

# REVISTA DA FACULDADE DE DIREITO DA UNIVERSIDADE DE LISBOA

---

LISBON LAW REVIEW



Número Temático: Tecnologia e Direito

ANO LXIII

2022

NÚMEROS 1 E 2

REVISTA DA FACULDADE DE DIREITO  
DA UNIVERSIDADE DE LISBOA  
Periodicidade Semestral  
Vol. LXIII (2022) 1 e 2

LISBON LAW REVIEW

---

#### COMISSÃO CIENTÍFICA

Alfredo Calderale (Professor da Universidade de Foggia)  
Christian Baldus (Professor da Universidade de Heidelberg)  
Dinah Shelton (Professora da Universidade de Georgetown)  
Ingo Wolfgang Sarlet (Professor da Pontifícia Universidade Católica do Rio Grande do Sul)  
Jean-Louis Halpérin (Professor da Escola Normal Superior de Paris)  
José Luis Díez Ripollés (Professor da Universidade de Málaga)  
José Luís García-Pita y Lastres (Professor da Universidade da Corunha)  
Judith Martins-Costa (Ex-Professora da Universidade Federal do Rio Grande do Sul)  
Ken Pennington (Professor da Universidade Católica da América)  
Marc Bungenberg (Professor da Universidade do Sarre)  
Marco Antonio Marques da Silva (Professor da Pontifícia Universidade Católica de São Paulo)  
Miodrag Jovanovic (Professor da Universidade de Belgrado)  
Pedro Ortego Gil (Professor da Universidade de Santiago de Compostela)  
Pierluigi Chiassoni (Professor da Universidade de Génova)

---

#### DIRETOR

M. Januário da Costa Gomes

---

#### COMISSÃO DE REDAÇÃO

Paula Rosado Pereira  
Catarina Monteiro Pires  
Rui Tavares Lanceiro  
Francisco Rodrigues Rocha

---

#### SECRETÁRIO DE REDAÇÃO

Guilherme Grillo

---

#### PROPRIEDADE E SECRETARIADO

Faculdade de Direito da Universidade de Lisboa  
Alameda da Universidade – 1649-014 Lisboa – Portugal

---

#### EDIÇÃO, EXECUÇÃO GRÁFICA E DISTRIBUIÇÃO LISBON LAW EDITIONS

Alameda da Universidade – Cidade Universitária – 1649-014 Lisboa – Portugal

---

ISSN 0870-3116

---

Depósito Legal n.º 75611/95

Data: Outubro, 2022

- 
- M. Januário da Costa Gomes  
9-16 Editorial

## ESTUDOS DE ABERTURA

- 
- Guido Alpa  
19-34 On contractual power of digital platforms  
*Sobre o poder contratual das plataformas digitais*
- 
- José Barata-Moura  
35-62 Dialéctica do tecnológico. Uma nótula  
*Dialectique du technologique. Une notule*

## ESTUDOS DOUTRINAIS

- 
- Ana Alves Leal  
65-148 Decisões, algoritmos e interpretabilidade em ambiente negocial. Sobre o dever de explicação das decisões algorítmicas  
*Decisions, Algorithms and Interpretability in the Context of Negotiations. On the Duty of Explanation of Algorithmic Decisions*
- 
- Ana María Tobío Rivas  
149-215 Nuevas tecnologías y contrato de transporte terrestre: los vehículos automatizados y autónomos y su problemática jurídica  
*Novas tecnologias e contrato de transporte terrestre: veículos automatizados e autónomos e seus problemas jurídicos*
- 
- Aquilino Paulo Antunes  
217-236 Avaliação de tecnologias de saúde, acesso e sustentabilidade: desafios jurídicos presentes e futuros  
*Health technology assessment, access, and sustainability: present and future legal challenges*
- 
- Armando Sumba  
237-270 *Crowdinvesting* e proteção do investidor: vantagens e limites do financiamento colaborativo de empresas em Portugal  
*Crowdinvesting and investor protection: the advantages and limits of business crowdfunding in Portugal*
- 
- Diogo Pereira Duarte  
271-295 O Regulamento Europeu de *Crowdfunding*: risco de intermediação e conflitos de interesses  
*The European Crowdfunding Regulation: intermediation risk and conflicts of interests*
- 
- Eduardo Vera-Cruz Pinto  
297-340 Filosofia do Direito Digital: pensar juridicamente a relação entre Direito e tecnologia no ciberespaço  
*Digital Law Philosophy: thinking legally the relation between Law and Technology in the Cyberspace*

- 
- Francisco Rodrigues Rocha**  
341-364 O «direito ao esquecimento» na Lei n.º 75/2021, de 18 de Novembro. Breves notas  
*Le « droit à l'oubli » dans la loi n. 75/2021, de 18 novembre. Brèves remarques*
- 
- Iolanda A. S. Rodrigues de Brito**  
365-406 The world of shadows of disinformation: the emerging technological caves  
*O mundo das sombras da desinformação: as emergentes cavernas tecnológicas*
- 
- João de Oliveira Geraldes**  
407-485 Sobre a proteção jurídica dos segredos comerciais no espaço digital  
*On the Legal Protection of Trade Secrets in the Digital Space*
- 
- João Marques Martins**  
487-506 Inteligência Artificial e Direito: Uma Brevíssima Introdução  
*Artificial Intelligence and Law: A Very Short Introduction*
- 
- Jochen Glöckner | Sarah Legner**  
507-553 Driven by Technology and Controlled by Law Only? – How to Protect Competition  
on Digital Platform Markets?  
*Von Technologie getrieben und nur durch das Recht gebremst? – Wie kann Wettbewerbschutz auf  
digitalen Plattformmärkten gelingen?*
- 
- Jones Figueirêdo Alves | Alexandre Freire Pimentel**  
555-577 Breves notas sobre os preconceitos decisoriais judiciais produzidos por redes neurais  
artificiais  
*Brief notes about the judicial decisional prejudices produced by artificial neural networks*
- 
- José A. R. Lorenzo González**  
579-605 Reconhecimento facial (FRT) e direito à imagem  
*Facial recognition (FRT) and image rights*
- 
- José Luis García-Pita y Lastres**  
607-661 Consideraciones preliminares sobre los llamados *smart contracts* y su problemática  
en el ámbito de los mercados bursátiles y de instrumentos financieros [Las órdenes  
algorítmicas y la negociación algorítmica]  
*Considerações preliminares sobre os chamados smart contracts e os seus problemas no domínio dos  
mercados bolsistas e dos instrumentos financeiros [As ordens algorítmicas e a negociação  
algorítmica]*
- 
- Mariana Pinto Ramos**  
663-727 O consentimento do titular de dados no contexto da *Internet*  
*The consent of the data subject in the Internet*
- 
- Neuza Lopes**  
729-761 O (re)equilíbrio dos dois pratos da balança: A proteção dos consumidores perante  
os avanços no mundo digital – Desenvolvimentos recentes no direito europeu e  
nacional  
*(Re)balancing the scale: Consumer protection in the face of advances in the digital world – Recent  
developments in European and national law*

- 
- Nuno M. Guimarães**  
763-790 Sistemas normativos e tecnologias digitais: formalização, desenvolvimento e convergência  
*Normative systems and digital technologies: formalization, development, and convergence*
- 
- Paulo de Sousa Mendes**  
791-813 Uma nota sobre Inteligência Artificial aplicada ao Direito e sua regulação  
*A Note on Artificial Intelligence in Legal Practice and Its Regulation*
- 
- Renata Oliveira Almeida Menezes | Luís Eduardo e Silva Lessa Ferreira**  
815-838 *Cyberbullying* por divulgação de dados pessoais  
*Cyberbullying by doxxing*
- 
- Rui Soares Pereira**  
839-865 Sobre o uso de sistemas de identificação biométrica (e de tecnologias de reconhecimento facial) para fins de segurança pública e de aplicação coerciva da lei: reflexões a propósito da proposta de regulamento europeu sobre a inteligência artificial  
*On the use of biometric data systems (and facial recognition technologies) for security and law enforcement purposes: reflections on the proposal for the european regulation on artificial intelligence*
- 
- Rute Saraiva**  
867-930 Segurança Social, Direito e Tecnologia – Entre *Rule-as-Code* e a personalização  
*Social Security, Law and Technology – Between rule-as-Code and personalization*

## VULTOS DO(S) DIREITO(S)

- 
- Alfredo Calderale**  
933-969 Augusto Teixeira de Freitas (1816-1883)

## JURISPRUDÊNCIA CRÍTICA

- 
- A. Barreto Menezes Cordeiro**  
973-981 Anotação ao Acórdão *Meta Platforms* – TJUE 28-abr.-2022, proc. C-319/20  
*Commentary to the Meta Platforms Judgment – CJEU 28-apr.-2022 proc. C 310/20*
- 
- Rui Tavares Lanceiro**  
983-999 2020: um ano histórico para a relação entre o Tribunal Constitucional e o Direito da UE – Um breve comentário aos Acórdãos do Tribunal Constitucional n.º 422/2020 e n.º 711/2020  
*2020: A landmark year for the relationship between the Constitutional Court and EU law – A brief commentary on the Constitutional Court judgments 422/2020 and 711/2020*

## VIDA CIENTÍFICA DA FACULDADE

- 
- J. M. Sérvulo Correia**  
1003-1007 Homenageando o Doutor Jorge Miranda  
*Homage to Professor Dr. Jorge Miranda*

- **Jorge Miranda**  
1009-1016 Nótula sobre os direitos políticos na Constituição portuguesa  
*Notice about Political Rights in the Portuguese Constitution*

**LIVROS & ARTIGOS**

- **M. Januário da Costa Gomes**  
1019-1024 Recensão à obra *L'intelligenza artificiale. Il contesto giuridico*, de Guido Alpa

# Decisões, algoritmos e interpretabilidade em ambiente negocial. Sobre o dever de explicação das decisões algorítmicas

## *Decisions, Algorithms and Interpretability in the Context of Negotiations. On the Duty of Explanation of Algorithmic Decisions*

---

Ana Alves Leal\*

**Resumo:** A formação de declarações negociais através de processos decisórios baseados em algoritmos vem desafiar noções e esquemas de decisão estabilizados pela dogmática do direito do negócio jurídico. Tendo como pano de fundo o quadro geral do direito civil, debate-se, neste artigo, se existe um dever de explicação, cominado ao declarante, da decisão algorítmica que baseia ou que constitui a sua declaração negocial. Particularizam-se, por subjazerem ao grosso das preocupações hodiernas com a sindicabilidade dos algoritmos, as decisões algorítmicas de rejeição ou de desvinculação do negócio, sobretudo quando tidas como irracionais ou injustas. Este debate não é só provocado pelo espaço que determinados regimes setoriais – que incentivam ou impõem essa explicação – têm vindo a ocupar no tráfico negocial (como os da proteção de dados). O debate impõe-se também para discutir a associação da opacidade dos algoritmos a problemas de ininteligibilidade

**Abstract:** The formation of declarations of negotiation through algorithmic decision-making processes challenges notions and decision schemes stabilized by the dogmatic of legal transaction law. Within the general framework of civil law, this article explores whether the declarant has a duty to explain the algorithmic decision underlying their declaration of negotiation. Close attention is paid to the algorithmic decisions of proposal rejection or of exit from the contract, especially when considered irrational or unfair, as these decisions lie behind the current concerns about algorithmic accountability. This debate is encouraged by the increasing significance of specific legal regimes, which promote or impose this explanation (such as data protection), in the role of private law modernization. Nonetheless, this debate is mainly justified as a tool to discuss the linkage between the opacity of algorithms and the problems of the unintelligibility

---

\* Assistente Convidada da Faculdade de Direito da Universidade de Lisboa. Doutoranda em Ciências Jurídicas. Investigadora do Centro de Investigação de Direito Privado. Email: [analeal@fd.ulisboa.pt](mailto:analeal@fd.ulisboa.pt)

da declaração negocial, cujo tratamento deve ser feito nos domínios da interpretação e dos vícios do negócio jurídico.

**Palavras-chave:** decisões algorítmicas; inteligência artificial explicável; direito dos algoritmos; declaração negocial; interpretação; deveres de informação.

of the declaration of negotiation, which shall be addressed in the fields of contractual interpretation and validity.

**Keywords:** algorithmic decisions; explainable artificial intelligence (XAI); algorithm law; declaration of negotiation; interpretation; information duties.

**Sumário:** § 1.º Entre as técnicas corretivas e informacionais e as técnicas proibitivas no direito dos algoritmos; 1.1. A explicação dos algoritmos no Regulamento Geral sobre a Proteção de Dados; 1.2. Justificação e delimitação do tema; 1.3. A proibição de decisões algorítmicas como novo discurso; § 2.º Escalonamento de riscos e modelos regulativos da Proposta de Regulamento sobre Inteligência Artificial; 2.1. Transparência na cadeia de valor. Ambiguidades da ausência de um dever de explicação das decisões algorítmicas; § 3.º O perigo das decisões algorítmicas: a amplificação do enviesamento decisório humano; 3.1. Intuições e preconceito algorítmico: tentativa de desconstrução; 3.2. Graus de intervenção humana no processo decisório: da assistência algorítmica às decisões automatizadas; 3.3. O efeito discriminatório, o efeito lesivo e o efeito multiplicador das decisões algorítmicas; particularidades da *big data*; 3.4. Boas decisões algorítmicas, *alternative data* e a lógica de perigo abstrato; § 4.º Explicação da declaração negocial; 4.1. Decisão algorítmica como declaração negocial; 4.2. Opacidade do percurso decisório e do discurso decisório – por um critério de unificação das decisões humanas e das decisões algorítmicas; 4.3. Processo formativo e declaração negocial: a explicação como descrição e a tendencial irrelevância da motivação.

## § 1.º Entre as técnicas corretivas e informacionais e as técnicas proibitivas no direito dos algoritmos

1. Com o advento da inteligência artificial e com a multiplicação dos “processos decisórios baseados em algoritmos” (em tradução de “*algorithmic decision-making*”, e aos quais, por simplicidade, nos referiremos como “decisões algorítmicas”<sup>1/2</sup>), o

---

<sup>1</sup> Não existe uma definição legal de «algoritmo», pese embora a profusa utilização deste termo em instrumentos normativos – o que pode ser ilustrado com uma pesquisa rápida desta palavra no motor de busca do Diário da República Eletrónico (<<https://dre.pt>>): à data em que se escrevem estas linhas, existem 263 diplomas em vigor no ordenamento português em que o termo é utilizado, sem contar com os diplomas de Direito da União Europeia, em particular, com aqueles que têm aplicabilidade direta. Recorre-se, para os propósitos deste estudo, à noção ampla de «algoritmo» oriunda da matemática e da



---

ciência da computação: o conjunto ou a sequência finita de instruções que podem ser executadas quando acionadas e que visam obter uma solução para um determinado tipo de problema. Não nos interessam *todos* os algoritmos enquanto produto do raciocínio matemático ou mesmo enquanto produto das nossas ações quotidianas – por exemplo, o processo de lavar os dentes ou o processo de utilizar certos transportes para viajar diariamente para o nosso local de trabalho são descritíveis por algoritmos (ou *algoritmizáveis*), como bem assinala THOMAS H. COORMEN, *Algorithms Unlocked*, Cambridge, Massachusetts, The MIT Press, 2013, p. 1. Centramos o nosso interesse nos algoritmos que podem ser executados por computadores – ou seja, aqueles cujas instruções e cujos problemas a que visam responder “[are] described precisely enough that a computer can run [them]” – THOMAS H. COORMEN, *Algorithms Unlocked*, cit., p. 1. A definição precisa dos componentes do algoritmo é, assim, um dos elementos distintivos dos algoritmos computacionais (o que é ilustrado por THOMAS H. COORMEN, *Algorithms Unlocked*, cit., p. 1: “What distinguishes an algorithm that runs on a computer from an algorithm that you run? You might be able to tolerate it when an algorithm is imprecisely described, but a computer cannot. For example, if you drive to work, your drive-to-work algorithm might say “if traffic is bad, take an alternate route.” Although you might know what you mean by “bad traffic,” a computer does not.”). Esta circunscrição é frequentemente adotada pelos autores que se debruçam sobre os deveres de explicação da decisão algorítmica (*vide*, entre outros, REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, disponível em SSRN <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4066781](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4066781)> [consultado em 20.07.2022], na versão de 25-mar.-2022, a publicar em 2023 na *Vanderbilt Law Review*, p. 3).

Questões diferentes, aqui não tratadas, mas que também ocupam os juristas, são as da *criação* do algoritmo, da *imputação dessa criação*, e a da *tutela jurídica* do algoritmo criado, designadamente, ao abrigo dos regimes positivos de proteção da propriedade intelectual. Para este propósito, limitamo-nos a reconhecer que existem diferenças fundamentais entre algoritmos que descrevem um processo da natureza, algoritmos que são criados por programadores e algoritmos que são o produto de técnicas de *machine-learning* (com mais detalhes, cf. ANDREW CHIN, “Software Patenting and Section 101’s Gatekeeping Function”, in Woodrow Barfield (ed.), *The Cambridge Handbook of the Law of Algorithms*, Cambridge, Cambridge University Press, 2021 (374-390), bem como NIKLAS MAAMAR, *Computer als Schöpfer, Der Schutz von Werken und Erfindungen künstlicher Intelligenz*, Tübingen, Mohr Siebeck, 2021). Esclarecemos que este estudo, centrado na interpretabilidade das decisões algorítmicas e na explicação das mesmas, pressupõe que os algoritmos que baseiam esses processos decisórios dependam de criação humana (ainda que seja em descrição de um processo natural) e sejam, em maior ou menor medida, controláveis por seres humanos. O que não garante, em todo o caso, que esses algoritmos sejam compreensíveis por seres humanos, à luz dos critérios hermenêuticos com relevância no caso concreto. Tanto assim é que se tornou comum, com algum humor, definir «algoritmo» como “Word used by programmers when they do not want to explain what they did” – cf. MAJA BRKAN / GRÉGORIE BONNET, “Legal and Technical Feasibility of the GDPR’s Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas,” in *European Journal of Risk Regulation*, 11, 2020 (18-50), p. 18.

<sup>2</sup> Este estudo abrange quaisquer decisões algorítmicas, nos termos circunscritos na nota anterior, embora não necessariamente aquelas que são o produto de sistemas de inteligência artificial. Ainda a propósito das dificuldades definitórias e conceptuais destas novas realidades, convém esclarecer que, diversamente do que sucede com os algoritmos, menos obstáculos enfrentam os juristas na definição de «inteligência artificial» ou de «sistemas de inteligência artificial». Ainda que se trate de uma realidade conceptual oriunda de outras disciplinas, intui-se hoje, com relativa facilidade, que a inteligência artificial é composta por grupos de algoritmos que são modificáveis a todo o tempo em função de uma certa aprendizagem do sistema (aprendizagem,

debate jurídico em torno da admissibilidade do recurso a estas decisões em contexto negocial, por uma ou por ambas as partes envolvidas no negócio, atingiu, nos

---

por sua vez, também algoritmizada). Para além desta *aprendizagem* do sistema, são também comumente atribuídas à inteligência artificial as características de *reconhecimento de padrões* (sejam padrões visuais e sensoriais, sejam padrões de comportamento), de *raciocínio* e de *inferência* (seguir e aplicar regras lógicas a um conjunto de dados disponíveis para chegar a uma conclusão).

A favor desta intuição muito contribui a universalização da expressão “inteligência artificial” para designar esta realidade de base tecnológica: por um lado, esta designação assenta numa metáfora evidente que fomenta a associação dos sistemas de inteligência artificial a um símile das pessoas humanas (mesmo que não nos refiramos a hipóteses, porventura ainda de ficção científica, de sistemas autoconscientes e sencientes); por outro lado, favorece essa *personalização* a ocorrência de atos de linguagem, seja no nosso quotidiano, seja na própria lei, que se referem ao “raciocínio”, às “capacidades” ou à “atuação” dos sistemas de inteligência artificial, ou mesmo imputando-lhes determinados estados subjetivos.

Tome-se como exemplo o Regulamento sobre Inteligência Artificial (hoje, ainda uma proposta – a Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial e altera determinados atos legislativos da União, de 21.04.2021, COM(2021) 206 final, 2021/0106 [COD]), que define os sistemas de inteligência artificial como programas informáticos desenvolvidos com uma ou várias das técnicas e abordagens enumeradas no seu Anexo I, e que são *capazes* de, tendo em vista um determinado conjunto de objetivos definidos por seres humanos, *criar* resultados, tais como conteúdos, previsões, recomendações ou decisões, que influenciam os ambientes com os quais *interage* (artigo 3.º, n.º 1). Esta definição – ainda que, por agora, contida em projeto de Regulamento – é útil para os juristas não só por descrever uma realidade intuída por estes em moldes que se julgam coincidentes com os da linguagem que baseia essa intuição, mas sobretudo por materializar uma (futura) regra definitiva que fixa – pelo menos, para efeitos de aplicação do Regulamento – o significado de um conceito jurídico. Contudo, como bem se alerta no Considerando da Proposta de Regulamento, “[a] definição de «sistema de IA» deve ser inequívoca, para assegurar a segurança jurídica, concedendo em simultâneo a flexibilidade suficiente para se adaptar a futuras evoluções tecnológicas. A definição deve basear-se nas principais características funcionais do software [...]. Os sistemas de IA podem ser concebidos para operar com diferentes níveis de autonomia e ser utilizados autonomamente ou como componente de um produto, independentemente de o sistema estar fisicamente incorporado no produto (integrado) ou servir a funcionalidade do produto sem estar incorporado nele (não integrado)”. A necessidade de a lei acompanhar a evolução tecnológica é uma preocupação manifesta do legislador europeu, que optou pela delegação da atualização da regulação: esclarece-se na parte final do referido Considerando (6) que “a definição de «sistema de IA» deve ser completada por uma lista de técnicas e abordagens específicas utilizadas para o seu desenvolvimento, que deve ser atualizada face à evolução do mercado e da tecnologia, mediante a adoção de atos delegados da Comissão que alterem essa lista”.

Para um quadro geral destinado a juristas sobre estes e outros conceitos relacionados com as decisões algorítmicas, encontra-se utilidade no estudo de REMBRANDT DEVILLÉ / NICO SERGEYSSELS / CATHERINE MIDDAG, “Basic Concepts of AI for Legal Scholars”, in Jan De Bruyne / Cedric Vanleenhove (ed.), *Artificial Intelligence and the Law*, Cambridge, Intersentia, 2021 (1-22).

últimos anos, um estado de maturidade no discurso jurídico que torna obsoleta qualquer descrição e discussão sobre as vantagens e sobre os riscos que lhes estão associados. Ao mesmo tempo, considerado o espaço europeu, não será apressado falar – ainda que com a devida prudência metodológica – de um proto-“direito dos algoritmos”: consente-o o desenvolvimento (em curso) do seu sistema externo, proporcionado pelo complexo de fontes dirigido à regulação das decisões algorítmicas<sup>3/4</sup>.

---

<sup>3</sup> Segue-se de perto a conceção de «sistema externo» de CLAUS W. CANARIS, *Pensamento sistemático e conceito de sistema na Ciência do Direito*, 4.<sup>a</sup> ed. (trad. de António Menezes Cordeiro da 2.<sup>a</sup> ed., 1983), Lisboa, Fundação Calouste Gulbenkian, 2008, p. 26. A ideia exposta em texto, mais do que uma apologia tímida em torno da (futura) autonomia dogmática do “direito dos algoritmos”, é uma constatação e uma afirmação do espaço que os problemas jurídicos sobre algoritmos têm vindo a ocupar no discurso jurídico e no sistema de fontes do direito. Não cabe, neste estudo, desenvolver a relevância lógico-jurídica de um “direito dos algoritmos” como disciplina normativa, nem os moldes em que o atual e o projetado sistema de fontes facilita a tarefa do intérprete-aplicador, conferindo a praticabilidade necessária ao processo aplicativo. Assinala-se, não obstante, que essa autonomização tem sido vaticinada: *vide*, entre outros, MARIO MARTINI, “Algorithmen als Herausforderung für die Rechtsordnung”, *JuristenZeitung*, 21, 72/2017 (1017-1025); HORST EIDENMÜLLER / GERHARD WAGNER, *Law by Algorithm*, Tübingen, Mohr Siebeck, 2021, pp. 1-7; JAN-FREDERICK GÖHSL / JOHANNES ROTTMANN / MATTHIAS SCHAUT, “Regulierung für Algorithmen”, *JuristenZeitung*, 1, 76/2021 (34-36); WOODROW BARFIELD / JESSICA BARFIELD, “An Introduction to Law and Algorithm”, in Woodrow Barfield (ed.), *The Cambridge Handbook of the Law of Algorithms*, cit. (3-15).

Em qualquer caso, e mesmo sem fazer apelo aos regimes setoriais que contêm normas dirigidas à regulação dos sistemas de inteligência artificial, não se pode ignorar que, sendo aprovada a Proposta de Regulamento sobre Inteligência Artificial, este Regulamento dará resposta à agregação, ordenação e codificação das fontes do direito de acordo com critérios racionais – sem prejuízo da necessidade de aprovação, pelos Estados-membros, de atos de execução de direito interno. As razões que subjazem a essa agregação e ordenação estão plasmadas na exposição de motivos da Proposta de Regulamento e assentam nas ideias de *harmonização espacial* (no território europeu) e de *centralização e sistematização material* da regulação dos algoritmos. Entre outros aspetos, afirma-se, na exposição de motivos, tratar-se de “uma proposta legislativa relativa a uma abordagem europeia coordenada às implicações humanas e éticas da inteligência artificial” (p. 2), e que a “natureza horizontal da proposta exige a plena coerência com a legislação da União em vigor aplicável aos setores [já regulados de inteligência artificial]” (em particular, é “garantida coerência com a Carta dos Direitos Fundamentais da UE e a legislação derivada da União em vigor em matéria de proteção de dados, defesa dos consumidores, não discriminação e igualdade de género”) (p. 4). Acresce que “[n]o que diz respeito aos sistemas de IA de risco elevado que são componentes de segurança de produtos, a [...] proposta será integrada na legislação de segurança setorial em vigor para assegurar a coerência, evitar as duplicações e minimizar os encargos adicionais” (p. 4).

A Proposta de Regulamento faz parte de um pacote abrangente de medidas que visa resolver os problemas decorrentes do desenvolvimento e da utilização da inteligência artificial, tal como examinados no *Livro Branco da Comissão Europeia sobre a inteligência artificial – Uma abordagem*

---

*européia virada para a excelência e a confiança* [COM(2020) 65 final]. No Livro Branco, apesar de se reconhecer que “a legislação da UE continua em princípio, a ser plenamente aplicável independentemente do envolvimento da IA” (p. 15), assumiu-se não estar esta legislação setorial totalmente adequada para fazer face aos riscos gerados pelos sistemas de inteligência artificial. Razão pela qual se apontou, nesse Livro, a necessidade de um quadro regulador *especialmente e materialmente comum* (p. 11). O apontamento desta necessidade num Livro Branco da Comissão Europeia – instrumento que, por excelência, é propulsor de procedimentos legislativos específicos – favorece uma visão holística do legislador europeu em relação às várias iniciativas legislativas em gênese e em curso. Desta forma, como se afirma na exposição de motivos da Proposta de Regulamento, “é garantida a coerência e a complementaridade com outras iniciativas em curso ou planeadas da Comissão que também visam responder a esses problemas [de inteligência artificial], incluindo a revisão da legislação setorial em matéria de produtos (por exemplo, a Diretiva Máquinas e a Diretiva Segurança Geral dos Produtos) e as iniciativas que abordam questões de responsabilidade associadas às novas tecnologias, incluindo os sistemas de IA” (p. 5). Importa também reter que o esforço codificador materializado na Proposta de Regulamento “dá resposta a pedidos explícitos do Parlamento Europeu e do Conselho Europeu, que têm apelado, repetidamente, para a ação legislativa, com vista a assegurar o bom funcionamento do mercado interno de sistemas de inteligência artificial, no qual os benefícios e os riscos da IA sejam abordados de forma adequada a nível da União” (p. 2).

O trabalho que o Parlamento Europeu tem desenvolvido no domínio da inteligência artificial é igualmente merecedor de nota. Desde finais de 2020 que o Parlamento adotou um conjunto de resoluções neste domínio: em matéria de ética (cf. *Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre o regime relativo aos aspetos éticos da inteligência artificial, da robótica e das tecnologias conexas* [2020/2012(INL)]), em matéria de responsabilidade (cf. *Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre o regime de responsabilidade civil aplicável à inteligência artificial* [2020/2014(INL)]), em matéria de direitos de autor (cf. *Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre os direitos de propriedade intelectual para o desenvolvimento de tecnologias ligadas à inteligência artificial* [2020/2015(INI)]), em matéria penal (cf. *Projeto de relatório do Parlamento Europeu sobre a inteligência artificial no direito penal e a sua utilização pelas autoridades policiais e judiciais em casos penais* [2020/2016(INI)]), nas áreas da educação, da cultura e do setor audiovisual (cf. *Projeto de relatório do Parlamento Europeu sobre a inteligência artificial na educação, na cultura e no setor audiovisual* [2020/2017(INI)]), bem como *Educação Digital 2021-202 – Reconfigurar a educação e a formação para a era digital* [COM(2020) 624 final], que prevê o desenvolvimento de orientações éticas em matéria de inteligência artificial e utilização de dados no ensino).

<sup>4</sup> A atenção dedicada ao sistema de fontes vigente (ou a vigorar) no ordenamento jurídico português não nos desonera de observar que no ordenamento estado-unidense se assiste a um movimento regulador de natureza similar, embora mais sinuoso. Em fevereiro de 2022, foi apresentado no Senado (S.3572) e na Câmara dos Deputados (H.R.6580) a proposta do *Algorithmic Accountability Act of 2022* (disponível em <<https://www.congress.gov/bill/117th-congress/senate-bill/3572/text?r=1&s=1>> e <<https://www.congress.gov/bill/117th-congress/house-bill/6580/text?r=37&s=1>> [consultados em 20.07.2022]).

Trata-se de um projeto legislativo cuja primeira versão foi apresentada em abril de 2019 – não tendo, nessa data, reunido o consenso necessário para ser aprovada. Desde então, sofreu alterações (essencialmente, aperfeiçoamentos técnicos e clarificação dos seus âmbitos material e subjetivo) e foi recentemente reintroduzido num contexto em que a regulação da inteligência artificial ocupa uma posição privilegiada nas agendas legislativas dos Estados Unidos da América e de outras jurisdições. As alterações que o projeto sofreu neste *interim* não o desviaram das finalidades a que

2. Tal não significa, porém, que se assista a uma estabilização das soluções normativas e dos posicionamentos doutrinários ou jurisprudenciais sobre o uso de decisões algorítmicas na formação ou na execução de negócios jurídicos. Pelo contrário: estamos na viragem daquela que é considerada a primeira fase do discurso dos juristas sobre algoritmos, essencialmente centrada nos problemas do erro, da distorção e da discriminação algorítmica e na necessidade de transparência, controlo e aperfeiçoamento dos sistemas tecnológicos existentes para mitigação desses problemas (fazendo-se apelo a “melhores dados, melhores procedimentos, melhores sistemas”)<sup>5</sup><sup>6</sup>.

