento Europeu e do Conselho, de 5 de abril de 2017, relativo aos dispositivos médicos. Como se lê no considerando 19 deste regulamento "*É necessário clarificar que o software, por si só, é qualificado como dispositivo médico quando especificamente destinado pelo fabricante a ser utilizado para um ou vários fins médicos indicados na definição de dispositivo médico, ao passo que o software de uso geral, mesmo quando utilizado num contexto de saúde, ou o software previsto para fins relacionados com o estilo de vida e o bem-estar, não são um dispositivo médico. A qualificação de um software, quer como dispositivo quer como acessório, deverá ser independente da localização do software ou do tipo de interconexão entre este e um dispositivo*". Sobre este regulamento, cf. Peter Feldschreiber (ed.), *The Law and Regulation of Medicines and Medical Devices* (OUP 2021).

([35]) Michael Veale/Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach' Computer Law Review International (2021), p. 98. Nesse sentido o AI Act faz copiosas referências ao Regulamento (UE) 2019/1020 relativo à fiscalização do mercado e à conformidade dos produtos.

([36]) As regras da marcação CE constam do Regulamento (CE) n.º 765/2008 do Parlamento Europeu e do Conselho, de 9 Julho de 2008, que estabelece os requisitos de acreditação e fiscalização do mercado relativos à comercialização de produtos. As normas sobre standards constam do Regulamento (UE) n.º 1025/2012 do Parlamento Europeu e do Conselho de 25 de outubro de 2012 relativo à normalização europeia.

do Regulamento de Standards(37), beneficiando desse modo de uma presunção de conformidade (arts. 40.º/1 e 42.º/2)(38).

Apesar disso, o Regulamento não pode deixar de ter presente a complexidade (e sofisticação) da Inteligência Artificial. Para usar as palavras de Laura Caroli, "[um sistema de] IA não é uma torradeira"(39). Por esse motivo, o AI Act apresenta consideráveis desvios face ao paradigma regulatório de segurança dos produtos, nomeadamente a imposição de deveres aos utilizadores dos sistemas (art. 26.º) e, nalguns casos, uma avaliação de impacto sobre os direitos fundamentais (art. 27.º). Nesse sentido, estamos perante um híbrido, que conjuga uma abordagem típica do RGPD com outra, própria do Direito Regulatório, que é utilizada em instrumentos como o Regulamento (UE) n.º 2017/745 relativo a dispositivos médicos. No entanto, à exceção do direito a apresentar queixa (art. 85.º) e a uma explicação do papel do sistema de IA em determinadas decisões (art. 86.º), este Regulamento não consagra direitos subjetivos.

Não obstante estar em causa um regulamento de "Inteligência Artificial" parece-me que muitas destas práticas e atuações, em especial as práticas proibidas, já estariam cobertas por outro quadro regulatório, nomeadamente o Regulamento dos Serviços Digitais(40), o Regulamento Geral de Proteção de Dados(41), as regras de lealdade na concorrência e proteção do consumidor, incluindo o Direito da Publicidade e, de um modo mais geral, as normas de proteção de direitos de personalidade e Direitos Fundamentais(42). Não nos podemos esquecer que a regulamentação da Inteli-

---

(37) Regulamento (UE) n.º 1025/2012 do Parlamento Europeu e do Conselho de 25 de outubro de 2012 relativo à normalização europeia, que altera as Diretivas 89/686/CEE e 93/15/CEE do Conselho e as Diretivas 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE e 2009/105/CE do Parlamento Europeu e do Conselho e revoga a Decisão 87/95/CEE do Conselho e a Decisão 1673/2006/CE do Parlamento Europeu e do Conselho.

(38) Michael Veale/Frederik Zuiderveen Borgesius, *ob. cit.*, p. 105 assinalam que este será provavelmente o caminho seguido pela generalidade dos produtores.

(39) <https://iapp.org/news/a/will-the-eu-ai-act-work-lessons-learned-from-past-legislative-initiatives-future-challenges/>. Na mesma linha, cf. Marco Almada/Nicolas Petit, *ob. cit.*

(40) Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho de 19 de outubro de 2022 relativo a um mercado único para os serviços digitais e que altera a Diretiva 2000/31/CE (Regulamento dos Serviços Digitais).

(41) Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 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 (Regulamento Geral sobre a Proteção de Dados).

(42) Cf. Stefan Scheurer, 'Artificial Intelligence and Unfair Competition — Unveiling an Underestimated Building Block of the AI Regulation Landscape' GRUR Int, Vol. 70(9) (2021), pp. 834-845.

gência Artificial não começa nem acaba neste Regulamento, não obstante a sua inegável importância(43).

3. Conceitos

O Regulamento adotou uma abordagem maximalista às definições, definindo termos que constam já do acervo europeu como "dados pessoais", "dados não pessoais", "definição de perfis", "dados biométricos", consagrando definições pouco úteis como "literacia no domínio de IA" e termos que são autoexplicativos tais como "espaço acessível ao público", "dados de treino" ou "instruções de utilização"(44).

Em contrapartida, o termo "aplicação da lei" é enganador face ao seu sentido comum. Este termo, que surge 98 vezes no Regulamento, é definido no art. 3.º/46 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*", sendo a "autoridade responsável pela aplicação da lei" "*uma autoridade pública competente 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; ou qualquer outro organismo ou entidade designado pelo direito de um Estado-Membro para exercer autoridade pública e poderes públicos para efeitos de prevenção, investigação, deteção ou pelo exercício da ação penal relativo a infrações penais ou execução de sanções penais, incluindo a proteção contra ameaças à segurança pública e a prevenção das mesmas*;" (art. 3.º/45). Ou seja, quando o Regulamento se refere à aplicação da lei está a referir-se essencialmente à atividade policial.

---

(43) Mesmo em termos de segurança dos produtos no mercado interno, o considerando 166 assinala que "*é importante os sistemas de IA relacionados com produtos que não são de risco elevado, nos termos do presente regulamento e que, como tal, não são obrigados a cumprir os requisitos aplicáveis a sistemas de IA de risco elevado, sejam seguros quando são colocados no mercado ou colocados em serviço. A fim de contribuir para alcançar esse objetivo, o Regulamento (UE) 2023/988 do Parlamento Europeu e do Conselho deverá ser aplicado como uma rede de segurança*.".

(44) Cf., respetivamente, os números 50, 51, 52, 34, 56, 44, 29 e 15 do art. 3.º. Por outro lado, define-se "infração generalizada" (art. 3.º/61.º) e "falsificações profundas" (arts. 3.º/60.º), que são termos usados apenas uma vez no Regulamento (respetivamente arts. 73.º/3 e 50.º/4).

Apesar do que se disse, a definição de sistema de Inteligência Artificial e a análise das várias categorias de sujeitos são indispensáveis à boa compreensão do Regulamento.

3.1. Sistema de Inteligência Artificial

Um primeiro desafio da regulação passa por encontrar uma definição satisfatória de Inteligência Artificial. Muitas definições associam a inteligência à inteligência humana, à capacidade de utilizar raciocínio para alcançar objetivos. Outras perspetivas abordam o conceito pelas técnicas de programação utilizadas([45]). Depois de muitas discussões, o Regulamento acabou por adotar a definição de "sistemas de Inteligência Artificial", que replica a definição atualizada da OCDE: "*um sistema baseado em máquinas concebido para funcionar com níveis de autonomia variáveis, e que pode apresentar capacidade de adaptação após a implantação e que, para objetivos explícitos ou implícitos, e com base nos dados de entrada que recebe, infere a forma de gerar resultados, tais como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais*" (art. 3.º/1)([46]).

Esta noção parece particularmente ampla e quase coincidente com a noção de *software*. As notas distintivas são a existência de algum grau de

---

([45]) Era essa a abordagem, muito criticada, originalmente proposta pela Comissão Europeia.

([46]) A noção corresponde também à utilizada no art. 2.º da Convenção CoE. Sobre a atualização da definição OCDE veja-se o *Explanatory Memorandum on The Updated OECD Definition of an AI System* (Março 2024), que prefere utilizar, para efeitos regulatório, a noção de *sistemas* de IA. Esta perspetiva vai na linha da que consta da Ordem Executiva norte-americana sobre Inteligência Artificial (*Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)*, mas apresenta algumas diferenças assinaláveis. A Ordem presidencial, que generaliza a abordagem da Executive Order 13960 (essa dirigida apenas às agências federais), assenta sobretudo em exigências de cibersegurança, monitorização e qualidade técnica dos sistemas e define Inteligência Artificial, na section 3 *b*) (da EO 14110), como "a *machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action*". Como se pode perceber, a Executive Order norte-americana faz referência à definição humana de objetivos, que não se exige na definição da OCDE e do AI Act. Sobre o tema, cf. Luciano Floridi, 'On the Brussels-Washington Consensus About the Legal Definition of Artificial Intelligence'. Philosophy & Technology (2023), Vol. 36 (87). A definição de IA utilizada no standard ISO/IEC 22989:2022 (2022) é próxima: "*a technical and scientific field devoted to the engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of defined objectives*".

autonomia e a menção de **inferências**(47). Nesse sentido, o considerando 12 explicita que "*a definição deverá basear-se nas principais características dos sistemas de IA que o distinguem de sistemas de software ou abordagens de programação tradicionais mais simples e não dever abranger sistemas baseados nas regras definidas exclusivamente por pessoas singulares para executarem operações automaticamente*". Para o efeito, sublinha que o essencial é a capacidade de fazer *inferências*, isto é, a possibilidade de tratar ou gerar novos dados em contextos diferentes daqueles em que o sistema foi treinado(48). Por outras palavras, excluem-se automações simples, fórmulas, *software* estático, ou programação totalmente determinística (*se x, então y*)(49). Como a noção é ampla, em caso de dúvida dever-se-á considerar que o sistema analisado constitui um sistema de IA.

É importante sublinhar que a regulamentação incide essencialmente sobre os sistemas como um todo (incluindo *hardware*, isto é, computadores, sensores, periféricos e *software* que não constitua inteligência artificial). Os **sistemas** devem ser distinguidos de **modelos**. Como se assinala no considerando 97:

> *Embora os modelos de IA sejam componentes essenciais dos sistemas de IA, não constituem, por si só, sistemas de IA. Os modelos de IA exigem a adição de outros componentes, como, por exemplo, uma interface de utilizador, para se tornarem sistemas de IA. Os modelos de IA são tipicamente integrados e fazem parte integrante dos sistemas de IA*.

Enquanto o ChatGPT (da OpenAI) constitui um sistema de IA (incluindo várias camadas de *software*, uma interface gráfica, servidores, etc.), existem vários modelos (que funcionam como "motor" do sistema) que o podem integrar (à data e no caso do ChatGPT estão disponíveis três opções: GPT 3.5, GPT 4 e GPT 4.º). É possível utilizar um mesmo modelo e construir sistemas com aplicações, finalidades e modos de funcionamento muito diferentes(50).

---

(47) Francisco Andrade, 'Proposta de Regulamento Europeu para a Inteligência Artificial' ROA (2023), p. 491.

(48) Considerando 12: "*...a capacidade de um sistema de IA fazer inferências vai além do tratamento básico de dados, permitindo a aprendizagem, o raciocínio ou a modelização*.". Como se sublinha no *Explanatory Memorandum on The Updated OECD... cit.*, também há inferências na fase de treino, em especial no caso de *unsupervised machine learning*.

(49) Os casos de *Robotic Process Automation* (i.e., uma forma de automatizar processos repetitivos, normalmente num contexto empresarial) terão de ser analisados em concreto. Nalguns casos, poderão estar em causa agentes de IA, noutros casos, mera programação determinística. Em todo o caso, parece que a maior parte dos casos de RPA não cairá no âmbito de aplicação material do Regulamento por não apresentarem um risco relevante.

(50) É sobretudo nessa medida que, como veremos, o Regulamento se preocupa também com modelos. Há, porém, uma definição de "sistema de IA de finalidade geral", no art. 3.º/66, como "*um*

3.2. Sujeitos

Entre os vários sujeitos mencionados no Regulamento e que formam parte da cadeia de distribuição de sistemas de IA, temos o fornecedor, o importador, o distribuidor, o mandatário e o utilizador, todos eles abrangidos pela noção genérica de "**operador**". Como veremos melhor, o Regulamento aplica-se a qualquer disponibilização do sistema na UE, mesmo que a título gratuito([51]).

O sujeito central, o principal visado pelas regras do AI Act, é o **prestador** (*provider*), definido no art. 3.º/3 como "*pessoa singular ou coletiva, autoridade pública, agência ou outro organismo que desenvolva, ou mande desenvolver, um sistema de IA ou um modelo de IA de finalidade geral e o coloque no mercado, ou coloque o sistema de IA em serviço sob o seu próprio nome ou a sua própria marca, a título oneroso ou gratuito*". O traço central que define alguém como prestador é a circunstância de oferecer um sistema de IA em nome próprio([52]). Os prestadores, quando não

---

sistema de IA baseado num modelo de IA de finalidade geral, e com a capacidade de servir para diversas finalidades, tanto para utilização direta como para integração noutros sistemas de IA". Esta noção só é utilizada no caso de modificação de um sistema desta natureza para servir uma finalidade (específica) classificada como sendo de risco elevado [art. 25.º/1/*c*)].

([51]) O Regualmento refere-se ao território da União, não abrangendo outros países do Espaço Económico Europeu.

([52]) Isto incluirá também os chamados OEM (*Original Equipement Manufacturer*), que podem não ter tido qualquer papel no desenvolvimento do sistema, mas que o integram no seu produto e/ou apresentam o sistema de IA como seu. A qualificação não parece, no entanto, poder ser contornada defendendo que o "produtor" do sistema se limita a fornecer meios técnicos. É claro que poderemos ter situações de qualificação duvidosa: quando a empresa A fornece *middleware* para permitir aos seus clientes desenvolverem modelos, aplicações ou mesmo sistemas de IA e/ou permite que esses modelos e aplicações corram na sua infraestrutura (servidores) quem é que é o prestador? Creio que poderemos considerar esse sistema de *middleware* da empresa A como um sistema de IA e a empresa A como prestador desse sistema. No entanto, os sistemas desenvolvidos por cada cliente da empresa A e eventualmente disponibilizado a terceiros constituirão eventualmente sistemas autónomos dos quais os clientes da empresa A serão prestadores. A situação pode complicar-se se a empresa A fornecer um sistema de IA configurável/parametrizável. Nesse caso, à luz do art. 25.º saber se esses clientes são responsáveis pela implementação ou se passam a ser prestadores de um novo sistema dependerá da extensão das modificações feitas e/ou da aposição da marca desse cliente no sistema. Caso os seus clientes assumam a qualificação de prestadores (de um novo sistema), a empresa A (prestadora do sistema original) deve "*cooperar estreitamente com novos prestadores, disponibilizar as informações necessárias e facultar o acesso técnico e a assistência razoavelmente esperados e necessários para o cumprimento das obrigações estabelecidas no presente regulamento*" (art. 25.º/2). Está também prevista uma "contratação forçada": nos termos do art. 25.º/4 "*O prestador de um sistema de IA de risco elevado e o terceiro que disponibilize um sistema de IA, ferramentas, serviços, componentes ou processos que sejam utilizados ou integrados num sistema de IA de risco elevado devem, **mediante acordo escrito**, especificar as informações necessárias, as capacidades, o acesso técnico e demais assistência, com base no estado da arte geralmente reconhecido, a fim de permitir que o prestador do sistema de IA de risco elevado cumpra*

estejam estabelecidos na UE, deverão cumprir as suas obrigações por meio de **mandatários** estabelecidos na UE (definidos no art. 3.º/5), tal como preveem os arts. 22.º (no caso dos sistemas de IA de risco elevado) e 54.º (modelos de IA de finalidade geral)(53).

O utilizador, excetuando aqueles que usem o sistema no âmbito de uma atividade pessoal de caráter não profissional(54), é designado "**responsável pela implementação**" (*deployer*) (art. 3.º/4) e tem igualmente obrigações próprias, nomeadamente de supervisão do funcionamento do sistema (cf. arts. 26.º e 50.º/3 e /4).

Os importadores, isto é, aqueles sujeitos localizados na UE que coloquem um sistema de IA no mercado interno (art. 3.º/6) terão algumas obrigações de verificação e garantia de conformidade, bem como de colaboração com as autoridades (art. 23.º). O Regulamento reserva a expressão "colocação no mercado" para a primeira disponibilização de um sistema de IA no território da UE (art. 3.º/9), sendo disponibilização definida como qualquer fornecimento no âmbito de uma atividade comercial (art. 3.º/10). Assim, os importadores fazem uma "colocação no mercado", enquanto os distribuidores (art. 3.º/7) se dedicam à "disponibilização no mercado" subsequente à importação(55). Os distribuidores são sujeitos a obrigações de verificação e cooperação com as autoridades muito semelhantes às dos importadores (art. 24.º).

Uma outra noção, que não se encontra definida, mas que é incluída no conceito de operador, é a de "fabricante de produtos" [referida nos arts. 2.º/1/*e*) e 3.º/8]. Tendo em conta que o que está em causa é a disponibilização conjunta de um produto e um sistema de IA sob o próprio nome ou marca, os fabricantes de produtos devem ser considerados prestadores(56).

---

plenamente as obrigações estabelecidas no presente regulamento". Apesar de o considerando 88 poder dar uma impressão diferente não creio que o art. 25.º/4 seja aplicável a quem se limite a disponibilizar modelos e creio que o dever de contratação "obrigatória" previsto neste artigo deve ser interpretado restritivamente (caso contrário e por absurdo, o fornecedor de sistemas de arrefecimento para os computadores utilizados no treino de um sistema de IA ou até o fornecedor de refeições aos *data scientists* estariam abrangidos).

(53) Esta obrigação é semelhante ao previsto no art. 27.º RGPD.

(54) Antecipo que esta exceção seja interpretada de forma restritiva. Assim, a minha utilização de um sistema de IA para, na qualidade de professor ou advogado, gerar imagens para uma apresentação numa conferência não estará abrangida.

(55) Esta distinção será mais frequente quando os sistemas de IA integrem *hardware* do que em relação a puro *software*. Em todo o caso, existem frequentemente acordos de distribuição de *software*, incluindo revenda. Sobre o tema, cf. Nuno Sousa e Silva, 'Contratos (sobre bens) informáticos: notas sobre a formação, conteúdo, incumprimento e *open source'* Revista de Direito Civil n.º 4 (2024), pp. 695-729.

(56) Nesse sentido, cf. o art. 25.º/3.

Está previsto que alguém se possa tornar prestador caso venha a "*colocar o seu nome ou marca num sistema de IA de risco elevado já colocado no mercado*" [art. 25.º/1/*a*)], "*introduzir uma modificação substancial num sistema de IA de risco elevado que já tenha sido colocado no mercado ou colocado em serviço, de forma que o mesmo continue a ser um sistema de IA de risco elevado*" [art. 25.º/1/*b*)] ou "*modificar a finalidade prevista* (…) *de forma que o sistema de IA em causa se torne um sistema de IA de risco elevado*" [art. 25.º/1/*c*)](57). Embora a hipótese inversa não esteja expressamente contemplada, a solução contrária também deverá valer: a modificação da utilização visada para uma que não seja considerada de risco elevado permitirá que o novo sistema escape à aplicação de certas regras ou mesmo do Regulamento.

4. Âmbito (material, territorial e temporal) de aplicação

Apesar de se apresentar como um regulamento geral e, em princípio, de unificação total, o Regulamento ressalva a aplicação do restante quadro regulatório (art. 2.º, números 5, 7 e 9)(58) e permite que em certos domínios se adotem regras nacionais complementares, como normas mais favoráveis para proteção de trabalhadores (art. 2.º/11) ou regras relativas à utilização de sistemas de identificação biométrica à distância (art. 5.º/5 e /10). Além disso, continua ressalvada a aplicação de alguma legislação (art. 2.º/2, referindo-se à lista que consta da Seção B do Anexo I) e supervisão setorial (arts. 72.º e 74.º)(59). Também há matérias que ficam dependentes de

(57) A noção de modificação substancial é definida no art. 3.º/23 como "*uma alteração do sistema de IA após a sua colocação no mercado ou colocação em serviço, que não tenha sido prevista ou planeada pelo prestador na avaliação da conformidade inicial e que, consequentemente, afete a conformidade do sistema de IA com os requisitos estabelecidos no capítulo II, secção 2, do presente regulamento, ou modifique a finalidade prevista relativamente à qual o sistema de IA foi avaliado*". Esta definição aproxima-se da ideia de desvio de finalidade estabelecida no art. 6.º/4 do RGPD. Um sistema sujeito a uma modificação substancial é tratado no Regulamento como sendo um novo sistema (cf. art. 43.º/4).

(58) Além destas previsões existem normas, como o art. 87.º, que remetem expressamente para outra legislação europeia. Mesmo assim haverá algumas dúvidas. Por exemplo, o art. 33.º da Loi n.° 2019-222 du 23 mars 2019 (francesa) prevê uma sanção penal para quem utilize dados para efeitos de avaliação, análise, comparação ou previsão das práticas profissionais efectivas ou presumidas de um juiz. Esta lei ainda será compatível com o Regulamento ou dever-se-á considerar inaplicável à luz do princípio do primado?

(59) Assinale-se ainda que o Regulamento, nos arts. 102.º a 110.º, altera vários instrumentos de Direito da União Europeia.

medidas de execução a nível nacional, em especial a designação de autoridades nacionais e o quadro de supervisão (arts. 70.º, 74.º e 77.º), bem como o regime das sanções (art. 99.º/2). Por outro lado, a Comissão tem um amplo poder de adotar atos delegados, completando e atualizando o Regulamento (arts. 7.º e 97.º), e extensos deveres de avaliação e reexame (art. 112.º). Prevê-se igualmente que a Comissão elabore orientações sobre o Regulamento (art. 96.º) e que incentive a elaboração de códigos de boas práticas (art. 56.º).

O Regulamento de IA alinha com as tendências mais recentes em matéria de regulação do mercado único digital, procurando uma aplicação **extra-territorial**(60). De acordo com o art. 2.º/1 basta um mínimo ponto de contacto do utilizador ou do resultado do sistema de IA com o território da União para desencadear a aplicabilidade do Regulamento. Assim, se o resultado de um sistema de IA for utilizado na UE ou afetar pessoas localizadas na UE isso bastará para que o Regulamento seja aplicável(61). Em contrapartida, o Regulamento não se aplica a quem desenvolver sistemas de IA na UE, ainda que com fins proibidos pelo Regulamento, para serem utilizados em países terceiros (diz-se, por isso, que não há uma proibição de exportação).

Uma nota importante em termos de **jurisdição** diz respeito ao caráter descentralizado da supervisão. Exceto no caso de modelos de IA de finalidade geral, que serão fiscalizados pela Comissão Europeia, as autoridades nacionais competentes serão responsáveis por lidar com todas as infrações que tenham lugar no seu território e âmbito de competência. Nesse sentido, e ao contrário do que se prevê no RGPD, a mesma entidade e infração pode ser sujeita à jurisdição de várias entidades nacionais(62).

---

(60) Veja-se Nuno Sousa e Silva, 'Novas Regras para a Internet: Notas Breves sobre Iniciativas Europeias de Regulação de Plataformas Digitais' Revista de Direito Intelectual 1/2021, pp. 75-102. Especificamente sobre o RGPD *vide* Christopher Kuner, 'Protecting EU Data Outside EU Borders under the GDPR' Common Market Law Review 60 (2023), pp. 77-106. Esta abordagem da União Europeia tem contribuído para o designado "Efeito Bruxelas", termo cunhado e descrito por Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (OUP 2020). Esta expressão alude ao poder de influência do acervo regulatório da União Europeia em matérias tais como o Direito da Concorrência, Direito Ambiental, Direito Digital e proteção de dados. Nesses domínios, a UE tem sido pioneira na regulação e frequentemente é seguida como modelo noutros ordenamentos. Além disso, empresas multinacionais acabam por adotar as regras europeias como padrão global de *compliance*. No entanto, no caso específico do Regulamento de Inteligência Artificial está longe de ser claro se a abordagem seguida a nível da União Europeia terá esse efeito (no sentido negativo veja-se Ugo Pagallo, *Why the AI Act Won't Trigger a Brussels Effect* (2023) *in* <https://ssrn.com/abstract=4696148>). Sobre o fenómeno veja-se também Nuno Cunha Rodrigues, *A Globalização do Poder Regulatório da União Europeia* (Almedina 2024).

(61) Nessa mesma linha, prevê-se a obrigação dos prestadores estabelecidos em países terceiros designarem um mandatário (arts. 22.º e 54.º).

(62) É igualmente possível que ocorra um conflito positivo ou negativo de competências.

Estão excluídas do âmbito material de aplicação do Regulamento as atividades de investigação e desenvolvimento "em laboratório" (para teste em ambiente real prevê-se, nos arts. 57.º e ss., um esquema complexo) (art. 2.º/6), bem como as atividades anteriores à colocação do sistema no mercado ou em serviço (art. 2.º/8). O Regulamento também não se aplicará a sistemas desenvolvidos ou utilizados exclusivamente para fins militares, de segurança nacional ou de defesa (art. 2.º/3) ou à utilização por autoridades públicas de países terceiros e organizações internacionais desde que essas entidades salvaguardem adequadamente os direitos fundamentais (art. 2.º/4).

O tema de ***open source*** foi objeto de grandes debates(63). Estão excluídos os usos "domésticos", i.e., "*no âmbito de uma atividade pessoal de caráter não profissional*" (art. 2.º/10). No entanto, a disponibilização de *software* (incluindo os parâmetros de um modelo) ao abrigo de licenças de código aberto pode ser feito também num contexto profissional(64). A solução de compromisso passa por uma isenção limitada (art. 2.º/12)(65). A chave de leitura é a já referida diferença entre modelos e sistemas. A disponibilização de *modelos* de IA em *open source* goza de determinadas isenções previstas no Regulamento. Prevê-se que os *modelos* disponibilizados ao abrigo de licenças *open source* só tenham de cumprir duas obrigações (política de cumprimento de direito de autor e transparência sobre dados de treino)(66) exceto nos casos de modelos de finalidade geral com risco sistémico (arts. 25.º/4, 53.º/2 e 54.º/6). Em contrapartida, para os *sis-*

---

(63) Sobre a noção e história do *open source*, com referências adicionais, cf. Nuno Sousa e Silva, 'Contratos (sobre bens) informáticos… *cit.*, pp. 723-729. Quanto à IA o debate coloca-se a vários níveis. Alguns advogam a necessidade de restringir a circulação de informação (sendo proponentes daquilo a que em inglês se chama *security through obscurity*), chegando a comparar a disponibilização de código relativo a certos sistemas à disponibilização de instruções para produzir uma bomba atómica. Outros argumentam que a abertura é o meio mais eficaz de garantir diversidade, avanço e mesmo segurança. Há também uma divergência considerável quanto a saber o que é que quer dizer *open source* em IA: se basta a disponibilização da arquitetura e parâmetros de um modelo (v.g. *open weights*) ou se também deve ser disponibilizado o conjunto de dados (*dataset*) utilizado para o desenvolver. Sobre a discussão conceptual nesta matéria vide Andreas Liesenfeld/Mark Dingemanse, 'Rethinking open source generative AI: open-washing and the EU AI Act' FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (June 2024), pp. 1774-1787. Nos considerandos 102 e 103, o Regulamento parece adotar uma noção bastante restrita de *open source*.

(64) Aliás, nalguns contextos apenas empresas com muitos recursos é que terão capacidade para desenvolver determinados modelos (v.g. o LLama desenvolvido pela Meta).

(65) Na verdade, o texto do art. 2.º/12 é totalmente inútil: a exclusão prevista não se aplica aos três tipos de sistemas cobertos pelo Regualmento.

(66) Como se explica no considerando 104, o facto de um modelo estar em *open source* não significa que se tenha acesso aos dados de treino ou que tenha sido garantido o respeito pelos direitos de propriedade intelectual.

temas de IA abrangido pelo Regulamento (independentemente do nível de risco), o facto de serem disponibilizados em *open source* é irrelevante. Em palavras simples, a isenção é para modelos, não para sistemas.