---

está orientado: aumentar a transparência dos algoritmos utilizados em contextos de tomada de decisão, ou mesmo proibir a sua utilização, a par da responsabilização dos criadores e utilizadores dos algoritmos, com vista a reduzir resultados (decisões) discriminatórios, tendenciosos ou prejudiciais. Em traços largos, o *Algorithmic Accountability Act of 2022* comina um conjunto de deveres procedimentais e de proibições às empresas que utilizem algoritmos nos seus processos decisórios quando esteja implicada a tomada de “decisões críticas” (“*critical decisions*”). A este respeito, o projeto elenca várias categorias de decisões que têm um impacto significativo na vida de pessoas singulares, nomeadamente decisões relacionadas com o acesso ou o custo da educação, emprego, serviços essenciais, planeamento familiar, serviços financeiros, saúde, habitação e serviços jurídicos (Section 2(8) do projeto). O projeto exige que as empresas que tomem decisões críticas baseadas em algoritmos realizem avaliações de impacto dessas decisões (tendo em conta enviesamentos e outros fatores), e vem fundar um registo e repositório público dos algoritmos na Federal Trade Commission, criando várias dezenas de cargos nessa entidade reguladora funcionalizados à fiscalização do cumprimento do diploma.

Para uma descrição dos principais aspetos deste projeto, veja-se a comunicação do Senador Ron Wyden, um dos seus autores, disponível em <2022-02-03 Algorithmic Accountability Act of 2022 One-pager (senate.gov)> [consultado em 20.07.2022]. Para uma análise das fragilidades do diploma, com indicação dos aspetos a rever, cf. JAKOB MÖKANDER / LUCIANO FLORIDI, “From Algorithmic Accountability to Digital Governance”, *Nature Machine Intelligence*, 4/2022 (508-509). Destacamos, como principais críticas formuladas pelos Autores, (i) o facto de o mesmo apenas se aplicar a empresas de grande dimensão (que tenham uma faturação anual superior a 50 milhões US\$, cujo património exceda 250 milhões US\$ ou que processe informações de mais de 1 milhão de utilizadores), (ii) o facto de o conteúdo do diploma ser, em muitos casos, vago, pouco detalhado e programático, carente de densificação posterior pela Federal Trade Commission, ao abrigo do seu poder regulamentar, (iii) o facto de o aumento da transparência dos algoritmos não impedir que se mantenha a discriminação de certos subgrupos de pessoas, e (iv) o facto de ser um instrumento legislativo bem mais modesto do que o seu congénere europeu, a Proposta de Regulamento sobre Inteligência Artificial – desde logo, porque este último diploma não circunscreve o seu âmbito de aplicação ao “consumidor”, como sucede com o *Algorithmic Accountability Act of 2022*, aplicando-se a qualquer pessoa singular.<sup>5</sup> Na fórmula sintética de ARI EZRA WALDMAN, “Algorithmic Legitimacy”, in Woodrow Barfield (ed.), *The Cambridge Handbook of the Law of Algorithms*, cit. (107-120), p. 107.

<sup>6</sup> Demarcando a primeira fase (“*first wave*”) da segunda fase (“*second wave*”) do discurso dos juristas sobre o controlo algorítmico, cf. FRANK PASQUALE, “The Second Wave of Algorithmic Accountability”, *Law and Political Economy* (November 25, 2019), <<https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability/>> [consultado em 20.07.2022].

Nesta primeira fase, juristas, legisladores e reguladores empenharam-se em reclamar e em construir soluções normativas que assegurassem processos decisórios de base tecnológica assentes em procedimentos justos e equitativos (*due process*), que garantissem a neutralidade das plataformas em linha<sup>7</sup>, e que vedassem práticas discriminatórias na conceção e no funcionamento dos algoritmos. A preocupação, dirigida a uma dimensão procedimental, focou-se sobretudo na vigilância dos algoritmos utilizados e no aperfeiçoamento dos sistemas tecnológicos existentes para tomar decisões – incluindo, decisões em contexto negocial –, culminando com a defesa da “garantia” da intervenção humana para confirmação ou correção da decisão algorítmica (“*human in the loop*”<sup>8</sup>).

Destas exigências de vigilância algorítmica decantam-se duas técnicas distintas de regulação: uma técnica *corretiva*, orientada ao aperfeiçoamento da conceção e funcionamento dos algoritmos, e uma técnica *informativa*, construída sob a égide da “transparência” dos algoritmos, e que, em fórmula muito sintética, se traduz na cominação de deveres de informação sobre a composição dos algoritmos e sobre o seu modo de funcionamento.

---

<sup>7</sup> Cf. FRANK PASQUALE, “Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power”, *Theoretical Inquiries in Law*, 17/2016 (487-513), p. 487, enfatizando a ambivalência na gestão destas plataformas, que tanto marginalizam (ou bloqueiam totalmente) a divulgação de informações ou opiniões e potenciais conexões entre os seus utilizadores, ao abrigo daquilo que consideram ser o seu direito de controlar a informação que é disseminada através dessas plataformas, como não interferem no fluxo de informação gerado e por elas disseminado, demitindo-se de qualquer competência ou encargo nesse sentido, alegando que a informação gerada deve refletir, na realidade, os desejos e as preferências da sua base de utilizadores.

<sup>8</sup> Trata-se de expressão amplamente difundida entre os juristas que estudam o direito dos algoritmos para referir o direito à intervenção humana na decisão algorítmica, sem tradução perfeita para língua portuguesa. É também expressão disseminada no domínio interdisciplinar da inteligência artificial, sendo hoje considerado um ramo desta (geralmente, pelo acrónimo “HITL”), no âmbito do qual se desenvolvem modelos de *machine learning* que requerem intervenção humana em alguma das fases da decisão algorítmica. Pelo seu desenvolvimento e pela sua centralidade nesta matéria, remetemos para o estudo de REBECCA CROTOFF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit.

Numa dimensão jurídica, o tema divide-se em vários subproblemas, que essencialmente visam dar resposta às seguintes três questões: existe o direito à intervenção de um ser humano na formação de uma decisão algorítmica – ou, mais sinteticamente, o *direito a uma decisão humana*, na fórmula abreviada de AZIZ Z. HUQ, “A Right to a Human Decision”, *Virginia Law Review*, 106/2020 (611-688)? Se sim, *em que fase*, e *reunidos que pressupostos*, poderá exigir-se essa intervenção? E qual o conteúdo, e por que regras, se rege essa intervenção? Regressaremos ao tema adiante, abordando aspetos cujo desenvolvimento aí se afigura mais justificado, revisitando a obra agora citada.

## 1.1. A explicação dos algoritmos no Regulamento Geral sobre a Proteção de Dados

3. No direito positivo, é com o Regulamento Geral sobre a Proteção de Dados (abreviadamente designado “RGPD”)<sup>9</sup> que a técnica informacional tem sido especialmente testada. No catálogo dos “Direitos dos titulares dos dados” (Capítulo III) encontramos um conjunto de preceitos que, com certa ambivalência na linguagem das situações jurídicas (tanto referem o direito de acesso à informação, como o dever de fornecer informações), cominam ao responsável pelo tratamento deveres de informação sobre os algoritmos que baseiam uma decisão automatizada (tal como referida no artigo 22.º, n.º 1, do RGPD), quando a tomada dessa decisão implique o tratamento de dados pessoais. O quadro regulativo a considerar apresenta a seguinte estrutura:

- (i) Aquando da recolha dos dados pessoais, o responsável pelo tratamento tem o dever de fornecer ao titular dos dados “informações úteis” relativas à “lógica subjacente” da decisão automatizada (incluindo a definição de perfis) que venha a tomar quanto a este titular, bem como “a importância e as consequências previstas de tal tratamento para o titular dos dados” (artigo 13.º, n.º 2, alínea f), do RGPD). Ao responsável pelo tratamento não será cominado este dever na estrita medida em que o titular dos dados já tenha conhecimento dessas informações (artigo 13.º, n.º 4, do RGPD).
- (ii) Este dever de informação estende-se também aos casos em que o responsável pelo tratamento não recolhe os dados pessoais junto do titular (artigo 14.º, n.º 2, alínea g), do RGPD). Nesse caso, deve informar o titular dos dados nos prazos fixados nas alíneas do n.º 3 do artigo 14.º do RGPD.
- (iii) O titular dos dados tem o direito de obter do responsável pelo tratamento, a todo o tempo (designadamente, após o tratamento)<sup>10</sup>, as informações abrangidas por este dever (artigo 15.º, n.º 1, alínea h), do RGPD).

---

<sup>9</sup> Regulamento (EU) n.º 679/2016, 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.

<sup>10</sup> Ainda que a propósito de outro direito, assinalando que o artigo 15.º do RGPD não especifica um prazo para o exercício dos direitos nele consagrados, cf. MATTHIAS BÄCKER, em anotação a esse artigo, em KÜLLING / BUCHNER, *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz – Kommentar*, 2. Auflage, 2018, München, C.H.BECK, Rn. 21.

4. Estas informações deverão ser prestadas pelo responsável pelo tratamento em observância dos ditames do artigo 12.º do RGPD: o responsável pelo tratamento deve adotar as medidas adequadas para transmitir esta informação “de forma concisa, transparente e de fácil acesso, utilizando uma linguagem clara e simples”, podendo ser prestada por qualquer meio (incluindo, oralmente) (n.º 1); as informações “podem ser dadas em combinação com ícones normalizados a fim de dar, de uma forma facilmente visível, inteligível e claramente legível, uma perspetiva geral significativa do tratamento previsto” (n.º 7), e o responsável pelo tratamento deve fazê-lo de forma gratuita (n.º 5), Estando em causa pedidos manifestamente infundados ou excessivos (designadamente, por serem repetitivos), o responsável pelo tratamento pode recusar-se a responder ao pedido ou exigir o pagamento de uma taxa razoável, tendo em conta os custos associados à obtenção, estruturação e fornecimento da informação (n.º 5).

O Considerando 63 do RGPD – sem o correspondente seguimento no texto normativo – refere-se aos segredos comerciais e propriedade intelectual como uma exceção a este direito à explicação: “[e]sse direito não deverá prejudicar os direitos ou as liberdades de terceiros, incluindo o segredo comercial ou a propriedade intelectual e, particularmente, o direito de autor que protege o software. Todavia, essas considerações não deverão resultar na recusa de prestação de todas as informações ao titular dos dados”.

5. Perante estas imposições dirigidas à informação do titular dos dados, a doutrina tem identificado duas possíveis bases legais para o reconhecimento, na esfera deste, de um direito à explicação da decisão<sup>11</sup>: por um lado, o complexo de normas contidas nos artigos 13.º e 14.º do RGPD, versadas pelos Considerandos 60 a 62 desse diploma, que impõem um dever de prestação de informação *espontânea* ao responsável pelo tratamento e *previamente* à formação de qualquer decisão ao titular dos dados; por outro lado, o complexo de normas contidas no artigo 15.º do RGPD que atribuem ao titular dos dados o direito de acesso a um conjunto de informações – caso em que o correspondente dever de informação do responsável pelo tratamento apenas se constituirá *a pedido* do titular dos dados e, em princípio, *posteriormente* à formação da decisão<sup>12</sup>. O artigo 15.º, n.º 1, alínea h), do RGPD, não

---

<sup>11</sup> Assim, entre outros, cf. SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, in *International Data Privacy Law*, 7/2, 2017 (76-99), pp. 79 ss.

<sup>12</sup> Esta diferenciação entre a prestação de informação espontaneamente e a prestação de informação mediante solicitação do credor da informação convoca a usual distinção entre, respetivamente,



prevê apenas um mecanismo de reação do titular dos dados perante a falta de explicações devidas pelo responsável pelo tratamento à luz dos artigos 13.º, n.º 2, alínea f), e 14.º, n.º 2, alínea g), do RGPD. Recorrendo à diferenciação entre os *Informationspflichten* e os *Auskunftsrechte* convocada a propósito da triangulação destes preceitos do RGPD (e neles epigrafada na versão alemã do diploma), consideramos que o artigo 15.º, n.º 1, alínea h), do RGPD, mais do que uma resposta a omissões ilícitas, permite também o exercício de um direito à explicação independentemente do cumprimento, pelo responsável pelo tratamento, dos deveres de explicação que lhe são impostos<sup>13</sup>. Com a ressalva, claro, da paralisação de

---

deveres de esclarecimento (*Aufklärungspflichten*) e deveres de informação (*Auskunftspflichten*). Os deveres de esclarecimento (no caso, o complexo de normas contidas nos artigos 13.º e 14.º, comentadas pelos Considerandos 60 a 62, todos do RGPD) surgem na esfera do devedor (aqui, o responsável pelo tratamento) mal se verifiquem os seus pressupostos, não se exigindo qualquer atuação do credor da informação nesse sentido. Já os deveres de informação (no caso, o complexo de normas contidas no artigo 15.º do RGPD) implicam, para a sua constituição, o exercício de uma faculdade potestativa pelo titular do direito: apenas com a eficácia do pedido de informação se constitui na esfera do devedor uma situação jurídica obrigacional em sentido próprio. Sobre esta distinção entre deveres de esclarecimento e deveres de informação, cf., entre nós, JORGE SINDE MONTEIRO, *Responsabilidade por conselhos, recomendações ou informações*, Coimbra, Almedina, 1989, pp. 409 ss.; NUNO MANUEL PINTO OLIVEIRA, *Princípios de direito dos contratos*, Coimbra, Coimbra Editora, 2011, p. 188; MIGUEL BRITO BASTOS, “Deveres acessórios de informação. Em especial, os deveres de informação do credor perante o fiador”, *Revista de Direito das Sociedades*, V/1-2 (2013), pp. 191 ss. No espaço alemão, são inesgotáveis as referências doutrinárias a esta distinção. A título meramente exemplificativo, cf., por todos, GEBHARD M. REHM, *Aufklärungspflichten im Vertragsrecht*, München, C.H. Beck, 2003, pp. 148 ss.; ANDRÉ POHLMANN, *Die Haftung wegen Verletzung von Aufklärungspflichten. Ein Beitrag zur culpa in contrahendo und zur positiven Forderungsverletzung unter Berücksichtigung der Schuldrechtsreform*, Berlin, Duncker & Humblot, 2002, pp. 22 ss.

É justamente esta distinção entre a prestação de informação a título espontâneo e a prestação de informação a pedido que justifica a utilização, nos correspondentes enunciados normativos, de uma diferente linguagem deontica: a prestação de informação espontânea é objeto de um *dever* cominado ao responsável pelo tratamento (“o responsável pelo tratamento *fornece* ao titular as seguintes informações adicionais”, “o responsável pelo tratamento *fornece* ao titular as seguintes informações, necessárias para lhe garantir um tratamento equitativo e transparente” – cf. artigos 13.º, n.º 2, e 14.º, n.º 2, do RGPD); a prestação de informação espontânea a pedido é objeto de um *direito* do titular dos dados (“O titular dos dados tem [...] o *direito* de aceder [...] às seguintes informações” – cf. artigo 15.º, n.º 1, do RGPD).

<sup>13</sup> Expressamente neste sentido, cf. EUGEN EHMANN, em anotação ao artigo 15.º do RGPD, em EHMANN / SELMAYR, *Datenschutz-Grundverordnung Kommentar*, 2. Auflage, 2018, München, C.H.BECK., Rn. 5-9. O Autor dá como exemplo a informação prestada, em tempo, em cumprimento desses deveres de explicação pelo responsável pelo tratamento que já não se encontra disponível e à qual o titular dos dados pretende (voltar a) aceder. Consideramos que o entendimento do GRUPO DE TRABALHO DO ARTIGO 29.º, vertido nas *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, 17/PT, WP251rev.01, 03.10.2017

pretensões abusivas do titular dos dados, tal como sancionadas pelo disposto no artigo 12.º, n.º 5, do RGPD<sup>14</sup>.

Ainda que não seja base legal suficiente para dele se extrair autonomamente um direito à explicação, o artigo 22.º, n.º 3, do RGPD, ao qual alude o Considerando 71 desse diploma, é também frequentemente apontado como uma das “salvaguardas” do modelo informativo acolhido no RGPD que reforçam – e, numa dimensão teleológica, justificam – a existência desse direito<sup>15</sup>. Este preceito determina que mesmo

---

(rev. 06.02.2018), <<https://ec.europa.eu/newsroom/article29/items/612053/en>> [consultado em 20.07.2022] não prejudica esta conclusão. Afirma-se, aí, a propósito do artigo 15.º, n.º 1, alínea h), do RGPD, que “[o] responsável pelo tratamento deveria já ter facultado ao titular dos dados as referidas informações, em conformidade com as suas obrigações no âmbito do artigo 13.º” (p. 28). Mas tal não significa que esta explicação não se possa referir a *mais* informações do que as versadas nos artigos 13.º, n.º 2, alínea f), e 14.º, n.º 2, alínea g), do RGPD.

<sup>14</sup> EUGEN EHMANN, em anotação ao artigo 15.º do RGPD, em EHMANN / SELMAYR, *Datenschutz-Grundverordnung Kommentar*, cit., Rn. 7.

<sup>15</sup> Recuando ao processo legislativo, revelam-nos os trabalhos preparatórios do RGPD que o Parlamento Europeu apresentou algumas alterações ao então artigo 20.º do projeto, correspondente ao atual artigo 22.º do RGPD, em particular, quanto à estatuição de um direito à explicação da decisão automatizada. Propôs, para tal, a seguinte redação para o n.º 5 do artigo 20.º: “5. **Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject shall not be based solely or predominantly on automated processing and shall include human assessment, including an explanation of the decision reached after such an assessment. The suitable measures to safeguard the data subject’s legitimate interests referred to in paragraph 2 shall include the right to obtain human assessment and an explanation of the decision reached after such assessment.**” (alterações propostas assinaladas a negrito, conforme versão original) – cf. EUROPEAN PARLIAMENT – COMMITTEE ON CIVIL LIBERTIES, JUSTICE AND HOME AFFAIRS, *Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation, A7-0402/2013*, <[https://www.europarl.europa.eu/doceo/document/A-7-2013-0402\\_EN.html](https://www.europarl.europa.eu/doceo/document/A-7-2013-0402_EN.html)> [consultado em 20.07.2022]. Este direito à explicação não ficou, contudo, plasmado na versão final do artigo 22.º do RGPD, tendo sido transferido – após proposta do Conselho Europeu nesse sentido para o Considerando 71 (cf. *Comparison of the Parliament and Council Text on the General Data Protection Regulation*, European Digital Rights International, 2016, p. 140, <[https://edri.org/files/EPCouncil Comparison.pdf](https://edri.org/files/EPCouncil%20Comparison.pdf)> [consultado em 20.07.2022]. Este documento apresenta uma comparação lado a lado dos projetos e versões preliminares do RGPD aqui referidos (a versão da Comissão Europeia e a versão do Parlamento Europeu), que integraram o *iter* legislativo, bem como das alterações à versão da Comissão Europeia propostas pelo Conselho Europeu. O Considerando 71, que versa sobre o regime do artigo 22.º do RGPD, faz expressa alusão ao direito à explicação da decisão automatizada, através de texto que é decalcado do texto proposto pelo Parlamento Europeu para o n.º 5 do então artigo 20.º: “tal tratamento deverá ser acompanhado das garantias adequadas, que deverão incluir a informação específica ao titular dos dados e o direito de obter a intervenção humana, de manifestar o seu ponto de vista, *de obter uma explicação sobre a decisão tomada na sequência dessa avaliação* e de contestar a decisão” (itálico nosso). Assim deslocado

estando em causa decisões automatizadas “permitidas” (as elencadas nas alíneas do n.º 2 do artigo 22.º do RGPD, em particular, nas alíneas a) e c)), o responsável pelo tratamento tem o dever de aplicar “medidas adequadas para salvaguardar os direitos e liberdades e legítimos interesses do titular dos dados”, designadamente “o direito de [este], pelo menos, manifestar o seu ponto de vista e contestar a decisão” (provoque, ou não, a “intervenção humana” por parte desse responsável – direito também consagrado neste preceito). A este propósito, afirma-se que estes direitos (à expressão de “um ponto de vista” e à contestação da decisão automatizada) apenas serão de exercício *possível* se o titular dos dados compreender os termos em que essa decisão automatizada foi tomada – estando aqui implicada a compreensão da composição e do processo de funcionamento do algoritmo que baseia essa decisão<sup>16/17</sup>.

---

para um considerando do RGPD, o direito à explicação não goza, por esta via, de força normativa própria (sobre a ausência de força normativa dos considerandos nos instrumentos legislativos europeus, mas reconhecendo a sua relevância no processo de interpretação da lei, cf. TODAS KLIMAS / JURATE VAICIUKAITE, “The Law of Recitals In European Community Legislation”, *ILSA Journal of International & Comparative Law*, 15/1, 2008 (63-91), pp. 83 ss.; mais recentemente, no mesmo sentido, e criticando a utilização excessiva de considerandos nos diplomas da União Europeia, cf. MAARTEN DEN HEIJER / TEUN VAN OS VAN DEN ABELEN / ANTANINA MASLYKA, “On the Use and Misuse of Recitals in European Union Law”, in *Amsterdam Law School Research Paper No. 2019-31, Amsterdam Center for International Law No. 2019-15*, <<https://ssrn.com/abstract=3445372>> [consultado em 20.07.2022].

<sup>16</sup> A visão do direito à explicação como instrumental destes direitos atribuídos ao titular dos dados pelo artigo 22.º, n.º 3, do RGPD, encontra-se, desde logo, nas *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679* do GRUPO DE TRABALHO DO ARTIGO 29.º, cit. Fazendo apelo ao Considerando 71 e ao direito à explicação nele mencionado, realça o GRUPO DE TRABALHO DO ARTIGO 29.º que “[o] titular dos dados apenas poderá contestar uma decisão ou manifestar o seu ponto de vista se compreender plenamente como foi tomada e com que fundamento” (p. 30).

<sup>17</sup> A localização da técnica informacional no RGPD não leva a que se ignore que os enunciados normativos dos quais se extrai um dever de explicação (ou direito à explicação) das decisões automatizadas não beneficiaram, desde a Diretiva 95/46/CE do Parlamento Europeu e do Conselho, de 24 de outubro de 1995, relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados (mais conhecida por Diretiva sobre a Proteção de Dados), de evolução textual significativa. Como notam SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation”, cit., p. 81, no que se refere ao artigo 22.º do RGPD, “despite years of negotiations, the final wording of the GDPR concerning protections against profiling and automated decision-making hardly changed from the relevant Articles and Recitals of the Data Protection Directive 1995. As with the GDPR, a ‘right to explanation’ does not appear in Article 15 of the Directive [...], which addresses automated individual decisions”. Da mesma forma, assinalam os autores que o direito de acesso previsto no artigo 15.º, n.º 1, alínea h), do RGPD, que garante aos titulares dos dados pessoais uma explicação da decisão automatizada, não é uma novidade do RGPD, já existindo desde a Diretiva sobre a Proteção de Dados (no seu artigo

6. Partindo do regime descrito, a doutrina tem observado, por um lado, que a regulação contida no RGPD consente uma distinção entre explicações anteriores (*ex ante*) e explicações posteriores (*ex post*) à tomada da decisão<sup>18</sup>. Por outro lado, que essa regulação impõe dois tipos de explicação das decisões algorítmicas: uma explicação modelo-cêntrica (“*model-centric explanation*”) e uma explicação “sujeito-cêntrica” (“*subject-centric explanation*”).

A primeira reporta-se à composição e funcionamento do modelo ou algoritmo utilizado (nas palavras do RGPD, acima reproduzidas, à sua “lógica subjacente”) nos processos decisórios em geral, incluindo-se nessa explicação a indicação do contexto, das categorias de dados que alimentam o modelo e sua fonte, dos fatores (as “variáveis”) que têm a aptidão de influir no sentido da decisão, bem como do seu peso relativo. Trata-se de uma explicação numa dimensão sistémica e abstrata (ou, em linguagem de computação, numa dimensão “global”), procedendo para qualquer processo decisório que tenha por base aquele algoritmo.

A explicação “sujeito-cêntrica”, por sua vez, posiciona-se numa dimensão concreta (em linguagem de computação, numa dimensão “local”), reportando-se a um específico processo decisório, perante um dado sujeito<sup>19</sup>. A versão sujeito-cêntrica da explicação admite diferentes graus de concretização.

---

12.º, alínea a)) (pp. 85 ss.). Esclarecem também que a transposição deste regime não se revelou, nos vários Estados-membros, um mecanismo de transparência efetivo. Um dos obstáculos principais enfrentados respeitou à falta de consenso, entre os Estados-membros, sobre o tipo de informação que os responsáveis pelo tratamento deveriam prestar ao titular dos dados para lhe possibilitar “o conhecimento da lógica subjacente ao tratamento automatizado dos dados que lhe digam respeito”. Ainda que a Diretiva tenha estado em vigor por um período superior a 20 anos, este obstáculo acabou por não ser ultrapassado, muito por conta da falta de desenvolvimento jurisprudencial sobre o conteúdo deste direito de acesso (p. 85). Visão diferente tem o GRUPO DE TRABALHO DO ARTIGO 29.º, manifestada *passim* nas suas *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., ao afirmar que este regime do RGPD composto pela tríade de deveres de notificação (artigos 13.º e 14.º), deveres de acesso a informação (artigo 15.º) e pelas salvaguardas contidas no artigo 22.º, n.º 3, institui um verdadeiro sistema de controlo e sindicabilidade algorítmico, inexistente no regime precedente. Na mesma linha, desenvolvidamente sobre os avanços do RGPD face à Diretiva sobre a Proteção de Dados em matéria de explicação das decisões automatizadas, cf. MARGOT E. KAMINSKI, “The Right to Explanation, Explained”, *Berkeley Technology Law Journal*, 34, 2019 (189-218), pp. 206 ss. A Autora afirma até que “this regime, if enforced, has the potential to be a sea change in how algorithmic decision-making is regulated in the EU”.

<sup>18</sup> Cf., por todos, SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit., pp. 6-7.

<sup>19</sup> Acolhemos a distinção apresentada por LILIAN EDWARDS / MICHAEL VEALE, “Slave to the Algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for”, *Duke Law & Technology Review*, 16, 2017, (18-84), pp. 55 ss.

No seu grau mínimo de concretude, a explicação não tem de ser individualizada à luz dos específicos fatores atinentes a um determinado sujeito, podendo sê-lo à luz da categoria de sujeitos em que o mesmo se integra. Esta é a modalidade da versão sujeito-cêntrica que parece estar plasmada nos artigos 13.º, n.º 2, alínea f), e 14.º, n.º 2, alínea g), ambos do RGPD, na parte em que estes se referem à explicação sobre “a importância e as consequências previstas de tal tratamento para o titular dos dados”. Quer isto dizer que tanto a explicação centrada no modelo, quanto a explicação centrada no sujeito, tal como imposta nestes enunciados normativos, podem ser fornecidas ao sujeito antes de a decisão ser tomada (a referência, no texto legal, à *previsão* das consequências é um argumento sólido nesse sentido), embora essa explicação não se torne (fisicamente) impossível se for apresentada *ex post* – não se ignorando, porém, que tal constituiria uma inobservância dos prazos para o cumprimento do dever de explicação fixados nos artigos 13.º e 14.º do RGPD.

No seu grau máximo de concretude, a explicação centrada no sujeito foca-se essencialmente nos motivos e finalidades subjacentes *àquela* decisão, bem como nas circunstâncias individuais de um concreto sujeito que influíram no processo de tomada de decisão, bem como no seu peso relativo nesse processo. Trata-se de uma explicação sobre factos pretéritos e não sobre factos virtuais. De onde segue ser esta uma explicação posterior à tomada da decisão<sup>20</sup>, que entendemos ter acolhimento, conformada como um *direito à explicação*, no artigo 15.º, n.º 1, alínea h), do RGPD. Este entendimento assenta, por um lado, na ausência de um prazo para o exercício do direito à explicação pelo titular dos dados (assim se permitindo esse exercício *ex post*) e, por outro, no imperativo de conferir uma utilidade própria a esse direito, que extravase a utilidade que se retira da cominação dos deveres de explicação operada pelos artigos 13.º, n.º 2, alínea f), e 14.º, n.º 2, alínea g), ambos do RGPD. A este propósito, já nos referimos a um conteúdo próprio do direito à explicação consagrado no artigo 15.º, n.º 1, alínea h), do RGPD, de âmbito material e de âmbito temporal mais extensos do que os dos mencionados deveres de explicação. Existem, contudo, vozes adversárias a este entendimento: negam tal extensão de um direito à explicação, sustentando que uma explicação com esse conteúdo apenas resulta do Considerando 71, e não de qualquer enunciado

---

<sup>20</sup> Neste sentido, cf. SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit., p. 6. De acordo com os Autores, esta explicação reporta-se a uma decisão particular, não ao método ou sistema de tomada de decisão em si, apenas sendo possível após a decisão ter sido tomada.

normativo do RGPD, após não ter vingado a inclusão de um direito à explicação nesses moldes no texto do então n.º 5 do artigo 20.º da Proposta de Regulamento<sup>21/22</sup>.

7. Explorando algumas imprecisões e insuficiências do texto legal, bem como variações semânticas nas diferentes versões linguísticas do RGPD, os entendimentos doutrinários não convergem no que se refere à utilidade e ao conteúdo do dever de explicação da decisão algorítmica que implique o tratamento de dados pessoais. Depuradas das suas nuances, perfilam-se, a este propósito, duas teses distintas.

8. Uma primeira orientação, mais cética, foca-se nas limitações do dever de explicação das decisões algorítmicas, duvidando da sua exequibilidade e da sua utilidade nos moldes em que esse dever se encontra formulado no texto do RGPD. Esta tese, encabeçada por Sandra Wachter, Brent Mittelstadt e Luciano Floridi<sup>23</sup>, defende que o RGPD apenas consagra um direito à explicação *ex ante*, com assento nos artigos 13.º, n.º 2, alínea f), 14.º, n.º 2, alínea g), e até no artigo 15.º, n.º 1, alínea h), por a explicação *ex post*, apesar de proposta no decurso do processo legislativo, encontrar expressão unicamente no seu Considerando 71 do RGPD<sup>24</sup>.

---

<sup>21</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit., p. 6, nt. 12, afirmando que “[t]his is the type of explanation imagined in Recital 71 GDPR, which calls for “an explanation of the decision reached after such assessment.” The Recital explicitly refers to a singular decision that has been reached”.

<sup>22</sup> Sobre as vicissitudes do direito à explicação no procedimento legislativo, cf. nota 15 *supra*.

<sup>23</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit. Esta orientação desenvolveu-se em trabalhos posteriores dos Autores. *Vide*, por exemplo, SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, in *Harvard Journal of Law & Technology*, 31, 2018 (841-887); BRENT MITTELSTADT / CHRIS RUSSELL / SANDRA WACHTER, “Explaining Explanations in AI”, in *FAT\*19: Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, January 2019 (29-31), disponível em <ssrn.com> [consultado a 20.07.2022].

<sup>24</sup> Sobre o artigo 15.º, n.º 1, alínea h), do RGPD, entendem os Autores que este preceito atribui aos titulares dos dados um direito de acesso às *mesmas* informações referidas nos artigos 13.º, n.º 2, alínea f), 14.º, n.º 2, alínea g), uma vez que os respetivos enunciados partilham da mesma redação. Ainda que explorando a possibilidade de uma explicação *ex post* ao abrigo do artigo 15.º, n.º 1, alínea h) – que *prima facie* se justificaria pela ausência de um prazo para o exercício do direito de acesso –, os Autores assinalam, porém, que a replicação do mesmo texto nestas alíneas impede, em benefício da coerência, outra interpretação. E assim é porque esse texto é claro na referência à informação sobre as “consequências previstas de tal tratamento” (na versão inglesa do RGPD, “the envisaged consequences of such processing” [ênfase nosso]). Nas palavras dos Autores, “[t]he term “envisaged” limits these predictions to ex ante explanations of system functionality, for instance concerning the general

Trata-se de entendimento que vem desafiar a ideia, inicialmente difundida e aplaudida, de que o RGPD, de forma inovadora, atribui ao titular dos dados um verdadeiro direito à explicação<sup>25</sup>: para os defensores desta tese, o fornecimento, ao titular dos dados, das informações compreendidas nos referidos preceitos não possibilita uma explicação da decisão automatizada, uma vez que estas informações não permitem ao titular dos dados conhecer as razões subjacentes, as circunstâncias e o peso dos fatores individualmente considerados na formação da decisão automatizada que quanto a eles foi tomada. Em suma, de acordo com esta tese, o titular dos dados não tem um direito à explicação da decisão automatizada, mas tem o direito a ser informado da existência de uma decisão automatizada, bem como da funcionalidade do sistema algorítmico utilizado no processo decisório<sup>26</sup>.

Esta orientação vai mais longe: não só rejeita que exista um verdadeiro direito à explicação plasmado no RGPD, como sustenta que mesmo que esse direito à explicação existisse em termos normativamente vinculativos, o mesmo apenas valeria em casos muito circunscritos, porque reservado às decisões automatizadas contestáveis ao abrigo do disposto no artigo 22.º, n.º 3, do RGPD<sup>27</sup>. Acresce que, ainda que fosse devida, em muitos casos essa explicação, a um tempo, seria de prestação tecnicamente difícil ou mesmo impossível, e, a outro tempo, ainda que

---

purpose of the system, or the type of impact to be expected from the type of decision it makes. For instance, a credit agency could predict that the scores they produce will impact on credit worthiness (e.g. interest rates). If applied to decisions already made, the phrasing becomes incoherent” (p. 17). Em reforço desta conclusão, os Autores fazem ainda apelo à redação do artigo 15.º, n.º 1, alínea h), da versão alemã do RGPD, que estabelece o seguinte: “Tragweite und *angestrebten Auswirkungen* einer derartigen Verarbeitung für die betroffene Person” (ênfase nosso). Traduzido à letra, significa que devem ser transmitidas ao titular dos dados, a seu pedido, a importância e as consequências *pretendidas* (pelo responsável pelo tratamento) de tal tratamento para o titular dos dados. De acordo com esta redação, aduzem os Autores, ao responsável pelo tratamento não é exigido que ele preveja as consequências da decisão automatizada e comunique essa previsão ao titular dos dados, mas antes que lhe explique o âmbito, a intenção e o propósito de tal processamento automatizado de dados pessoais. Novamente recorrendo às palavras dos Autores: “[t]his suggests that the right of access is not addressing how an individual decision was reached, but rather the duty of the data controller to provide information about the existence, aims and consequences of such processing. This equates to an explanation of system functionality.” (p. 18).

<sup>25</sup> Cf. BRYCE GOODMAN / SETH FLAXMAN, “European Union regulations on algorithmic decision-making and a “right to explanation”,” in *AI Magazine*, 38, 2017 (50-57).

<sup>26</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit., p. 48;

<sup>27</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit., p. 41; SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., p. 842.

não fosse impossível, pela sua tendencial ininteligibilidade, seria marginal a utilidade que o titular dos dados retiraria dessa explicação<sup>28</sup>. Por fim, os defensores desta tese assumem ainda que os responsáveis pelo tratamento não têm interesse em partilhar detalhes relativos aos algoritmos que baseiam as suas decisões automatizadas, por três razões: para evitar a violação de segredos comerciais, para evitar a violação de direitos, liberdades e garantias de terceiros, e para evitar que os titulares dos dados, munidos de informação tão detalhada, possam manipular, a seu favor, o sistema em que assenta o processo decisório automatizado<sup>29</sup>.

Esta tese teve o mérito de lançar o debate sobre a existência de um direito à explicação com fonte no RGPD – debate que, até então, era inexpressivo, apesar de o direito consagrado no artigo 15.º, n.º 1, alínea h), do RGPD, não ser uma novidade deste diploma, por já resultar, embora com uma diferente redação, do artigo 12.º, alínea a), da Diretiva sobre a Proteção de Dados.

9. Em resposta a esta tese – que foi muito criticada por se basear num estudo que instiga uma conclusão que, em rigor, não alcança<sup>30</sup> – surgiram orientações que têm como denominador comum a ideia segundo a qual uma interpretação conjugada do Considerando 71, dos artigos 13.º a 15.º e do artigo 22.º, todos do RGPD, consente que deles se extraia um direito à explicação de decisões automatizadas que impliquem o tratamento de dados pessoais<sup>31</sup>.

Esta ideia assenta num conjunto de argumentos que, muito simplificada, se prendem (i) com a atribuição de relevância interpretativa ao Considerando 71, (ii) com o reconhecimento de um sentido *explicativo* (ainda que o texto legal não

---

<sup>28</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., p. 843.

<sup>29</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., p. 843.

<sup>30</sup> Dirigindo fortes críticas ao estudo de SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit., por considerar que o título do mesmo é enganoso e irresponsável por propiciar aproveitamentos perigosos “as it invites less scrupulous or more time-pressed advocates to cite the paper for the proposition that there is no right to explanation, which is not even what the paper argues in substance”, cf. ANDREW D. SELBST / JULIA POWLES, “Meaningful information and the right to explanation”, *International Data Privacy Law*, 7/4, 2017 (233-242), p. 238.

<sup>31</sup> Cf., entre outros, ANDREW D. SELBST / JULIA POWLES, “Meaningful information and the right to explanation”, cit., GIANCLAUDIO MALGIERI / GIOVANNI COMANDÉ; “Why a Right to Legibility of Automated Decision-Making exists in the General Data Protection Regulation”, *International Data Privacy Law*, 7/4, 2017 (243-265); MARGOT KAMINSKI, “The Right to Explanation, Explained”, cit.



mencione – nem precise de mencionar – a “explicação”) das exigências decorrentes dos artigos 13.º a 15.º relativas à prestação de “informações úteis” sobre a “lógica subjacente” ao titular dos dados<sup>32</sup>, e (iii) com a pronúncia, neste sentido, do Grupo de Trabalho do Artigo 29.º, nas suas *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, ao defender a existência de um dever de explicação da decisão automatizada com fonte nos artigos 13.º e 14.º do RGPD, e de um direito à explicação fundado no artigo 15.º do RGPD, com vista a assegurar as garantias necessárias a que se referem o artigo 22.º, n.º 3, e o Considerando 71 do RGPD<sup>33</sup>.

Apesar de defenderem a existência de um dever de explicação e de um direito à explicação da decisão automatizada, estas teses não desconsideram a complexidade – e, em muitos casos, a onerosidade excessiva que impende sobre o responsável pelo tratamento – do fornecimento de uma explicação *útil* ao titular dos dados. É no quadro deste reconhecimento que se compreende o apelo, feito por vários autores desta ala doutrinária, à *legibilidade*, *interpretabilidade* ou *inteligibilidade* da explicação sobre a decisão automatizada, em colisão com uma noção de explicação assente na comunicação assética do algoritmo ou do seu modo de funcionamento<sup>34</sup>. Em suporte desta ideia, o Grupo de Trabalho do Artigo 29.º invoca também, com exemplos (a decisão automatizada para classificar o risco de crédito de um cliente

---

<sup>32</sup> Afirmam lapidarmente ANDREW D. SELBST / JULIA POWLES, “Meaningful information and the right to explanation”, cit., p. 233, que “Articles 13-15 provide rights to ‘meaningful information about the logic involved’ in automated decisions. This is a right to explanation, whether one uses the phrase or not”.

<sup>33</sup> Cf. GRUPO DE TRABALHO DO ARTIGO 29.º, *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., pp. 27-28.

<sup>34</sup> O estudo de GIANCLAUDIO MALGIERI / GIOVANNI COMANDÉ; “Why a Right to Legibility of Automated Decision-Making exists in the General Data Protection Regulation”, cit., destaca-se neste contexto. Reconhecendo a falência, a este propósito, de uma noção “pura” de transparência algorítmica – que levaria a que ao titular dos dados fosse comunicado o modelo algorítmico ou o seu modo de funcionamento em linguagem o mais precisa possível (ou seja, em linguagem técnica) –, os Autores defendem um “modelo combinado de transparência e de compreensibilidade”, assente em três pilares: *readability*, *explanation* e *legibility* (p. 245). Propõem, também, com apoio no artigo 15.º, n.º 1, alínea h), do RGPD, a utilização de um teste de legibilidade do processo decisório algorítmico (“*legibility test*”) composto por um questionário dividido em dois passos: o primeiro, incidente sobre a *arquitetura* desse processo (o modelo e a lógica do seu funcionamento – nos dizeres do preceito, a “lógica subjacente”); o segundo, sobre a sua *implementação* (o seu uso contextual, importância e consequências – nos dizeres do preceito, “a importância e as consequências previstas de tal tratamento para o titular dos dados”) (p. 258). Para os Autores, este teste de legibilidade permite uma abordagem “legibility-by-design” das decisões automatizadas reguladas no RGPD, embora reconheçam eventuais dificuldades de diversificação e de adaptação do questionário proposto a todos processos decisórios algorítmicos (p. 265).

bancário; a decisão automatizada para definir os prémios de seguro automóvel com base na monitorização do comportamento ao volante dos clientes), a necessidade de compreensibilidade dessa explicação pelo titular dos dados, apontando, como características da mesma, a *simplicidade*, a *completude suficiente* e a *concretização*<sup>35</sup>, sem, contudo, exigir que se trate de uma explicação sobre a decisão individual, podendo ser prestada em moldes genéricos<sup>36</sup>. A necessidade de uma explicação inteligível funda-se numa interpretação do direito à explicação teleologicamente orientada a possibilitação do exercício dos direitos conferidos ao titular dos dados, destinatário da decisão automatizada, pelo artigo 22.º, n.º 3, do RGPD, e é nesses moldes que o seu conteúdo deve ser definido<sup>37</sup>.

---

<sup>35</sup> A propósito da exigência de prestação de “informações úteis” relativas à “lógica subjacente”, o GRUPO DE TRABALHO DO ARTIGO 29.º destaca que “o crescimento e a complexidade da aprendizagem automática poderão tornar difícil perceber o funcionamento do processo de decisão automatizada ou da definição de perfis. O responsável pelo tratamento deverá encontrar formas simples de comunicar ao titular dos dados a lógica subjacente, ou os critérios aplicados para tomar a decisão”, sendo que “[o] RGPD obriga o responsável pelo tratamento a fornecer informações úteis relativas à lógica subjacente, e não necessariamente uma explicação complexa sobre os algoritmos utilizados ou a divulgação do algoritmo na íntegra”. Em qualquer caso, salvaguarda que “[a]s informações prestadas devem, no entanto, ser suficientemente completas para permitir ao titular dos dados compreender os motivos da decisão” (cf. *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., p. 28). O GRUPO DE TRABALHO DO ARTIGO 29.º esclarece também, recorrendo ao Considerando 58 do RGPD, que quanto mais complexo é o algoritmo que baseia o processo decisório (e, admite-se, a sua explicabilidade), mais importante é o fornecimento dessa explicação: “A complexidade não serve de justificação para que não sejam fornecidas as informações ao titular dos dados. O Considerando 58 indica que o princípio da transparência é «especialmente relevante em situações em que a proliferação de operadores e a complexidade tecnológica das práticas tornam difícil que o titular dos dados saiba e compreenda se, por quem e para que fins os seus dados pessoais estão a ser recolhidos, como no caso da publicidade por via eletrónica»” (p. 28).

<sup>36</sup> Não é claro se o GRUPO DE TRABALHO DO ARTIGO 29.º se opõe ao fornecimento de explicações sobre a decisão individualmente tomada (ou a tomar) pelo responsável pelo tratamento. Nas suas *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., p. 28, é inequívoca a utilização do termo “dever” para caracterizar a situação jurídica do responsável pelo tratamento perante quem é exercido o direito consagrado no artigo 15.º, n.º 1, alínea h), do RGPD: “Ao exercer os seus direitos previstos no artigo 15.º, o titular dos dados pode tomar conhecimento de uma decisão que lhe diga respeito, incluindo se for baseada na definição de perfis. O responsável pelo tratamento *deve* facultar ao titular dos dados informações genéricas (nomeadamente, sobre os fatores tidos em conta no processo decisório e a «relevância» dos mesmos em termos globais) que também lhe sejam úteis para contestar a decisão” (ênfase nosso). O GRUPO DE TRABALHO DO ARTIGO 29.º parece, assim, conformar este dever de explicação, constituído na sequência do dito exercício pelo titular dos dados, como sendo um dever de explicação do processo decisório de *qualquer* decisão automatizada da mesma categoria.

<sup>37</sup> Segundo ANDREW D. SELBST / JULIA POWLES, “Meaningful information and the right to explanation”, cit., p. 242, “[t]he right to explanation should be interpreted functionally, flexibly, and should, at a

A esta justificação acrescentam vantagens de ordem prática evidentes: o entendimento das exigências decorrentes dos artigos 13.º a 15.º do RGPD relativas à prestação de “informações úteis” sobre a “lógica subjacente” como uma explicação da decisão algorítmica e não de uma divulgação integral do algoritmo permite afastar a maioria das reservas – inicialmente apresentadas contra essa explicação – relativas à suscetibilidade de manipulação do sistema algorítmico pelo destinatário da decisão e à proteção dos segredos comerciais que envolvem a produção ou invenção desse sistema<sup>38</sup>.

10. O quadro semântico em que se debate a existência de um dever – e de um (não necessariamente correlativo *prima facie*<sup>39</sup>) direito – de explicação da decisão automatizada que implica o tratamento de dados pessoais, ao apelar à utilidade da informação a transmitir ao titular dos dados (“informações úteis”), serve ainda de base à discussão sobre a *melhor* forma de explicar a decisão algorítmica.

Curiosamente, foi entre os defensores da orientação que nega que o RGPD seja fonte de um direito à explicação da decisão automatizada individualmente considerada que se recolheram aos contributos mais valiosos sobre a idealização de uma “boa” explicação. Mais concretamente, Sandra Wachter, Brent Mittelstadt e Chris Russell, partem das insuficiências do RGPD com vista a expandir o tema para dar resposta à “questão clássica de saber como podem as decisões algorítmicas ser explicadas aos especialistas [*experts*]”, bem como às partes não especialistas [*non-expert parties*] afetadas pela decisão”<sup>40</sup>. Consideram os Autores que a explicação

---

minimum, enable a data subject to exercise his or her rights under the GDPR and human rights law”.

<sup>38</sup> Notando e criticando a incoerência das teses que negam ou propugnam uma interpretação restritiva do direito à explicação da decisão automatizada *também* com fundamento na falta de normatividade do Considerando 71 do RGPD, e que aduzem, em benefício dessa posição, a necessidade de proteção dos segredos comerciais, invocando, para tal, o Considerando 63 do RGPD (“Esse direito [de acesso] não deverá prejudicar os direitos ou as liberdades de terceiros, incluindo o segredo comercial ou a propriedade intelectual e, particularmente, o direito de autor que protege o software. Todavia, essas considerações não deverão resultar na recusa de prestação de todas as informações ao titular dos dados”), cf. GIANCLAUDIO MALGIERI / GIOVANNI COMANDÉ; “Why a Right to Legibility of Automated Decision-Making exists in the General Data Protection Regulation”, cit., pp. 262 ss.

<sup>39</sup> Pois, como já referido, o direito à explicação atribuído pelo artigo 15.º, n.º 1, alínea h), do RGPD, ao titular dos dados não é apenas uma situação jurídica ativa constituída por reflexo do dever de explicação que se extrai dos artigos 13.º, n.º 2, alínea f), e 14.º, n.º 2, alínea g), do RGPD, tendo um conteúdo próprio, que extravasa a utilidade desses deveres. Usamos aqui a noção de *correlatividade* desenvolvida por WESLEY NEWCOMB HOHFELD, *Os conceitos jurídicos fundamentais aplicados na argumentação jurídica* (trad. Margarida Lima Rego), Lisboa, FCG, 2008, pp. 25 ss.

<sup>40</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., p. 861. Os Autores afirmam mesmo que “we therefore propose to move past the limitation of the GDPR and to use

de qualquer decisão algorítmica – seja, ou não, a decisão automatizada versada no RGPD – serve um de três propósitos: permite ao destinatário da decisão *compreendê-la*, permite *contestá-la* e permite conhecer o que deveria ser diferente para, no futuro, *alterar a realidade atual* para alcançar a decisão pretendida (desde que baseada nesse modelo decisório)<sup>41</sup>.

Para tal, apresentam um modelo explicativo das decisões algorítmicas, a que chamam “explicações contrafactuais incondicionais” (“*unconditional counterfactual explanations*”), baseado em proposições contrafactuais, ou seja, proposições fácticas que, por definição, no seu modelo abstrato, não só não se verificam no presente nem se verificaram no passado, como se assume que não se verificarão no futuro<sup>42</sup>. Muito simplificado, é contrafactual o evento ou facto que não se verificou, mas que se poderia ter verificado. Já o evento ou facto que se verificou é designado de atual. O evento ou facto contrafactual faz parte de um mundo possível que contradiz algo do mundo real, enquanto o evento atual faz parte do mundo real<sup>43</sup>. Ainda que reconhecendo que a explicação contrafactual “excede os objetivos e requisitos dos mecanismos de transparência, dos deveres de informação e do direito de acesso plasmados no RGPD”<sup>44</sup>, e que se distancia do conteúdo das múltiplas variações construtivas sobre o conteúdo da explicação que a doutrina tem vindo a formular, assentes essencialmente na transmissão do sistema algorítmico e do seu modo de funcionamento, aponta-se a esta proposta de modelo explicativo contrafactual as vantagens de ser facilmente convertido em linguagem computacional<sup>45</sup> e de não implicar a revelação

---

counterfactuals as unconditional explanations. These unconditional explanations should be given whenever requested, regardless of outcome (positive or negative decision), whether the decision was based on solely automated processes and their (legal or similar significant) effects” (p. 881).

<sup>41</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., p. 843.

<sup>42</sup> Os Autores apresentam o seguinte exemplo de explicação contrafactual de uma decisão algorítmica: “Foi-lhe recusada a concessão de crédito porque o seu rendimento anual é de £ 30.000. Se esse rendimento fosse de £ 45.000, o crédito ter-lhe-ia sido concedido” (cf. SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., p. 844). Neste caso, à comunicação da decisão (a recusa da concessão de crédito) segue-se um evento contrafactual, ou uma declaração sobre como “o mundo teria de ser diferente” para o resultado desejado ocorrer.

<sup>43</sup> Sobre este tipo de proposições, que define como proposições sobre *mundos possíveis ou ideais*, cf. DAVID LEWIS, *Counterfactuals*, Malden, Harvard University Press, 1973, pp. 1 ss.; RENAAT DECLERCK / SUSAN REED, *Conditionals: A Comprehensive Empirical Analysis*, New-York, De Gruyter Mouton, 2001, pp. 37 ss.

<sup>44</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., p. 871.

<sup>45</sup> Apresentando exemplos de modelos computacionais explicativos, incluindo com recurso a imagens ou símbolos, cf. SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual

desse sistema e do seu funcionamento ao destinatário da decisão (assim se protegendo segredos comerciais e, em geral, direitos de terceiros), bastando-se com a descrição de uma dependência da decisão de certos factos ou eventos externos que conduzem (ou poderiam ter conduzido) a essa ou a outra decisão<sup>46</sup>. Nas palavras dos Autores, possibilita-se o fornecimento de uma explicação “sem abrir a *blackbox*”.

11. O modelo de explicação da decisão algorítmica baseada em contrafactuais recebeu forte adesão, incluindo fora do espaço europeu. Desde logo, pela sua apontada similitude com as explicações exigidas, no direito estado-unidense, para a recusa de concessão de crédito bancário<sup>47</sup>. Promissora parecia ser, à partida, a explicação que não é contaminada pela complexidade do sistema algorítmico a que respeita, que não implica a divulgação desse sistema e que dá a conhecer ao destinatário da decisão o que precisa ou precisava de ter, ser ou fazer para obter uma decisão diferente.

Uma reflexão crítica a este modelo, ao apelidar estas explicações de “acionáveis” (*actionable explanations*) e de assinalar que as mesmas são alicerçadas num conjunto

---

explanations without opening the black box: automated decisions and the GDPR”, cit., pp. 854 e ss. Os Autores esclarecem ainda que havendo vários “mundos possíveis”, as explicações contrafactuais devem refletir isso mesmo: “[a]s multiple outcomes based on changes to multiple variables may be possible, a diverse set of counterfactual explanations should be provided, corresponding to different choices of nearby possible worlds for which the counterfactual holds” (p. 881).

<sup>46</sup> Cf. SANDRA WACHTER / BRENT MITTELSTADT / CHRIS RUSSELL, “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, cit., pp. 845, 871.

<sup>47</sup> Cf. SOLON BAROCAS / ANDREW D. SELBST / MANISH RAGHAVAN, “The hidden assumptions behind counterfactual explanations and principal reasons”, in Mireille Hildebrandt *et al.* (eds.), *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, ACM, 2020 (80-89), p. 80. Os Autores reportam-se, em particular, ao tipo de “*feature-highlighting explanation*” que há muito é exigida nos Estados Unidos da América – desde o *Fair Credit Reporting Act* (FCRA), de 1970, e desde o *Equal Credit Opportunity Act* (ECOA), de 1974, e, em particular, no seu *Regulation B*. Exige-se a todos os sujeitos que utilizem informação relevante para a concessão de crédito que forneçam aos consumidores uma explicação para as decisões que lhes são desfavoráveis (recusa de concessão de crédito, concessão de crédito em condições adversas); essa explicação caracteriza-se por ser composta por uma lista de razões principais ou razões específicas para essa decisão desfavorável. No *Appendix C to Part 1002 – Sample Notification Forms* do *Regulation B* é apresentada uma lista não exaustiva de razões tidas como aceitáveis para essa decisão desfavorável: “*Income insufficient for amount of credit requested*”, “*Excessive obligations in relation to income*”, “*No credit file*”, “*Poor credit performance with us*”, “*Delinquent past or present credit obligations with others*” ou “*Temporary or irregular employment*”. De acordo com o *Regulation B*, “The statement of reasons for adverse action required by paragraph (a)(2)(i) of this section must be specific and indicate the principal reason(s) for the adverse action. Statements that the adverse action was based on the creditor’s internal standards or policies or that the applicant, joint applicant, or similar party failed to achieve a qualifying score on the creditor’s credit scoring system are insufficient.” (32, § 1002.9(b)(2)).

de pressupostos determinados subjetivamente, tem exposto, porém, um conjunto “tensões inevitáveis”<sup>48</sup>. Essas críticas ancoram-se essencialmente no facto de as explicações contrafactuais, ao apresentarem ao destinatário da decisão algorítmica o cenário ou a simulação dos mundos possíveis, instigam – embora não imponham – alterações às circunstâncias relativas a esse destinatário, no que de uma atuação deste possam depender. Por outras palavras, as explicações contrafactuais levam a que o destinatário conforme a sua conduta ou modele as suas circunstâncias para obter uma decisão que considere favorável, nos termos que aquelas explicações apresentam como mundos possíveis *desejáveis*. Aqui, não só assoma o problema da escolha dos cenários tidos como desejáveis, como se revela o que é designado de “paradoxo da autonomia”: as explicações visam proteger o destinatário da decisão da posição mais forte dos decisores (força essa que lhes advém do facto de não estarem obrigados a decidir, nem a decidir num determinado sentido) e corrigir alguns desequilíbrios nessa relação de poder. Contudo, ao instigar uma conformação das circunstâncias do destinatário da decisão – e, com isso, à partida, uma alteração dos seus comportamentos –, em função de certos fatores ou condições que o decisor *valoriza*, as explicações contrafactuais mais do que corrigir, catalisam esse desequilíbrio. Sobretudo quando à eleição dos cenários ou mundos possíveis (e desejáveis) pelo decisor subjazem razões paternalistas ou fenómenos de *nudging* encobertos<sup>49</sup>.

Adiante – a propósito do conteúdo do dever de explicação da decisão algorítmica – se desenvolverá, com mais pormenor, o cabimento e a projeção destas críticas.

## 1.2. Justificação e delimitação do tema

12. O tema deste estudo situa-se no âmago do modelo informacional do direito dos algoritmos e é fundamentalmente instigado pelo regime aqui descrito do Regulamento Geral sobre a Proteção de Dados. Tem, contudo, uma vocação mais ambiciosa, por não se limitar às decisões algorítmicas que impliquem o tratamento de dados pessoais. O que se pretende saber é se o decisor algorítmico que atue em contexto negocial está adstrito a algum dever de informar o destinatário da decisão sobre os aspetos envolvidos na tomada dessa decisão. Em particular, (i)

---

<sup>48</sup> Cf. SOLON BAROCAS / ANDREW D. SELBST / MANISH RAGHAVAN, “The hidden assumptions behind counterfactual explanations and principal reasons”, cit., pp. 82 ss.

<sup>49</sup> Cf. SOLON BAROCAS / ANDREW D. SELBST / MANISH RAGHAVAN, “The hidden assumptions behind counterfactual explanations and principal reasons”, cit., p. 86. Os Autores afirmam, até, que “[i]ronically, respecting a decision subject’s autonomy will require making assumptions about which information will be valuable to a given decision subject”.

sobre a natureza algorítmica da decisão, (ii) sobre a composição e funcionamento do algoritmo utilizado, (iii) sobre os motivos subjacentes à escolha dessa composição, e (iv) sobre a relevância do resultado algorítmico na formação da decisão. Por simplicidade, e usando expressão já vulgarizada, falaremos da “explicação” da decisão algorítmica para nos referirmos à comunicação destes factos – e não da própria decisão – ao destinatário desta.

A explicabilidade das decisões algorítmicas, numa dimensão pragmática, tem sido alvo de desenvolvimentos assinaláveis: é hoje uma subdisciplina da inteligência artificial designada como “*explainable AI*” (ou abreviadamente “*xAI*”), focada na criação de processos e modelos tecnológicos que possibilitem a compreensão humana das decisões ou previsões resultantes do funcionamento de sistemas de inteligência artificial<sup>50</sup>.

13. Confrontados com a propalada ideia de “opacidade” dos algoritmos, o interesse dos juristas sobre a interpretabilidade e explicação das decisões algorítmicas acompanha estes desenvolvimentos. Ainda que de manifestação mais frequente no quadro dos processos decisórios algorítmicos que impliquem o

---

<sup>50</sup> Para uma visão geral – e também jurídica – dos fundamentos e finalidades deste ramo da inteligência artificial, remetemos para o estudo de BRENT MITTELSTADT / CHRIS RUSSELL / SANDRA WACHTER, “Explaining Explanations in AI”, cit., disponível em <ssrn.com> [consultado a 20.07.2022]. Os Autores identificam duas abordagens distintas na “Explainable AI”: a transparência (que se preocupa com a explicação da forma como o algoritmo funciona internamente) e a interpretação *post hoc* ou póstuma (que visa explicar quais os resultados ou consequências do funcionamento desse algoritmo e porquê). Todavia, a inteligência artificial explicável pode não ser a panaceia que se esperava: como afirma lapidariamente ZACHARY C. LIPTON “The Mythos of Model Interpretability: in Machine Learning, the Concept of Interpretability is both Important and Slippery”, in *Queue*, 16/3, 2018 (31-57), disponível em <queue.acm.org> [consultado em 20.07.2022], p. 33, “[s]ome suggest model interpretability as a remedy, but in the academic literature, few authors articulate what interpretability means or precisely how their proposed solution is useful”. A ilusoriedade da explicação inútil tem sido uma das principais reservas aos préstimos da inteligência artificial explicável. A este propósito, merece também nota o estudo de MARANKE WIERINGA, “What to account for when accounting for algorithms. A systematic literature review on algorithmic accountability”, in *FAT\* ’20, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 2020 (1-18), disponível em <ssrn.com> [consultado a 20.07.2022]. Numa revisão da literatura sobre o tema, a Autora foca-se na necessidade de identificar critérios para uma “boa” explicação: “[d]ifferent researchers have different ideas about the term explanation, so it is not yet clear how to objectively know which form of XAI is good or better/worse than others for a specific domain. To answer this, “goodness” criteria are needed. But there is a lack of literature about which form of explanation (e.g., global, local, counterfactual) is best and how much information is suitable for human data subjects. Thus researchers should attempt to theoretically and empirically develop goodness criteria for the practical use of AI. A core research problem is to understand the features that make for a beneficial explanation of an AI system”.

tratamento de dados pessoais e, nesse âmbito, das decisões tomadas por sujeitos de direito público em relação a particulares, ou de decisões judiciais algorítmicas<sup>51</sup>, encontram-se alguns estudos recentes que tematizam o processo decisório algorítmico em contexto negocial de direito privado – ou seja, em ambiente moldado pela autonomia privada dos sujeitos nele envolvidos, e que se pode referir a qualquer fase do negócio jurídico (formação, execução ou extinção) ou aos seus preliminares<sup>52</sup>.

14. A delimitação do tema nestes moldes impõe um esclarecimento, com vista à justificação da sua singularidade. O problema da interpretabilidade e da explicação de decisão algorítmica tomada em contexto negocial não é desafiante apenas por representar uma ampliação das questões discutidas no estrito quadro das decisões algorítmicas que impliquem o tratamento de dados pessoais (sobre as quais, como visto, já se debruçam a doutrina e a jurisprudência). É certo que, feita esta ampliação, enfrenta-se a dificuldade de identificação da fonte do dever de explicação de *qualquer decisão algorítmica*. O problema aqui tratado é desafiante também pelas dificuldades de identificação de um dever de explicação de *qualquer decisão, algorítmica ou não*, tomada em contexto negocial.

A complexidade do tema prende-se, a montante, com a ausência de regulação civil específica e, portanto, de regra da qual se possa extrair esse dever de explicação. A imposição, a qualquer decisor, da explicação da decisão tomada num determinado contexto negocial corresponde a uma compressão da autonomia privada deste – sendo nesse quadro que devem ser mobilizados os critérios para a descoberta da fonte de tal dever. O mesmo vale para as decisões algorítmicas tomadas em contexto negocial: a cominação de um dever de explicação ao decisor algorítmico corresponde igualmente a uma compressão da autonomia privada deste. Pelo que, na ausência de fonte legal (específica) ou contratual desse dever, também procede quanto a este tipo de decisões o mesmo modelo de mobilização de critérios para a descoberta da fonte de um dever de explicação.

---

<sup>51</sup> A este propósito, remetemos para a monografia recente de DAVID NINK, *Justiz und Algorithmen. Über die Schwächen menschlicher Entscheidungsfindung und die Möglichkeiten neuer Technologien in der Rechtsprechung*, Berlin, Duncker & Humblot, 2021.

<sup>52</sup> Cf., por exemplo, o estudo de DOMINIK SCHÄFFERS, “Rechtsgeschäftliche Entscheidungsfreiheit im Zeitalter von Digitalisierung, Big Data und Künstlicher Intelligenz Überlegungen am Beispiel der individualisierten Vertragsanbahnung”, *Archiv für die zivilistische Praxis*, 1-2/221, 2021 (32-67). Merece também destaque o texto de LAUREN HENRY SCHOLZ, “Algorithms and Contract Law”, in Woodrow Barfield (ed.), *The Cambridge Handbook of the Law of Algorithms*, cit. (141-152).