A **aplicação no tempo** deste Regulamento é faseada. A entrada em vigor ocorre no dia 1 de agosto de 2024, produzindo-se nessa data as alterações aos diplomas mencionados nos arts. 102.º a 110.º. A aplicação geral do Regulamento está prevista para dia 2 de agosto de 2026 (art. 113.º). Há, no entanto, partes do Regulamento que serão aplicáveis mais cedo. É o caso dos dois primeiros capítulos (relativos a práticas proibidas), aplicáveis a partir de 2 de fevereiro de 2025 [art. 113.º/*a*)], e das regras sobre o quadro institucional, aplicáveis a partir de 2 de agosto de 2025 [art. 113.º/*b*)].

Em contrapartida, as regras relativas aos sistemas de risco elevado que sejam componentes de segurança de produtos harmonizados (art. 6.º/1), terão uma *vacatio legis* de 36 meses, só sendo aplicáveis a partir de 2 de agosto de 2027 [art. 113.º/*c*)]. De um modo mais geral, as regras relativas aos sistemas de IA de risco elevado só serão aplicáveis a sistemas de IA que sejam colocados no mercado a partir dessa data. Os sistemas de IA já colocados no mercado, quando sejam considerados de risco elevado, estão isentos do cumprimento das regras do Regulamento a não ser que sejam objeto de alterações significativas (art. 111.º/2)([67]). Os modelos de IA de finalidade geral colocados no mercado antes de 2 de agosto de 2025 só estarão obrigados a cumprir o Regulamento a partir de 2 de agosto de 2027 (art. 111.º/3)([68]).

([67]) E, nesse sentido, já não se trate do mesmo sistema. Não é claro se a noção de "alterações significativas em termos de conceção" difere de "modificação substancial" utilizada nos arts. 25.º e 43.º/4. O considerando 128 indicia que os conceitos não são coincidentes. Em qualquer caso, esta regra, que confere uma vantagem signifcativa aos operadores já instalados, explica-se pela proibição de retroatividade (lembre-se que o que despoleta a aplicação da generalidade das regras do Regulamento é a colocação no mercado). Em contrapartida, as proibições do art. 5.º, que se referem às práticas proibidas (e não a requisitos de sistemas), ou seja, que regulam condutas, podem ser plenamente aplicáveis aos sistemas que já se encontram no mercado. Procurando dar o exemplo, no caso de certos "sistemas informáticos de grande escala" da União Europeia já em utilização, como o sistema informático de Schengen ou de informação de vistos e viagens (a lista consta do anexo X), que já estão em funcionamento, prevê-se que estes devem ser tornados conformes com o Regulamento até 31 de dezembro de 2030 (art. 111.º/1).

([68]) Em contrapartida, os modelos que sejam colocados no mercado a partir de 2 de Agosto de 2025 terão de cumprir as regras "imediatamente" [art. 113.º/1/*b*)].

5. Princípios

Apesar de não estar previsto na proposta inicial, que se dirigia essencialmente a determinar as práticas proibidas e a regular aplicações de risco elevado, o legislador chegou a ponderar consagrar um conjunto de princípios gerais aplicáveis a todos os operadores e a todos os sistemas sujeitos ao Regulamento([69]). Na versão final, o único dever com essa amplitude é a obrigação imposta aos prestadores e responsáveis pela implementação de garantir que as pessoas que operem ou usem os sistemas de IA "*dispõem de um nível suficiente de* ***literacia*** *no domínio da IA*" (art. 4.º)([70]).

Não obstante, muitos dos princípios discutidos continuam subjacentes às exigências colocadas essencialmente sobre os sistemas de risco elevado (arts. 8.º a 15.º) e os seus operadores (arts. 16.º a 27.º).

Está em causa um conjunto de preocupações desenvolvidas no campo interdisciplinar conhecido como segurança em Inteligência Artificial (*AI safety*) ou *FATE AI (*do inglês *Fairness* (jusiça), *Accountability* (responsabilidade), *Transparency* (transparência), *Ethics* (ética), incluindo preocupações de controlo, transparência, alinhamento, não discriminação, robustez e segurança.

A enumeração destes princípios, em especial a **equidade** (*fairness*), traz à memória as palavras do juiz inglês Justin Jacob: "…*We are all against misappropriation, just as we are all in favour of mother and apple pie*"([71]). Como é evidente, somos todos a favor de justiça (*fairness*). A grande dificuldade, que é aliás o campo da filosofia e depois da política, traduzindo na escolha comprometida de cada sociedade num certo tempo e lugar através do Direito positivo, está em definir o que é que é justo e equitativo, e qual a melhor composição de interesses. Este problema é simulta-

---

([69]) Nomeadamente num art. 4.º apresentado em Maio de 2023 [COM(2021)0206 — C9 0146/2021 — 2021/0106(COD)], que enunciava os seguintes princípios: "*a*) supervisão e controlo humano; *b*) robustez técnica e segurança; *c*) privacidade e governação de dados; *d*) transparência; *e*) diversidade, não discriminação e equidade (*fairness*); *f*) bem estar social e ambiental". O considerando 27 alude ainda a este sete princípios como base para a elaboração de códigos de conduta. Os arts. 7.º a 13.º da Convenção CoE também enunciam os seguintes princípios: dignidade humana e autonomia, transparência e controlo, responsabilidade (*accountability and responsibility*), igualdade e não discriminação, proteção da privacidade e dados pessoais, fiabilidade e inovação segura. Muitos destes princípios coincidem com aqueles que são enumerados no art. 5.º do RGPD, os quais serão plenamente aplicáveis quando os sistemas de IA processem dados pessoais.

([70]) O art. 20.º da Convenção CoE também estabelece um princípio de promoção da literacia digital.

([71]) *L'Oreal SA & Ors v Bellure NV & Ors* [2007] EWCA Civ 968, §160.

neamente conceptual e técnico-matemático(72). Por isso, em termos práticos não se pode extrair muito deste princípio.

Existem dificuldades semelhantes em relação à chamada **discriminação** (ou enviesamento) algorítmica (*algorithmic bias*) sendo que alguns dos problemas conhecidos surgem da falta de qualidade dos dados utilizados (nomeadamente falta de representatividade ou insuficiência quantitativa ou qualitativa) ou de erros de programação(73). Em contrapartida, um grande número de situações problemáticas resulta, simplesmente, de o sistema ter sido otimizado para atingir um dado objetivo benéfico ou inócuo. Por exemplo, se um algoritmo for desenhado para privilegiar aquilo a que um internauta dedica mais atenção, pode acabar a recomendar-lhe bebida alcoólicas (tendo detetado, indiretamente, que é um alcoólico), ou a promover discurso ofensivo ou agressivo (uma vez que é aquilo a que generalidade das pessoas prestará mais atenção). Estes desafios, em especial aqueles colocados pelos sistemas de recomendação, são já parcialmente abordados no Regulamento dos Serviços Digitais ("DSA")(74). Em todo o

---

(72) Sorelle A. Friedler/Carlos Scheidegger/Suresh Venkatasubramanian, 'The (Im)possibility of fairness: different value systems require different mechanisms for fair decision making' Communications of the ACM. 64 (4) (2021), pp. 136-143.

(73) Os exemplos multiplicam-se, nomeadamente sistema de reconhecimento facial do Google Photos classificar indivíduos negros como gorilas (em 2015), a ferramenta de recrutamento da Amazon prejudicar mulheres (2018) e, mais recentemente, em 2023, a ferramenta iTutorGroup, utilizada em recrutamento, rejeitar automaticamente candidaturas de mulheres com mais de 55 anos e homens com mais de 60. O problema da discriminação algorítmica é disseminado e atinge grande escala como demonstraram Z. Obermeyer, *et al.*, 'Dissecting racial bias in an algorithm used to manage the health of populations' Science, (2019) 366(6464), pp. 447-453 a propósito do sistema de saúde nos EUA. Sobre o tema, cf. Hilde Weerts, *et al.*, 'Algorithmic unfairness through the lens of EU non-discrimination law: Or why the law is not a decision tree'. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (2023), pp. 805-816 e Philipp Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law' Common Market Law Review 55 (2018), pp. 1143-1186.

(74) O DSA define «Sistema de recomendação» como "um *sistema total ou parcialmente automatizado utilizado por uma plataforma em linha para sugerir na sua interface em linha informações específicas aos destinatários do serviço ou conferir prioridade a essa informação, nomeadamente como resultado de uma pesquisa iniciada pelo destinatário do serviço, ou que determine de outra forma a ordem relativa ou a proeminência das informações apresentadas"* [art. 3.º/*s*)] e impõe, apenas aos fornecedores de plataformas em linha, obrigações de transparência desses sistemas (art. 27.º). No caso dos fornecedores de plataformas em linha ou motores de pesquisa em linha de muito grande dimensão existem ainda deveres de avaliação de risco sistémico, incluindo a avaliação da "*conceção dos seus sistemas de recomendação e de qualquer outro sistema algorítmico pertinente*" [art. 34.º/2/*a*)] e adotar medidas de atenuação dos riscos identificados sobre esses sistemas [art. 35.º/1/*d*)]. Nos termos do art. 38.º do DSA as plataformas em linha de muito grande dimensão e os motores de pesquisa em linha de muito grande dimensão devem permitir aos utilizadores configurar os sistemas de recomendação para que estes não façam definição de perfis (noção definida no art. 4.º/4 do RGPD). Prevê-se ainda que os fornecedores destes sistemas tenham de explicar aos reguladores "*a conceção, a lógica, o funcionamento*

caso, o Regulamento coloca ênfase relevante na diversidade e prevenção de discriminação e enviesamento([75]). Acabar com essas ocorrências é impossível, mas existe uma obrigação de colocar esforços adequados, seguir as melhores práticas, com visto a prevenir erros facilmente evitáveis.

Em relação à **transparência** esta pode ser entendida como referente a vários conceitos diferentes([76]). Numa primeira aceção, constante do art. 50.º, pode referir-se tão somente à identificação da origem de um determinado conteúdo ou agente como sendo ou provindo de sistemas de IA. A transparência está também contemplada na obrigação de disponibilização e manutenção de documentação técnica (arts. 11.º, 18.º, 20.º e Anexo IV), manutenção de registos (arts. 12.º e 19.º), prestação de informações (art. 13.º) e cooperação com autoridades (art. 21.º).

Quando a transparência diz respeito à caraterística do sistema de IA, este conceito pode aludir à descrição das tarefas humanas de conceção, configuração e disponibilização do sistema, ainda que este seja em si (i.e. no seu funcionamento) opaco. Por vezes utiliza-se transparência para referir **interpretabilidade**, isto é, a capacidade de perceber como é que um sistema de IA funciona([77]), e/ou

---

e a testagem dos seus sistemas algorítmicos, incluindo os seus sistemas de recomendação" (art. 40.º/3 DSA). Sobre o tema dos sistemas de recomendação, cf. Sergio Genovesi/Katharina Kaesling /Scott Robbins (eds.), *Recommender Systems: Legal and Ethical Issues* (Springer 2023) e Mireille Hildebrandt, 'The issue of proxies and choice architectures. Why EU law matters for recommender systems.' Frontiers in Artificial Intelligence 5 (2022): 789076.

([75]) Em especial no art. 10.º a propósito da governação de dados e no art. 15.º/4 relativo a cibersegurança. Na documentação técnica exigida aos prestadores de modelos de finalidade geral também está prevista "*uma descrição pormenorizada dos elementos do modelo* (…) *e as informações pertinentes sobre o processo de desenvolvimento, incluindo* (…) *Informações sobre os dados utilizados para treino, testagem e validação, se aplicável, incluindo o tipo e a proveniência dos dados e as metodologias de curadoria (p. ex., limpeza, filtragem, etc.), o número de pontos de dados, o seu âmbito e as principais características; a forma como os dados foram obtidos e selecionados, bem como todas as outras medidas para detetar a inadequação das fontes de dados e dos métodos para detetar enviesamentos identificáveis, se aplicáve*l" [Anexo XI, Secção 1 (2)]. Por outro lado, o Regulamento confere poderes de fiscalização dos sistemas de IA de risco elevado às autoridades ou organismos públicos nacionais que supervisionam ou asseguram o respeito das obrigações previstas na legislação da União que protege os direitos fundamentais, incluindo o direito à não discriminação (art. 77.º).

([76]) O RGPD também usa o conceito de transparência no art. 5.º/1 e no considerando 58, referindo-se à comunicação clara de informação. Assinalando a "marcada polissemia" do conceito de transparência, cf. Lorenzo Cotino Hueso, 'Transparencia y explicabilidad de la inteligencia artificial y "compañía" (comunicación, interpretabilidad, integilibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad…). Para qué, para quién y cuánta.' *in* Lorenzo Cotino Hueso/Jorge Castellanos Claramunt (eds.), *Transparencia y explicabilidad de la inteligencia artificial* (Tirant lo Blanch 2022) p. 25, ss. Em 2017 Zachary C. Lipton, *The Mythos of Model Interpretability*, arXiv:1606.03490 (2017) afirmava mesmo: "*the term interpretability holds no agreed upon meanin*g". A citada norma técnica de 2020 utilizada neste texto parece contribuir para uma maior certeza terminológica.

([77]) Esta é a definição da norma técnica ISO/IEC TR 29119-11:2020(en), 3.1.42.

explicabilidade (*explainability*), ou seja, a elucidação da razão pela qual se obteve um determinado resultado pela operação do sistema([78]). Um sistema pode ser interpretável, mas produzir resultados concretos que não são explicáveis([79]). Por exemplo, sabemos quais os parâmetros utilizados e os passos seguidos pelo sistema para atribuir um valor do prémio no contrato de seguro, mas em concreto não conseguimos explicar porque é que o individuo A tem um prémio mais elevado que o individuo B. Existem, porém, técnicas de inteligência artificial que geram sistemas totalmente opacos (v.g. os grandes modelos de linguagem, como GPT); não sabemos quase nada acerca do seu funcionamento interno([80]). Em relação a esses, a interpretabilidade e explicabilidade não são tecnicamente possíveis([81]).