15. Admite-se, porém, que se possa estranhar esta equiparação entre decisões humanas e decisões algorítmicas tomadas em contexto negocial. A desconfiança dirigida aos algoritmos – adensada pelo discurso dos juristas e pelos avanços normativos nesta matéria – explica essa estranheza. Epidermicamente, as decisões algorítmicas são vistas como sendo “mais perigosas” do que as decisões humanas. Veremos se esta ideia merece ser desconstruída, e se a explicação da decisão em contexto negocial é matéria que deva convergir num esquema unitário de regulação.

### 1.3. A proibição de decisões algorítmicas como novo discurso

16. O que questionamos neste estudo não fica comprometido ou prejudicado pelo novo rumo que o discurso jurídico sobre algoritmos está a tomar.

A abordagem corretiva e informacional que caracteriza a primeira fase desse discurso – e que é o berço da ideia de explicação dos algoritmos – tem sido criticada por *legitimar* a utilização de decisões algorítmicas em qualquer ambiente negocial, desde que o procedimento de formação das mesmas obedeça aos parâmetros regulativos aplicáveis, cujas fontes vão desde a *hard law* à (discutível) *soft law*<sup>53</sup>.

---

<sup>53</sup> Entre as vozes mais críticas desta legitimação, cf. ARI EZRA WALDMAN, “Algorithmic Legitimacy”, cit., pp. 107 ss. Para sustentar estas críticas, o Autor parte das fragilidades das concepções de autoridade (do direito) baseadas na correção procedimental dos atos praticados pelas instituições de autoridade (por exemplo, a sentença de um tribunal ou a decisão de aplicação de uma coima por uma entidade administrativa), como as de TOM. R. TYLER, *Why People Obey the Law*, 2<sup>nd</sup> ed., New Jersey, Princeton University Press, 2006, pp. 3-8 (que associa a legitimidade de um ato de autoridade, do prisma do destinatário desse ato, à justiça procedimental [“*procedural justice*”, “*fair decision-making procedures*”] que envolve a formação desse ato, e não ao seu concreto resultado [“*outcome*”]), invocando ainda o liberalismo de JOHN RAWLS, em *A Theory of Justice. Original Edition*, Cambridge, MA, Harvard University Press, 2005 (1971), “Justice as Fairness: Political Not Metaphysical”, *Philosophy and Public Affairs*, 14/1985 (223-252), *Political Liberalism*, New York, Columbia University Press, 1993, ou, mais significativamente, *Justice as Fairness – A Restatement* (ed. Erin I. Kelly), Cambridge, MA, Harvard University Press, 2001 (em particular, o postulado de que a justiça dos atos de autoridade é alcançada não com base em *standards* ou num conceito de justiça formulado *a priori*, mas através da garantia de uma “justiça procedimental perfeita” [“*perfect procedural justice*”]). Segundo ARI EZRA WALDMAN, embora reconhecendo que “[i]n certain contexts, fair procedures are sufficient to determine lawfulness” (p. 119), no contexto algorítmico não procede a ideia de que “*fair processes lead to fair results*”: a legitimidade das decisões algorítmicas através do aperfeiçoamento dos sistemas tecnológicos e do fortalecimento das garantias dos destinatários das decisões falha por assentar e recorrer a instrumentos corretivos do processo decisório humano, que são desadequados para processos decisórios automatizados (“put another way, they are trying to bring the things that make human decision-making legitimate – fair procedure, respect for persons, and the opportunity to be heard – into the algorithmic context”) (p. 110). Para o Autor, a dissemelhança entre o processo decisório humano e o processo decisório algorítmico que justifica este seu entendimento (e que determina a desadequação

Na base destas críticas está a ideia de que a correção procedimental de uma decisão algorítmica, ainda que seja uma “pátina de legitimidade” da mesma<sup>54</sup>, não garante aprioristicamente a “justiça” do seu resultado<sup>55</sup>. Tem-se intensificado, por isso, a

---

das soluções da primeira fase) encontra-se fundamentalmente na virtude da *eficiência* (e não da *justiça*) que condiciona de forma determinante todo o processo decisório algorítmico.

<sup>54</sup> A expressão é de FRANK PASQUALE, “The Second Wave”, cit.

<sup>55</sup> Ainda que a “justiça” (ou “injustiça”) da decisão algorítmica seja questão implicada num conjunto de aspetos abordados adiante, não cabe aqui discutir, numa dimensão moral ou ética, nem sequer numa dimensão normativa, os diferentes conceitos e as diferentes concepções de «justiça» – enquanto *valor* e enquanto *virtude* – das decisões negociais em geral, e das decisões negociais algorítmicas em particular. Muito menos nos debruçaremos sobre a ideia de «bem», de «bom» e suas derivações. Seria tarefa ciclópica e interdisciplinar e, por isso, necessariamente incompleta, para além de não ter lugar adequado num texto desta natureza. Para uma visão de conjunto das relações entre o direito e a justiça, numa perspetiva histórica e jusfilosófica, limitamo-nos a remeter para os estudos de ARTHUR KAUFMANN, *Theorie der Gerechtigkeit. Problemgeschichtliche Betrachtungen*, Würzburger Vorträge zur Rechtsphilosophie, Rechtstheorie und Rechtssoziologie, Heft 1, Frankfurt am Main, 1984, ou, entre nós, de ANTÓNIO PEDRO BARBAS HOMEM, *O justo e o injusto*, Lisboa, AAFDL Editora, 2001. Num esforço singular, exclusivamente dirigido aos propósitos deste estudo – e descomprometido de quaisquer ideais de neutralização axiológica do direito – parte-se de duas ideias mais básicas: em primeiro lugar, a ideia de que a justiça corresponde a uma qualidade de uma concreta decisão, não sendo uma qualidade atribuída ao ordenamento jurídico positivo enquanto tal – cf. ANTÓNIO CASTANHEIRA NEVES, *Questão-de-facto – questão-de-direito ou o problema metodológico da juridicidade (ensaio de uma reposição crítica)*, vol. I, Coimbra, Almedina, 1967, pp. 586 ss.; em segundo lugar, admitindo que “o direito não é pensável sem a referência ao valor da justiça” (cf. MIGUEL TEIXEIRA DE SOUSA, *Introdução ao Direito*, Coimbra, Almedina, 2012, p. 81), e concedendo que “o direito é a justiça pensada” (cf. WOLFGANG FIKENTSCHER, *Methoden des Rechts in vergleichender Darstellung*, vol. IV, *Dogmatischer Teil*, Tübingen, Mohr Siebeck, 1977, p. 6), considerar-se-á que certa decisão algorítmica tomada em contexto negocial constitui um ato de injustiça na estrita medida da ilicitude da conduta subjacente à tomada dessa decisão, ou da invalidade que fere essa decisão *qua tale*. Circunscritos a um contexto negocial, modelado pela autonomia privada dos sujeitos que nele atuam, não estamos, aqui, a falar das decisões (judiciais) que “dizem o direito” (*iuris-dictio*): quanto a estas, releva antes “a «justiça» no sentido que atende, nomeadamente, à factualidade e historicidade das fontes, à limitação do conhecimento dos factos (e à proibição de *non liquet*), à diversidade das razões (p. ex., distributivas, ou políticas em geral) relevantes para cada questão (p. ex., comutativa e privada) litigiosa, e à contingência do processo (jurídico-processual) decisório” (cf. PEDRO FERREIRA MÚRIAS, *A análise axiológica do direito civil*, Dissertação de doutoramento, polic., Lisboa, FDUL, 2021, p. 306, nt. 788). Não se ignora, porém, o cariz demasiado redutor desta nossa pretensão de simplificação, por deixar de fora uma discussão profunda sobre o conceito e concepções de «justiça» nas decisões tomadas por sujeitos de direito privado, em particular, em ambiente negocial – para uma apresentação recente do lugar da «justiça» na dogmática do direito privado moderno, vide MATTHIAS WENDLAND, *Vertragsfreiheit und Vertragsgerechtigkeit. Subjektive und objektive Gestaltungskräfte im Privatrecht am Beispiel der Inhaltskontrolle Allgemeiner Geschäftsbedingungen im unternehmerischen Geschäftsverkehr*, Tübingen, Mohr Siebeck, 2019, em especial, pp. 105 ss., merecendo ainda destaque o conhecido modelo alocativo do direito civil desenvolvido por CHRISTIANE WENDEHORST na sua

defesa de modelos proibitivos, que assentam no reconhecimento da necessidade de refrear ímpetus expansionistas da utilização de sistemas algorítmicos em qualquer domínio negocial, apesar dos desenvolvimentos dos últimos anos dirigidos ao aperfeiçoamento dos algoritmos e à cominação de deveres de informação ao “decisor algorítmico” no quadro da referida primeira fase<sup>56</sup>.

17. Na viragem para a segunda fase do discurso dos juristas sobre algoritmos, a questão medular passou a ser a seguinte: deverá proibir-se ou limitar-se a utilização de decisões algorítmicas em contextos negociais dirigidos à regulação de determinadas matérias, ou no qual intervenham determinadas categorias de pessoas? A resposta é, hoje, maioritariamente positiva, embora (ainda) sem manifestações inequívocas de direito positivo.

O artigo 22.º, n.º 1, do Regulamento Geral da Proteção de Dados Pessoais, contém uma *proibição fraca e, em qualquer caso, ilusória* de utilização de decisões algorítmicas que impliquem o tratamento de dados pessoais. De acordo com este preceito, “o titular dos dados tem o direito de não ficar sujeito a nenhuma decisão

---

*Habilitationsschrift*, publicada como *Anspruch und Ausgleich. Theorie einer Vorteils- und Nachteilsausgleichung im Schuldrecht*, Tübingen, Mohr Siebeck, 1999.

<sup>56</sup> Como afirma JULIA POWLES, em “The Seductive Diversion of ‘Solving’ Bias in Artificial Intelligence”, *One Zero*, 7-dez.-2018, disponível em <<https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53a>> [consultado em 20.07.2022], “[t]rying to ‘fix’ A.I. distracts from the most urgent questions about technology”, sobretudo quando a resolução pensada se cinge à dimensão procedimental referida em texto: assegurar um procedimento decisório algorítmico justo e equitativo. Mais detalhadamente, identifica a Autora dois obstáculos a esta abordagem. Em primeiro lugar, considerar o viés algorítmico um problema essencialmente *computacional* obscurece a sua causa, pois o viés algorítmico é, antes de tudo, um problema social (humano), e procurar resolvê-lo dentro da lógica da automação conduzirá inevitavelmente a resultados desadequados. Em segundo lugar, o combate aos enviesamentos algorítmicos, ainda que seja bem-sucedido, comporta consequências perversas. A Autora, partindo de um caso concreto, dá como exemplo um sistema tecnológico de reconhecimento facial que apresenta falhas de funcionamento quando aplicado a mulheres negras devido à subrepresentação desta categoria de pessoas nos modelos de análise utilizados. Ora, a correção deste problema, através da “equalização” da representação desta categoria, tem um efeito óbvio: os modelos de análise são alimentados com mais e melhores dados, proporcionando o aperfeiçoamento dos sistemas de vigilância e de classificação (*scoring*) das pessoas. Todavia, se esse aperfeiçoamento decorrer num contexto em que questões sistémicas subjacentes permanecem fundamentalmente intocadas, o combate ao preconceito algorítmico através destas manobras reparatórias simplesmente torna os seres humanos mais “legíveis” e mais “classificáveis” por máquinas, expondo as minorias (ou outros sujeitos tipicamente subrepresentados) a desvantagens adicionais e inicialmente não ponderadas. A estes efeitos perversos voltaremos adiante, a propósito das fragilidades da *alternative data*.

tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar”.

Partindo da letra da lei, em causa não está uma proibição genérica de tomada deste tipo de decisões, ainda que se tenha popularizado, na literatura, essa ideia<sup>57</sup>: trata-se antes da atribuição, ao titular dos dados, de uma situação jurídica ativa – a exercer *antes* ou *depois* da tomada de decisão – pela qual impede que, quanto a si,

---

<sup>57</sup> Entre nós, referindo-se a uma proibição, cf. A. BARRETO MENEZES CORDEIRO, “Anotação ao artigo 22.º”, in A. BARRETO MENEZES CORDEIRO (coord.), *Comentário ao Regulamento Geral de Proteção de Dados e à Lei n.º 58/2019*, Coimbra, Almedina, 2021, p. 220. Também MAFALDA MIRANDA BARBOSA, “Discriminação algorítmica: reflexões em torno da responsabilidade civil”, *Revista de Direito da Responsabilidade Civil*, 3/2021 (1200-1231), pp. 1220 ss., encontra neste regime uma proibição genérica de decisões automatizadas que preencham os pressupostos elencados no artigo 22.º, n.º 1, do RGPD. A Autora parte do entendimento do GRUPO DE TRABALHO DO ARTIGO 29.º, nas suas *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*. Segundo o GRUPO DE TRABALHO, a expressão direito é usada no artigo 22.º em sentido impróprio: “[o] termo «direito» nesta disposição não significa que o artigo 22.º, n.º 1, seja aplicável somente quando ativamente invocado pelo titular dos dados. O artigo 22.º, n.º 1, estabelece uma proibição geral da tomada de decisões com base exclusivamente no tratamento automatizado. Esta proibição aplica-se independentemente de o titular dos dados adotar uma medida relativa ao tratamento dos seus dados pessoais” (p. 21). A arquitetura do regime contido no artigo 22.º revela-se, de acordo com este entendimento, em três níveis distintos de regulação: “i) em regra, existe uma proibição geral das decisões individuais totalmente automatizadas, incluindo a definição de perfis com efeitos jurídicos ou similarmente significativos; ii) há exceções a essa regra; iii) sempre que se aplique uma dessas exceções, devem existir medidas para salvaguardar os direitos e liberdades e os legítimos interesses do titular dos dados” (p. 21). Para alcançar esta conclusão, o GRUPO DE TRABALHO recorre a uma interpretação integrada com o Considerado 71, que se refere expressamente, a propósito da solução consagrada no n.º 2 do artigo 22.º, a um tratamento *permitido*. Acrescenta que “a interpretação do artigo 22.º como uma proibição, em vez de um direito que pode ser invocado, significa que as pessoas estão automaticamente protegidas dos possíveis efeitos deste tipo de tratamento” (p. 22).

Retomando a MAFALDA MIRANDA BARBOSA, é à luz do contributo do GRUPO DE TRABALHO DO ARTIGO 29.º que a Autora defende ser esta proposta interpretativa autorizada pela articulação do n.º 1 do artigo 22.º do RGPD, com o teor do Considerando 71 e com a solução consagrada no n.º 2 do referido artigo. Daí conclui pelo reconhecimento, ao titular dos dados, de um direito em termos genéricos, que exclui *a priori* a possibilidade de uma tomada de decisão exclusivamente automatizada, e não um qualquer direito unicamente exercitável por reação ao comportamento do responsável pelo tratamento de dados. Assim, para a Autora, “teleologicamente, a solução impõe-se com meridiana clareza. A utilização de algoritmos (quer na definição de perfis, quer na tomada de decisões automatizadas) caracteriza-se pela sua opacidade [...]. Nessa medida, porque a incolumidade dos direitos dos titulares dos dados pessoais não se pode garantir com o mero cumprimento de deveres de informação prestados pelo responsável pelo tratamento, há que proibir que certas decisões sejam tomadas por algoritmos, de forma totalmente automatizada” (p. 1221).

sejam tomadas essas decisões<sup>58</sup>. Diversamente do direito ao apagamento ou do direito de oposição – que se configuram técnico-juridicamente como *direitos* –, aqui, do que se trata, é de uma situação de *imunidade* do titular dos dados à tentativa de outra pessoa (o decisor) alterar a sua esfera jurídica através de uma ação (a tomada da decisão), a que correlativamente corresponde a ausência de um poder ou a incompetência dessa pessoa tendo em vista a alteração em causa<sup>59</sup>. Mas tem em comum com esses

---

<sup>58</sup> Mais abertamente, reconhecem ISAK MENDOZA / LEE A. BYGRAVE, “The Right Not to be Subject to Automated Decisions Based on Profiling”, in TATIANI SYNODINOU / PHILIPPE JOUGLEUX / CHRISTIANA MARKOU / THALIA PRASITTOU (ed.), *EU Internet Law. Regulation and Enforcement*, Cham, Springer, 2017 (77-98), que tanto é defensável um resultado interpretativo no sentido da vigência de uma proibição geral das decisões individuais automatizadas, como um resultado interpretativo no sentido de um direito de oposição a esse tratamento a exercer pelo titular dos dados, destinatário da decisão automatizada. Da mesma forma, referindo-se a duas interpretações do artigo 22.º do RGPD, esclarecem SANDRA WACHTER / BRENT MITTELSTADT / LUCIANO FLORIDI, “Why a Right to Explanation of Automated Decision-Making does not exist in the General Data Protection Regulation”, cit., pp. 94-95, que “this ambiguity has existed since the Data Protection Directive 1995. The wording of Article 15 of the Directive allowed the ‘right not to be subject of an automated decision’ referred to in Section 1 [...] to be interpreted as a prohibition or a right to object. The ambiguity led Member States to implement this right and associated protections differently”.

Neste quadro, concluem os Autores que, devido às similitudes de linguagem e conteúdo entre o artigo 15.º da Diretiva sobre a Proteção de Dados e o artigo 22.º do RGPD, e à alteridade na implementação, pelos Estados-membros, do então artigo 15.º, seja como uma proibição, seja como direito de oposição, consentem-se estes dois resultados interpretativos, embora tão díspares nas suas consequências de regime – incluindo, no que respeita à amplitude e conteúdo do direito à explicação da decisão automatizada. Em particular, se a situação jurídica do titular dos dados for perspectivada como um direito de oposição, ampliando-se o universo de decisões automatizadas permitidas, o direito à explicação da decisão automatizada não se conteria aos casos do artigo 22.º, n.º 2, por via do entendimento dominante do artigo 22.º, n.º 3, do RGPD, em articulação com o Considerando 71: o direito do titular dos dados à explicação aplicar-se-ia a qualquer decisão automatizada de acordo com a definição fornecida no artigo 22.º, n.º 1, do RGPD, mesmo que não se subsumível a alguma das alíneas do n.º 2 desse artigo (p. 95).

<sup>59</sup> Neste sentido escrevemos em “Aspetos jurídicos da análise de dados na Internet (*big data analytics*) nos setores bancário e financeiro: proteção de dados pessoais e deveres de informação”, in António Menezes Cordeiro / Ana Perestrelo de Oliveira / Diogo Pereira Duarte (eds.), *FinTech – Desafios da Tecnologia Financeira*, 2.ª ed., Coimbra, Almedina, 2019 (89-220), pp. 90-91. Em sentido convergente, esclarecendo que “[o]s casos de proibições de tratamento de dados no Regulamento são claros. Pense-se no artigo 9.º, no qual se lê que é proibido o tratamento de dados pessoais que revele a origem racial ou étnica, as opiniões políticas, as convicções religiosas ou filosóficas, bem como o tratamento de dados genéticos, dados biométricos para identificar uma pessoa de forma inequívoca, dados relativos à saúde ou dados relativos à vida sexual ou orientação sexual de uma pessoa. A redação do artigo 22.º é diametralmente diferente: não se proíbem as decisões automatizadas, concedendo-se apenas ao titular dos dados o direito a não ficar sujeito a uma decisão desse tipo. A redação da norma parece aproximar-se do direito de *opt out* que consta do artigo 7.º/3, pensado para casos em que o titular consentiu no tratamento dos dados e, posteriormente, retira o seu consentimento, tornando,

direitos o facto de também funcionar, num primeiro momento, de acordo com uma lógica potestativa: o titular dos dados, através de declaração unilateral dirigida ao responsável pelo tratamento ou decisor, constitui na esfera deste a incompetência (ou, noutra perspetiva, extingue da esfera deste a competência) de *assim* decidir<sup>60</sup>. Nestes termos, esvaziada a competência para decidir algorítmicamente, o que se pode questionar é *se a conduta cujos efeitos são impedidos (decidir algorítmicamente) é conduta proibida*. A resposta é positiva e tem relevância no plano indemnizatório: sempre que a conduta proibida (a decisão algorítmica impedida) cause danos ao titular dos dados – *apesar* da ineficácia dessa decisão – serão desencadeados os mecanismos próprios da responsabilidade civil.

Esta imunidade não é, porém, totalmente protetora do titular dos dados, pois de pouco vale nos casos em que a decisão algorítmica tenha como resultado a sua *exclusão* de certos negócios jurídicos (por exemplo, a decisão algorítmica de recusa de concessão de crédito imputada a uma instituição de crédito e dirigida a um cliente, pessoa singular)<sup>61</sup>. Sem que exista um motivo injuntivo – *maxime*, que faça corresponder as prestações integrantes de um contrato a um serviço essencial (aquele que proporciona condições de dignidade à vida humana) –, não pode o aplicador do direito impor ao decisor a celebração de um contrato ou substituir-se à conformação do negócio através da autonomia privada<sup>62</sup>.

A solução já será diferente no que toca aos casos de modificação unilateral ou de cessação unilateral de um contrato por parte do decisor, porque em causa estará o exercício, por este, de um direito potestativo: admitindo que essa modificação ou

---

a partir dessa data, o tratamento dos dados ilícito, a não ser que seja invocado outro fundamento de licitude para além do consentimento. A lógica subjacente parece ser a mesma que preside à norma do artigo 22.º/1: as decisões automatizadas são permitidas, com base em qualquer dos fundamentos que legitimam o tratamento de dados, mas se o titular se opuser a uma decisão com impacto na sua esfera jurídica, deverá ser desconsiderada a decisão”, *vide* MADALENA PERESTRELO DE OLIVEIRA, “Definição de perfis e decisões automatizadas no Regulamento Geral sobre a Proteção de Dados Pessoais”, in António Menezes Cordeiro / Ana Perestrelo de Oliveira / Diogo Pereira Duarte (eds.), *FinTech II – Novos estudos sobre tecnologia financeira*, Coimbra, Almedina, 2019 (61-88), pp. 80-81.

<sup>60</sup> Recorre-se aos conceitos hohfeldianos dos correlativos jurídicos (*jural correlatives*) «imunidade» (*immunity*) – «não faculdade» ou «incompetência» (*disability*) – cf. WESLEY NEWCOMB HOHFELD, *Os conceitos jurídicos fundamentais aplicados na argumentação jurídica*, cit., pp. 36 ss.

<sup>61</sup> Cf. ANA ALVES LEAL, “Aspetos jurídicos da análise de dados na Internet (*big data analytics*) nos setores bancário e financeiro: proteção de dados pessoais e deveres de informação”, cit., p. 91.

<sup>62</sup> Cf. ANA ALVES LEAL, “Aspetos jurídicos da análise de dados na Internet (*big data analytics*) nos setores bancário e financeiro: proteção de dados pessoais e deveres de informação”, cit., p. 91. Como aí referimos, assim será mesmo que, ultrapassadas as normais dificuldades no plano probatório, seja possível ao titular dos dados provar os factos constitutivos do juízo de causalidade segundo o qual a decisão de não celebração assentou exclusivamente (ou foi influenciada de forma determinante) no tratamento automatizado dos seus dados pessoais.

cessação unilateral era decidida algoritmicamente de forma válida, pode o titular dos dados, lançando mão deste meio de defesa, *paralisar* os efeitos do exercício de tal poder, determinando-se o retorno ao *status quo* com a consequente subsistência da relação jurídica negocial nos moldes vigentes antes da decisão.

## § 2.º Escalonamento de riscos e modelos regulativos da Proposta de Regulamento sobre Inteligência Artificial

18. O exemplo mais recente desta técnica proibitiva encontra-se na Proposta de Regulamento sobre Inteligência Artificial (doravante, abreviadamente referida como a “Proposta de Regulamento”). Seguindo aquilo a que designa por uma “abordagem proporcionada baseada no risco”<sup>63</sup> – pela qual se diferenciam, de forma escalar, as utilizações de sistemas de inteligência artificial que criam um *risco inaceitável*, um *risco elevado*, e um *risco baixo* ou *mínimo* –, elenca-se no Título II (artigo 5.º) da Proposta de Regulamento um conjunto de “práticas de inteligência artificial proibidas”<sup>64</sup>. Salvo circunscritas exceções (artigo 5.º, n.º 3, 2.º parágrafo, e n.º 4), estão em causa proibições de “práticas”. Este regime é composto, portanto, por normas regulativas. A proibição não se dirige *ao resultado* a cujo alcance está funcionalizado o sistema de inteligência artificial (tais como conteúdos, previsões, recomendações ou decisões), mas antes à *utilização* desse mesmo sistema, quando o funcionamento do mesmo implique o emprego de certas técnicas (por exemplo, “técnicas subliminares que contornem a consciência de uma pessoa” – alínea a) do n.º 1 do artigo 5.º), produza resultados tidos como indesejados (por exemplo, “a classificação social conduz [ao] tratamento prejudicial ou desfavorável de certas pessoas singulares ou grupos inteiros das mesmas em contextos sociais não relacionados com os contextos nos quais os dados foram originalmente gerados ou recolhidos” – inciso i) da alínea c) do n.º 1 do artigo 5.º) ou que seja apto a produzir esse resultado indesejado, ainda que o mesmo não se concretize (por exemplo, “técnicas subliminares [...] para distorcer substancialmente o [...] comportamento de uma forma que cause ou seja suscetível de causar danos físicos ou psicológicos a essa ou a outra pessoa” – alínea a) do n.º 1 do artigo 5.º).

---

<sup>63</sup> Cf. ponto 1.1 da Exposição de Motivos.

<sup>64</sup> Com detalhe sobre esta abordagem, apresentando críticas certas ao “*harm requirement*”, veja-se MICHAEL VEALE / FREDERIK ZUIDERVEEN BORGESIU, “Demystifying the Draft EU Artificial Intelligence Act”, *Computer Law Review International*, 4/2021 (97-112), p. 99. Entre nós, cf. MAFALDA MIRANDA BARBOSA, “Dos *expert systems* aos *data systems AI*: impacto ao nível da proteção de dados”, *Julgar*, 45/2021 (13-33), pp. 32-33.

É comum afirmar-se que esta lista de práticas proibidas inclui os sistemas de inteligência artificial cuja utilização seja considerada inadmissível face os valores da União Europeia (por exemplo, por violar direitos fundamentais) – *vide* o Considerando 15 da Proposta de Regulamento. Todavia, o elenco apresentado no artigo 5.º da Proposta de Regulamento é, apesar da textura aberta dos conceitos jurídicos nele descritos ou referidos, um elenco fechado. As proibições abrangem práticas com potencial significativo para manipular as pessoas por meio de técnicas subliminares que lhes passam despercebidas ou para explorar as vulnerabilidades de grupos específicos, como as crianças ou as pessoas com deficiência, para distorcer substancialmente o seu comportamento de uma forma que seja suscetível de causar danos psicológicos ou físicos a essa ou a outra pessoa<sup>65</sup>. A Proposta de Regulamento também proíbe a classificação social assente na inteligência artificial para uso geral por parte das autoridades públicas. Proíbe-se igualmente a utilização de sistemas de identificação biométrica à distância em tempo real em espaços acessíveis ao público para efeitos de manutenção da ordem pública, a não ser que se apliquem determinadas exceções limitadas.

19. Abandonando o modelo proibitivo, a regulação dos sistemas de inteligência artificial de risco elevado consagra um modelo corretivo e eminentemente informacional. Os enunciados normativos relevantes, plasmados no Título III (artigos 6.º e seguintes) da Proposta de Regulamento, evidenciam a heterogeneidade das soluções acolhidas por este regime.

Em primeiro lugar, atende-se à classificação dos sistemas de inteligência artificial como sendo de “risco elevado”, remetendo-se para as matérias e para os sistemas listados nos Anexos II e III da Proposta de Regulamento. Estão em causa essencialmente sistemas de inteligência artificial que criam um risco elevado para a saúde e para a segurança ou para os direitos fundamentais de pessoas singulares. Dividem-se em duas categorias: a categoria dos sistemas de inteligência artificial destinados a ser utilizados como componente de segurança de um produto ou se for ele próprio um produto, abrangido pela legislação de harmonização enumerada no Anexo II;

---

<sup>65</sup> Todavia, outras práticas manipuladoras ou exploratórias que são possibilitadas pelos sistemas de inteligência artificial e que afetam os adultos podem ser abrangidas pela legislação em matéria de proteção de dados, de defesa dos consumidores e de serviços digitais (ponto 5.2.2. da Exposição de Motivos). Essa legislação setorial deve “garant[ir] que as pessoas singulares sejam devidamente informadas e tenham a liberdade de decidir não se sujeitar a uma definição de perfis ou a outras práticas que possam afetar o seu comportamento” (ponto 5.2.2. da Exposição de Motivos). O artigo 22.º do RGPD confirma essa regulação setorial.



e que estão sujeitos a uma avaliação da conformidade *ex ante* por terceiros (como os relativos a brinquedos ou a dispositivos médicos); e a categoria dos sistemas de inteligência artificial com implicações em matéria de direitos fundamentais que são explicitamente mencionados no Anexo III da Proposta de Regulamento<sup>66</sup>, numa das seguintes oito áreas: (i) na identificação biométrica e categorização de pessoas singulares, (ii) na gestão e funcionamento de infraestruturas críticas (as relativas ao trânsito rodoviário e às redes de abastecimento de água, gás, aquecimento e eletricidade), (iii) na educação e formação profissional (por exemplo, os sistemas concebidos para serem utilizados para fins de determinação do acesso ou da afetação de pessoas singulares a instituições de ensino e de formação profissional, ou para serem utilizados para fins de avaliação de estudantes em instituições de ensino), (iv) no emprego, gestão de trabalhadores e acesso ao emprego por conta própria (por exemplo, sistemas concebidos para serem utilizados no recrutamento ou na seleção de pessoas singulares, ou na tomada de decisões sobre promoções ou cessações de relações contratuais de trabalho), (v) no acesso a serviços privados e a serviços e prestações públicas essenciais (em que se destacam os sistemas concebidos para serem utilizados para avaliar a elegibilidade de pessoas singulares quanto a prestações e serviços públicos de assistência, bem como para conceder, reduzir, revogar ou recuperar tais prestações e serviços [por exemplo, para uma operadora de telecomunicações decidir não celebrar um contrato de prestação de serviços com um determinado cliente, dado o histórico de incumprimento deste], ou os sistemas concebidos para serem utilizados para avaliar a capacidade de endividamento de pessoas singulares e, com isso, estabelecer o seu grau de merecimento de crédito ou a sua classificação de crédito [*credit scoring*]), (vi) na manutenção da ordem pública (domínio que tem merecido atenção por parte da doutrina, especialmente no que respeita a sistemas concebidos para serem utilizados por autoridades policiais em avaliações individuais de riscos relativamente a pessoas singulares, a fim de determinar o risco de uma pessoa singular cometer infrações ou voltar a cometer

---

<sup>66</sup> Este Anexo lista, em função da matéria, um conjunto fechado de sistemas de inteligência artificial cujos riscos já se materializaram ou são suscetíveis de se materializar num futuro próximo. A fim de garantir que o Regulamento acompanha o devir tecnológico da inteligência artificial, a Comissão pode alargar essa lista de sistemas de inteligência artificial de risco elevado desde que utilizados nos domínios pré-definidos nesse Anexo, mediante a aplicação de um conjunto de critérios e de uma metodologia de avaliação de riscos, tal como vertido no artigo 7.º da Proposta de Regulamento. Quer isso dizer, como notam MICHAEL VEALE / FREDERIK ZUIDERVEEN BORGESIUS, “Demystifying the Draft EU Artificial Intelligence Act”, cit, p. 102, que à Comissão é atribuída a competência de atualizar esse Anexo adicionando sub-áreas aos domínios pré-definidos, mas não áreas totalmente novas.

infrações ou o risco para potenciais vítimas de infrações penais), (vii) na gestão da migração, do asilo e do controlo das fronteiras (tendo vindo a merecer especial atenção dos juristas os sistemas concebidos para auxiliar autoridades públicas competentes na análise dos pedidos de asilo, de visto e de autorização de residência e das queixas relacionadas, no que toca à elegibilidade das pessoas singulares que requerem determinado estatuto), e (viii) na administração da justiça e processos democráticos (por exemplo, os sistemas concebidos para auxiliar um tribunal na investigação e na interpretação de factos e do direito e na aplicação da lei).

O “risco elevado” tem como referente o sistema de inteligência artificial, e não a prática ou a conduta associada à utilização desse sistema (como sucede com as “práticas” cuja proibição é estatuída no artigo 5.º da Proposta de Regulamento). Claro que este risco é qualificado em função dos possíveis resultados nefastos causados pelo funcionamento desse sistema. Como regra geral, a utilização dos sistemas de inteligência artificial de risco elevado é autorizada no espaço europeu. Quer isto dizer, por exemplo, que a decisão de uma instituição de crédito relativa à concessão de crédito a um cliente, antecedida de uma avaliação do grau de merecimento de crédito desse cliente, pode ser uma decisão algorítmica, formada com recurso a sistemas de inteligência artificial (ainda que estes sejam, como visto, qualificados como sendo de risco elevado).

A utilização autorizada desses sistemas implica, contudo, o cumprimento de um complexo de deveres de natureza procedimental e de deveres de informação, bem uma avaliação *ex ante* (antes da colocação no mercado) da conformidade desses sistemas, nos termos do artigo 43.º<sup>67</sup>. Em traços largos, esses deveres estão

---

<sup>67</sup> A avaliação da conformidade, como resulta de uma interpretação conjugada do artigo 19.º e do artigo 43.º, é prévia à colocação do sistema de risco elevado no mercado ou à colocação em serviço. E assim que a conformidade dos sistemas com os requisitos estabelecidos na Proposta de Regulamento fique demonstrada, os fornecedores devem elaborar uma declaração de conformidade EU, ao abrigo do disposto no artigo 48.º, e apor a marcação de conformidade CE, de acordo com o artigo 49.º. Regimes especiais ou setoriais de certificação dos sistemas ou de avaliação da conformidade não são desconsiderados por este regime geral contido na Proposta de Regulamento: de forma meramente exemplificativa, é o que sucede em relação aos sistemas de risco elevado funcionalizados à avaliação do merecimento de crédito de cliente bancários, quando o fornecedor desse sistema seja a instituição de crédito e esta seja regulamentada pela Diretiva 2013/36/UE, do Parlamento Europeu e do Conselho, de 26 de junho de 2013: neste caso, tal como dispõe o artigo 19.º, n.º 2, a avaliação da conformidade deve ser realizada no âmbito do procedimento a que se referem os artigos 97.º a 101.º dessa diretiva. Este procedimento de avaliação da conformidade poderá, não obstante, ser dispensado em casos excecionais. De acordo com o artigo 47.º, qualquer autoridade de fiscalização do mercado pode autorizar a colocação no mercado ou a colocação em serviço de determinados sistemas de risco elevado no território do Estado-Membro em causa, por motivos excecionais de segurança pública ou de proteção da vida e da saúde das pessoas, de proteção do ambiente e de proteção de ativos

orientados (i) à criação de um sistema de gestão de riscos, nos termos do artigo 9.º, que deve consistir num processo iterativo contínuo, executado ao longo de todo o ciclo de vida do sistema de inteligência artificial, o que requer atualizações regulares sistemáticas, (ii) à observância de um conjunto de práticas adequadas de governação e de gestão de dados, tal como enunciadas no artigo 10.º, (iii) ao cumprimento de regras específicas no que respeita à elaboração *ex ante* e a todo o tempo de documentação técnica sobre o sistema, nos termos do artigo 11.º, (iv) à manutenção de registos das operações decorrentes do funcionamento desse sistema, nos termos do artigo 12.º, (v) à promoção da transparência e à prestação de informações aos utilizadores sobre o funcionamento do sistema, em observância do disposto no artigo 13.º, (vi) à supervisão humana, a qual é pormenorizadamente regulada no artigo 14.º, e (vii) à exatidão, solidez e cibersegurança do sistema, seguindo-se os ditames do artigo 15.º<sup>68</sup>. Estes deveres, bem como os listados no catálogo de deveres do Capítulo III, são dirigidos, em primeira linha, aos *fornecedores* dos sistemas de risco elevado (artigo 16.º e seguintes)<sup>69</sup>. O mesmo sucede quanto

---

industriais e infraestruturas essenciais. Tendo em conta a excecionalidade desta dispensa, tal autorização deve ser concedida por um período limitado, enquanto os procedimentos de avaliação da conformidade necessários estiverem a ser executados, e cessa assim que esses procedimentos tiverem sido concluídos. A conclusão desses procedimentos deve ser realizada sem demora injustificada. Em rigor, mantém-se o dever de iniciar o procedimento de avaliação da conformidade; o que se admite é que, nas condições excecionais referidas e mediante autorização, a colocação no mercado ou em serviço ocorra sem que tenham sido alcançados os resultados típicos dessa avaliação da conformidade – como seja a declaração de conformidade UE e a marcação de conformidade CE.

<sup>68</sup> Como se afirma no ponto 5.2.3. da Exposição de Motivos, estes deveres impõem condutas que são já habituais para muitos operadores que procuraram adaptar-se, nos últimos anos, a um complexo de práticas plasmadas em instrumentos jurídicos não vinculativos (*soft law*). Na base do catálogo de deveres apresentado em texto está um trabalho preparatório de dois anos, decorrente das *Orientações éticas para uma IA de confiança*, de 2019, do Grupo de Peritos de Alto Nível em Inteligência Artificial (disponível em <<https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pt/format-PDF>> [consultado em 20.07.2022]), que foi aplicado numa fase-piloto por mais de 350 organizações. Esclarece-se ainda, nesse ponto da Exposição de Motivos, que estes deveres também são amplamente coerentes com outras recomendações e princípios internacionais, com vista a assegurar que o quadro para inteligência artificial proposto seja compatível com os adotados por parceiros comerciais internacionais da União Europeia. É também consentida alguma flexibilidade na seleção das soluções técnicas específicas para garantir o cumprimento dos deveres referidos: os fornecedores dos sistemas de inteligência artificial podem escolher como cumprir esses deveres, tendo em conta os progressos científicos e tecnológicos nos domínios a que esses deveres se referem.

<sup>69</sup> Na aceção do artigo 3.º, n.º 2, o fornecedor corresponde à pessoa, entidade ou organismo, de natureza pública ou privada, que desenvolve um sistema de inteligência artificial ou que tem um sistema de inteligência artificial desenvolvido com vista à sua colocação no mercado ou colocação em serviço sob o seu próprio nome ou marca, a título oneroso ou gratuito.

ao complexo de deveres *ex post* elencados no Título VIII (artigo 61.º e seguintes), direcionados para o acompanhamento pós-comercialização e para a partilha de informações sobre acidentes e anomalias). O universo de sujeitos vinculados a estes deveres estende-se, em segunda linha, e com diferentes configurações e graus de vinculação, aos *fabricantes dos produtos* (artigo 24.º), aos *importadores* (artigo 26.º)<sup>70</sup> e aos *distribuidores* (artigo 27.º)<sup>71</sup>. Os deveres dos *utilizadores* dos sistemas de risco elevado encontram-se também especificados na Proposta de Regulamento<sup>72</sup>: desde logo, o dever de utilizar os sistemas de risco elevado de acordo com as instruções de utilização que acompanham os sistemas (artigo 29.º, n.º 1)<sup>73</sup>. O que, contudo, não afasta outras obrigações do utilizador previstas na legislação da União Europeia ou nacional nem prejudicam o poder discricionário do utilizador para organizar os seus próprios recursos e atividades para efeitos de aplicação das medidas de supervisão humana indicadas pelo fornecedor (artigo 29.º, n.º 2). Acresce que o utilizador é equiparado ao fornecedor – ficando vinculado ao complexo de deveres cominado a este último no artigo 16.º<sup>74</sup> – se colocar no mercado ou colocar em

<sup>70</sup> Tal como definido no artigo 3.º, n.º 6, o importador corresponde à pessoa singular ou coletiva com estabelecimento num Estado-membro da União Europeia que coloca no mercado ou coloca em serviço um sistema de inteligência artificial que ostenta o nome ou a marca de uma pessoa singular ou coletiva estabelecida fora da União Europeia

<sup>71</sup> Nos termos do artigo 3.º, n.º 7, é distribuidor a pessoa singular ou coletiva inserida na cadeia de abastecimento, distinta do fornecedor e do importador, que disponibiliza um sistema de inteligência artificial no mercado da União Europeia sem alterar as suas propriedades.

<sup>72</sup> Considera-se utilizador, de acordo com a norma definitiva do artigo 3.º, n.º 4, a pessoa singular ou coletiva, autoridade pública, agência ou outro organismo que utilize, sob a sua autoridade, um sistema de inteligência artificial, salvo se esse sistema for utilizado no âmbito de uma atividade pessoal de caráter não profissional.

<sup>73</sup> Poder-se-á estranhar a positivação deste dever no contexto da regulação da utilização de sistemas de inteligência artificial de risco elevado, por a utilização de um dado produto de acordo com as instruções do fornecedor ou do fabricante ser conduta devida ao abrigo de outros regimes setoriais e, em geral, para qualquer produto (no caso, para todos os sistemas, independentemente da sua classificação). A menção legal expressa ao dever de utilização do produto de acordo com as instruções do fabricante evidencia, porém, duas vantagens. Por um lado, deixa claro que se trata de um *dever* do utilizador, e não de mero um *onus* (onus que, por exemplo, se extrai do artigo 40.º, n.º 2, alínea a), do Decreto-Lei 84/2021, de 18 de outubro, que regula os direitos do consumidor na compra e venda de bens, conteúdos e serviços digitais: de acordo com este preceito, o produtor pode opor-se ao exercício dos direitos de reparação ou substituição pelo consumidor caso a não conformidade do produto resulte de má utilização). Por outro lado, essa deontificação torna essa norma impositiva numa norma de proteção para o sujeito destinatário da decisão (ou de outro resultado) algorítmico, desencadeando-se a correspondente tutela indemnizatória em caso de violação da mesma, se preenchidos os respetivos pressupostos.

<sup>74</sup> Tanto pode ser uma *extensão* da posição jurídica passiva do fornecedor ao utilizador (nos casos da alínea a) do n.º 1 do artigo 28.º), como uma *substituição* do fornecedor pelo utilizador que é

serviço um sistema de risco elevado sob o seu nome ou marca, se modificar a finalidade prevista de um sistema de risco elevado já colocado no mercado ou colocado em serviço, ou se provocar nesse sistema uma modificação substancial (artigo 28.º, n.º 1).

20. Merecem ainda nota os sistemas de inteligência artificial de baixo risco, regulados no Título IV da Proposta de Regulamento, quanto aos quais se impõe apenas determinadas obrigações de transparência (artigo 52.º). Estas obrigações reportam-se, em traços largos, aos sistemas de inteligência artificial que (i) interagem com seres humanos, que (ii) são utilizados para detetar emoções ou determinar a associação com categorias (sociais) com base em dados biométricos, ou que (iii) geram ou manipulam conteúdos de imagem, áudio ou vídeo que sejam consideravelmente semelhantes a pessoas, objetos, locais ou outras entidades ou acontecimentos reais e que, falsamente, pareçam ser autênticos e verdadeiros a uma pessoa («falsificação profunda»). Será o caso, por exemplo, dos sistemas de conversação automática ou dos videojogos.

21. Num balanço geral sobre a Proposta de Regulamento sobre Inteligência Artificial, identifica-se o hibridismo do modelo de regulação adotado, moldado em função da classificação e do escalonamento dos riscos em torno dos quais essa regulação foi desenhada.

Assim, a técnica proibitiva está reservada às práticas de inteligência artificial descritas no artigo 5.º, cujo risco é classificado como inaceitável. A técnica proibitiva traduz-se deonticamente na adoção de um conjunto de normas que vedam a realização de condutas – a *colocação em mercado* e a *utilização* de sistemas de inteligência artificial nos casos descritos, aí se incluindo a tomada de decisões algorítmicas. No texto do diploma estabelecem-se proibições, acrescentando-se uma punição para o caso de incumprimento, mas não se identifica a quem são estas proibições dirigidas. O titular do dever será identificável em função da conduta vedada (simpliciter, o *fornecedor*, quando a conduta proibida é a colocação no mercado do sistema de inteligência artificial, e o *utilizador*, quando a conduta proibida é a sua utilização). Esta técnica proibitiva corresponde a uma das principais (e mais debatidas) inovações introduzidas pela Proposta de Regulamento: o conteúdo e a extensão das proibições plasmadas neste regime não têm equivalente em nenhum

---

equiparado a fornecedor (nos casos das alíneas b) ou c) do n.º 1 do artigo 28.º, atento o disposto no n.º 2 desse artigo).

regime de direito positivo – nem mesmo, como visto anteriormente, no regime contido no artigo 22.º do Regulamento Geral da Proteção de Dados Pessoais. A violação destas proibições terá consequências que não estão integralmente densificadas na Proposta de Regulamento – que se limita a balizar, no plano contraordenacional, a moldura das coimas, delegando nos Estados-membros a estabelecimento do regime de sanções (artigo 71.º). Numa dimensão civilística, em abstrato, estas consequências situar-se-ão no plano da ineficácia negocial (impedem-se os efeitos da conduta proibida, quando esta vise uma declaração negocial) e no plano da responsabilidade civil, caso essa conduta cause danos e estejam preenchidos os correspondentes pressupostos.

### 2.1. Transparência na cadeia de valor. Ambiguidades da ausência de um dever de explicação das decisões algorítmicas

22. Na regulação dos sistemas de inteligência artificial de risco elevado não se encontra expressamente a cominação de um dever de explicação da decisão algorítmica ao seu destinatário, em moldes semelhantes aos que encontramos no Regulamento Geral da Proteção de Dados Pessoais, para as decisões automatizadas que impliquem o tratamento de dados pessoais. Em coerência, também não se encontra a previsão desse dever quando esteja em causa a utilização de sistemas de risco mínimo.

O artigo 13.º da Proposta de Regulamento, sob a epígrafe “Transparência e prestação de informações aos utilizadores”, fornece, porém, um contributo nesse sentido: limitando-se à relação entre fornecedor e utilizador, estabelece, no seu n.º 1, o dever dirigido ao fornecedor de este “conceb[er] e desenvolv[er] [o sistema de risco elevado] de maneira que assegure que o seu funcionamento seja suficientemente transparente para permitir aos utilizadores interpretar o resultado do sistema e utilizá-lo corretamente”. Em causa, à primeira vista, está um *dever de conceber algoritmos interpretáveis*. Adicionalmente, no seu n.º 2, determina que estes sistemas “devem ser acompanhados de instruções de utilização, num formato digital ou outro adequado, que incluam informações concisas, completas, corretas e claras que sejam pertinentes, acessíveis e compreensíveis para os utilizadores”, apresentando, no n.º 3, um extenso conjunto de informações a transmitir. Entre elas, destacam-se, por integrar o conteúdo da explicação de uma decisão algorítmica, as informações referidas alínea b): a especificação das características, capacidades e limitações de desempenho do sistema de risco elevado (que inclui a finalidade prevista do sistema, o desempenho do sistema no tocante às pessoas ou grupos de pessoas em que o sistema se destina a ser utilizado, as especificações para os dados de entrada, ou quaisquer outras informações importantes em termos dos conjuntos de dados de treino, validação e teste usados, tendo em conta a finalidade prevista

do sistema, e as medidas de supervisão humana, incluindo as soluções técnicas adotadas para facilitar a interpretação dos resultados dos sistemas pelos utilizadores). Em causa está um *dever de informar com vista a assegurar essa interpretabilidade*.

23. A timidez do regime não permite nele encontrar um dever de explicação cominado ao decisor algorítmico: a montante, temos apenas a cominação, ao fornecedor, de deveres de informação sobre o funcionamento do algoritmo, sendo o beneficiário dessa informação o utilizador (tipicamente, o decisor algorítmico). Todavia, este dever não visa apenas esclarecer o utilizador sobre o funcionamento algorítmico: visa, com esse esclarecimento, assegurar a possibilidade de este cumprir os deveres que lhe são cominados na Proposta de Regulamento, como seja os plasmados no artigo 29.º (onde não se inclui, porém, expressamente um dever de explicação da decisão algorítmica ao seu destinatário). Esta nota finalística de segunda ordem extrai-se da parte final do n.º 1 do artigo 13.º: “Deve ser garantido um tipo e um grau adequado de transparência, que permita cumprir as obrigações que incumbem ao utilizador e ao fornecedor por força do capítulo 3 do presente título [da Proposta de Regulamento]”.

No âmbito da sua técnica informacional, a Proposta de Regulamento favorece inequivocamente uma lógica de transparência dos algoritmos na “cadeia de valor da inteligência artificial”: seja pelas exigências quanto à concepção dos algoritmos interpretáveis (artigo 13.º, n.º 1), seja pela elaboração de documentação técnica (artigo 11.º), seja pela manutenção de um sistema de registos que deve assegurar um nível de rastreabilidade do funcionamento do sistema ao longo do seu ciclo de vida (artigo 12.º), seja pela disponibilização de instruções de utilização (artigo 13.º, n.º 2), seja pela prestação de informações específicas que acompanham essas instruções (artigo 13.º, n.ºs 2 e 3).

Circunscreve essa lógica de transparência, todavia, aos participantes a montante dessa cadeia – aos *operadores* – sem nela abranger, na cauda dessa cadeia, o destinatário final (no caso, o destinatário da decisão algorítmica). Este Regulamento não constitui, portanto, a fonte primária de um dever de explicação da decisão algorítmica ao seu destinatário, nem de um direito à explicação de tal decisão – seja esta o resultado de um sistema de inteligência artificial de risco elevado ou de risco mínimo. Esta conclusão não impede, todavia, que reconheça que a abordagem informacional que estrutura o Regulamento está também funcionalizada ao fornecimento dessa explicação nas relações entre o decisor algorítmico e o destinatário da decisão. A identificação da fonte de um direito subjetivo à explicação (e do correspondente dever jurídico), não podendo esta ser extraída do escopo de proteção das normas do Regulamento que impõem esta lógica de transparência, terá de provir de outros indicadores.

### § 3.º O perigo das decisões algorítmicas: a amplificação do enviesamento decisório humano

#### 3.1. Intuições e preconceito algorítmico: tentativa de desconstrução

24. Uma visão geral do quadro regulador das decisões algorítmicas – independentemente do modelo regulativo aplicável – faz surgir um conjunto de representações sobre estas realidades que intuitivamente se pressupõem como desvaliosas, perigosas ou mesmo censuráveis. Perante este quadro, percebe-se que ao mesmo subjaz a ideia de que as decisões algorítmicas, por alguma razão, apresentam uma perigosidade superior às decisões humanas. Em particular, gera-se a percepção de que há decisões algorítmicas em que essa perigosidade é tão acentuada que o ordenamento jurídico não consente a sua produção e existência.

Este preconceito algorítmico, fruto da intuição assinalada, reconduz-se à ideia de que decisão humana é “melhor” para o destinatário da decisão, por assegurar uma decisão “mais justa” (embora não necessariamente “favorável” a esse destinatário). Compreender-se-ia, neste plano intuitivo, que a exigência de intervenção humana na decisão algorítmica fosse, em qualquer caso, a saída natural para combater a perigosidade destas decisões. A garantia de uma revisão humana da decisão algorítmica – tal como já advogada na primeira fase do discurso dos algoritmos acima descrita, positivada no artigo 22.º, n.º 3, do Regulamento Geral da Proteção de Dados Pessoais, e prevista no artigo 14.º da Proposta de Regulamento sobre Inteligência Artificial – bastaria, por isso, para superar esse entrave. Todavia, esta intervenção humana *ex post* não é sempre vista como resposta suficiente: pense-se nos modelos proibitivos dos algoritmos que postulam o impedimento à produção de efeitos de certas decisões algorítmicas, não cedendo, em prol da eficácia dessas decisões, à revisão humana destas. Nestes casos, terão de existir razões para não se reconhecer uma aptidão corretiva à revisão humana. Estes casos são também aqueles que contrariam a ideia intuída de que a decisão humana é sempre melhor do que a decisão algorítmica.

25. Por maioria de razão, também não colhe o argumento de que as decisões algorítmicas são mais perigosas pela opacidade do seu processo decisório, quando este é desconhecido ou incompreensível para o destinatário da decisão. Também as decisões humanas poderão resultar de um processo decisório que não seja dado a conhecer ao seu destinatário, ou que este não compreenda, ainda que o *iter* decisório lhe seja revelado. A opacidade é uma propriedade natural de qualquer decisão tomada no contexto de relações jurídicas privadas, e uma propriedade tendencial e não exclusiva das decisões algorítmicas.



26. A regulação das decisões algorítmicas encontra justificação na ideia segundo a qual *os sistemas algorítmicos possibilitam uma multiplicação, tendencialmente incomensurável e – pela sua tendencial opacidade – incontrolável dos vieses decisórios humanos*. Por outras palavras, consideram-se as decisões algorítmicas mais perigosas do que as decisões humanas justamente porque apresentam a aptidão de amplificar aquelas que possam ser tidas como “más” decisões humanas, projetando-se, com essa amplificação, na esfera de mais sujeitos<sup>75</sup>. Na eleição entre um modelo regulador informacional, corretivo ou proibitivo, o critério mobilizado é, do menos para o mais intenso, o do «grau de perigosidade» da decisão algorítmica.

Estas afirmações implicam que se esclareçam alguns aspetos. Em primeiro lugar, porque a pressupõe, a distinção entre decisão humana e decisão algorítmica; em segundo lugar, a explicação do que são o efeito discriminatório, o efeito lesivo, e o efeito multiplicador das decisões algorítmicas; em terceiro lugar, se estas conclusões procedem nos casos em que a decisão algorítmica, num juízo de prognose, não constitui uma “mã” decisão, no sentido que acima lhe é atribuído.

### 3.2. Graus de intervenção humana no processo decisório: da assistência algorítmica às decisões automatizadas

27. A par da inexistência de uma definição legal de «algoritmo», não é possível isolar uma definição legal de «decisão algorítmica», o que poderá explicar uma certa inexatidão dos contornos da figura. Tal não significa, porém, que as fontes sejam omissas a propósito desta realidade:

O artigo 22.º, n.º 1, do RGPD, apresenta como «decisão individual automatizada» a “decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua [do destinatário, titular dos dados pessoais] esfera jurídica ou que o afete significativamente de forma similar”. A partir desta *pista*, a doutrina, a par do Grupo de Trabalho do Artigo 29.º, tem construído a noção de «decisão individual automatizada» em torno de um critério negativo: a

---

<sup>75</sup> Em sentido próximo, ainda que sem reconduzir o problema a um juízo de perigo, cf. IFEOMA AJUNWA, “The Paradox of Automation as Anti-Bias Intervention”, *Cardozo Law Review*, 41, 2020 (1671-1742), pp. 1698-1699. O Autor aponta a irrelevância da discussão que gira em torno de saber se as decisões algorítmicas são, quando comparadas com as humanas, mais discriminatórias. Basta que se considere, para concluir por essa irrelevância, que os algoritmos, quando produzam um resultado discriminatório, se limitam a replicar e a perpetuar enviesamentos estruturais dos processos decisórios humanos.

ausência de intervenção humana no processo decisório<sup>76</sup>. Trata-se de um critério material e não formal<sup>77</sup>, que conduz a uma interpretação restritiva do preceito<sup>78</sup>: a decisão automatizada, neste sentido, não deixa de o ser quando o decisor humano, apesar de formalmente interveniente, (i) não tem competência ou possibilidade de influenciar o resultado decisório final, ou (ii) ainda que tenha essa competência ou possibilidade, aplica de forma rotineira e totalmente acrítica as decisões produzidas automaticamente, mesmo que estas careçam da prática de um específico ato confirmativo por parte do decisor humano<sup>79</sup>. Nestes casos, assiste-se apenas a uma simulação ou “fabricação” de intervenção humana no processo decisório<sup>80</sup>. Por isso, para que se considere haver uma intervenção humana relevante para a “desqualificação” de uma decisão como «automatizada», “o responsável pelo tratamento tem de garantir que qualquer supervisão da decisão seja relevante, e não um mero gesto simbólico. Essa supervisão deve ser levada a cabo por alguém com autoridade e competência para alterar a decisão e que, no âmbito da análise, deverá tomar em consideração todos

---

<sup>76</sup> Cf. GRUPO DE TRABALHO DO ARTIGO 29.º, *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., pp. 22-23.

<sup>77</sup> Sustenta, neste sentido, A. BARRETO MENEZES CORDEIRO, “Decisões individuais automatizadas à luz do RGPD e da LGPD”, in MAFALDA MIRANDA BARBOSA *et al.* (ed.), *Direito Digital e Inteligência Artificial: Diálogos entre Brasil e Europa*, São Paulo, Editora Foco, 2021, p. 267, que se “trata [...] de um critério material e não de um critério formal”, não relevando a intervenção humana que “assuma contornos burocráticos, meramente confirmadores ou acríticos”.

<sup>78</sup> Afirmamos em “Aspetos jurídicos da análise de dados na Internet (*big data analytics*) nos setores bancário e financeiro: proteção de dados pessoais e deveres de informação”, cit., pp. 149 ss., que esta leitura corresponde a uma interpretação restritiva da letra da lei: isto porque para a aplicação do regime constante do artigo 22.º do RGPD não é necessário que o processo decisório seja *totalmente* automatizado, nem é necessário que a decisão assente *unicamente* nos resultados de um tratamento automatizado de dados (podem esses resultados ser o fator único e exclusivo da tomada de decisão, ou o fator principal ou determinante, que coexiste com outros fatores). Esclarece-se, aí, que “esta leitura encontra apoio na ideia de que este mecanismo de defesa deve ser possibilitado ao titular dos dados sempre que o tratamento automatizado seja ou tenha sido a influência *real* no resultado de um concreto processo decisório”.

<sup>79</sup> Cf. GRUPO DE TRABALHO DO ARTIGO 29.º, *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., p. 23. O exemplo aí fornecido é o seguinte: “[o] responsável pelo tratamento não pode eximir-se do disposto no artigo 22.º fabricando uma intervenção humana. Por exemplo, se alguém aplicar de forma sistemática perfis gerados automaticamente a pessoas sem ter qualquer influência efetiva no resultado, tratar-se-á, ainda assim, de uma decisão tomada exclusivamente com base no tratamento automatizado”. Assim, na doutrina portuguesa, para além dos Autores acima referenciados, cf. MADALENA PERESTRELO DE OLIVEIRA, “Definição de perfis e decisões automatizadas no Regulamento Geral sobre a Proteção de Dados Pessoais”, cit., p. 75; MAFALDA MIRANDA BARBOSA, “Discriminação algorítmica: reflexões em torno da responsabilidade civil”, cit., p. 1222.

<sup>80</sup> A expressão é, como visto, a utilizada pelo GRUPO DE TRABALHO DO ARTIGO 29.º, *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., pp. 22-23.

os dados pertinentes”<sup>81</sup>. Desta forma, como já defendemos, a adoção deste critério material da intervenção humana deixa claro que em jogo está um juízo de causalidade: apenas está preenchida a previsão da norma contida no artigo 22.º, n.º 1, do RGPD, quando o ser humano possa (ainda que não o faça) influenciar o sentido da decisão – o que, no caso de a decisão já ter sido tomada, obriga à reconstrução da situação que provavelmente existiria se o tratamento automatizado não tivesse ocorrido ou os seus resultados conhecidos pelo decisor<sup>82</sup>.

O artigo 3.º, n.º 1, da Proposta de Regulamento sobre Inteligência Artificial, define como «sistema de inteligência artificial» “um programa informático desenvolvido com uma ou várias das técnicas e abordagens enumeradas no anexo I, capaz de, tendo em vista um determinado conjunto de objetivos definidos por seres humanos, criar resultados, tais como conteúdos, previsões, recomendações ou decisões, que influenciam os ambientes com os quais interage”. Não está aqui espelhada, como no RGPD, uma definição de decisão automatizada, mas associa-se a *decisão* (e não necessariamente a «decisão individual automatizada» com os requisitos da apresentada no RGPD) a um dos possíveis resultados e finalidades do sistema de inteligência artificial. Essencial é que o sistema de inteligência artificial, em função da remissão operada para o anexo I da Proposta de Regulamento, opere de acordo com alguma das técnicas ou abordagens aí listadas<sup>83</sup>.

---

<sup>81</sup> Cf. GRUPO DE TRABALHO DO ARTIGO 29.º, *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, cit., p. 23. De acordo com as *Orientações*, compete ao responsável pelo tratamento, no âmbito da avaliação de impacto, identificar e registar o grau de intervenção humana no processo decisório e a fase em que essa intervenção ocorre.

<sup>82</sup> Cf. ANA ALVES LEAL, “Aspetos jurídicos da análise de dados na Internet (*big data analytics*) nos setores bancário e financeiro: proteção de dados pessoais e deveres de informação”, cit., p. 152.

<sup>83</sup> Essas técnicas ou abordagens correspondem a (i) “abordagens de aprendizagem automática, incluindo aprendizagem supervisionada, não supervisionada e por reforço, utilizando uma grande variedade de métodos, designadamente aprendizagem profunda”, a (ii) “abordagens baseadas na lógica e no conhecimento, nomeadamente representação do conhecimento, programação (lógica) indutiva, bases de conhecimento, motores de inferência e de dedução, sistemas de raciocínio (simbólico) e sistemas periciais”, e (iii) “abordagens estatísticas, estimação de Bayes, métodos de pesquisa e otimização”. Esta lista de técnicas e abordagens específicas deve ser atualizada face à evolução do mercado e da tecnologia, mediante a adoção de atos delegados da Comissão Europeia que alterem essa lista. No Considerando 6 da Proposta de Regulamento, assume-se o cuidado para que a “definição de «sistema de IA» [seja] inequívoca, para assegurar a segurança jurídica”, mas “concedendo em simultâneo a flexibilidade suficiente para se adaptar a futuras evoluções tecnológicas”. Pelo que “a definição deve basear-se nas principais características funcionais do software, em particular a capacidade, tendo em vista um determinado conjunto de objetivos definidos pelos seres humanos, de criar resultados, tais como conteúdos, previsões, recomendações ou decisões que influenciam o ambiente com o qual o sistema interage, quer numa dimensão física, quer digital”. Esclarece-se também a matriz gradativa dos sistemas de inteligência artificial, os quais “podem ser concebidos para operar com diferentes níveis de autonomia e ser utilizados autonomamente ou como componente

28. Encontra-se relativamente estabilizada a polarização entre decisões humanas e decisões automatizadas, intermediadas – por esta ordem – por um grau decrescente de intensidade da intervenção humana na formação da decisão. Num dos polos, encontramos decisões puramente humanas, não baseadas em qualquer algoritmo executado por computadores<sup>84</sup>; no outro polo, encontramos decisões baseadas em algoritmos totalmente executadas por computadores, sem intervenção humana no *iter* decisório. Falta, porém, simetria perfeita entre estes polos: diversamente de uma decisão humana, que pode não sofrer qualquer influência de algoritmos criados de acordo com certa linguagem computacional, o algoritmo que baseia a decisão algorítmica não é (mesmo que totalmente automatizado, e mesmo que gerado automaticamente, em sistemas de *machine learning*, desenvolvidos através de *black box*) alheio, na sua gênese, a uma qualquer intervenção humana – pelo menos, à intervenção do seu (primitivo) criador<sup>85</sup>. Para além disso, o paradoxo na escolha evidencia-se nas decisões automatizadas: ao escolher a utilização do sistema algorítmico que forma a decisão automatizada, o utilizador decide voluntariamente retirar-se do processo de formação da decisão, *delegando* nesse sistema a tomada de decisão<sup>86</sup>.