O AI Act não impõe uma obrigação geral de gerar modelos ou decisões explicáveis. No entanto, no caso dos sistemas de risco elevado estabelece-se um direito à explicação do papel do sistema (arts. 13.º e 86.º), e a perceber os grandes princípios do seu funcionamento e da decisão tomada (arts. 14.º e 86.º). O texto do art. 86.º (e o considerando 171) não é totalmente claro quanto a saber se é necessário explicar a decisão concreta ou se basta uma explicação genérica([82]). Por outro lado, as referência às capacidades técnicas pertinentes para explicar os resultados [art. 13.º/3/*b*)/*iv*)] e "*se for caso disso, informações que permitam aos responsáveis pela*

([78]) Cf. a definição utilizada na norma técnica ISO/IEC TR 29119-11:2020(en), 3.1.31. Como explica Tiago Sérgio Cabral, 'Regulamento sobre a Inteligência Artificial na União Europeia: potenciais impactos nas entidades públicas' Revista de Direito Administrativo 12 (2021) pp. 98-99 apesar de o art. 9.º da Carta Portuguesa de Direitos Humanos na Era Digital (Lei n.º 27/2021, de 17 de maio), relativo a "uso da inteligência artificial e de robôs" falar em "explicabilidade" e "sistema auditáveis" não parece que essa norma tenha real impacto e, em qualquer, caso terá de ser sujeita a interpretação conforme ao Direito da UE.

([79]) O art. 14.º/4/*c*) prevê que o sistema deve permitir a um ser humano "*interpretar corretamente os resultados do sistema de IA de risco elevado, tendo em conta, por exemplo, as ferramentas e os métodos de interpretação disponíveis*". Esta formulação parece admitir a utilização da chamada *black-box AI*, pois, nesses casos, não há ferramentas ou métodos de interpretação disponíveis.

([80]) Este é um domínio de investigação cientifica. Recentemente um conjunto vasto de investigadores da Anthropic publicou um paper "Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet" (<https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html>) em que o tema é discutido em detalhe e demonstrado avanços na possibilidade de interpretação de modelos de linguagem e da utilização desta técnica para efeitos de segurança.

([81]) Apesar de o considerando 71 e, numa certa medida o art. 15.º/1/*h*) do RGPD poderem dar a impressão de que existiria um direito a uma explicação das decisões automatizadas, essa não parece ser a interpretação mais correta. Cf. *supra* nota 4 e ainda L. Edwards./M. Veale, 'Enslaving the algorithm: From a "right to an explanation" to a "right to better decisions"?' IEEE Security & Privacy, 16(3) (2018), pp. 46-54.

([82]) As diferentes versões linguísticas (em inglês "meaningful explanation", em português "explicação clara e pertinente", em espanhol "claras y significativas", em francês "claires et pertinentes", em italiano "chiare e significative" e em alemão "klare und aussagekräftige") não são conclusivas.

implantação interpretar os resultados do sistema de IA de risco elevado e utilizá-los adequadamente" [art. 13.º/3/*b*)/*vii*)] são feitas no contexto da documentação técnica, o que parece indiciar que está em causa uma explicação genérica e abstrata (*interpretabilidade*) e não uma verdadeira *explicabilidade*. Além disso, mesmo que se estabelecesse um direito à explicação da decisão em concreto, a proteção de dados pessoais, de segredos comerciais e de outros tipos de sigilo, funcionaria como limite ao exercício desse direito([83]). Nesse sentido, a meu ver, as técnicas de IA que não permitem gerar explicações (v.g. redes neuronais de aprendizagem profundas ou *support vector machines*) continuam a ser legalmente admissíveis, mesmo no caso de sistemas de risco elevado.

A **supervisão e controlo humano** traduzem-se na obrigação de o prestador adotar um sistema de gestão de riscos (art. 9.º), controlo de qualidade (art. 17.º), de acompanhar o seu funcionamento pós comercialização (art. 72.º), comunicar incidentes graves (art. 73.º) e de conceber os sistemas de risco elevado de um modo que permita a compreensão e intervenção sobre o respetivo funcionamento (art. 14.º), nomeadamente a existência de um meio de paragem imediato ("kill switch") [art. 14.º/4/*e*)]([84]). Estes aspetos cruzam-se com as preocupações de cibersegurança e robustez (art. 15.º) —, à qual está associado um quadro legislativo importante, designadamente a Diretiva NIS 2 (Dir. 2022/2555 de 14 de dezembro de 2022 relativa a medidas destinadas a garantir um elevado nível comum de cibersegurança na União) — e com a norma do RGPD que restringe a possibilidade de decisões automatizadas a certos casos (art. 22.º RGPD)([85]).

A implementação destes princípios e do Regulamento como um todo será em grande medida densificada através de normas técnicas (*standards*) e orientações da Comissão, o que contribuirá para aumentar a segurança jurídica.

([83]) Em sentido próximo veja-se o art. 25.º/5. Estão aliás previstos deveres de sigilo e confidencialidade (art. 78.º). Sobre uma tensão análoga, cf. Filipa Pinto de Oliveira, *Os Segredos de Negócio como limite aos Titulares de Direitos de Proteção de Dados. Em especial, as interações dos segredos de negócio com o direito de acesso e o direito de portabilidade de dados pessoa*is, FDUCP (Porto), 2022 (publicada em <https://openbooks.ucp.pt/ucp/catalog/view/174/175/2017>) e Gianclaudio Malgieri, 'Trade Secrets v Personal Data: A Possible Solution for Balancing Rights' International Data Privacy Law, Vol. 6(2) (2016), pp. 102-116.

([84]) Criticando esta previsão, cf. Francisco Andrade, *ob. cit.*, p. 491.

([85]) Sobre esta norma e os problemas associados, cf. Mafalda Miranda Barbosa, *ob. cit.*, p. 131, ss., e Diogo Morgado Rebelo, *ob. cit.*, p. 280, ss. Sobre as intereseções entre IA e Proteção de dados mais amplamente, cf. Federico Marengo, *Privacy and AI: Protecting Individual's Rights in the Age of AI* (2023).

6. Práticas proibidas

Numa fase inicial a Comissão propôs o estabelecimento de quatro práticas proibidas, ditas de risco inaceitável, que podiam ser sumariamente descritas como sistemas de manipulação subliminar, sistemas que explorassem vulnerabilidades causando distorção de comportamento e danos, sistemas de pontuação (*scoring*) social e sistemas de identificação biométrica em tempo real (e.g. reconhecimento facial). Estas previsões tinham algumas exceções e utilizavam linguagem particularmente vaga([86]). Depois de intensas discussões e negociações, a linguagem foi apurada, a lista de práticas proibidas foi alargada, mas o resultado não é muito melhor. Proíbem-se agora:

— Manipulação e exploração de vulnerabilidades — art. 5.º/1/*a*) e *b*);

— Pontuação social (*social scoring*) geral — art. 5.º/1/*c*);

— Policiamento preditivo — art. 5.º/1/*d*);

— Criação de bases de dados de reconhecimento facial — art. 5.º/1/*e*);

— Sistemas de reconhecimento de emoções no local de trabalho ou ensino — art. 5.º/1/*f*);

— Classificação biométrica de categorias protegidas — art. 5.º/1/*g*);

— Casos especiais de identificação biométrica em tempo real — art. 5.º/1/*h*).

Esta lista não é exaustiva. Como é evidente outras práticas poderão ser proibidas ou ilícitas com base noutros fundamentos (art. 5.º/8). Por exemplo, os sistemas que gerem falsificações profundas (*deep fakes*) não são normalmente vistos como sendo de risco elevado, estando sujeitos apenas a obrigações de transparência (art. 50.º/4). No entanto, quando esse sistema seja configurado ou preparado para gerar pornografia de menores, estará em causa a prática do crime previsto no art. 176.º do Código Penal([87]).

([86]) Muito críticos dessa abordagem inicial veja-se Michael Veale/Frederik Zuiderveen Borgesius, *ob. cit.*, pp. 98-99: "*In briefings on the prohibitions, the Commission has presented an example for each. They border on the fantastical* (…) *A cynic might feel the Commission is more interested in prohibitions' rhetorical value than practical effect*".

([87]) Curiosamente o mesmo não acontecerá com o chamado "face swap porn" de adultos. Em Portugal essa prática não tem, à data, enquadramento penal evidente. Para menores o art. 176.º basta-se com a existência de uma "representação realista de menor", independetemente de estar em causa uma falsificação. No caso de um adulto, é difícil dizer que há uma ofensa à privacidade (visto que não

6.1. Manipulação e exploração de vulnerabilidades

Um pressuposto da liberdade em geral, em especial a liberdade de pensamento, de escolha e de expressão é a adequada perceção/representação da realidade. A autonomia privada e o livre desenvolvimento da personalidade requerem-no. Por isso, o sistema jurídico nacional torna anuláveis os negócios jurídicos celebrados com base em vícios da vontade, proíbe e pune as práticas comerciais desleais e a publicidade enganosa. A autonomia da vontade, como reflexo da dignidade da pessoa humana, reflete-se também na proibição de experimentação em pessoas e na exigência de consentimento livre e informado, em especial no caso de limitação voluntária de direitos de personalidade.

Alguns sistemas de IA têm potencial de manipulação e engano, interferindo na livre formação de pensamentos, opiniões e escolhas([88]). Nesse sentido o texto do art. 5.º/1/*a*) do Regulamento proíbe "*A colocação no mercado, a colocação em serviço ou a utilização de um sistema de IA que empregue técnicas subliminares que contornem a consciência de uma pessoa, ou técnicas manifestamente manipuladoras ou enganadoras, com o objetivo ou o efeito de distorcer substancialmente o comportamento de uma pessoa ou de um grupo de pessoas prejudicando de forma considerável a sua capacidade de tomar uma decisão informada e levando, assim, a que tomem uma decisão que, caso contrário, não tomariam, de uma forma que cause ou seja razoavelmente suscetível de causar danos significativos a essa ou a outra pessoa, ou a um grupo de pessoas*". Trata-se de uma formulação que utiliza conceitos indeterminados e linguagem qualificada ("manifestamente", "substancialmente", "de forma considerável", "razoavelmente suscetível"), aumentando as dificuldades do intérprete-aplicador([89]). Ape-

---

houve real captação de imagens verdadeiras), podendo colocar-se o problema numa dupla perspetiva: ofensa à imagem e ao bom nome. No entanto, não parece enquadrar-se nos crimes de difamação (art. 180.º Código Penal), nem na devassa através de meio de comunicação social, da Internet ou de outros meios de difusão pública generalizadao (art. 193.º), nem nas gravações e fotografias ilícitas (art. 199.º Código Penal). O art. 5.º/1/*b*) da Diretiva 2024/1385 relativa ao combate à violência contra as mulheres e à violência doméstica parece prever a criminalização desta prática. Entretanto pode igualmente ponderar-se a aplicação dos tipos penais da proteção de dados (arts. 46.º e ss. da Lei 58/2019 de 8 de agosto).

([88]) A Convenção CoE no art. 5.º/2 refere-se mesmo à liberdade de formação de opinião. Para uma visão crítica deste artigo e dos conceitos que lhe subjazem, cf. Mark Leiser, 'Psychological Patterns and Article 5 of the AI Act: AI-Powered Deceptive Design in the System Architecture and the User Interface' Journal of AI Law and Regulation, Vol. 1(1) (2024), pp. 5-23.

([89]) Existe aliás uma controvérsia sobre o fundamento científico da influência subliminar (isto é, aquela que fica abaixo do limiar de perceção consciente) do sujeito. No sentido, da sua demonstração e consequente necessidade de regulação, cf. Rostam J. Neuwirth, 'Prohibited artificial intelligence

sar disso, a existência desses qualificativos parece indiciar um grau elevado de exigência — a proibição não abrangerá toda e qualquer técnica publicitária ou prática oculta ou enganosa([90]). Na verdade, creio que os critérios do Direito da Publicidade e da proteção do consumidor serão menos exigentes, ou seja, certas condutas qualificadas como publicidade e/ou práticas comerciais agressivas ou enganosas não caberão no art. 5.º/1/*a*) do Regulamento. Nesses casos, o sistema de IA não será proibido, mas as atividades em causa, independentemente do recurso a um sistema informático, serão abrangidas pelas regras já existentes.

Por sua vez, o art. 5.º/1/*b*) proíbe "*a colocação no mercado, a colocação em serviço ou a utilização de um sistema de IA que explore vulnerabilidades de uma pessoa singular ou de um grupo específico de pessoas devidas à sua idade, incapacidade ou situação socioeconómica específica, com o objetivo ou o efeito de distorcer substancialmente o comportamento dessa pessoa ou de uma pessoa pertencente a esse grupo de uma forma que cause ou seja razoavelmente suscetível de causar danos significativos a essa ou a outra pessoa*". Estas condutas, no contexto dos negócios jurídicos, já são proibidas pelo sistema jurídico, especialmente nas relações de consumo. Também aqui parece que os qualificativos utilizados e limitação a certas caraterísticas tornam a norma do Regulamento mais exigente que a legislação em vigor e, nessa medida, o Regulamento terá um impacto reduzido([91]).

Pense-se, por exemplo, nos sistemas de preços personalizados que tenham em conta que um potencial cliente está numa situação que o dispõe a pagar um preço mais elevado (v.g. o telemóvel tem pouca bateria ou os dados biométricos indicam desidratação ou fadiga). Creio que essas situa-

---

practices in the proposed EU artificial intelligence act (AIA)' Computer Law & Security Review, 48 (2023), que aliás propõe a utilização do termo *transliminar* (em vez de subliminar), uma vez que a manipulação se dá habitualmente entre o plano da consciência e o da inconsciência.

([90]) Fala-se a este propósito de *dark patterns* (formas de interface com utilizadores que promovem uma ação ou escolha que os utilizadores provavelmente não fariam). O DSA, no considerando 67 define "padrões obscuros", como "*práticas que distorcem ou prejudicam de forma substancial, intencional ou de facto, a capacidade dos destinatários do serviço de fazerem escolhas ou decisões autónomas e informadas*". Sobre o tema *vide* Harry Brignull, *Deceptive patterns — exposing the tricks tech companies use to control you* (Testimonium Ltd 2023); Inge Graef, 'The EU Regulatory Patchwork for Dark Patterns: An Illustration of an Inframarginal Revolution in European Law?' (2023) <https://ssrn.com/abstract=4411537> e Mark Leiser/Cristiana Santos, 'Dark Patterns, Enforcement, and the Emerging Digital Design Acquis: Manipulation beneath the Interface' European Journal of Law and Technology, Vol. 15, N.º 1 (2024): BILETA Special Issue.

([91]) Rostam J. Neuwirth, *ob. cit.*, pp. 6-7. Aparentemente em sentido próximo, cf. Vera Lúcia Raposo, 'Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence' International Journal of Law and Information Technology, Vol. 30 (2022), pp. 93-94.

ções não caberiam no âmbito deste artigo do Regulamento, não obstante poderem ser consideradas ilícitas com outro fundamento(92).

6.2. Pontuação (*scoring*) social

A prática de pontuação (*scoring*), isto é, a atribuição de valores numéricos a indivíduos, apesar de não estar definida, tem já um enquadramento no RGPD, na medida em que envolve quase sempre definição de perfis e frequentemente também uma decisão automatizada(93). Essa operação é muitas vezes necessária para que os sistemas computacionais possam desempenhar as suas funções. No entanto, levanta preocupações, sobretudo à luz do que certos países, como a Índia e a China vieram a implementar: sistemas de classificação social, que tomam em conta a generalidade dos comportamentos dos cidadãos para atribuição de uma classificação que determine ou influencie o seu tratamento em vários contextos(94).

O Regulamento apenas proíbe sistemas de IA "*para avaliação ou classificação de pessoas singulares ou grupos de pessoas durante um certo período com base no seu comportamento social ou em características de personalidade ou pessoais, conhecidas, inferidas ou previsíveis, em que a classificação social conduza a uma das seguintes situações* (…) *tratamento prejudicial ou desfavorável* (…) *em contextos sociais não relacionados com os contextos nos quais os dados foram originalmente gerados ou recolhidos ou tratamento prejudicial ou desfavorável de* (…) *que seja injustificado ou desproporcionado face ao seu comportamento social ou à gravidade do mesmo"* [art. 5.º/1/*c*)]. Está em causa o chamado *social scoring*, ou seja, a avaliação global dos comportamentos de uma pessoa sin-

---

(92) A Diretiva (UE) 2019/2161 do Parlamento Europeu e do Conselho de 27 de novembro de 2019 que altera a Diretiva 93/13/CEE do Conselho e as Diretivas 98/6/CE, 2005/29/CE e 2011/83/UE do Parlamento Europeu e do Conselho a fim de assegurar uma melhor aplicação e a modernização das regras da União em matéria de defesa dos consumidores, transposta pela Lei n.º 10/2023, de 3 de março, veio impor a obrigação de informação de que os preços são determinados de forma automatizada [agora constante do art. 4.º/1/*l*) do Decreto-Lei n.º 24/2014, de 14 de fevereiro]. Sobre o tema, numa perspetiva sobretudo económica, cf. Mateusz Grochowski/Fabrizio Esposito/Antonio Davola, *Price 'Personalization vs. Contract Terms Personalization: Mapping the Complexity* (2024), *in* <https://ssrn.com/abstract=4791124>.

(93) *Vide* Pinto Monteiro/Sandra Passinhas, 'Definição algorítmica de perfis e não discriminação dos consumidores' RLJ, n.º Vol. 152 (4041) (2023), pp. 368-379.

(94) Cf. Ralph Schroeder, 'Aadhaar and the Social Credit System: Personal Data Governance in India and China' International Journal of Communication, Vol. 16 (2022), pp. 2370-2386.

gular([95]). Em contrapartida, sistemas de IA que façam *scoring* mais restrito, como aqueles que se dedicam a classificação de crédito (*credit scoring*), avaliação de solvabilidade ou avaliações de risco e na fixação de preços de seguros de vida ou de saúde serão classificados como sendo de risco elevado [Anexo III, 5/*b*) e *c*)]([96]). Por último, sistemas que façam *scoring* para efeitos de deteção de fraude financeira ou para fixação de preços num seguro automóvel nem sequer estarão abrangidos pelo Regulamento. Também neste contexto, o que determina a classificação de risco do sistema é a finalidade da avaliação quantitativa e não a prática de *scoring*.

Como se assinalou, o facto de haver *scoring* estará habitualmente associado a uma decisão automatizada, a qual, quando implicar tratamento de dados pessoais e produzir efeitos na esfera jurídica ou afetar significativamente o titular de dados pessoais, poderá, à partida, ser proibida por este nos termos do art. 22.º RGPD([97]). No entanto, é importante notar que o art. 22.º do RGPD só se aplica a decisões *totalmente* automatizadas([98]). Assim, pelo menos no caso dos sistemas de alto risco, em que o regulamento impõe supervisão humana (art. 14.º do Regulamento), poder-se-á escapar à aplicação dessa norma do RGPD.

([95]) Nizan Geslevich Packin, 'Disability Discrimination Using Artificial Intelligence Systems and Social Scoring: Can We Disable Digital Bias?' Journal of International Comparative Law (2021), p. 496: "*Social scoring, however, attempts to systematically rate people in their entirety (and not just their creditworthiness) based on social, reputational and even behavioural features (as opposed to credit history)*". Sobre o fenómeno veja-se Danielle Keats Citron/Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' Washington Law Review 89 (2014), pp. 1-33.

([96]) Vera Lúcia Raposo, 'Ex machina… *cit.*, p. 94, aponta que a referência a "um certo período de temo" excluirá a pontuação episódica.

([97]) No acórdão C-634/21, *Schufa*, (EU:C:2023:957), §44-46 o TJ adotou um conceito amplo de decisão, dizendo que um cálculo de solvabilidade (*credit score*) se qualificava como tal.

([98]) A norma exige "*três requisitos cumulativos, a saber, primeiro, deve existir uma «decisão», segundo, essa decisão deve ser «tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis», e, terceiro, deve produzir «efeitos jurídicos na [esfera jurídica do titular dos dados]» ou afetá-lo «significativamente de forma similar»*". (C-634/21, *Schufa*, §43). As EDPB, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (2018), p. 21, assinalam que uma intervenção humana meramente simbólica não basta. Uma decisão não se considera totalmente automatizada quando existem medidas organizativas que asseguram um envolvimento humano substancial e estruturado. Na jurisprudência, cf. decisão do Rechtbank Amsterdão de 11.III.2021 (ECLI:NL:RBAMS:2021:1018,) em que estava em causa a exigência de um consenso entre várias pessoas), decisão do Rechtbank Den Haag, de 11.II.2021 (NL:RBDHA:2020:1013) em que se previa um direito de veto e decisão do Bundesverwaltungsgericht austríaco de 18.XII.2020 (AT:BVWG:2020:W256.2235360.1.00) em que existiam formações e linhas orientadoras para lidar com a recomendação produzida pelo sistema. Para mais jurisprudência cf. Sebastião Barros Vale/Gabriela Zanfir-Fortuna, *Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities* (Future of Privacy Forum 2022).

6.3. Identificação e classificação biométrica, incluindo deteção de sentimentos

Os sistemas de identificação biométrica, em especial reconhecimento facial e reconhecimento de emoções geraram uma controvérsia significativa durante o processo legislativo. Estes sistemas constituem desde logo um vetor de ataque à privacidade e liberdade individuais, com um elevado potencial discriminatório([99]). Nesse sentido, empresas como a Clearview.AI, que se dedicavam a fazer *scraping* (i.e., extração automática de dados) sistemático da Internet (em especial de redes sociais) para gerar uma base de dados de reconhecimento facial tinham sido já sancionadas por violação do RGPD([100]). Em todo o caso, agora o Regulamento proíbe expressamente esta prática [art. 5.º/1/*e*)].

De igual modo, o uso de sistemas de reconhecimento de emoções tem sido atacado desde logo nos seus fundamentos técnicos. Argumenta-se que as expressões são variáveis a nível individual e dependem do contexto social e cultural, pelo que estes sistemas não funcionam corretamente (estaremos no domínio da pseudociência). Além disso, têm um elevado potencial discriminatório([101]). Paradoxalmente, o Regulamento só proíbe a utilização destes sistemas de reconhecimento de emoções no contexto laboral e do ensino([102]). Nos restantes casos os sistemas de reconhecimento de emoções são considerados sistemas de risco elevado [Anexo III/1/*c*)]. A referência a "local" tem de ser habilmente entendida,

---

([99]) Rui Soares Pereira, 'Sobre o uso de sistemas de identificação biométrica (e de tecnologias de reconhecimento facial) para fins de segurança pública: reflexões a propósito da proposta de Regulamento Europeu sobre a Inteligência Artificial' RFDUL (2022), pp. 846-848, referido também os receios quanto à fiabilidade. Em contexto penal, um falso positivo poderá revelar-se especialmente danoso para um cidadão.

([100]) A empresa foi sujeita a coimas de 20 milhões de euros em França (2021, tendo havido também uma sanção pecuniária compulsória no valor de cinco milhões em 2023), Grécia (2022) e Itália (2022). Em 2023 autoridade austríaca também considerou a atividade desta empresa contrária ao RGPD, mas não aplicou qualquer coima ou outra medida. Em 2021 a autoridade de controlo sueca aplicou uma coima às entidades policiais por terem recorrido aos serviços da Clearview. Em contrapartida, no Reio-Unido a mesma empresa conseguiu, numa decisão judicial de 17.X.2023, anular a coima aplicada, com base numa questão de jurisdição e de direito aplicável, sobretudo à luz do *Brexit* — [2023] UKFTT 00819 (GRC).

([101]) Cf. considerando 44.

([102]) A proibição de sistemas de detação de sentimentos não abrange "*os sistemas de IA colocados no mercado exclusivamente por razões médicas ou de segurança, como os sistemas destinados a utilização terapêutic*a". Isto poderá levantar dúvidas no casos em que se utilizem sistemas de segurança ou por razões médicas no contexto de trabalho ou de ensino. Nesses casos, parece-me que a utilização deve ser admitida.

uma vez que tanto o ensino como a atividade laboral pode ser feita à distância/remotamente e nem por isso tais situações deixarão de estar abrangidas pela proibição.([103])

Por outro lado, a própria noção de reconhecimento de emoções tem de ser lida restritivamente. Nesse sentido o considerando 18 explica:

> *O conceito refere-se a emoções ou intenções como a felicidade, a tristeza, a raiva, a surpresa, a repugnância, o embaraço, o entusiasmo, a vergonha, o desprezo, a satisfação e o divertimento. Não inclui estados físicos, como dor ou fadiga, incluindo, por exemplo, sistemas utilizados para detetar o estado de fadiga dos pilotos ou motoristas profissionais para efeitos de prevenção de acidentes. Também não inclui a mera deteção de expressões, gestos ou movimentos rapidamente visíveis, a menos que sejam utilizados para identificar ou inferir emoções. Essas expressões podem ser expressões faciais básicas, tais como franzir a testa ou sorrir, ou gestos como o movimento das mãos, dos braços ou da cabeça, ou características da voz de uma pessoa, como levantar a voz ou sussurrar*.

Se é certo que a identificação biométrica em espaços públicos pode servir para a obtenção de fins louváveis (encontrar pessoas desaparecidas ou fugitivos), o seu funcionamento implica a compressão da privacidade dos cidadãos e a criação de um estado de vigilância constante, intolerável numa democracia com os valores europeus([104]). Nesse sentido, em 2023, numa decisão unânime, o TEDH confirmou o caráter ilícito (por violação do art. 8.º da CEDH) da utilização de tecnologia de reconhecimento facial para identificar, localizar e deter um individuo num processo contraordenacional([105]).

A solução adotada no art. 5.º/1/*h*) do Regulamento vai no sentido de proibir a utilização destes sistemas de "*identificação biométrica à distância em «tempo real» em espaços acessíveis ao público para efeitos de aplicação da lei*", exceto quando estritamente necessário para um de três fins:

([103]) Os sistemas de entrevista automática colocam um problema de qualificação. À luz do Anexo III, 4 parece que deverão ser vistos como sistemas de risco elevado. No entanto, caso estes sistemas incluam uma componente de deteção de emoções, parece-me que a noção de "local de trabalho" pode ser interpretada de modo a abranger também o recrutamento. Sobre estes sistemas, o seu mau funcionamento e potencial discriminatório, cf. Ifeoma Ajunwa, 'Automated video interviewing as the new phrenology' Berkeley Technology Law Journal, Vol. 36 (2021), pp. 1173-1225.

([104]) Esta matéria já é regulada pela Diretiva 2016/680, transposta em Portugal pela Lei n.º 59/2019, de 8 de Agosto. Sobre o tema, com uma perspectiva comparativa, *vide* Maria Luiza Mezzomo, *O uso do Reconhecimento Facial na Investigação Criminal e na Segurança Pública* (Almedina 2024) e Vera Lúcia Raposo 'Look at the camera and say cheese': the existing European legal framework for facial recognition technology in criminal investigations' Information & Communications Technology Law, 33(1) (2024), pp. 1-20.

([105]) *Glukhin v. Russia*, 11519/20 (decisão de 4.VII.2023).

i) busca de pessoas desaparecidas ou busca seletiva de vítimas específicas de rapto, tráfico de seres humanos ou exploração sexual de seres humanos;

ii) prevenção de uma ameaça específica, substancial e iminente à vida ou à segurança física de pessoas singulares ou de uma ameaça real e atual ou real e previsível de um ataque terrorista; e

iii) localização ou identificação de uma pessoa suspeita de ter cometido uma infração penal referida no Anexo II e punível no Estado-Membro em causa com pena ou medida de segurança privativa de liberdade de duração máxima não inferior a quatro anos.

Nesses casos o art. 5.º/2 impõe a realização de uma avaliação de impacto sobre direitos fundamentais (art. 27.º) e registo do sistema (art. 49.º) e o art. 5.º/4 obriga a que as autoridades de fiscalização do mercado pertinente e de proteção de dados sejam notificadas desta utilização.

Deve sublinhar-se que a identificação biométrica à distância para outros fins ou em diferido não está proibida(106), sendo geralmente classificada como uma utilização de risco elevado, exceto no caso de simples sistemas de reconhecimento e verificação de identidade [Anexo III, 1, *a*)](107).

O Regulamento trata também da *categorização* biométrica, que se distingue da *identificação* biométrica. Enquanto na identificação o objetivo é determinar quem é que a pessoa é a partir de determinadas caraterísticas físicas, psicológicas ou comportamentais (os dados biométricos — art. 3.º/34); a categorização biométrica visa classificar o sujeito — saber se alguém tem uma dada caraterística(108). Assim, na *identificação* biomé-

---

(106) O art. 26.º/10 prevê que, no caso de sistema de identificação biométrica à distância em diferido (definidos no art. 3.º/43, em contrapasição com os "em tempo real" definidos no art. 3.º/42), "o *responsável pela implantação* (…) *deve solicitar uma autorização, prévia ou sem demora injustificada e no prazo máximo de 48 horas, a uma autoridade judiciária ou uma autoridade administrativa cuja decisão seja vinculativa e esteja sujeita a controlo jurisdicional, para a utilização desse sistema"*. Caso a autorização seja rejeitada a utilização deve cessar e os dados devem ser destruídos. Proíbe-se também a sua utilização indiscriminada ("de forma não seletiva") e admite-se que os Estados-Membros venham a adotar legislação mais restritiva.