29. Esta distinção permite, em qualquer caso, servir de base à explicitação de diferenças entre o processo decisório humano e o processo decisório algorítmico.

Não nos deteremos na forma como os seres humanos decidem, em contraposição com a forma como as decisões algorítmicas são produzidas<sup>87</sup>. Não só é vasta e crescente a literatura que explora o processo de tomada de decisão humana, como essa é matéria que se situa muito para lá das fronteiras da ciência jurídica: as teorias da decisão têm um cunho marcadamente interdisciplinar, pela quais se interessam economistas, filósofos, matemáticos, psicólogos ou especialistas em inteligência

---

de um produto, independentemente de o sistema estar fisicamente incorporado no produto (integrado) ou servir a funcionalidade do produto sem estar incorporado nele (não integrado)”.

<sup>84</sup> Relembramos que centramos o nosso interesse nos algoritmos que podem ser executados por computadores.

<sup>85</sup> Como afirma AZIZ Z. HUQ, a “Right to a Human Decision”, cit., p. 646, “all machine-learning tools are at their origin the fruit of specific human design and engineering choices. There is simply no such thing as a wholly endogenous algorithm”. O Autor sustenta ainda que a intervenção humana não está limitada o *design* inicial do algoritmo: ao criador do algoritmo caberá também selecionar os dados que irão alimentar o modelo de análise criado (p. 647).

<sup>86</sup> Cf. MICHAL S. GAL, “Algorithmic Challenges to Autonomous Choice”, *Michigan Technology Law*, 25, 2018 (59-104), p. 62.

<sup>87</sup> Para um confronto entre a decisão humana e a decisão algorítmica, AZIZ Z. HUQ, a “Right to a Human Decision”, cit., pp. 636 ss.

artificial<sup>88</sup>. No campo jurídico, excetuando a análise económica do direito<sup>89</sup> e os fenómenos de adaptação da argumentação jurídica à inteligência artificial<sup>90</sup>, é francamente reduzido o interesse pelo processo de tomada de decisão<sup>91</sup>.

Assinala-se apenas que a decisão humana é tida como mais flexível e contextual, por o decisor humano, em função do concreto contexto – neste particular, em função de fatores ou de variáveis quer teriam sido desconsiderados pelo sistema autónomo –, do cruzamento de vários contextos, ou da generalização do contexto, poder escolher desviar-se das regras específicas que conformam o processo decisório ou exercer, com menos predição, a discricionariedade decisória que lhe é atribuída. Em alguns casos, tal poderá consistir numa violação, pelo decisor humano, das regras procedimentais que presidem à formação da decisão; noutros, poderá tratar-se apenas do exercício de uma competência dentro da margem de livre decisão que, independentemente da fonte habilitante, lhe é atribuída. Sustenta-se que esta flexibilidade deriva do *conhecimento tácito* ou *intuição* – o conhecimento que não pode ser traduzido para certa linguagem computacional – do decisor humano<sup>92/93</sup>.

---

<sup>88</sup> Para uma teoria geral da decisão, cf. DANIEL KAHNEMAN / AMOS TVERSKY, “Choices, values, and frames”, *American Psychologist*, 39/4, 1984 (341-350); MICHAEL RESNIK, *Choices: An introduction to decision theory*, Minneapolis, Univ. of Minnesota Press, 1987; RICHARD C. JEFFREY, *The Logic of Decision*, 2<sup>nd</sup> ed., Chicago, Univ. of Chicago Press, 1990; HELMUT LAUX, *Entscheidungstheorie*, 5. Auflage, Heidelberg, Berlin, Springer, 2003; LEIGH BUCHANAN / ANDREW O’CONNELL, “A Brief History of Decision Making”, *Harvard Business Review*, 84/1, 2006 (32-41); PETER DÓRSAM, *Grundlagen der Entscheidungstheorie*, 6. Auflage, Heidenau, PD-Verlag, 2013; MARTIN PETERSON, *An introduction to decision theory*, 2<sup>nd</sup> ed., Cambridge, Cambridge University Press, 2017. Essencial, para uma visão geral das teorias da decisão, é a leitura da entrada “Decision Theory”, na *Stanford Encyclopedia of Philosophy*, da autoria de KATIE STEELE e de H. ORRI STEFÁNSSON, na sua última versão de 2020, EDWARD N. ZALTA *et al.* (ed.), disponível em <<https://plato.stanford.edu/entries/decision-theory/>> [consultado em 20.07.2022].

<sup>89</sup> Com interesse no domínio das decisões no contexto negocial, cf. por todos, ANTHONY T. KRONMAN/RICHARD A. POSNER, *The economics of contract law*, Boston, Little Brown & Co Law & Business, 1979; FERNANDO ARAÚJO, *Teoria económica do contrato*, Coimbra, Almedina, 2007, pp. 47 ss.

<sup>90</sup> Cf., entre outros estudos clássicos, RICHARD E. SUSSKIND, “Expert systems in law: a jurisprudential approach to artificial intelligence and legal reasoning”, *The Modern Law Review*, 49/2, 1986 (168-194), ANNE VON DER LIETH GARDNER, *An Artificial Intelligence Approach to Legal Reasoning*, 2<sup>nd</sup> ed., Cambridge, MA, The MIT Press, 1987; CASS R. SUNSTEIN, “Of Artificial Intelligence and Legal Reasoning”, *University of Chicago Public Law & Legal Theory Working Paper No. 18*, 2001, disponível em <[https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1206&context=public\\_law\\_and\\_legal\\_theory](https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1206&context=public_law_and_legal_theory)> [consultado em 20.07.2022].

<sup>91</sup> Dando conta do fraco interesse dos juristas pela teoria da decisão (não judicial), MARGARIDA LIMA REGO, “Decisões em ambiente de incerteza: probabilidade e convicção na formação das decisões judiciais”, *Julgar*, 21, 2013 (119-147), pp. 126-127.

<sup>92</sup> Desde a introdução do conceito de «*tacit knowledge*» por MICHAEL POLANYI, *The Tacit Dimension*, London, Routledge and Kegan Paul, 1966, até ao seu desenvolvimento com HUBERT DREYFUS /

Por seu turno, a principal fragilidade da decisão humana reconduz-se aos reflexos desta flexibilidade decisória: a inconsistência da decisão (que tanto pode ser aferida no quadro das várias decisões tomadas pelo mesmo decisor, como no quadro de várias decisões tomadas por diferentes decisores), seja porque se consideram fatores ou variáveis anteriormente desconsiderados, seja porque lhes foi atribuído um peso distinto na formação da decisão. A atribuição do valor e reconhecimento do peso relativo desses fatores tanto pode resultar de uma escolha consciente do próprio decisor, como de um enviesamento não percepcionável por este<sup>94</sup>.

Às decisões algorítmicas são comumente atribuídas características tidas como vantajosas, desde logo por serem de formação rápida e por considerarem um universo de dados ou de informações bem mais alargado do que aquele que a mente humana pode processar. A principal força das decisões algorítmicas encontra-se justamente naquilo que falha nas decisões humanas: a consistência. Os algoritmos são notoriamente consistentes e os seus resultados replicáveis: considerados os mesmos dados (*inputs*), e aplicável o mesmo modelo de análise, produz-se-á, à partida, a mesma decisão sem variações significativas<sup>95</sup>. Esta é a razão pela qual se defende que as decisões algorítmicas, quando comparadas com as decisões humanas, apresentam um risco consideravelmente inferior de produzir resultados discriminatórios<sup>96</sup>. Já as fragilidades usualmente apontadas às decisões algorítmicas prendem-se, a um tempo, com a suscetibilidade de enviesamento do processo decisório que lhes subjaz (enviesamento esse que, em maior ou menor medida, se

---

STUART DREYFUS, *Mind over Machine*, New York, The Free Press, 1986, e HUBERT DREYFUS, *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, MA, The MIT Press, 1992, que se disseminou a ideia de que o conhecimento científico é também o produto de um conhecimento tácito ou implícito, derivado de experiências pessoais, tradições, práticas ancestrais, valores implícitos, preconceitos, *sensatez e insights*. No contexto do texto, afirma FRANK PASQUALE, *New Laws of Robotics – Defending Human Expertise in the Age of AI*, Cambridge, Massachusetts, Belknap Press, Harvard University Press, 2020, p. 24, que “*we know more than we can explain*”.

<sup>93</sup> Seguimos de perto a síntese de características da decisão humana apresentada em REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., pp. 29-30.

<sup>94</sup> Cf. REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., pp. 30-31.

<sup>95</sup> Cf. REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., pp. 31-32.

<sup>96</sup> Cf. STEPHANIE BORNSTEIN, “Antidiscriminatory Algorithms”, *Alabama Law Review*, 70, 2018 (519-572); ALEX P. MILLER, “Want Less-Biased Decisions? Use Algorithms”, *Harvard Business Review*, 26, 2018, disponível em <<https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms>> [consultado em 20.07.2022]; IFEOMA AJUNWA, “The Paradox of Automation as Anti-Bias Intervention”, cit., p. 1671.

deverá à intervenção humana) e, a outro tempo, com a tendencial opacidade ou incompreensibilidade, para o homem médio, do *iter* decisório, dificultando a deteção – seja pelo destinatário da decisão, seja por terceiros – de eventuais enviesamentos. Radica aqui a ideia de que os algoritmos obscurecem e mascaram as intenções dos decisores humanos, dificultando a responsabilização destes<sup>97</sup>.

30. A distinção entre decisões humanas e decisões automatizadas convive, como visto, com realidades intermédias moldadas pelo grau de intervenção humana na formação da decisão. Quando esta distinção baseia a aplicação, a algum destes tipos de decisões, de um determinado regime legal ou contratual, a gradação da intervenção humana constitui um entrave significativo à tarefa do intérprete-aplicador. Sobretudo se o regime convocado não explicita a medida da intervenção humana (ou da ausência dessa intervenção) necessária para a sua aplicação.

Partindo da já reconhecida impossibilidade de produção de uma decisão algorítmica totalmente arredada de intervenção humana, o critério aventado, desenvolvido a propósito da noção de decisão individual automatizada apresentada no artigo 22.º do RGPD, faz apelo à materialidade dessa intervenção na formação da decisão: será uma decisão humana, e não uma decisão algorítmica, aquela em que o decisor humano, no âmbito da sua liberdade decisória, não se tenha limitado a *transmitir* ao destinatário a decisão formada algorítmicamente. Intui-se o paralelismo entre a figura do nuncio e a figura do representante e a similitude do critério que preside à distinção destas figuras com o critério aqui adotado<sup>98</sup>. Não significa isto, porém, que a qualificação de uma decisão como humana pressuponha que o decisor humano não siga ou não adira à pré-decisão formada algorítmicamente, ou que, fazendo-o, se exija a formulação de um juízo de causalidade hipotética pela qual se afira se o decisor humano, não fosse a pré-decisão formada algorítmicamente, teria tomado a *mesma* decisão. Tal tornaria qualquer decisão humana apoiada em sistemas de assistência algorítmica numa decisão algorítmica, ainda que, de facto, fosse conferida margem de liberdade decisória final ao ser humano.

---

<sup>97</sup> Cf. REBECCA CROTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., p. 34; IFEOMA AJUNWA, “The Paradox of Automation as Anti-Bias Intervention”, cit., p. 1692; SOLON BAROCAS / ANDREW D. SELBST BAROCAS, “Big Data’s Disparate Impact”, *California Law Review*, 104, 2016 (671-732), p. 693.

<sup>98</sup> O representante tem uma margem de autonomia na qual, pelo menos, se reconhece o poder de decidir o *se* do negócio e frequentemente, mais do que isso, o seu *conteúdo*. Nas palavras de CARLOS MOTA PINTO, *Teoria Geral do Direito Civil*, 4.ª ed., Coimbra, Coimbra Editora, p. 543, “o representante emite uma declaração em nome de outrem. O nuncio *transmite* uma declaração de outrem. O nuncio é um mero *longamanus*, um braço mais comprido. O representante consuma; o nuncio *transmite o já consumado*”.

O critério de primeira ordem a mobilizar para uma distinção, neste contexto, entre decisões humanas e algorítmicas assenta, assim, na *suscetibilidade* de alteração, pelo decisor, da pré-decisão produzida algorítmicamente e não na sua efetiva alteração. Nos casos em que o decisor humano adira a essa pré-decisão, convoca-se a aplicação de um critério de segunda ordem: há intervenção humana relevante para qualificar a decisão como humana quando a conduta do decisor humano no processo de tomada de decisão, apesar da sua adesão, seja reveladora do exercício de autonomia decisória – ainda que essa conduta seja, no caso concreto, desconhecida do destinatário da decisão<sup>99</sup>. A exteriorização desse exercício pelo decisor humano assentará em indícios próprios, combinados ou não entre si:

Entre outros indícios, considere-se (i) o conhecimento, pelo decisor humano, dos parâmetros que norteiam o funcionamento do algoritmo, (ii) a intervenção deste no processo decisório (por exemplo, o decisor humano tem de aceitar a pré-decisão através de ato especificado), (iii) a possibilidade de o decisor humano tomar uma decisão rejeitando a pré-decisão ou desviando-se do conteúdo desta (por exemplo, permitindo o *software* utilizado essa rejeição ou desvio, sem impedir ou bloquear o seguimento do procedimento decisório, ou sem que sejam criados mecanismos organizacionais desincentivadores desta rejeição ou desvio), (iv) a justificação especificada apresentada pelo decisor humano para aderir à pré-decisão (que pode acompanhar a referida aceitação), (v) a transmissão da decisão ao seu destinatário (sem ser um envio *per relationem* ou uma reprodução do enunciado da pré-decisão apresentado pelo sistema algorítmico), (v) bem como a explicação da mesma ao destinatário.

A fixação indiciária da exteriorização de uma intervenção humana leva a que estes indícios apenas sejam valorados como depondo no sentido dessa intervenção se, geralmente em combinação, forem suficientemente fortes para demonstrar, com a adesão à pré-decisão, um exercício efetivo (não simbólico) da liberdade decisória – *i.e.*, da *decisão de aderir ou de manter* a pré-decisão algorítmica.

31. A ideia de que a decisão humana, para o ser, nunca poderá, em alguma medida, estar pré-determinada, merece reparo. No plano negocial, mesmo sem entrar nas sobejamente conhecidas discussões em torno da liberdade de celebração e liberdade de estipulação concitadas por regimes como o das cláusulas contratuais

---

<sup>99</sup> Regressando ao paralelismo com o critério distintivo entre nuncio e representante, cf., porém, o entendimento segundo o qual a qualidade de representante ou de nuncio se afere à luz do comportamento do sujeito em face da contraparte e não em face da vontade do comitente (CARLOS MOTA PINTO, *Teoria Geral do Direito Civil*, cit., p. 543).



gerais – em que ainda se reconhece ao sujeito aderente um espaço de liberdade decisória<sup>100</sup> –, basta considerar hipóteses em que parâmetros decisórios de fonte legal condicionam a decisão negocial humana – ou mesmo a decisão algorítmica – que venha a ser produzida.

Pense-se na apresentação, por uma instituição de crédito a um cliente, de uma proposta de crédito à habitação. A decisão de concessão de crédito é hoje modelada por um conjunto de parâmetros – cuja normatividade é discutível, por serem definidos em Recomendação do Banco de Portugal<sup>101</sup> – que interferem, em grande medida, com o *se* e com o *como* da proposta. Entre esses parâmetros, encontra-se a não concessão de crédito à habitação destinado à aquisição ou à construção de habitação própria e permanente sempre que apresente um *loan-to-value* superior a 90%, ou sempre que *debt service-to-income* (taxa de esforço) dos mutuários seja superior a 50%. São também estabelecidos limites temporais ao contrato de crédito à habitação, em função da idade dos mutuários: a maturidade máxima destes créditos deve ser de 40 anos, para mutuários com idade inferior ou igual a 30 anos; de 37 anos, para mutuários com idade superior a 30 anos e inferior ou igual a 35 anos; e de 35 anos, para mutuários com idade superior a 35 anos<sup>102</sup>.

32. Isolamos, a propósito da distinção entre decisão humana e decisão automatizada, dois casos-difíceis: o da assistência algorítmica ao decisor humano e o da revisão humana das decisões algorítmicas.

Pese embora um se reporte a decisões humanas *tout court* e o outro a decisões algorítmicas *dobradas* por decisões humanas, suscitam estes dois casos o mesmo problema nuclear: o da presunção de que, quanto a certas matérias, ou em certos contextos, as “máquinas decidem melhor”, com base nisso se justificando a adesão e confirmação da decisão algorítmica por parte do decisor humano, incluindo quando, formulado um juízo hipotético, o decisor humano, só por si, não teria decidido daquela forma. Esta presunção, também designada por “*automation bias*”, parte da alocação de competências decisórias entre o ser humano e a máquina

---

<sup>100</sup> Vide, por todos, JOAQUIM DE SOUSA RIBEIRO, *O Problema do Contrato – As Cláusulas Contratuais Gerais e o Princípio da Liberdade Contratual*, Coimbra, Almedina, 1999, pp. 275 ss.

<sup>101</sup> Cf. *Recomendação do Banco de Portugal no âmbito dos novos contratos de crédito celebrados com consumidores*, disponível em <[https://www.bportugal.pt/sites/default/files/recomendacao\\_contratoscredito\\_consolidada\\_pt.pdf](https://www.bportugal.pt/sites/default/files/recomendacao_contratoscredito_consolidada_pt.pdf)> [consultado em 20.07.2022].

<sup>102</sup> Cf. *Comunicado do Banco de Portugal sobre convergência da maturidade média dos novos contratos de crédito à habitação para 30 anos*, de 31 de janeiro de 2022, disponível em <<https://www.bportugal.pt/comunicado/comunicado-do-banco-de-portugal-sobre-convergencia-da-maturidade-media-dos-novos-0>> [consultado em 20.07.2022].

sugerida pela ideia de “*MABA-MABA trap*” (acrónimos de “*Men Are Better At*” e de “*Machines Are Better At*”)<sup>103</sup>, geralmente com base em regras da experiência, pugnando por essa alocação em função das capacidades requeridas para a formação de uma dada decisão (por exemplo, as máquinas são melhores na tomada de decisões que impliquem a execução de tarefas repetitivas, enquanto o ser humano é melhor a tomar decisões que impliquem a elaboração de análises contextuais). É na base dessa alocação de competências decisórias que se forma a presunção ou rasteira (“*trap*”) que leva a que o decisor humano  *siga* as decisões algorítmicas, não as contrariando, quando considera aprioristicamente que, dada a matéria ou o contexto, a máquina decidirá “melhor”.

33. Por assistência algorítmica referimo-nos a sistemas algorítmicos de apoio à escolha e decisão humana – sem que a utilização desses sistemas, pelo menos *prima facie*, desvirtue a natureza humana da decisão<sup>104</sup>.

Entre as formas de assistência algorítmica mais conhecidas encontram-se, no seu nível menos intenso, a sugestão, ao consumidor digital, de produtos similares aos pesquisados ou adquiridos através de plataformas digitais, ou a apresentação, a esse consumidor, de uma comparação (quanto a características técnicas, preço, etc.) de produtos similares aos pesquisados ou adquiridos.

Num nível mais avançado, tem-se popularizado a utilização de assistentes algorítmicos que oferecem ao seu utilizador – o decisor humano – uma decisão final ou quase final (a “decisão ótima” ou a “melhor decisão”), bem como decisões alternativas (que surgem, geralmente, ordenadas pelo seu nível de otimização face às preferências – ou outros parâmetros considerados pelo algoritmo – do destinatário final da decisão). No domínio financeiro, pense-se, por exemplo, no *software* de *robo-advising* utilizado por um consultor para investimento para – ele próprio – poder apresentar ao seu cliente uma ou várias recomendações para investimento, ou nesse mesmo *software* utilizado por um gestor de carteiras para – também ele próprio – tomar decisões de investimento, em nome e por conta dos seus clientes, em relação à composição da carteira destes. Ainda no domínio financeiro, considere-se o perfil do investidor apresentado ao intermediário financeiro pelo *software* de assistência algorítmica que este

---

<sup>103</sup> A expressão deve-se a MEG L. JONES, “The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles”, *Vanderbilt Journal of Entertainment and Technology Law*, 18, 2015 (77-134), p. 104, tendo sido amplificada por REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., p. 8.

<sup>104</sup> Apresentando uma taxonomia de assistentes algorítmicos que cruza o critério dos parâmetros decisórios utilizados pelo algoritmo e o critério do grau de liberdade decisória deixada ao ser humano, cf. MICHAL S. GAL, “Algorithmic Challenges to Autonomous Choice”, cit., pp. 66 ss.

utiliza, após o investidor preencher o questionário de avaliação da adequação que lhe é exigido, em cumprimento de ditames legais, pelo intermediário financeiro. No domínio bancário, pense-se no *software* utilizado pelo colaborador do banco para aferir o risco de crédito do cliente com quem o banco pretende celebrar um contrato de crédito à habitação, e com base no qual o banco decidirá se aquele cliente é elegível para a concessão de crédito e, se sim, ao abrigo de que condições contratuais.

O grau de assistência algorítmica dependerá da medida da liberdade decisória deixada ao seu utilizador humano: mesmo nos casos em que o assistente algorítmico apresente uma decisão que poderia ser final (ou seja, com completude suficiente para a decisão ser processada pelo decisor humano, ou ser transmitida pelo utilizador humano ao seu destinatário), e mesmo que o decisor humano adira à pré-decisão formada algoritmicamente, não se trata, como visto, só por essa razão, de uma decisão automatizada. Para além dos indícios acima enunciados que, geralmente combinados entre si, evidenciam o exercício, pelo decisor humano, de uma autonomia decisória, na assistência algorítmica cabe ainda considerar outro índice, frequentemente desconsiderado, mas que é revelador de um espaço decisório óbvio aquando da intervenção humana: as decisões respeitantes à dimensão pessoal e afetiva da pessoa humana.

Retomando o exemplo das modalidades de assistência algorítmica que atuam em função das preferências do seu utilizador, considere-se o caso das *dating apps* ou plataformas de encontro *online*, que apresentam ao decisor humano uma ou várias pessoas compatíveis com as preferências por aquele transmitidas. Mesmo que o utilizador de uma *dating app* siga intuitivamente a recomendação que lhe é apresentada, encetando contacto com a pessoa que lhe é sugerida, é incontestável o exercício de autonomia decisória pelo utilizador humano.

Outro problema da assistência algorítmica, que aqui não cabe explorar, prende-se com a possível diluição ou apagamento das qualidades pessoais ou profissionais do decisor humano que fundaram a decisão de contratar de um terceiro ou, em geral, que são relevantes na conformação de uma qualquer relação jurídica *intuitu personae*.

Ilustra esta preocupação a hipótese – que tem vindo a ser debatida na disciplina da *corporate governance* – de um administrador de uma sociedade comercial, que foi designado pelos acionistas em função dos seus conhecimentos e experiência para desempenhar atividades de considerável complexidade técnica e de correspondente responsabilidade, tomar decisões compreendidas no exercício dessas funções através de

uma mera adesão às pré-decisões que lhe são apresentadas pelo assistente algorítmico que utiliza para o efeito. Em domínio marcadamente avesso à substituição dos administradores (nos termos do artigo 392.º, n.º 7, do Código das Sociedades Comerciais, não é permitido aos administradores fazerem-se representar no exercício do seu cargo), e com especiais exigências quanto à delegação de poderes nos casos previstos na lei (artigo 407.º do Código das Sociedades Comerciais), a assistência algorítmica utilizada nestes moldes pode representar um meio de superação das referidas limitações legais, comprometendo os valores por estas tutelados.

Questão diferente é o da designação de um sistema algorítmico como administrador de uma sociedade comercial, com a consequente atribuição, a esse sistema, de direito a votar no conselho de administração. A hipótese é *quase* real, remontando, no seu caso pioneiro, à designação de um sistema algorítmico (chamado Vital) como “observador” (mas não administrador) de um conselho de administração de uma sociedade comercial de Hong Kong. O caso tornou-se conhecido porque, mais do que um assistente algorítmico, os membros do conselho de administração aceitaram não tomar decisões de investimento sem a corroboração dessas decisões por parte do sistema algorítmico – reconhecendo-se a este sistema, em termos materiais, um poder de veto em relação às decisões de investimento dos administradores<sup>105</sup>.

34. A revisão humana de decisões algorítmica apresenta desafios particulares, desde logo pela sua consagração, no artigo 22.º, n.º 3, do RGPD, como um direito subjetivo do titular dos dados destinatário de uma decisão individual automatizada. Também a Proposta de Regulamento sobre Inteligência Artificial regula, no seu artigo 14.º, a matéria da supervisão humana, quando em causa estejam resultados de sistemas de inteligência artificial de risco elevado. Diversamente do que sucede com a revisão humana de decisões algorítmicas, em que a decisão humana *dobra* – confirmando ou alterando – a decisão algorítmica consumada, a supervisão humana proposta neste diploma é conformada como um dever, e não como um direito do destinatário da decisão, e reporta-se também ao *iter* decisório algorítmico e não apenas ao controlo humano de um resultado algorítmico já produzido.

---

<sup>105</sup> Cf. MARTIN PETRIN, “Corporate Management in the age of AI”, *Columbia Business Law Review*, 3, 2019 (965-1030), pp. 967-968. Entre nós, já desenvolveu o tema MADALENA PERESTRELO DE OLIVEIRA, em conferência intitulada “Automated Decisions: Self Driving Governance?”, na *International Conference on Autonomus Systems and the Law*, CIDP – FDUL, Lisboa, 20-21 de abril de 2022.

Esta conclusão apoia-se no artigo 14.º, n.º 4, da Proposta de Regulamento: determina este preceito que as medidas de supervisão humana devem possibilitar às pessoas incumbidas dessa supervisão, em função das circunstâncias, que estas “compreendam completamente as capacidades e limitações do sistema [...] e sejam capazes de controlar devidamente o seu funcionamento, de modo que os sinais de anomalias, disfuncionalidades e desempenho inesperado possam ser detetados e resolvidos o mais rapidamente possível”, que “estejam conscientes da possível tendência para confiar automaticamente ou confiar excessivamente no resultado produzido pelo sistema”, que “sejam capazes de interpretar corretamente o resultado do sistema”; que “sejam capazes de decidir, em qualquer situação específica, não usar o sistema [...] ou ignorar, anular ou reverter o resultado do sistema”; que possam “intervir no funcionamento do sistema [...] ou interromper o sistema por meio de um botão de «paragem» ou procedimento similar”. Tratando-se de sistemas de inteligência artificial de risco elevado concebidos para serem utilizados para a identificação biométrica, exige-se, como medida de supervisão, que o resultado produzido pelo sistema seja verificado e confirmado por, pelo menos, duas pessoas singulares (artigo 14.º, n.º 5, da Proposta de Regulamento).

35. Têm subido de tom as vozes que propalam a revisão humana das decisões algorítmicas, no contexto do já referido modelo corretivo de regulação algorítmica. A solução apresentada na Proposta de Regulamento sobre Inteligência Artificial sinaliza a atenção e acompanhamento, pelo legislador da União Europeia, destes movimentos. A defesa da revisão humana encontra justificação no papel que o revisor pode *ex post* desempenhar.

São apontadas seis razões para a exigência de revisão humana da decisão algorítmica<sup>106</sup>. Em primeiro lugar, a possibilidade de *correção* da decisão algorítmica: seja corrigindo os dados considerados pelo algoritmo, seja corrigindo a decisão em função de um contexto desconsiderado pelo algoritmo (“*situational tailoring*”), seja corrigindo enviesamentos detetados no modo de funcionamento ou nos resultados do algoritmo. Em segundo lugar, apela-se à revisão do ser humano para que este *explique* a decisão automatizada ao seu destinatário. Razões de transparência efetiva (com vista a combater a ilusoriedade da transparência que poderá resultar da mera comunicação do sistema algorítmico, com todas as suas especificidades técnicas, ao destinatário) e razões de perceção de legitimidade da decisão (independentemente da sua correção) justificam este papel explicativo da revisão

---

<sup>106</sup> Seguimos as razões apresentadas em REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., pp. 50 ss.

humana<sup>107</sup>. Em terceiro lugar, fala-se da revisão humana como *garantia da dignidade* de cada ser humano: invoca-se, a este propósito, que a sujeição do ser humano a decisões algorítmicas leva a que os destinatários da decisão não sejam tratados de forma individual, mas categorial, sendo – com apelo ao conhecido imperativo categórico kantiano – meros meios para alcançar um fim. O combate a esta objetificação do destinatário da decisão agita até pergaminhos constitucionais, falando-se de um direito fundamental à revisão humana de decisões automatizadas<sup>108</sup>. Em quarto lugar, a revisão humana reforça a *imputabilidade* da decisão algorítmica, evitando a ideia de delegação, nos sistemas algorítmicos, das decisões difíceis e, com isso, a subversão dos esquemas de responsabilidade legalmente previstos. A intervenção do ser humano *ex post* serve o propósito de garantir que alguém “dá a cara” perante o destinatário da decisão, independentemente da aplicabilidade de mecanismos de imputação e de responsabilidade jurídica. Mas não só. É a este propósito que se fala do revisor humano como um escudo – o “*fall guy*”<sup>109</sup>, a “*liability sponge*”<sup>110</sup> ou mesmo a “*moral crumple zone*”<sup>111</sup> – da organização criadora

---

<sup>107</sup> De acordo com um conjunto de estudos explanados em KIRSTEN MARTIN / ARI EZRA WALDMAN, “Perceptions of the Legitimacy of Automated Decision-Making”, *Big Data & Society*, 9/1, 2022 (1-16), a opacidade decisória das decisões algorítmicas leva a uma diminuição da percepção de legitimidade da decisão, sendo essa diminuição contrariada com a explicação humana dessa decisão. Não obstante, resgatando as conclusões desse estudo, a um tempo, “[t]he perceived legitimacy decreased significantly when algorithmic systems made decisions based on general data about an individual gathered from the Internet”, mas, a outro tempo, “[i]nterestingly, we found that human-led decisions were perceived as less legitimate than algorithmic decisions when general data was used for the decision. There could be multiple explanations for this. For example, respondents may not believe humans could adequately analyze a larger, more heterogeneous dataset as compared to a computer. [...] The responses may also reflect the recognition, well known in the literature, that humans are biased, too” (p. 11).

<sup>108</sup> Cf. AZIZ Z. HUQ, a “Right to a Human Decision”, cit., p. 677; REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., p. 56.

<sup>109</sup> Cf. REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., p. 57.

<sup>110</sup> Cf. MADELEINE. C. ELISH / TIM HWANG, “Praise the Machine! Punish the Human! The Contradictory History of Accountability in Automated Aviation, Comparative Studies in Intelligent Systems”, *Working Paper #1 V2* (May 18, 2015), disponível em <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2720477](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2720477)> [consultado em 20.07.2022], p. 15.

<sup>111</sup> Cf. MADELEINE. C. ELISH, “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction”, in *Engaging Science, Technology, and Society*, 5, 2019 (40-60). A Autora diz que introduz este conceito “to describe how responsibility for an action may be misattributed to a human actor who had limited control over the behavior of an automated or autonomous system. Just as the crumple zone in a car is designed to absorb the force of impact in a crash, the human in a highly complex and automated system may become simply a component – accidentally or intentionally – that bears the brunt of the moral and legal responsibilities when the overall system malfunctions” (p. 40). No mesmo sentido, em tom

ou utilizadora do sistema algorítmico. Em quinto lugar, fala-se da intervenção humana como mera presença ou “*warm bodies*”<sup>112</sup>, para valorizar e manter a necessidade e o significado da força de trabalho humano em áreas dominadas pela inteligência artificial, em que essa força de trabalho é facilmente substituível<sup>113</sup>. Por fim, à revisão humana das decisões automatizadas aponta-se ainda o papel de *interface* entre o destinatário da decisão e o revisor humano. A intermediação da decisão pelo ser humano, mesmo que não diretamente dirigida à explicação da decisão algorítmica ao seu destinatário, é geralmente mais útil, mais confiável e mais expedita para esse destinatário do que a remissão para um *interface* algorítmico de resposta a questões do destinatário.

### 3.3. O efeito discriminatório, o efeito lesivo e o efeito multiplicador das decisões algorítmicas; particularidades da *big data*

36. Tanto no quadro de um discurso ético, como no quadro de um discurso jurídico<sup>114</sup>, é hoje lugar-comum a ideia de que a utilização dos algoritmos na

---

crítico, notam REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., pp. 57-58, que não só o revisor humano incluído da fase final do circuito é um escudo protetor do próprio sistema, como também protegem dos mecanismos de imputação uma série de decisores que contribuíram ou poderiam ter contribuído para evitar o evento negativo verificado – *i.e.*, os decisores humanos que projetaram, programaram, fabricaram, compraram ou implantaram o sistema. Neste quadro, os Autores dão conta, fornecendo as devidas referências, que os tribunais norte-americanos, chamados a pronunciar-se em litígios de responsabilidade civil extracontratual relativos a acidentes que envolveram robôs controlados *in loco* por uma pessoa humana, consideraram, em numerosos casos, ser exclusivamente responsável essa pessoa (mesmo quando o acidente se deveu a falha do sistema algorítmico), sem estender a responsabilidade às pessoas singulares ou coletivas que criaram ou utilizaram, em seu proveito, o sistema, e que poderiam ter contribuído para evitar o acidente.

<sup>112</sup> A expressão é de REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., p. 58.

<sup>113</sup> Com desenvolvimentos sobre esta ideia, invoca FRANK PASQUALE, *New Laws of Robotics – Defending Human Expertise in the Age of AI*, cit., p. 3, aquilo que considera ser um princípio fundamental da robótica: os sistemas robóticos e a inteligência artificial devem complementar os seres humanos profissionais, mas não substituí-los (ou, noutras palavras do Autor, a inteligência artificial deve *aumentar*, mas não substituir a inteligência humana). O Autor sustenta que mesmo em contextos totalmente automatizados, deve reter-se propositalmente trabalho significativo para o ser humano, porque o exercício da força de trabalho o valoriza enquanto indivíduo e enquanto membro de uma determinada comunidade.

<sup>114</sup> Analisando os fenômenos de enviesamento algorítmico no contexto dos *big data* pela lente do discurso ético e pela lente do discurso jurídico, cfr. KLAUS WIEGERLING / MICHAEL NERURKAR / CHRISTIAN WADEPHUL, “Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien”, in BARBARA KOLANY-RAISER / REINHARD HEIL / REINHARD ORWAT / THOMAS

formação de uma decisão apresenta o risco de serem tomadas decisões discriminatórias em relação a um determinado sujeito. Sem esgotar a fenomenologia de disfunções possíveis, o conteúdo discriminatório de uma decisão algorítmica pode ancorar-se em enviesamentos da amostra de dados sujeitos a tratamento algorítmico (causados pela insuficiência desses dados, seja em quantidade, seja em qualidade<sup>115</sup>), em enviesamentos introduzidos pela correlação estatística, ou mesmo em práticas

---

HOEREN (eds.), *Big Data und Gesellschaft. Eine multidisziplinäre Annäherung*, Wiesbaden, Springer VS, 2018 (1-32), pp. 7 e 8.

<sup>115</sup> Isto se admitirmos que (i) os dados assim sujeitos a tratamento não são “falsos”, e que (ii) esse tratamento não é, à luz de um específico regime normativo (designadamente, os regimes da proteção de dados pessoais), *ilícito*. Note-se que entre os diversos problemas apontados às decisões algorítmicas não se encontra, por regra, a “falsidade” dos dados que alimentam os modelos de análise. Numa perspetiva jurídica, os problemas deslocam-se, antes, para os enviesamentos que envolvem as inferências estabelecidas a partir desses dados (“verdadeiros”) ou na seleção desses dados. Referimo-nos, aqui, a dados “verdadeiros” ou “falsos” com cautela, cientes da impropriedade do discurso se pensarmos nos dados gerados eletronicamente e nos dados em larga escala extraídos da Internet (que correspondem à “pegada digital” do utilizador) com recurso à tecnologia *big data*. Os dados são uma representação (descrição) da realidade (dos factos, brutos ou institucionais). Como nota JOHN SEARLE (*The construction of social reality*, London, Penguin Books, 1996, p. 2), “the fact stated needs to be distinguished from the statement of it”. Sendo os dados descrições, as mesmas reconduzir-se-ão aos valores de verdade ou de falsidade. Todavia, há dados que, sendo descrições, são *meros registos* de factos, frequentemente gerados de forma automática, sem dependência de uma concreta intervenção humana. Isto significa que são representações ou descrições (quantitativas ou qualitativas) neutras de factos, cuja veracidade não pode ser posta em causa através de uma outra interpretação ou leitura e que servem de matéria-prima a um conjunto de inferências (as resultantes do seu processamento e análise) – são estes os “*raw data*”, “*atomic data*”, “*primary data*” ou, como designados por CHARLES TAYLOR, *Philosophy and the Human Sciences. Philosophical Papers, Volume 2*, Cambridge, Cambridge University Press, 1985, p. 19, os “*brute data*”, que os define nos seguintes termos: “By ‘brute data’ I mean here and throughout data whose validity cannot be questioned by offering another interpretation or reading, data whose credibility cannot be founded or undetermined by further reasoning. If such a difference of interpretation can arise over [brute] data, then it must be possible to structure the argument so as to distinguish the basic, brute data from the inferences made on the basis of them”. No mesmo sentido, convincentemente, após percorrer a noção de «dados», partindo da sua etimologia e à luz de vários ramos do conhecimento, afirma DANIEL ROSENBERG, “Data before the Fact”, in LISA GITELMAN (ed.), “*Raw Data*” *Is an Oxymoron*, Cambridge, Massachusetts, London, The MIT Press, 2013, p. 37, que “*Data has no truth. Even today, when we speak of data, we make no assumptions at all about veracity. Electronic data, like the data of the early modern period, is given*”. Pense-se nos seguintes dados recolhidos da Internet: o utilizador A publicou numa rede social uma fotografia sua a mergulhar numa piscina; o utilizador B navegou em cinco páginas de jornais digitais; o utilizador C realizou uma transferência bancária através do seu *homebanking*, tendo ficado imediatamente registado no seu perfil de utilizador essa operação bancária; o utilizador D votou numa plataforma *online* num concorrente de um determinado concurso televisivo.



discriminatórias de quem tem o domínio da criação do algoritmo<sup>116</sup>. As disfunções algorítmicas – intencionais ou não – são imputáveis à atuação humana – à partida, mas sem excluir outros sujeitos intervenientes no circuito da decisão algorítmica, ao criador do algoritmo, que é quem tem o domínio da sua conceção, ainda que cumprindo instruções de terceiro.

O fenómeno discriminatório das decisões algorítmicas é amplamente difundido na literatura jurídica com exemplos que sugerem enviesamentos verificáveis *em qualquer fase* do processo decisório: desde a produção da fórmula algorítmica à obtenção do resultado decisório.

Casos paradigmáticos – cuja proliferação, na prática, torna inviável a enunciação dos contornos de casos concretos – são os do recrutamento *online* através de sistemas algorítmicos que discriminam (*i.e.*, não selecionam para entrevista de emprego, ou não contratam) candidatas que sejam mulheres em idade reprodutiva, candidatos que sejam pais e mães solteiros, candidatos que sejam pais de crianças pequenas, ou candidatos com residência numa determinada área da cidade, para determinadas funções ou categorias profissionais. Na concessão de crédito, tornaram-se conhecidos os casos de discriminação algorítmica de sujeitos pertencentes a comunidades imigrantes, da dificuldade do acesso ao crédito por parte de mulheres empreendedoras no lançamento do seu primeiro negócio, de casais em que o cartão de crédito do marido, associado a uma conta conjunta do casal, tem um *plafond* superior ao cartão de crédito da mulher, também associado a essa conta conjunta, ou mesmo a recusa de concessão de crédito à habitação a um casal que frequentava semanalmente sessões de terapia de casal (cujo pagamento constituía um facto evidenciado no extrato de conta bancária, disponibilizado pelo casal à instituição de crédito, em fase pré-negocial)<sup>117</sup>.

A associação de resultados discriminatórios às decisões algorítmicas carece de dois esclarecimentos, atrás antecipados.

Por um lado, o esclarecimento de que as decisões humanas também são discriminatórias, não sendo este um atributo das decisões algorítmicas: as decisões algorítmicas discriminatórias mais não são, em rigor, do que uma *projeção* da

---

<sup>116</sup> Em sentido próximo, sustentando que o viés algorítmico pode ser decomposto em três componentes: viés na escolha dos dados (*inputs*), viés na escolha do resultado (*outputs*) e viés na construção do algoritmo, cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, *Journal of Legal Analysis*, 2019 (1-62), disponível em <<https://inequality.hks.harvard.edu/files/inequality/files/ludwig19.pdf>>, p. 52.

<sup>117</sup> Para mais exemplos de discriminação algorítmica, *vide*, entre nós, MAFALDA MIRANDA BARBOSA, “Discriminação algorítmica: reflexões em torno da responsabilidade civil”, cit., pp. 1200 ss.

discriminação que envolve os processos decisórios humanos<sup>118</sup>. O uso de um algoritmo é, pois, nestes casos, uma forma sucedânea de materialização e de revelação de um viés do processo decisório humano<sup>119</sup>. Sucede, porém, que a discriminação resultante de uma decisão humana é mais difícil de detetar: seja porque os decisores humanos não estão frequentemente cientes dessa discriminação, seja porque as decisões humanas sobre matérias mais elementares padecem, regra geral, de uma opacidade natural, tanto externa como interna, derivada da não problematização nem decomposição analítica do processo decisório, seja porque – como se verá de seguida – as decisões humanas, por regra, não se projetam e replicam em tão larga escala como as decisões algorítmicas<sup>120</sup>.

Estas conclusões conduzem-nos ao segundo esclarecimento. Uma das principais preocupações dos juristas relativas a práticas discriminatórias que sejam consideradas antijurídicas prende-se com a *prova* dessa discriminação ou das motivações que subjazem à decisão humana. A associação de resultados discriminatórios a decisões algorítmicas é potenciada justamente porque os algoritmos não constituem um desafio epistemológico tão espinhoso como as decisões humanas: com mais ou menos dificuldades de compreensão para o intérprete, a fórmula algorítmica utilizada, comportando a matriz dessa discriminação, exporá os elementos discriminatórios que são relevados, facilitando a prova dessa discriminação. Com isto, dir-se-á que as decisões humanas são até potencialmente *mais* discriminatórias do que as decisões algorítmicas; simplesmente, essa discriminação não é revelada pelo acesso à fórmula mental utilizada pelo decisor humano<sup>121</sup>.

A revelação algorítmica da discriminação humana é ilustrada por um caso estudado por investigadores da Carnegie Mellon University: descobriu-se que o Google tende a dirigir anúncios de emprego com salários mais baixos para mulheres, quando comparados com aqueles direcionados a homens. Embora os investigadores nunca tenham comprovado o motivo para esta disparidade, admitiram que os algoritmos do Google aprenderam com as desigualdades existentes na sociedade: as mulheres estão mais associadas a trabalhos de

---

<sup>118</sup> Referindo-se mesmo a escolhas humanas discriminatórias que “infetam” os algoritmos, cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, cit., p. 26.

<sup>119</sup> Cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, cit., p. 42. Conforme assinalam os Autores, “[a]lgorithms can reveal our own biases”, cit., p. 49.

<sup>120</sup> Cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, cit., p. 2.

<sup>121</sup> Cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, cit., p. 2.

salários mais baixos e tendem a clicar em anúncios de empregos de salários mais baixos. Simplesmente, o algoritmo extrapolou esse comportamento e continuou o padrão<sup>122</sup>.

37. Os algoritmos podem ser utilizados para que determinados sujeitos – pessoas singulares ou coletivas – tomem decisões em contextos jurídico-negociais sobre matérias fundamentais da vida de uma pessoa singular, como seja, entre exemplos já abordados, a concessão de crédito, o acesso ao emprego ou à educação<sup>123</sup>.

---

<sup>122</sup> Cf. A. DATTA / M. C. TSCHANTZ / A. DATTA, “Automated Experiments on and Privacy Settings: a Tale of Opacity, Choice, and Discrimination”, in *Proceedings on Privacy Enhancing Technology* (PoPETs), De Gruyter Open, 2015, disponível em < <https://arxiv.org/pdf/1408.6491.pdf> > [consultado em 20.07.2022].

<sup>123</sup> “Fundamentais”, neste sentido, não por respeitarem (necessariamente) a direitos fundamentais de consagração e proteção constitucional, mas pela sua aptidão, segundo as regras da experiência, para determinar e condicionar o percurso de vida de uma pessoa singular. Mesmo que perspectivado de um prisma “puro” de direito privado (alheado de qualquer consideração sobre a eficácia dos direitos fundamentais nas relações entre particulares), estas são matérias marcadas pela “oportunidade” que representa a celebração de um dado negócio, ou a celebração de um dado negócio em certos termos. Ainda que a avaliação da oportunidade do negócio em face das condicionantes que o envolvem caiba, regra geral, a cada uma das partes envolvidas, não é arriscado afirmar que a oportunidade, nestas matérias, não se cinge à expectativa de aproveitamento das utilidades emergentes daquele concreto negócio, antes radicando na ideia de que “oportunidade gera oportunidade”, *i.e.*, na suscetibilidade de multiplicar e fazer frutificar, em cadeia, resultados tidos como “bons” para o sujeito. Pela perspectiva da teoria de «igualdade de oportunidades» – cuja relevância atravessa os planos político, filosófico, sociológico e económico e reclama, com vista à sua *legal enforceability*, um lugar próprio no campo jurídico, em particular, no direito constitucional –, as matérias indicadas em texto (o crédito, o emprego, a educação), entre outras, respeitam a fatores-chave de igualação de oportunidades num universo ideal de sujeitos que, perante uma certa conceção de justiça distributiva, devem, a partir do mesmo *level playing field*, poder ser alcançados por esses sujeitos *sem* (ou com *os mesmos*) obstáculos. Para umas das primeiras vozes sobre o ideal do “*level playing field*”, cfr. o estudo exordial de JOHN ROEMER, “Equality and Responsibility”, *Boston Review*, April-May, 1995 (3-7), a que se seguiu a sua obra seminal *Equality of Opportunity*, Cambridge, Cambridge University Press, 1998. O tema tem recebido amplo tratamento interdisciplinar, destacando-se, com críticas, a análise mais recente de MATTHEW T. JEFFERS (“Luck and the limits of equality”, *Philosophical Papers*, 49/3, 2020 [397-429]) às teses do *luck egalitarianism*, corrente filosófica assim cunhada por ELIZABETH ANDERSON (em “What is the point of Equality?”, *Ethics*, 109, 1999 [287-337]), para expressar a ideia de que o alcance da justiça requer, numa fase inicial, o nivelamento do *playing field*, tornando iguais (*equalizing*) as oportunidades de todos os sujeitos de um dado universo, permitindo, em seguida, que as escolhas individuais destes sujeitos e seus efeitos ditem os resultados finais – de forma a que se possa defender que o sucesso conseguido por um determinado sujeito ao longo da sua vida resulta apenas das decisões e escolhas que lhe são imputáveis durante esse percurso e não de condicionamentos ou interferências de “*brute luck*”. Para um quadro geral da teoria de «igualdade de oportunidades» e suas projeções nos diversos domínios das ciências sociais, cfr. RICHARD ARNESON, “Equality of Opportunity”, in EDWARD N. ZALTA (ed.), *The Stanford*

Nestes casos, uma decisão algorítmica discriminatória poderá conduzir a um resultado indesejado – à lesão de um interesse (mais dificilmente, de um direito

---

*Encyclopedia of Philosophy* (Summer Edition, 2015), disponível em <<https://plato.stanford.edu/archives/sum2015/entries/equal-opportunity/>> [consultado em 20.07.2022]. As patologias associadas a erros, vieses e discriminações que afetem os algoritmos que baseiam decisões em contextos jurídico-negociais nestas matérias “fundamentais”, mesmo que não sejam causadas intencionalmente pelo ser humano (ou mesmo que não sejam, de todo, causadas pelo ser humano), constituem obstáculos ao referido *level playing field*. E, como tal, poderão conduzir ao desperdício (e subsequente perda) de oportunidades dos sujeitos utilizadores ou destinatários dessas decisões algorítmicas. Recorrendo à noção de oportunidade formulada por PETER WESTEN (em “The Concept of Equal Opportunity”, in *Ethics*, 95/4, 1985 (837-850), pp. 838-839), é curial enfatizar o papel dos “obstáculos” nessa concetualização: de acordo com o Autor, a oportunidade consiste numa relação tridimensional entre (i) a pessoa a quem a oportunidade pertence, (ii) um objetivo (ou conjunto de objetivos) desejado, e (iii) determinados obstáculos. Uma pessoa só tem uma oportunidade se tiver a possibilidade, embora não a garantia, de atingir aquele objetivo (“*while an opportunity is something less than a guarantee, it is something more than a mere possibility*”, p. 839). Não existe oportunidade quando o sujeito em causa enfrenta, pelo caminho, obstáculos intransponíveis (“*insurmountable*”) ou obstáculos excessivos que tornem difícil (“praticamente impossível”) alcançar aquele objetivo. O exemplo fornecido por WESTEN, que aqui recuperamos, é o da oportunidade de uma mulher ser recrutada para exercer funções de hospedeira numa determinada companhia aérea. Imagine-se que, inicialmente, essa companhia aérea elegera como requisito determinante para a contratação de hospedeiras que estas fossem mulheres não casadas, e que, posteriormente, havia dispensado esse requisito, aceitando contratar como hospedeiras *também* mulheres casadas. Ora, o estatuto conjugal de casada não é (não seria) um obstáculo intransponível, na aceção de WESTEN, como seria o da raça, da etnia ou do género, se esses fossem critérios para a contratação de hospedeiras. E não o é (não seria) porque o estatuto de casada pode ser alterado unilateralmente pela própria candidata àquele emprego; ainda assim, não tivesse o requisito do estado civil de “não-casada” sido removido, não usaríamos um discurso de oportunidade para descrever a contratação de hospedeiras por essa companhia aérea. O que sugere que o conceito de oportunidade se encontra justamente entre uma garantia e uma mera possibilidade – “somewhere between the absence of all possible obstacles between X and Y, on the one hand, and the possible absence of all obstacles between X and Y, on the other hand” (p. 840).

Esclareça-se que a “oportunidade” que aqui releva não se ancora na visão rawlsiana do contrato (ou do negócio) como *veículo de justiça* (cfr. JOHN RAWLS, *A Theory of Justice*, Cambridge, MA, Harvard University Press, 1999, pp. 262 ss.). Pois, como alerta FERNANDO ARAÚJO (*Teoria económica do contrato*, Coimbra, Almedina, 2007, pp. 495-496) em alusão a esta teoria, “os contratos não são o veículo adequado promoção social da justiça. Conquanto sejam capazes de produzir externalidades, tanto negativas como positivas, conquanto eles sejam suscetíveis de gerar «efeitos de contágio» e «efeitos de rede», normalmente a sua eficácia é puramente bilateral – diríamos mesmo, assumida e *exemplarmente* bilateral, esgotando-se nesse plano, isto ainda que indiretamente o somatório de relações contratuais possa sedimentar um costume com grande relevância social e jurídica”. As “oportunidades” emergentes dos contextos jurídico-negociais aqui versados muito menos derivam, já no plano da eficiência paretiana, de conceções utilitaristas do contrato, que o identificam como instrumento maximizador de bem-estar.

subjetivo) – para o destinatário afetado pela decisão: a exclusão ou a perda de oportunidades negociais. Não sendo as decisões algorítmicas mais discriminatórias do que as decisões humanas, não seria correto afirmar, sem mais, que o risco de desperdício de oportunidades negociais é mais intenso em decisões algorítmicas. Esta afirmação convive, não obstante, com a constatação de que as decisões algorítmicas se projetam em larga escala (o efeito multiplicador de que se fala adiante), replicando e amplificando as consequências negativas que delas resultem. Só dentro destas coordenadas é que se encontra justificação a ideia de um efeito lesivo próprio das decisões algorítmicas.

A perda de oportunidades negociais (oportunidade de contratar, ou de contratar em certos moldes) – seja numa perspectiva abstrata e sistémica do tráfico jurídico contratual, seja na perspectiva individualizada da lesão dos interesses do sujeito destinatário da decisão – apenas releva enquanto efeito lesivo das decisões algorítmicas na estrita medida em que tal perda se ancore numa decisão algorítmica formada em violação dos parâmetros principiológicos e normativos aplicáveis. Noutros casos, como se explora adiante (3.4.), a verificação de uma perda negocial será, com elevada probabilidade, na perspectiva do sujeito destinatário da decisão, uma lesão dos seus interesses, mas se não se lhe pode apontar qualquer antijuridicidade que torne esse sujeito merecedor de tutela específica. Se muito, poderá até concluir-se que o sistema algorítmico decidiu melhor – considerados certos critérios de racionalidade económica, a decisão algorítmica é mais racional (mais otimizada) do que a decisão humana. As consequências dessa otimização poderão consistir num aumento das decisões de recusa em contratar (quando comparado com as decisões humanas), mas daí não se extrai que o processo decisório ou que a decisão padeça de algum enviesamento. Nestes casos, é noutro plano que se situa o debate: a recolha, a natureza e a extensão dos dados utilizados pelo sistema algorítmico.

38. A era dos *big data* – em que vivemos – fornece o ambiente paradigmático para a tomada de decisões algorítmicas otimizadas de acordo com critérios de racionalidade económica<sup>124</sup>. A utilização de dados em larga escala – na grande maioria dos casos, dados pessoais *ab initio* ou desanonimizados<sup>125</sup> –, cuja principal fonte é

---

<sup>124</sup> Sobre a noção de *big data*, remetemos para o que escrevemos em ANA ALVES LEAL, “Aspetos jurídicos da análise de dados na Internet (*big data analytics*) nos setores bancário e financeiro: proteção de dados pessoais e deveres de informação”, cit., pp. 93-96.

<sup>125</sup> As técnicas próprias da *big data* propiciam a re-identificação de dados pessoais que foram anonimizados. E, desta forma, os dados que não eram, à luz do RGPD, «dados pessoais» (mais corretamente: que deixaram de o ser, por aplicação de técnicas de anonimização), voltam, num plano

superveniente, a sê-lo. Claro que o problema da re-identificação e as *nuances* da re-identificabilidade dos dados, que assumem maior expressão nas técnicas da *big data*, têm, no RGPD, um campo fértil de discussão: basta que se considere a noção de «dado pessoal» plasmada no artigo 4.º, n.º 1, do RGPD, em que se autonomiza o elemento da *identificabilidade* da pessoa singular, que prescinde da identificação do sujeito pelo nome (para uma enunciação dos critérios que concretizam este elemento definidor, cfr., por todos, A. BARRETO MENEZES CORDEIRO, “Anotação ao artigo 4.º”, in A. BARRETO MENEZES CORDEIRO (coord.), *Comentário ao Regulamento Geral de Proteção de Dados e à Lei n.º 58/2019*, Coimbra, Almedina, 2021, p. 83).

Como desenvolvemos em “Aspetos jurídicos da análise de dados na Internet”, cit., pp. 119-127, devido à interseção e combinação de largos conjuntos de dados (dados pessoais anonimizados e dados não pessoais) que é característica dos modelos de análise da *big data*, torna-se possível, através de certas correlações, identificar o titular dos dados, mesmo não sabendo o seu nome – assim, cfr. FREDERIK J. ZUIDERVEEN BORGESIJUS, “Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation”, *Computer Law & Security Review*, 32/2, 2016 [256-271]; com critérios para determinar em que casos se verifica a *identificação* da pessoa, com recurso à ideia de *individualidade*, remete-se, pela sua transversalidade, para o estudo de KARL-NIKOLAUS PEIFER, *Individualität im Zivilrecht. Der Schutz persönlicher, gegenständlicher und wettbewerblicher Individualität im Persönlichkeitsrecht, Immaterialgüterrecht und Recht der Unternehmen*, Tübingen, Mohr Siebeck, 2001, *passim*, que apresenta, fora do regime da proteção dos dados pessoais, um quadro geral da proteção dessa individualidade no contexto do ordenamento jurídico alemão. Considera-se hoje assente, à luz das técnicas conhecidas de anonimização, que os dados pessoais não são irreversivelmente anónimos. Concluindo neste sentido, e tendo alcançando grande visibilidade, veja-se o estudo de LUC ROCHER / JULIEN M. HENDRICKX / YVES-ALEXANDRE DE MONTJOYE, “Estimating the success of re-identifications in incomplete datasets using generative models”, *Nature Communications*, 10/3069, 2019, disponível em <https://www.nature.com/articles/s41467-019-10933-3/> [consultado em 20.07.2022]: este estudo versa sobre um método de desanonimização de dados que permite, com base na combinação de quinze critérios demográficos, a re-identificação de 99,98% das pessoas singulares cujos dados são anonimizados e depois incorporados em bases de dados. Pelo que advertem os Autores que “[o]ur results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR and seriously challenge the technical and legal adequacy of the de-identification release-and-forget model”. Esta conclusão, apesar de impressionar, não surpreende. Note-se, a título de exemplo, que um dos Autores deste estudo já havia demonstrado, em estudo anterior, que partindo, de metadados de cartões de crédito, bastariam apenas quatro informações aleatórias sobre os compradores utilizadores do cartão para re-identificar 90% desses compradores (“*We study 3 months of credit card records for 1.1 million people and show that four spatiotemporal points are enough to uniquely reidentify 90% of individuals.*”) – cfr. YVES-ALEXANDRE DE MONTJOYE *et al.*, “Unique in the shopping mall: On the reidentifiability of credit card metadata”, *Science*, 347/6221, 2015 (536-539), p. 536.

Neste cenário, as fragilidades da anonimização de dados pessoais não são ignoradas pelas entidades reguladoras do espaço europeu, mesmo que em estrita observância das exigências do RGPD (sobre estas exigências, são conhecidos – mas porventura já ultrapassados – os critérios apresentados pelo GRUPO DE TRABALHO DO ARTIGO 29.º, na sua *Opinion 05/2014 on Anonymisation Techniques*, de 10-04-2014, disponível *online* em [https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216\\_en.pdf](https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf) [consultado em 20.07.2022], os quais devem orientar a operação de anonimização de dados para que esta seja completa e – tendencialmente – irreversível e, por isso, seja apta a afastar a aplicação do RGPD no tratamento posterior desses dados). Perante as dúvidas do que seja a anonimização relevante para dispensar a aplicação do RGPD ao tratamento

---

de dados, os reguladores europeus têm sido cautelosos em considerar seguros os processos de anonimização de dados. Destacamos, em particular, as preocupações da AEPD – Agencia Española de Protección de Datos e do EDPS – European Data Protection Supervisor, vertidas numa comunicação conjunta intitulada “10 Misunderstandings related to Anonymisation”, de 27-04-2021, disponível *online* em <[https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related\\_en](https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en)> [consultado em 20.07.2022]. O alerta que mais se evidencia nesta comunicação prende-se, como expectável, com a ideia de que a anonimização não é “para sempre”, que não reduz “para zero” a probabilidade de re-identificação e que não é “um conceito binário que não possa ser medido” (pp. 10-11). Em rigor, no RGPD reconhece-se que a irreversibilidade da anonimização é uma propriedade contingencial, porque dependente da evolução tecnológica: dispõe o Considerando 26 que na aferição da impossibilidade de identificação da pessoa singular haverá que considerar “todos os meios suscetíveis de ser razoavelmente utilizados, tais como a seleção, quer pelo responsável pelo tratamento quer por outra pessoa, para identificar direta ou indiretamente a pessoa singular”, e na determinação da probabilidade razoável de os meios serem utilizados para identificar a pessoa singular, “importa considerar todos os fatores objetivos, como os custos e o tempo necessário para a identificação, tendo em conta a tecnologia disponível à data do tratamento dos dados e a evolução tecnológica”.

A suscetibilidade de re-identificação dos dados pessoais tratados através de tecnologia *big data* tem suscitado sérias reservas na comunidade jurídica, seja quanto à não integração do tratamento de dados anonimizados no âmbito de aplicação material do RGPD e legislação complementar, seja quanto às técnicas combinatórias utilizadas. Com mais desenvolvimentos, cfr. MATTHIAS LEISTNER / LUCIE ANTOINE / THOMAS SAGSTETTER, *Big Data. Rahmenbedingungen im europäischen Datenschutz- und Immaterialgüterrecht und übergreifende Reformperspektive*, Tübingen, Mohr Siebeck, 2021, pp. 214-227; BERTRAM RAUM, § 89, in EHMANN / SELMAYR, *Datenschutz-Grundverordnung*, 2. Auflage, München, C.H.Beck, 2018, Rn. 42-45; BENJAMIN SCHÜTZE / STEFANIE HÄNOLD / NIKOLAUS FORGÓ, “Big Data – Eine informationsrechtliche Annäherung”, in Barbara Kolany-Raiser / Reinhard Heil / Reinhard Orwat / Thomas Hoeren (eds.), *Big Data und Gesellschaft*, cit. (233-308), pp. 253 ss.; DANIEL SUPRIYADI, “Personal and non-personal data in the context of big data”, Tilburg Institute for Law, Technology and Society, janeiro de 2017, pp. 30 ss, disponível *online* em <<http://arno.uvt.nl/show.cgi?fid=142300>> [consultado em 20.07.2022]; MANON OOSTVEEN, “Identifiability and the applicability of data protection to big data”, *International Data Privacy Law*, 6/4, 2016 (299-309).