(107) Cf. considerandos 15, 17 e 52 e a definição de verificação biométrica (art. 3.º/36).

(108) Sistema de categorização biométrica é definido no art. 3.º/40 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*" (para exemplos de categorização acessória, cf. considerando 16), já identificação biométrica diz respeito ao "*reconhecimento automatizado de características humanas físicas, fisiológicas, comportamentais ou psicológicas para efeitos de determinação da identidade de uma pessoa singular, comparando os dados biométricos dessa pessoa com os dados biométricos de pessoas armazenados numa base de dados*" (art. 3.º/35).

trica, a partir da minha cara ficam a saber que eu sou o Nuno Sousa e Silva, na *categorização* biométrica, a partir da minha forma de andar determinam se tenho risco de desenvolver Alzheimer ou, pela análise do meu rosto, avaliam se serei um perigoso anarcossindicalista.

De acordo com o Regulamento os sistemas de categorização biométrica "*que classifiquem individualmente as pessoas singulares com base nos seus dados biométricos para deduzir ou inferir a sua raça, opiniões políticas, filiação sindical, convicções religiosas ou filosóficas, vida sexual ou orientação sexual*" estão proibidos [art. 5.º/1/*g*)](109). Assim, não serão admissíveis sistemas como a controversa rede neuronal que alegadamente detetava a orientação sexual das pessoas a partir de fotografias(110). Há, no entanto, uma ressalva para processamento e categorização de dados biométricos no domínio da aplicação da lei, que continua a ser admitida(111). Por outro lado, os "*sistemas de IA concebidos para serem utilizados para categorização biométrica, de acordo com atributos ou características sensíveis ou protegidos com base na inferência desses atributos ou características*" não são proibidos, sendo classificados como sistemas de risco elevado [Anexo III, 1, *b*)].

6.4. Policiamento preditivo

A definição de perfis tem como premissa a repetibilidade e padronização de comportamentos. Parte-se da ideia de que o passado se repete no futuro ("cesteiro que faz um cesto, faz um cento…") e de que há certas caraterísticas com capacidade preditiva. A aplicação destas técnicas no contexto criminal levanta especiais preocupações, desde logo atendendo às potenciais consequências de um erro ou injustiça e à existência da presunção de inocência(112).

---

(109) Esta previsão pode ser criticada por ser demasiado restrita nas "categorias protegidas". Como explica Catarina Botelho, 'Algoritmos discriminatórios' *in* Anabela Miranda Rodrigues/Susana Aires de Sousa (eds.), *I Congresso… cit.*, pp. 29-30, em geral as categorias suspeitas no direito antidiscriminação são mais vastas e as listas tendem a ser não taxativas.

(110) O controverso estudo original foi repetido por John Leuner, 'A replication study: Machine learning models are capable of predicting sexual orientation from facial images' *arXiv:1902.10739* (2019) que sustenta que estes modelos atendem a outros factores e não à fisionomia/estrutura facial.

(111) Cf. considerando 30.

(112) Como se pode ler no considerando 42: "*Em conformidade com a presunção de inocência, as pessoas singulares na União deverão ser sempre avaliadas em função do seu comportamento real. As pessoas singulares nunca poderão ser julgadas com base no comportamento previsto pela IA*

Assim, o Regulamento proíbe prática de policiamento preditivo *(predictive policing)* que recorram a sistemas de IA que avaliem 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*" [art. 5.º/1/*d*)]. No entanto, "*esta proibição não se aplica aos sistemas de IA utilizados para apoiar a avaliação humana do envolvimento de uma pessoa numa atividade criminosa, que já se baseia em factos objetivos e verificáveis diretamente ligados a uma atividade criminos*a". Por outras palavras, o sistema deve atender ao comportamento concreto e traços particulares de uma pessoa específica, e não apenas à sua pertença a determinadas categorias ou grupos. Esta exceção reconhece a utilidade potencial da IA no contexto da investigação e prevenção criminal, assegurando, simultaneamente, que a avaliação se baseia em dados concretos e não *exclusivamente* na definição (necessariamente especulativa) de perfis.

Na verdade, o policiamento preditivo pode ser orientado para prever crimes, prever ou identificar criminosos e/ou prever ou identificar potenciais vítimas de crime([113]). A generalidade desses sistemas, quando não assente exclusivamente na definição de perfis, será abrangida pela classificação de risco elevado (Anexo III, 6). Nessa linha, o considerando 42 esclarece que a proibição do art. 5.º/1/*d*) não abrange "*sistemas de IA que utilizam análises de risco para avaliar a probabilidade de fraude financeira por parte de empresas com base em transações suspeitas, ou ferramentas de análise de risco para prever a probabilidade de localização de estupefacientes ou mercadorias ilícitas pelas autoridades aduaneiras, por exemplo, com base em rotas de tráfico conhecidas*".

Um exemplo conhecido de sistema de IA para efeitos preditivos no contexto penal, é o sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), utilizado nalguns tribunais norte-americano para calcular o risco de reincidência e, com base nisso, definir a moldura penal([114]). Ferramentas como essa, desde que não assen-

---

com base exclusivamente na definição do seu perfil, nos traços ou características da sua personalidade, como a nacionalidade, o local de nascimento, o local de residência, o número de filhos, o nível de endividamento ou o tipo de automóvel que têm, sem que exista uma suspeita razoável do seu envolvimento numa atividade criminosa com base em factos objetivos verificáveis, e sem uma avaliação humana dos mesmos".

([113]) Cf. Walter Perry, *et al.*, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (RAND Corporation 2013).

([114]) Objeto de grande discussão académica e judicial. No conhecido acórdão *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S. Ct. 2290 (2017) o Supremo Tribunal do Wiscon-

tem exclusivamente na definição de perfis, não são cobertas pela proibição, sendo antes consideradas sistemas de IA de risco elevado [Anexo III, 6 *d*) e *e*) e 8].

7. Sistemas de risco elevado

7.1. Qualificação

A definição de sistemas de risco elevado é feita através da remissão para dois Anexos([115]).

O Anexo I inclui legislação relativa a determinadas categorias de produtos (como brinquedo, veículos, explosivos, elevadores ou dispositivos médicos) e, de acordo com o art. 6.º/1. quando os sistemas de IA sejam utilizados como componentes de segurança nestes produtos (ou os sistemas de IA sejam eles mesmos produtos)([116]) sujeitos a uma obrigação de avaliação de conformidade estará em causa um sistema de risco elevado([117]).

É importante notar que dentro do Anexo I existem duas seções, a secção A, que diz respeito a um conjunto de legislação ao abrigo do "Novo quadro legislativo" (*New Legislative Framework*)([118]), e a secção B, relativa a legislação anterior. Ora, de acordo com o art. 2.º/2, estes últimos, da secção B, estão praticamente excluídos do âmbito de aplicação do Regulamento. Além disso, os requisitos do art. 6.º(1) são limitados pela interação

---

sin rejeitou o recurso de um indivíduo que tinha sido considerado pelo referido *software* como tendo um risco elevado de reincidência e, assim, condenado a 6 anos de prisão. De acordo com o tribunal o *due process* não tinha sido violado apesar de a sentença ter sido determinada com recurso ao COMPAS, cujo algoritmo e modo de funcionamento não são conhecido. A natureza discriminatória deste sistema foi objeto de uma reportagem por parte da ProPublica.

([115]) A razão para esta definição ser feita por remissão consiste na flexibilização da atualização destes anexos em processo simplificado (atos delegados da Comissão Europeia) previstos nos arts. 6.º/6, /7 e /8 e 7.º

([116]) Pode acontecer nomeadamente no caso de brinquedos ou dispositivos médicos.

([117]) "Componente de segurança" é definido no art. 3.º/14 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"*. Esta definição é ampla, mas parece-me que a ideia de colocação em risco de saúde e segurança deve ser vista com base num critério de normalidade/previsibilidade (no mesmo sentido, cf. considerando 46). Sublinhando os eventuais conflitos entre a noção de componente de segurança do Regulamento e aquela que consta de legislação especial, cf. Emilija Leinarte, *ob. cit.*, pp. 271-273. Uma outra dúvida diz respeito a saber se o art. 6.º/1 só se aplica a produtos acabados (cf. Emilija Leinarte, *ob. cit.*, pp. 270-271).

([118]) Cf. *supra* nota 31.

com a legislação específica, nomeadamente no que diz respeito à definição de produto e de componente de segurança([119]).

Por sua vez o art. 6.º/2 remete para o anexo III em que se especificam determinadas utilizações de um sistema (autónomo) de IA([120]), como identificação biométrica, gestão de infraestruturas críticas, admissão e classificação em estabelecimentos de ensino, entrevistas de emprego, monitorização de trabalhadores, acesso e utilização de serviços essenciais (públicos e privados), utilização no controlo de fronteiras, em contexto judicial ou por entidades policiais. Como sublinha Philip Hacker([121]), mais importante do que o contexto de utilização é a finalidade — um sistema utilizado para operação médicas ou triagem não tem o mesmo risco do que um sistema que gere as marcações de consultas médicas.

É igualmente importante ler com atenção e à luz dos considerandos do Regulamento as várias hipóteses, estando igualmente previsto que a Comissão venha a adotar orientações que especifiquem a aplicação prática deste artigo "*juntamente com uma lista exaustiva de exemplos práticos de casos de utilização de sistemas de IA de risco elevado e de risco não elevado*" (art. 6.º/5)([122]). O sistema funciona com autoclassificação, ou seja, cada operador é que irá determinar a classificação de risco do seu sistema

A classificação de risco é feita com base na utilização prevista, mas existem algumas ressalvas. Por exemplo os sistemas de identificação biométrica à distância são geralmente de risco elevado, mas está prevista uma exclusão para sistemas de verificação de identidade [Anexo III(1)(*a*)]. De igual modo, os sistemas de avaliação de solvabilidade de pessoas singulares ou que fazem uma classificação de crédito (*credit scoring*) são sistemas de risco elevado, exceto quando esses sistemas sejam usados para deteção de fraude financeira [Anexo III(5)(*b*)]([123]).

Além de exceções específicas, está prevista uma derrogação. De acordo com o art. 6.º/3 é possível afastar a classificação de risco elevado

---

([119]) Emilija Leinarte, *ob. cit.*, p. 274: "[art. 6(1)] *covers a limited group of AI systems due to significant sectoral carve outs, limitations to sector-specific definitions of products and safety components of a product and a significant harm condition*".

([120]) Emilija Leinarte, *ob. cit.*, p. 275 sublinha que o art. 6.º/2 só se aplica a *stand-alone systems*.

([121]) *AI Regulation in Europe: From the AI Act to Future Regulatory Challenges* (2023) arXiv:2310.04072, p. 7.

([122]) Este artigo surpreende o jurista médio que ficará à espera de uma "lista exaustiva" num tema em vertiginosa evolução… Proeza que possivelmente só está ao alcance da Comissão Europeia. Ou então trata-se de um erro de tradução (até porque a versão inglesa utiliza a palavra "*comprehensive*" e não "*exhaustive*").

([123]) Sobre esses sistemas e o seu enquadramento jurídico, cf. Diogo Morgado Rebelo, *ob. cit.*

em relação a um sistema cujo uso previsível conste do anexo III "*se não representar um risco significativo de danos para a saúde, a segurança ou os direitos fundamentais das pessoas singulares, nomeadamente se não influenciarem de forma significativa o resultado da tomada de decisões*" e desde que não realize definição de perfis de pessoas singular (último parágrafo deste art. 6.º/3)([124]).

O art. 6.º/3 apresenta circunstâncias em que esses sistemas de IA, não obstante terem uma finalidade prevista no Anexo III, não representarão risco significativo:

a) quando desempenharem uma tarefa processual restrita;

b) quando se destinem a melhorar uma atividade humana previamente concluída;

c) quando visem detetar padrões de tomada de decisões ou desvios em relação a padrões de tomada de decisões anteriores e não se destinem a substituir nem influenciar uma avaliação humana previamente concluída; ou

d) quando o sistema de IA se destine a executar (apenas) uma tarefa preparatória.

Para que o sistema de IA, apesar da sua finalidade, não seja considerado de risco elevado, basta preencher uma dessas alíneas e que o sistema não realize definição de perfis (tal como definida no art. 4.º/4 RGPD).

O considerando 53 apresenta alguns exemplos de sistemas deste tipo, nomeadamente sistemas de IA destinados a melhorar a linguagem utilizada em documentos redigidos anteriormente, por exemplo em relação ao tom profissional, ao estilo académico ou ao alinhamento do texto com uma determinada mensagem de marca, sistemas que são utilizados para verificar *ex post* se um professor se pode ter desviado do seu padrão habitual de atribuição de notas, de modo a assinalar potenciais incoerências ou anomalias, soluções inteligentes para o tratamento de ficheiros que incluem várias funções, como a indexação, a pesquisa, o processamento de texto e de voz ou a ligação de dados a outras fontes de dados, ou sistemas de IA utilizados para a tradução de documentos.

Em qualquer caso, quem queira invocar essa derrogação tem de documentar essa avaliação (art. 6.º/4), proceder ao registo do seu sistema

---

([124]) Isto não quer dizer que todos os sistemas que realizam definições de perfis devam ser classificados como sendo de risco elevado. A única decorrência desta previsão é que um sistema que realize definição de perfis e seja classificado como sendo de risco elevado não pode beneficiar da isenção.

(art. 49.º/2), sujeitando-se a que uma autoridade de fiscalização do mercado venha a discordar, exigindo medidas corretivas (art. 80.º).

7.2. Regulação

De forma simplificada podemos dizer que em relação aos sistemas de risco elevado, o Regulamento exige que estes sejam bem feitos, adequadamente mantidos e controlados e que haja documentação adequada capaz de comprovar o respeito pelas normas do Regulamento.

Nesse sentido, em relação a sistemas de aprendizagem automática (*machine learning*) impõem-se requisitos de qualidade de dados, nomeadamente em termos de representatividade, aplicação de medidas para deteção e mitigação de enviesamentos (*biases*) (art 10.º). O n.º 5 do art. 10.º cria mesmo uma nova base de licitude para processamento de dados sensíveis (a acrescer às que constam do art. 9.º RGPD) estabelecendo que, em condições particulares, será possível tratar categorias especiais de dados pessoais "*para assegurar a deteção e a correção de enviesamentos em relação aos sistemas de IA de risco elevado*"([125]). Por outro lado, a generalidade das previsões do Regulamento legitimará o processamento de dados não sensíveis uma vez que este ocorrerá para o cumprimento de obrigações legais [art. 6.º/1/*c*) RGPD]([126]).

A nível dos sujeitos, os prestadores são os responsáveis pela conceção e implementação dos requisitos dos arts. 8.º a 15.º (art. 16.º), bem como por garantir a existência de um sistema de gestão de qualidade (art. 17.º), pela conservação da documentação pelo período de 10 anos subsequentes à data de colocação no mercado ou de colocação em serviço do sistema (art. 18.º), pela manutenção dos *logs* (art. 19.º). Está também previsto um dever de colaboração com autoridades competentes (arts. 20.º/2, 21.º e 73.º), a adoção de medidas corretivas (art. 20.º/1), e um acompanhamento pós-comercialização (art. 72.º). Esse acompanhamento inclui um dever de informação das autoridades no caso de um incidente grave (art. 73.º), definido no art. 3.º/49 como "*qualquer incidente ou anomalia num sistema de IA que, direta ou indiretamente, tenha alguma das*

([125]) Isto poderá tornar difícil aplicar medidas de mitigação de vieses a sistemas que não sejam de risco elevado uma vez que para esses não existirá base de licitude para o tratamento de dados sensíveis (Michael Veale/Frederik Zuiderveen Borgesius, *ob. cit.*, p. 103). Também está previsto um tratamento adicional de dados pessoais em determinadas condições de salvaguarda de interesse público (art. 59.º).

([126]) Sendo que não existe fundamento equivalente para os dados sensíveis.

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 ponto de vista mais burocrático, além de um dever de documentação e manutenção de registos, os prestadores de sistemas de IA de risco elevado estão obrigados a identificar-se como tal [art. 16.º/*b*)] e a seguir um procedimento de avaliação de conformidade (art. 43.º)([127]), incluindo a elaboração de uma declaração de conformidade (art. 47.º), a aposição da marca CE (art. 48.º) e o registo do sistema de risco elevado (arts. 49.º e 71.º)([128]).

Embora os deveres mais importantes recaiam sobre os *fornecedores* de sistemas de IA, os seus *utilizadores* ("responsáveis pela implementação") também estão sujeitos a um conjunto de obrigações, previstas no art. 26.º. Na medida em que controlam o sistema, os responsáveis pela implementação terão de respeitar as instruções de utilização do sistema, garantir a sua supervisão humana e a qualidade e adequação dos dados de entrada, colaborar com as autoridades, manter os registos de funcionamento do sistema e informar as pessoas singulares de que estão sujeitas à utilização do sistema de IA de risco elevado.

Nalguns casos, os organismos de direito público, ou as entidades privadas que prestam serviços públicos, e as entidades bancárias e seguradoras estão obrigados a realizar uma avaliação de impacto sobre direitos fundamentais (art. 27.º). Esta avaliação não se confunde com a obrigação de realizar uma avaliação de impacto sobre a proteção de dados (*DPIA*) estabelecida no art. 35.º do RGPD, ainda que o próprio Regulamento reconheça a existência de sobreposições parciais (art. 27.º/6).

8. Obrigação de transparência para certos sistemas

O art. 50.º do Regulamento, único do Capítulo IV, lida com certos sistemas, definidos à luz da sua finalidade, impondo requisitos mínimos de trans-

---

([127]) Está prevista a derrogação desse procedimento, nomeadamente em casos de urgência (art. 46.º).

([128]) Tendo em conta os princípios do país de origem e reconhecimento mútuo, esta operação só precisa de ser feita num Estado-Membro.

parência/informação([129]). Os dois primeiros números deste artigo impõem deveres aos prestadores, enquanto os números 3 e 4 dizem respeito aos deveres dos responsáveis pela implementação desses sistemas de IA([130]). Como sublinha Tiago Sérgio Cabral([131]). estes deveres aplicam-se aos sistemas referidos no art. 50.º independentemente da sua classificação de risco.

No art. 50.º/1 disciplinam-se os sistemas de IA "*destinados a interagir diretamente com pessoas singulares*", ou seja, os chamados *chatbots* ou sistemas "conversacionais". Estes sistemas devem ser concebidos de modo que se torne claro para as pessoas singulares "*que estão a interagir com um sistema de IA, salvo se tal for óbvio do ponto de vista de uma pessoa singular razoavelmente informada, atenta e advertida, tendo em conta as circunstâncias e o contexto de utilização*"([132]).

Os sistemas de IA generativa ("*que geram conteúdos sintéticos de áudio, imagem, vídeo ou texto*") são abordados no art. 50.º/2. Prevê-se a obrigação de identificar esses conteúdos sintéticos com uma "marca de água" digital "*num formato legível por máquina e detetáveis como tendo sido artificialmente gerados ou manipulados*"([133]).

Os responsáveis pela implementação de sistemas de reconhecimento de emoções ou de categorização biométrica estão sujeitos a um dever de revelação da sua utilização (art. 50.º/3)([134]). De igual modo, quem criar

---

([129]) Os deveres de transparência/revelação previstos no art. 50.º não se aplicam quando o sistema esteja legalmente autorizado "*para detetar, prevenir, investigar ou reprimir infrações penais, sob reserva de garantias adequadas dos direitos e liberdades de terceiros*". Sobre o tema *vide* David Silva Ramalho, *Métodos Ocultos de Investigação Criminal em Ambiente Digital* (Almedina 2017).

([130]) Não é claro se os fabricante destes sistemas estão cobertos pela isenção de responsabilidade do art. 6.º do DSA. Desde logo é discutível se podemos classificar os fornecedores de modelos ou sistemas de IA de finalidade geral ou com capacidade generativa como "serviço intermediário" [conforme previsto no art. 3.º/*g*) do DSA]. O considerando 119 do AI Act parece apontar para uma avaliação casuística.

([131]) *Ob. cit.*, p. 95.

([132]) Michael Veale/Frederik Zuiderveen Borgesius, *ob. cit.*, p. 106, falam em "bot disclosure".

([133]) As soluções técnicas de implementação da "marca de água" têm de ser "*eficazes, interoperáveis, sólidas e fiáveis, na medida em que tal seja tecnicamente viável, tendo em conta as especificidades e limitações dos vários tipos de conteúdos, os custos de aplicação e o estado da arte geralmente reconhecido, tal como estiver refletido em normas técnicas pertinentes*". Esta obrigação não se aplica a ferramentas de apoio à edição (como um corretor ortográfico ou editor de fotografia) e em geral aquelas "*não alterem substancialmente os dados de entrada disponibilizados pelo responsável pela implantação ou a semântica dos mesmos*" (art. 50.º/2). O âmbito desta exclusão não é totalmente claro: se a informação é apresentada em simples tópicos e o sistema escreve as frases e estrutura a apresentação da informação, dever-se-á considerar que não há uma alteração substancial dos dados de entrada? Criticando a natureza vaga da exceção *vide* Mateusz Łabuz, 'Deep fakes and the Artificial Intelligence Act — An important signal or a missed opportunity?' Policy & Internet (2024), p. 10.

([134]) Como vimos este tipo de sistemas podem ser proibidos ou classificados como sendo de risco elevado. Em todo o caso, como assinalam Michael Veale/Frederik Zuiderveen Borgesius, *ob. cit.*, p. 107, este dever não parece acrescentar nada face ao que resulta já do RGPD.

falsificações profundas (*deepfakes*) deve "*revelar que os conteúdos foram artificialmente gerados ou manipulado*" (art. 50.º/4, 1.º parágrafo)([135]). Este dever pode ser reduzido "*sempre que os conteúdos façam parte de um programa ou obra de natureza manifestamente artística, criativa, satírica, ficcional ou análoga*", bastando que seja cumprido "*de uma forma adequada que não prejudique a exibição ou a fruição da obra*". O dever de revelação também existe no caso de notícias ("*texto publicado com o objetivo de informar o público sobre questões de interesse público*"), exceto "*se os conteúdos gerados por IA tiverem sido objeto de um processo de análise humana ou de controlo editorial e se uma pessoa singular ou coletiva for responsável editorial pela publicação do conteúdo*" (art. 50.º/4/2.º parágrafo)([136]).

9. Modelos de finalidade geral

Quando, em Abril de 2021, a Comissão Europeia apresentou a proposta de Regulamento, já existiam alguns modelos de IA com capacidades diversificadas, mas o termo "modelos fundacionais", utilizado para indicar aqueles modelos treinados com grandes quantidades de dados e com potencial para servir várias aplicações, não tinha ainda sido cunhado. Foi só em Agosto de 2021 que um *paper* de investigadores de Stanford utilizou pela primeira vez esta noção([137]). A verdadeira explosão de modelos fundacionais, que incluem os GPTs da empresa OpenAI e os concorrentes PALM., BERT e Gemini (Google), Claude (Anthropic), Luminous (Aleph Alpha), Mistral 7B e LlaMA (Meta), estes dois últimos disponibilizados em *open source*, deu-se já em 2023.

Esta tecnologia tem particularidades que são especialmente desafiantes. Por um lado, apresentam elevados custos de desenvolvimento, o que gera consideráveis barreiras à entrada. Ao contrário dos sistemas especia-

---

([135]) Para usar a expressão de Michael Veale/Frederik Zuiderveen Borgesius, *ob. cit.*, p. 108, estará em causa um "direito mínimo à realidade". No entanto, como sublinham, um entendimento teleológico desta obrigação devia excecionar utilizações em contextos nos quais não haja um risco de engano (como no caso de imagens genéricas utilizadas para efeitos de *marketing* ou apresentações). Os considerandos 132 e 133 parecem comportar essa interpretação.

([136]) O que não exige que haja um autor humano. Basta que tenha ocorrido um controlo humano do conteúdo.

([137]) Rishi Bommasani, *et al.*, *On the Opportunities and Risks of Foundation Models*, arXiv:2108.07258 [cs.LG].

lizados, para os quais o Regulamento estava inicialmente vocacionado, estes modelos têm uma capacidade de generalização e com frequência serão disponibilizados mediante interfaces de programação (APIs)([138]) para que terceiros possam otimizar e adaptar a aplicações específicas. Nesse sentido, estes modelos, como explica Andrej Karpathy([139]), aproximam-se de sistemas operativos, gerando consideráveis dependências. Estas considerações são tipicamente abordadas pelo Direito da Concorrência([140]), mas o Regulamento dedicou-lhe um capítulo. Nesse sentido prevê-se, nos arts. 89.º/2 e 93.º, uma proteção dos prestadores a jusante, isto é, aqueles que integrem um modelo ou sistema de finalidade geral no seu sistema e que ficam dependentes de um sistema de finalidade geral que não controlam.

Por outro lado, estes grandes modelos de finalidade geral são frequentemente opacos: constituem um conjunto vastíssimos de números (os chamados parâmetros e pesos de uma rede neuronal) que interagem de formas que escapam à compreensão dos seres humanos([141]). Esta falta de compreensão levanta questões preocupantes de segurança, controlo e alinhamento.

Além disso, para o desenvolvimento desses modelos recorre-se a quantidades maciças de dados, grande parte dos quais retirados da Internet e que incluem dados pessoais e dados protegidos por direitos de propriedade intelectual. Ao contrário do que se pensava inicialmente, estes modelos retêm alguns dos dados em "memória"([142]). Esta circunstância torna ainda mais complexa a avaliação da licitude dessas utilizações.

Por último, a generalidades dos modelos fundacionais têm capacidades "criativas", cabendo igualmente na categoria de IA generativa coberta pelo art. 50.º([143]).

([138]) Sobre a noção de APIs e o seu estatuto jurídico veja-se Nuno Sousa e Silva, 'Ligações perigosas? — Reflexões sobre APIs e Direito de Autor a partir do acórdão Google v. Oracle do Supremo Tribunal dos EUA', *in* Revista de Direito Intelectual 1/2022, pp. 213-227.

([139]) Esta afirmação é feita em várias conferência públicas disponíveis no Youtube. Sugiro especialmente o vídeo "[1hr Talk] Intro to Large Language Models".

([140]) Cf. Hou Liyang, 'The Essential Facilities Doctrine — What was Wrong in Microsoft?' IIC 43(4) [2012], pp. 251-271.

([141]) Sobre o seu possível enquadramento na Propriedade Intelectual, cf. Nuno Sousa e Silva, *Are AI models' weights protected databases? in* <https://copyrightblog.kluweriplaw.com/2024/01/18/are-ai-models-weights-protected-databases/>.

([142]) Milad Nasr, *et al.*, *Scalable Extraction of Training Data from (Production) Language Models*, arXiv:2311.17035 [cs.LG].

([143]) Nem todos os sistemas de IA generativa são modelos fundacionais; existem uma série de aplicações especializadas na criação de música, imagens, texto, etc. que, sendo generativas, não são fundacionais.

O Regulamento trata de todos os modelos de IA de finalidade geral (nos arts. 53.º e 54.º) e impõe deveres adicionais (no art. 55.º) para os chamados modelos de IA de finalidade geral com risco sistémico([144]). De acordo com o art. 51.º existirá risco sistémico se o modelo tiver "*capacidades de elevado impacto avaliadas com base em ferramentas e metodologias técnicas adequadas, incluindo indicadores e parâmetros de referência*" [51.º/1/*a*)] ou "*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*" [51.º/1/*b*)]. "Capacidades de elevado impacto" é definido como "*capacidades que correspondem ou excedem as capacidades registadas nos modelos de IA de finalidade geral mais avançados*" (art. 3.º/64). Ou seja, nesta matéria o legislador remete essencialmente para critérios técnico-científicos, estabelecidos no anexo XIII e que serão concretizados pela Comissão em atos delegados (art. 51.º/3). Em qualquer caso, o n.º 2 do art. 51.º estabelece uma presunção (ilidível) de o modelo ter capacidades de elevado impacto quando a quantidade acumulada de cálculo utilizado para o seu treino, medido em operações de vírgula flutuante por segundo (FLOPS), for superior a 10^{25}([145]).

Está previsto no art. 52.º o procedimento de qualificação de um modelo como tendo risco sistémico em que o prestador "*pode apresentar* (…) *argumentos suficientemente fundamentados para demonstrar que,*

([144]) Como já referido, modelos de IA de finalidade geral estão definidos no art. 3.º/63. Por sua vez, "Risco sistémico" é definido 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*" (art. 3.º/65).

([145]) Operações de vírgula flutuante são definidas no art. 3.º/67 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*". Neste contexto, este valor é uma medida do desempenho e capacidade computacional do *hardware* utilizado no treino de um dado modelo de IA. Quanto mais alto, maior a complexidade dos modelos e os respetivos custos de treino. Curiosamente a Executive Order norte-americana utiliza 10^26 FLOPS como limiar, ou seja, dez vezes mais. Muito crítica deste limite, cf. Sandra Wachter, '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 and Technology 26.3 (2024), pp. 696-698. Como sublinha, com razão, o risco é até maior nos sistemas gerais de menor tamanho, porque terão pior qualidade e alguns dos sistemas mias comummente utilizados não estão abrangidos. A Autora sugere mesmo usar o número de utilizadores como critério (p. 715).

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 risco sistémico" (art. 52.º/2).

Os prestadores de modelos de IA de finalidade geral estão sujeitos essencialmente a quatro deveres previstos no art. 53.º:

i) manter documentação técnica adequada e atualizada [n.º 1/*b*) e Anexo XI];

ii) facilitar integração e interoperabilidade com o seu sistema [n.º 1/*b*), Anexo XII];

iii) 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*(146); e

iv) 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 caso de modelos com risco sistémico, além dos deveres aplicáveis a todos os modelos de finalidade geral, o n.º 1 do art. 55.º prevê que os respetivos prestadores devem:

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.

---

(146) O Regulamento dedica os considerandos 105 a 108 ao tema do Direito de Autor. Sobre isso veja-se Alexander Peukert, 'Copyright in the Artificial Intelligence Act — A Primer' GRUR-Int, Vol. 73(6) (2024), pp. 497-509. Também sobre o tema, com indicações adicionais, *vide* Nuno Sousa e Silva, 'Inteligência Artificial e Propriedade Intelectual: está tudo bem?' *in* Anabela Miranda Rodrigues/Susana Aires de Sousa (eds.), *I Congresso… cit.*, pp. 201-220 e Thomas Margoni/Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' GRUR-Int., Vol. 71(8) (2022), pp. 685-701.

Se os prestadores de modelos estiverem estabelecidos em países terceiros (fora da UE) os referidos deveres serão, em grande medida, cumpridos por mandatário, conforme estabelece o art. 54.º.

10. Certificação, supervisão e tutela

O Regulamento estabelece medidas de prevenção e de repressão, não obstante focar-se essencialmente na introdução no mercado ou colocação em serviço de sistemas de IA de risco elevado. Apesar de a responsabilidade civil não ser abordada diretamente([147]), algumas das normas do Regulamento poderão funcionar como normas de proteção, as quais, se violadas, poderão dar origem a uma obrigação de indemnização nos termos do art. 483.º do Código Civil([148]). Além disso, há uma referência à possibilidade de recurso a ações coletivas para proteção dos interesses coletivos dos consumidores nos termos da Diretiva 2020/1828, transposta em Portugal pelo Decreto-Lei n.º 114-A/2023, de 5 de dezembro (art. 110.º).

Uma vez que está em causa legislação de segurança dos produtos, está previsto, nos arts. 28.º e ss., um esquema de certificação e controlo. Existirá pelo menos uma autoridade nacional notificadora (em Portugal, previsivelmente, o Instituto Português de Qualidade, IP)([149]) e uma autoridade nacional de fiscalização de mercado (art. 70.º), que serão as autoridades nacionais competentes nos termos do Regulamento([150]).

A autoridade notificadora é aquela que avalia, designa e fiscaliza os organismos de avaliação de conformidade: tipicamente entidades privadas

---

([147]) Como referido, este tema é abordado em duas Diretivas ainda em fase de proposta: COM(2022)495 final e COM(2022)496 final.

([148]) Sobre as normas de proteção veja-se Adelaide Menezes Leitão, *Normas de Protecção e Danos Puramente Patrimoniais* (Almedina 2009). Como explica (p. 752, ss.) estas normas têm de caber no conceito de lei em sentido material.

([149]) Cuja orgânica foi aprovada pelo Decreto-Lei 71/2012, de 21 de março. A definição de autoridade notificadora ("*a autoridade nacional responsável por estabelecer e executar os procedimentos necessários para a avaliação, designação e notificação de organismos de avaliação da conformidade e pela fiscalização destes*") consta do art. 3.º/19 do Regulamento.

([150]) Arts. 3.º/48 e 74.º. Os diversos Estados-Membros têm seguido abordagens diferentes. Alguns, como Espanha, criaram uma autoridade nova. Outros têm preferido um sistema descentralizado, recorrendo apenas aos reguladores sectoriais. Há ainda quem procure atribuir estas competências a autoridades já existentes, como as autoridades de controlo no âmbito da proteção de dados ou os coordenadores de serviços digitais no âmbito do DSA. No caso das atividades da UE sujeitas ao Regulamento, a entidade de supervisão será a Autoridade Europeia para Proteção de dados (art. 74.º/9), que terá também competência para aplicação de coimas (art. 100.º).

independentes (v.g. APCER) que realizam atividades de testagem, certificação e inspeção dos sistemas com vista a garantir que estes cumprem a exigência do Regulamento. Os organismos notificados são uma categoria especial de organismos de avaliação de conformidade oficialmente designados e com competência para marcação CE(151).

Como preveem os arts. 40.º e ss., as organizações europeias de normalização desenvolverão *standards* que serão adotados pela Comissão Europeia nos termos do Regulamento (UE) n.º 1025/2012. A aplicação desses standards a um sistema de IA de risco elevado fará operar a presunção de conformidade (arts. 40.º/1 e 42.º)(152).

As autoridades nacionais de fiscalização de mercado lidarão com queixas (art. 85.º), incidente graves (art. 73.º) e exercerão os poderes previstos no Regulamento 2019/1020 (art. 74.º), incluindo avaliações de risco, imposição de medidas corretivas (art. 79.º), deteção de não conformidades (art. 83.º) e a supervisão de testes em condições reais (art. 76.º). É também expectável que sejam estas as autoridades com competência sancionatória.

A nível europeu será a Comissão, através do seu Serviço para IA (*AI Office*) (arts. 3.º/47 e 64.º)(153), a fiscalizar os modelos de IA de finalidade geral, funcionando, para esse efeito, como autoridade de fiscalização de mercado (arts. 75.º), com vastos poderes de fiscalização (arts. 88.º a 94.º) e competência para aplicação de coimas (art. 101.º). Além do Serviço de IA, está prevista a existência de um Comité Europeu para a IA (*AI Board*) (art. 65.º), composto por um representante de cada Estado-Membro e que terá essencialmente funções de coordenação da aplicação do Regulamento entre os vários Estados (art. 66.º). O Serviço de IA e o Comité serão assis-

---

(151) Veja-se art. 3.º/21 e /22 e em mais detalhe o *Guia Azul de 2022 sobre a aplicação das regras da UE em matéria de produtos* (2022/C 247/01). A Comissão Europeia mantem uma lista organimos notificados, conhecida como NANDO (<https://webgate.ec.europa.eu/single-market-compliance-space/#/notified-bodies>).

(152) As entidades europeias de standardização são o European Committee for Standardization (CN), o European Committee for Electrotechnical Standardization (CENELEC) e o European Telecommunications Standards Institute (ETSI). Está também prevista a hipótese de a Comissão adotar especificações comuns se estas organizações falharem (art. 41.º). O pedido de emissão de standards relativos a este Regulamento já foi apresentado pela comissão ao CN e CENELEC em Maio de 2023 a Comissão (C(2023)3215 — Standardisation request M/593). Sobre o processo e o papel dos standards no Regulamento *vide* Marta Cantero Gamito/Christopher T Marsden, 'Artificial intelligence co-regulation? The role of standards in the EU AI Act' International Journal of Law and Information Technology, Vol. 32 (1) (2024).

(153) Este departamento da Comissão Europeia foi criado por Decisão da Comissão de 24 de janeiro de 2024 [C(2024) 390 final].

tidos por um fórum consultivo (art. 67.º) e um painel científico de peritos independentes (art. 68.º)([154]).

As sanções variam conforme o tipo de infração, devem ter em conta as circunstâncias concretas (art. 99.º/7) e podem incluir advertências e medidas não pecuniárias (art. 99.º/1)([155]). Do ponto de vista das coimas estão previstas coimas até 7% do volume de negócios mundial ou 35 milhões de euros no caso de práticas proibidas (art. 99.º/3), até 3% do volume de negócios ou 15 milhões de euros para a generalidades das infrações (art. 99.º/4) e até 1% ou 7,5 milhões de euros no caso de fornecimento de informações "incorretas, incompletas ou falaciosas" aos organismos notificados e às autoridades competentes (art. 99.º/5)([156]). No caso dos modelos de IA de finalidade geral, a moldura sancionatória, a aplicar pela Comissão Europeia, irá até 3 % do seu volume de negócios anual total a nível mundial ou 15 milhões de euros (art. 101.º).

O facto de uma entidade ser sancionada ao abrigo do Regulamento não impede que se apliquem outras coimas, nomeadamente por violação do RGPD ou do DSA([157]).

11. Conclusão

Na minha perspetiva o Regulamento contém soluções geralmente equilibradas e razoáveis. No entanto, atenta a sua extensão, complexidade

---

([154]) Sobre este quadro institucional, cf. Claudio Novelli, *et al.*, 'A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities' (2024) *in* <https://ssrn.com/abstract=4817755>.

([155]) Apesar de o Regulamento não o referir expressamente, parece que deve valer o entendimento amplo de "empresa" proveniente do Direito da Concorrência e que tem vindo a ser usado em sede de regulação, nomeadamente no direito da proteção de dados e nas plataformas digitais, especialmente para efeitos sancionatórios. A definição é "*qualquer entidade que exerça uma atividade económica, independentemente do estatuto jurídico dessa entidade e do seu modo de financiamento*" (v.g. C-138/11, *Compass-Datenbank GmbH*, EU:C:2012:449, §35).

([156]) Em relação aos prestadores de modelos de IA de finalidade geral a moldura é equivalente (art. 101.º). Curiosamente no caso das autoridades europeias o valor máximo é de apenas 1,5 milhões de euros para as práticas proibidas (art. 100.º/2) e 750 mil euros nos restantes casos (art. 100.º/3). Mais importante é a possibilidade dada aos Estados-Membros de "*definir regras que permitam determinar em que medida podem ser aplicadas coimas às autoridades e organismos públicos estabelecidos nesse Estado-Membro*" (art. 99.º/8). Por outras palavras, tal como no RGPD, parece legalmente admissivel isentar as entidades públicas da aplicação de coimas. O melhor exemplo vem de cima…

([157]) Tiago Sérgio Cabral, *ob. cit.*, p. 97.

e fraca qualidade legislativa, tornar-se-á algo difícil de aplicar([158]). Por isso, existe um risco real de a União Europeia afetar negativamente a inovação e investimento no domínio da Inteligência Artificial. É aliás possível que venha a ocorrer a uma diminuição da oferta e/ou divergência dos produtos ou serviços, o público europeu receberá versões diferentes e menos avançadas([159]). Como escreve Miquel Peguera Poch([160]), o Regulamento é um instrumento de notável complexidade e com efeitos imprevisíveis.

A principal esperança reside no recurso a standards, cuja adoção massificada pode reduzir significativamente os custos de *compliance* e diminuir a considerável incerteza que este instrumento legislativo inevitavelmente gerará([161]). Um outro contributo para a superação das limitações deste diploma terá de vir dos juristas, seus intérpretes e aplicadores. Por isso, estou, modestamente, a dar o meu.

([158]) Michael Veale/Frederik Zuiderveen Borgesius, *ob. cit.*, p. 112.

([159]) Luciano Floridi, *ob. cit.*: "*fridges, dishwashers, washing machines and even vehicles may need to remain on the safe side of "artificial stupidity" to avoid having to comply with the AI Act (CP version). A scenario becomes plausible in which companies start dumbing down ("de-AI-ing") or at least stop smartening up their products in order not to be subject to the AI Act*". Isto não parece ser ficção – veja-se o anúncio recente da Apple no sentido de não oferecer a tecnologia de IA ("Apple Intelligence") na União Europeia por temer violar o Regulamento dos Mercados Digitais — Regulamento (UE) 2022/1925 (<https://www.theverge.com/2024/6/21/24183251/apple-eu-delay-ai-screen-mirroring-shareplay-dma>) e o da Meta de não disponibilizar um modelo mais avançado tendo em conta a "natureza demasiado imprevisível" do ambiente regulatório europeu (<https://www.theverge.com/2024/7/18/24201041/meta-multimodal-llama-ai-model-launch-eu-regulations>).

([160]) 'La propuesta de Reglamento de IA: una intervencióin legislativa insoslayable en un contexto de incertidumbre' in Miquel Peguera Poch (coord.), *Perspectivas Regulatios de La Inteligencia Artifical em La Unión Europea* (Reus 2023), p. 179.

([161]) O próprio Regulamento reconhece-o no considerando 121, onde se pode ler "*A normalização deverá desempenhar um papel fundamental, disponibilizando aos prestadores soluções técnicas que assegurem a conformidade com o presente regulamento, em consonância com o estado da arte, a fim de promover a inovação, a competitividade e o crescimento no mercado único*". Para uma lista não exaustiva de *standards* aplicáveis neste contexto *vide* Federico Marengo, *Privacy…cit.*, p. 196, ss., e Alessio Tartaro, 'Regulating by standards: current progress and main challenges in the standardisation of Artificial Intelligence in support of the AI Act' European Journal of Privacy Law and Technologies (2023), pp. 147-174. Alguns autores, incluindo Emilija Leinarte, *ob. cit.*, e Sandra Wachter, *ob. cit.*, consideram que o AI Act é uma versão diluída do que deveria ser, com um âmbito de aplicação demasiado restrito. Embora concorde parcialmente, isso não impede a geração de incerteza.