A discussão desloca-se, assim, para a dificuldade em estabelecer uma fronteira entre «dados pessoais» e «dados não pessoais» no contexto da *big data* – cfr. JAN-PETER OHRTMANN / SEBASTIAN SCHWIERING, “Big Data und Datenschutz – Rechtliche Herausforderungen und Lösungsansätze”, NJW, 41/2014 (2984-2990); ARNOLD ROOSENDAL, *Digital Personae and Profiles in Law. Protecting Individuals’ Rights in Online Contexts*, WLP, Oisterwijk, 2013, pp. 31 ss, *passim*, disponível *online* em <<https://ssrn.com/abstract=2313576>> [consultado em 20.06.2022]. DANIEL SUPRIYADI, “Personal and non-personal data in the context of big data”, cit., p. 36, refere-se, a este propósito, a um “*in-between type of data*”. Por seu turno, advertindo para o risco de o próprio responsável pelo tratamento dos *big data* não conseguir discernir se está, ou não, a tratar dados pessoais, ou quando é que estes se tornaram pessoais, cfr. BENJAMIN SCHÜTZE / STEFANIE HÄNOLD / NIKOLAUS FORGÓ, “Big Data – Eine informationsrechtliche Annäherung”, cit., p. 254. Reconhecendo a inevitável “comodização” dos dados, vozes menos conservadoras apontam, num plano de *jure condendo*, no sentido de o RGPD e respetiva regulamentação pelos Estados-membros não oferecer proteção adequada aos titulares dos dados anonimizados, defendendo a proteção desses dados através do regime jurídico dos segredos comerciais (cfr. MARIA BOTTIS / GEORGE BOUCHAGIAR, “Personal Data v. Big Data: Challenges of Commodification of

a Internet<sup>126</sup>, que alimentam modelos de análise preditivos e modelos de análise prescritivos<sup>127</sup>, introduz, apesar dessa pretensão de otimização, desafios específicos.

Por um lado, pela forma disruptiva se como gera informação e conhecimento: o *modus operandi* das tecnologias *big data* é marcadamente indutivo (os dados que alimentam os modelos algorítmicos, ainda que possam ser *muitos* dados, não são *todos* os dados, sendo a sua seleção o resultado de uma escolha humana) e mais focado nas correlações entre eventos do que na sua explicação causal (*correlation over causation*), sendo o conhecimento produzido por esta via inevitavelmente refutável (falsificável)<sup>128/129</sup>. Quer isto dizer, antes de mais, que o sentido da decisão algorítmica assente em dados recolhidos via *big data*, referentes ao sujeito destinatário da decisão (ou à categoria de sujeitos em que este se inclui), dependerá do critério subjacente à seleção do universo de dados tratado (seleção essa que não é neutra) e estará necessariamente condicionado pelas insuficiências quantitativas ou qualitativas desses dados<sup>130</sup>.

---

Personal Data”, *Open Journal of Philosophy*, 8/3, 2018 (206-215), pp. 207 ss.), ou mesmo a criação de um regime jurídico especial de “propriedade” dos dados, como direito conexo (“*Leistungsschutzrechts an Daten*”) dos direitos de propriedade intelectual (cfr. BENJAMIN SCHÜTZE / STEFANIE HÄNOLD / NIKOLAUS FORGÓ, “Big Data – Eine informationsrechtliche Annäherung”, cit., pp. 248-249).

<sup>126</sup> Cf. ANA ALVES LEAL, “Aspectos jurídicos da análise de dados na Internet”, cit., p. 94.

<sup>127</sup> Para uma distinção entre modelos de análise preditivos e modelos de análise prescritivos, Cf. ANA ALVES LEAL, “Aspectos jurídicos da análise de dados na Internet”, cit., pp. 95-96. Como aí se afirma, a análise preditiva é o processo de tratamento de dados mais utilizado pela *big data*, mediado, na maioria das vezes, por decisões automatizadas. A análise preditiva corresponde ao processo de extração e identificação de padrões a partir de conjuntos de dados, de forma a prever resultados ou tendências no futuro; a informação assim obtida será útil ao decisor – que, querendo, se pode adaptar no *presente* àquilo que serão comportamentos, eventos, e preferências no *futuro*. Numa decomposição simplista deste modelo prescritivo, encontramos as seguintes fases: (i) a identificação do problema ou da finalidade a que a análise pretende dar resposta; (ii) a delimitação do universo de dados a analisar, (iii) a identificação de padrões relevantes a partir desses dados, (iv) a associação, numa base probabilística, desses padrões a certos resultados, eventos ou tendências futuras, (v) a utilização desses resultados na tomada de decisões. À análise preditiva cabe, assim, a tarefa de construir e utilizar modelos que permitam fazer previsões com base em dados históricos.

<sup>128</sup> Falsificabilidade, aqui, na aceção de KARL POPPER, *The Logic of Scientific Discovery*, Routledge Classics, 2002 (ed. *e-book* de 1959), pp. 57 ss., 95 ss.

<sup>129</sup> Com desenvolvimentos sobre a produção de conhecimento científico através de *big data*, pelo prisma das inferências abduativas introduzidas por CHARLES SANDERS PIERCE, e discutindo as valências, no conhecimento gerado via *big data*, do argumento da “inferência a favor da melhor explicação” (com base na obra de PETER LIPTON, *Inference to the Best Explanation*, 2<sup>nd</sup> ed., London-New York, Routledge, 2004), *vide* KLAUS WIEGERLING / MICHAEL NERURKAR / CHRISTIAN WADEPHUL, “Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien”, in Barbara Kolany-Raiser / Reinhard Heil / Reinhard Orwat / Thomas Hoeren (eds.), *Big Data und Gesellschaft*, cit., pp. 20 ss.

<sup>130</sup> Criticando o discurso da irrelevância da falta de qualidade dos dados quando, no contexto global, a maioria dos dados sujeitos a tratamento são qualitativa e quantitativamente suficientes, considerando



Por outro lado, porque – num quadro patológico, que não se verifica necessariamente – as consequências dos vieses algorítmicos (desde o enviesamento do algoritmo causado pelo seu criador, bem como o enviesamento que resulte de correlações estatísticas e que é detetado com o “normal” processamento massivo de dados) são amplificadas com a tecnologia *big data*. Reforce-se, como se tem vindo a afirmar, que os vieses não são problemas privativos das decisões algorítmicas baseadas e produzidas com recurso à *big data*: são problemas clássicos de qualquer processo decisório, humano ou algorítmico. Sucede que, em ambiente *big data*, estas consequências são, em proporção, aumentadas, seja considerada uma fase inicial do circuito decisório (por exemplo, sendo tratados dados de forma massiva, detetar-se-ão mais correlações estatísticas enviesadas), seja na fase final desse circuito (por exemplo, sendo frequente o recurso a decisões automatizadas aplicáveis, em massa, a vários destinatários). Aqui encontra arrimo a ideia de um efeito multiplicador das decisões algorítmicas. A facilidade de reprodução e replicação, com consistência, de decisões algorítmicas atribui-lhes uma natural incidência coletiva, *com* ou *sem* atendimento das variáveis individualizantes dos sujeitos delas destinatários.

Ainda que seja de uso apontar às decisões tomadas em contexto *big data* um problema metodológico de generalização por partirem da categorização dos seus destinatários, há que desfazer o mito segundo o qual as decisões algorítmicas padecem de uma natural cegueira, por estarem, em maior ou menor medida, alheadas das características individuais do destinatário da decisão. Operando a *big data* através do tratamento de grandes volumes de dados, quando estes se referam a um sujeito, e pese embora a falsificabilidade das teorias ou conclusões alcançadas por esta via, possibilita-se – como dificilmente se verá num processo decisório estritamente humano – um conhecimento “microscópico” daquele sujeito. Tão microscópico que pode incidir sobre factos ou elementos do sujeito que ele próprio desconhece, ou para os quais não estava desperto. O tratamento de dados massivos favorece a personalização (*tailorização*) da oferta de produtos ou serviços ao sujeito, seja com publicidade dirigida (pense-se nos anúncios publicitários em barra lateral na navegação *online* em motores de busca), seja na adaptação de produtos ou serviços ao sujeito (pense-se na adaptação de um serviço financeiro às preferências do cliente, como a gestão de carteiras personalizada, ou na adaptação de um produto segurador às

---

que em certas áreas (como na medicina), essa falta de qualidade nunca é “negligenciável” (“*Vernachlässigbarkeit der Datenqualität*”), cfr. KLAUS WIEGERLING / MICHAEL NERURKAR / CHRISTIAN WADEPHUL, “Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien”, in BARBARA KOLANY-RAISER / REINHARD HEIL / REINHARD ORWAT / THOMAS HOEREN (eds.), *Big Data und Gesellschaft*, cit., p. 13.

características do segurado que influem na avaliação do risco). Esta personalização é gradativa, podendo ser mais ou menos intensa<sup>131</sup>: poderá consistir na integração do sujeito numa categoria ou grupo de sujeitos que partilhem características ou preferências comuns, poderá consistir na integração do sujeito em subcategorias ou subgrupos correspondentes a perfis mais individualizados, como pode revelar-se pela atribuição de um perfil único e exclusivo ao sujeito (hiperpersonalização).

39. Em suma: (i) pode falar-se de um efeito discriminatório das decisões algorítmicas *cum grano salis*, uma vez que as decisões humanas são tão ou mais discriminatórias do que as decisões algorítmicas, pese embora as dificuldades naturais à deteção dessa discriminação no processo decisório humano (e, em contraste, a facilidade de revelação desses enviesamentos no processo decisório algorítmico); (ii) o efeito lesivo das decisões algorítmicas (em particular, da perda de oportunidades negociais ou da exclusão de certos contextos contratuais) apenas se distancia do efeito lesivo das decisões humanas do ponto de vista da grandeza ou extensão subjetiva desses efeitos; (iii) o efeito multiplicador das decisões algorítmicas é catalisado em ambiente *big data*, com a análise de dados em larga escala e com a produção, de forma massificada, de decisões algorítmicas, com ou sem consideração das variáveis individualizantes dos sujeitos delas destinatários.

Alcançadas estas conclusões intercalares, retomamos a ideia apresentada na abertura deste capítulo: as decisões algorítmicas são tidas como *mais perigosas* do que as decisões humanas porque, podendo ser uma projeção e um produto de decisões humanas enviesadas, têm a aptidão de amplificar os efeitos lesivos dessas decisões junto de um universo mais alargado de sujeitos.

### 3.4. Boas decisões algorítmicas, *alternative data* e a lógica de perigo abstrato

40. Os casos em que a decisão algorítmica apresenta um resultado *bom* para o seu utilizador (à luz da respetiva racionalidade económica e da vantagem por elas gerada na esfera dos interesses do seu utilizador), e na medida em que a decisão e o processo decisório não padeçam de qualquer enviesamento que contrarie normas ou princípios jurídicos, parecem não ser captados pelo critério de perigosidade mobilizado para justificar a regulação jurídica dos algoritmos.

Indaga-se: nos casos em que a decisão algorítmica, num juízo de prognose, não constitui uma “má” decisão, o que poderá justificar as limitações legais dirigidas

---

<sup>131</sup> Cf. ANA ALVES LEAL, “Aspetos jurídicos da análise de dados na Internet”, cit., p. 99.

à formação e utilização destas decisões algorítmicas, com a conseqüente limitação da autonomia privada do seu utilizador?

41. As boas decisões algorítmicas – não só de um prisma de racionalidade económica, mas também do prisma de combate à discriminação embutida em determinados processos decisórios humanos – têm sido especialmente escrutinadas. O alcance dessas boas decisões explica-se essencialmente pela utilização e consideração de dados que tradicionalmente não são utilizados nos processos decisórios humanos. Centremo-nos nas decisões algorítmicas que são boas para o seu destinatário porque lhe permitem aceder a oportunidades negociais que, com grande probabilidade, através de um processo decisório humano, lhe estariam vedadas. Estão geralmente em causa destinatários integrados em grupos ou categorias desfavorecidas em certos contextos negociais, como seja o acesso ao emprego ou o acesso ao crédito bancário ou a boas condições contratuais (por exemplo, imigrantes, pessoas de etnia cigana, ou mulheres em idade reprodutiva). Nestes casos, as decisões algorítmicas concederão um “*disparate benefit*” a estes sujeitos<sup>132</sup>.

42. Fala-se, a este propósito, do uso de “*alternative data*” (dados alternativos) para tomar decisões tidas como justas para os seus destinatários. O decisor considera um conjunto de dados, não considerados nos processos decisórios clássicos, que permitem, com base em certos modelos preditivos também alternativos, inverter inferências que ditariam a exclusão negocial ou piores condições contratuais para o destinatário. Entre esses dados alternativos, contam-se as redes sociais de que este faz parte, bem como as interações nessas plataformas, a pegada digital em motores de busca, o histórico de utilização do cartão de crédito, ou mesmo os dados de geolocalização do destinatário da decisão. Pressuposto, logicamente, o consentimento do destinatário, titular destes dados alternativos, quase sempre dados pessoais. A inversão das inferências que ditariam a exclusão negocial ou piores condições contratuais para o destinatário dá-se pelo incremento de confiança do decisor naquele destinatário, suportada pelo modelo preditivo alimentado com os dados alternativos.

Retomando o contexto da concessão de crédito bancário, apela-se à importância do círculo social do potencial mutuário para o incremento da confiança da instituição de crédito: tornaram-se populares avisos como “*your deadbeat Facebook friends could*

---

<sup>132</sup> Cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, cit., p. 160.

*cost you a loan*”, “*your Facebook friends could be the ticket to your next loan*”<sup>133</sup>, ou mesmo “*lot of contacts in your mobile phone may get you loans*” (uma vez que “people are more likely to be better borrowers if they have friends who pay back their loans on time”)<sup>134</sup>.

Os dados alternativos poderão ser também uma forma de corrigir resultados algorítmicos procedimentalmente justos, mas materialmente injustos. Dos espaço estado-unidense chegam-nos exemplos de correções do *credit scoring* de pessoas vítimas de “*coerced debt*” – pessoas que geralmente tinham uma boa classificação na avaliação de risco de crédito, mas cujo cartão de crédito passou a ser utilizado apenas pelo respetivo cônjuge, de quem são vítimas de uma relação abusiva, com registos repetidos de ultrapassagem do limite mensal de crédito e de incumprimento (ainda que nem todos os casos de “*coerced debt*” se reduzam a hipóteses de fraude, apropriação indevida da identidade, coação física ou coação moral)<sup>135</sup>.

Os problemas de violação da privacidade concitados pelo uso de dados alternativos esbarram com a licitude da sua utilização, fundada no consentimento prestado pelo titular dos dados. Os incentivos à prestação desse consentimento decorrem da pretensão, pelo seu emitente, de alcançar uma oportunidade negocial, ou uma dada oportunidade negocial. Se olhado da perspectiva do direito civil, para lá da regulação específica do RGPD, alguns remédios são ponderáveis para este aproveitamento da tendencial situação de inferioridade em que o titular dos dados se encontrará no momento da prestação do consentimento. O regime da usura (artigos 282.º a 284.º do Código Civil), construído a partir da exploração ou do aproveitamento censurável de uma situação de inferioridade ou de necessidade, encabeça os regimes candidatos à resolução de um possível conflito entre o titular dos dados e o seu utilizador-decisor. Sublinhe-se que a disponibilização destes dados alternativos ao decisor constitui uma vantagem própria para este (o que permitirá, em muitos casos, o preenchimento do requisito da “promessa ou concessão de benefícios excessivos ou injustificados” da usura – artigo 282.º, n.º 1, *in fine*, do Código Civil), que extravasa a vantagem decorrente da utilidade da relação negocial encetada ou mantida com o titular dos dados em virtude da sua consideração no processo decisório algorítmico. Do que

---

<sup>133</sup> Cf. NIZAN GESLEVICH PACKIN / YAFIT LEV-ARETZ, “On Social Credit and the Right to be Unnetworked”, *Columbia Business Law Review*, 2, 2016 (339-425), p. 375.

<sup>134</sup> Cf. SANGWON YOON, “Lot of Contacts in Your Mobile Phone May Get You Loans”, in *Bloomberg* (12 de novembro de 2015), disponível em <<http://www.bloomberg.com/news/articles/2015-11-15/lot-of-contacts-in-your-mobile-phone-you-may-qualify-for-a-loan>> [consultado em 20.07.2022].

<sup>135</sup> Cf. ANTHONY GREEN, “Podcast: Can AI fix your credit? Ripple effects of automation in credit scoring extend beyond finances”, in *MIT Technology Review* (12 de maio de 2022), disponível em <<https://www.technologyreview.com/2021/05/12/1024837/podcast-can-ai-fix-your-credit-report-score/>> [consultado em 20.07.2022].

se trata é do *valor* dos dados alternativos obtidos com o consentimento do seu titular, geralmente através de técnicas da *big data*, em caso de comercialização dos mesmos – ainda que esses dados sejam, à partida, sujeitos a um processo de anonimização, por a sua comercialização não se encontrar, por regra, inscrita entre as finalidades do tratamento dos dados consentidas pelo seu titular.

O relato dos contextos de tendencial inferioridade dos sujeitos que consentem o tratamento de dados alternativos para obter certas oportunidades negociais tem feito soar alarmes sobre os perigos de *datificação* do ser humano (o ser humano não datificado de acordo com estes critérios poderá tornar-se inegável para a celebração de determinados contratos) e de *comodização da confiança* gerada entre as partes e que funda a celebração de um negócio entre estas – agora aferida não em função de concretos comportamentos das partes, mas através do tratamento algorítmico de certos tipos de dados pessoais<sup>136</sup>.

As exigências e espartilhos dos regimes clássicos de direito civil para dar resposta a eventuais aproveitamentos do decisor responsável pelo tratamento destes dados alternativos, a par da legitimação destas hipóteses no RGPD através da prestação do consentimento livre e esclarecido, explica algumas abordagens mais musculadas de oposição ao aproveitamento destes dados. Entre elas, conta-se a defesa do “direito a estar socialmente desconectado” (“*right to be unnetworked*”)<sup>137</sup>, bem como o regime proposto na Proposta de Regulamento sobre Inteligência Artificial. Neste, o uso algorítmico de dados alternativos deve, pelo menos, ser escrutinado no âmbito da fiscalização de sistemas de inteligência artificial de risco elevado – a qual deverá obedecer aos parâmetros acima explanados (§ 2.º). No quadro de relações jurídico-privadas, o regime não parece consentir remédio mais forte do que este: o tratamento dos dados alternativos nos moldes descritos apenas será proibido se realizado por autoridades públicas (artigo 5.º, n.º 1, alínea c), da Proposta de Regulamento) – o que não é, nos exemplos trabalhados, o caso.

43. O excursus agora terminado sobre o tratamento de dados alternativos não teve o propósito de insinuar uma estreita dependência entre o alcance de boas decisões algorítmicas, sobretudo do prisma do seu destinatário, e o

---

<sup>136</sup> Assim, apontando que a confiança é uma questão de perceção e predição, e que nestes casos a confiança é gerada sem o desempenho de um comportamento confiável, cf. STEFEN KRÜGER / CHRISTOPHER WILSON, “The problem with trust: on the discursive commodification of trust in AI”, in *AI & Society*, 2022, disponível em < <https://link.springer.com/content/pdf/10.1007/s00146-022-01401-6.pdf> > [consultado em 20.07.2022].

<sup>137</sup> Cf. NIZAN GESLEVICH PACKIN / YAFIT LEV-ARETZ, “On Social Credit and the Right to be Unnetworked”, cit., p. 416.

consentimento deste para o tratamento dos seus dados alternativos. Também não teve o propósito de debater possíveis estados de acrasia do destinatário da decisão, nem intervenções paternalistas do legislador neste particular, que vedem qualquer operação de tratamento dos dados alternativos, ainda que com previsíveis vantagens para o destinatário. Teve o modesto propósito de sinalizar que também nas “boas” decisões algorítmicas se poderá reconhecer, por outras razões, *a latere*, algum grau de perigosidade.

Quando se questiona se as limitações legais dirigidas à formação e utilização destas decisões algorítmicas, com a consequente restrição da liberdade do decisor, deverão aplicar-se a decisões algorítmicas tidas como boas ou vantajosas para ambas as partes (sobretudo, para o seu destinatário), sugere-se implicitamente que quanto a estas não vale o critério, atrás introduzido, de perigosidade de amplificação de más decisões humanas. Todavia, para além dos casos relatados de boas decisões algorítmicas perigosas, entendemos que essas limitações legais são também aplicáveis a estas decisões, em claro apelo à lógica de funcionamento do perigo abstrato: o único fator decisivo para a aplicação dessas limitações é a possibilidade, em abstrato, e aferida *ex ante*, de a decisão algorítmica reproduzir, pelas razões atrás analisadas, uma qualquer decisão humana enviesada, com projeção, provavelmente lesiva, num espetro alargado de sujeitos destinatários da mesma. Assim não será, obviamente, quando disposições específicas circunscrevam certas limitações apenas a decisões ou contextos decisórios que apresentem uma perigosidade mais intensa (basta pensar no escalonamento dos modelos de regulação que se encontram na Proposta de Regulamento sobre Inteligência Artificial). À semelhança do que alguma doutrina defende em contextos marcados pela fraqueza negocial de uma das partes (como os consumidores financeiros), onde também se aponta a presença de uma lógica de proteção do perigo abstrato, o objetivo da regulação não é o de proteger os destinatários da decisão “*from unfair harms, but rather from unfair odds*”<sup>138</sup>.

## § 4.º Explicação da declaração negocial

### 4.1. Decisão algorítmica como declaração negocial

44. Fez-se equivaler, até este ponto, e sem problematizar, a decisão algorítmica “em contexto negocial”. É amplamente aceite na doutrina, independentemente das teorias

---

<sup>138</sup> Cf. JOANNA BENJAMIN, “The Narratives of Financial Law,” in *Oxford Journal of Legal Studies*, 30, 2010 (787-814), p. 800. Como sustenta a Autora, a pré-proteção dos bens jurídicos a preservar é realizada pela regulação do procedimento, e não pela regulação do resultado.

que se acolham para justificar a eficácia da declaração negocial<sup>139</sup>, que também as decisões algorítmicas poderão ser qualificadas como declarações negociais. É, aliás, sintomática desta aceitação no espaço germânico a utilização do termo “*Computererklärungen*”, em vez de “*Willenserklärungen*”, para designar este tipo de declaração negocial<sup>140</sup>.

A declaração negocial corresponde a um comportamento humano voluntário, simples ou complexo, destinado, pelo seu autor, a produzir efeitos jurídico-privados de acordo com o seu sentido. A emissão de uma declaração negocial é sempre o produto de uma *decisão* (algorítmica, ou não) do declarante – dirija-se, ou não, essa declaração à formação de um negócio jurídico (nem todas as declarações negociais são propostas ou aceitações e nem todas as declarações negociais se destinam a integrar contratos<sup>141</sup>). Assim, no contexto que nos ocupa, é num exercício de metonímia que falamos de decisões algorítmicas como decisões negociais. Não se ignora, porém, que a formação de declarações negociais através de processos decisórios baseados em algoritmos vem desafiar noções e esquemas de decisão estabilizados pela dogmática do direito do negócio jurídico.

Não cabe, aqui, ponderar transversalmente esses desafios.

Também não nos interessam as discussões já conhecidas em torno da utilização de máquinas (por exemplo, as máquinas automáticas de vendas) em várias fases da relação negocial. O contexto em que nos situamos difere dessa utilização: o uso de inteligência artificial em ambiente negocial refere-se à utilização de sistemas algorítmicos ou agentes de *software* capazes de aprendizagem automática, tecnologias que mimetizam as capacidades cognitivas e de raciocínio humano – o que leva a que o sistema possa, desenvolvendo-se a si próprio, escolher diversas opções e mesmo alterar e adaptar o seu próprio funcionamento de acordo com factos pretéritos nele ou por ele registados (aprendizagem automática)<sup>142</sup>. O principal

---

<sup>139</sup> Vide, por todos, ambos com amplas referências neste sentido, SEBASTIAN HERRLER, “Vorbem zu §§ 116 ff”, in *J. von Staudingers Kommentar zum Bürgerlichen Gesetzbuch: Staudinger BGB – Buch I: Allgemeiner Teil*, Otto Schmidt-De Gruyter, 2021, Rn. 57; LISA FINKE, *Vertragsschluss und Willenserklärung durch automatisierte und autonome Systeme*, GRIN, 2020; ANDREAS RUSTER, “Willenserklärungen mittels autonomer Systeme? Zum Vertragsschluss bei Einsatz künstlicher Intelligenz (Teil 2)”, *Juristische Rundschau*, 12, 2020 (639-649).

<sup>140</sup> Cf. THOMAS KRÜGER / MICHAEL BÜTTER, “Elektronische Willenserklärungen im Bankgeschäftsverkehr: Risiken des Online-Banking – zugleich Besprechung des Urteils des Landgerichts Nürnberg-Fürth vom 19. 5. 1999”, *Wertpapiermitteilungen*, 2001, pp. 221 ss.

<sup>141</sup> Como afirma CARLOS FERREIRA DE ALMEIDA, *Texto e Enunciado na Teoria do Negócio Jurídico*, Vol. I, Coimbra, Almedina, 1992, p. 313, “a declaração negocial é a *declaração* de um negócio jurídico, mas é essencialmente *declaração para um* negócio jurídico, podendo conceber-se, existir e existir validamente, sem que o negócio jurídico em vista do qual foi emitida se venha a concluir”.

<sup>142</sup> Assim, PAULO MOTA PINTO, ““Agentes de software inteligentes” e negócio jurídico – alguns problemas”, *Julgar*, 45, 2021 (35-64), p. 37.

critério distintivo entre esta realidade e a (já vetusta) contratação com máquinas corresponde, por isso, à *imprevisibilidade controlada*<sup>143</sup> dos resultados (para o que nos interessa, das decisões) produzidos pelo sistema algorítmico. Também não nos ocupam as hipóteses de contratação máquina-a-máquina (M2M) sem intervenção humana<sup>144</sup>, hoje cada vez mais frequentes – ainda que sejam hipóteses abordadas para ponderar e concretizar casos de desnecessidade da explicação da decisão algorítmica. Fora de crise, porque (por enquanto) estabilizada, está também a imputabilidade da decisão algorítmica a uma pessoa humana<sup>145</sup>, geralmente o seu utilizador – sendo que, em qualquer caso, a esfera de imputação do comportamento negocial que constitui a declaração algorítmica pode ser regulada negocialmente, tal como pode ser regulada negocialmente – entre criador, utilizador ou outro sujeito participante no processo decisório algorítmico – a repartição de riscos relativos a erros do sistema ou derivados da sua utilização.

45. Interessa-nos, para os propósitos deste estudo, esclarecer tão-só que uma decisão algorítmica *pode ser* uma declaração negocial e que aqui tomaremos como modelo a decisão algorítmica *como* declaração negocial. Mesmo que essa decisão corresponda às realidades intermédias situadas entre os polos das decisões humanas (intervenção humana no seu grau máximo) e das decisões automatizadas (intervenção humana no seu grau mínimo). Crucial é que a mesma seja, por alguém ou de alguma forma exteriorizada, seja em observância de um escopo notificativo próprio, seja através de aparências objetivas, exteriores, de uma declaração<sup>146</sup>. Todavia, as referências às vicissitudes que podem atingir o processo decisório algorítmico não encontram tratamento óbvio no direito do negócio jurídico dedicado à formação da declaração negocial.

Cabe ainda acrescentar que não prejudica a qualificação como «declaração negocial» o facto de as decisões algorítmicas se produzirem no referido espaço de

---

<sup>143</sup> Como afirmando anteriormente, mesmo que totalmente automatizado, e mesmo que gerado automaticamente através de sistemas de *machine learning*, o algoritmo que baseia a decisão algorítmica não é alheio à intervenção humana. Os resultados obtidos em sistemas de *machine learning* são sempre o resultado de uma parametrização humana, de primeiro ou de segundo grau – daí se falar, em óbvio oxímoro, de uma *imprevisibilidade controlada* dos resultados algorítmicos.

<sup>144</sup> Sobre estas hipóteses, cf. PAULO MOTA PINTO, “Agentes de software inteligentes” e negócio jurídico – alguns problemas”, cit., p. 40.

<sup>145</sup> Ainda que com ponderação do recurso, por analogia, a esquemas de interposição pessoal na formação e transmissão da declaração negocial (paradigmáticamente, representante e nuncio) – cf. LISA FINKE, *Vertragsschluss und Willenserklärung durch automatisierte und autonome Systeme*, cit., 5.2.2.-5.2.3.

<sup>146</sup> Cf. PAULO MOTA PINTO, *Declaração tácita e comportamento concludente no negócio jurídico*, Coimbra, Almedina, 1995, p. 557.



imprevisibilidade controlada do algoritmo – que alguns autores, com vista a salvaguardar a sua eficácia e nexos de imputação, aproximam, fazendo o devido paralelismo, das declarações em branco (“*Blanketterklärungen*”) acompanhadas de um pacto de preenchimento, que serão completadas por terceiros<sup>147</sup>. O paralelismo tem a virtude de acentuar dois aspetos: por um lado, que esta imprevisibilidade não é suficiente para considerar a decisão algorítmica ferida de nulidade, com fundamento na alegada indeterminabilidade do seu conteúdo; por outro, que a pessoa a quem é imputada a decisão algorítmica não tem o poder de opor a terceiros quaisquer desvios que se venham a verificar entre o *design* do algoritmo e os resultados esperados e a decisão concretamente produzida pelo sistema algorítmico.

#### 4.2. Opacidade do percurso decisório e do discurso decisório – por um critério de unificação das decisões humanas e das decisões algorítmicas

46. É frequente afirmar-se que as decisões algorítmicas padecem de um maior grau de opacidade – sendo, por isso, mais resistentes à sua explicação – do que as decisões humanas<sup>148</sup>. Assim é, alega-se, porque os sistemas algorítmicos, sobretudo aqueles que se desenvolvem através de técnicas de *machine learning*, envolvem processos que não são explicáveis em linguagem humana<sup>149</sup>, envolvem o tratamento e combinação de um extenso universo de dados, de diferentes naturezas, e de fórmulas complexas, bem como uma lógica decisória imprevisível<sup>150</sup>.

---

<sup>147</sup> Cf. SEBASTIAN HERRLER, “Vorbem zu §§ 116 ff”, cit., Rn. 57; JÖRG NEUNER, *Allgemeiner Teil des Bürgerlichen Rechts*, 12. Auflage, München, C.H. Beck, 2020, § 32, Rn. 40; DIETER MEDICUS / JENS PETERSEN, *Allgemeiner Teil des BGB*, 11. Auflage, C.F. Müller, 2016, § 56, Rn. 913.

<sup>148</sup> Cf. AZIZ Z. HUQ, “A Right to a Human Decision”, cit., p. 640. O Autor descarta a ideia, muito difundida, de que as decisões algorítmicas não são transparentes por falta de vontade e de colaboração do criador e do utilizador do sistema algorítmico na divulgação dos detalhes desse sistema, com receio de sofrer danos com a divulgação de informação ou de obra que pretende manter confidencial. Esta pretensão de confidencialidade, como bem sustenta, não é critério distintivo nem critério explicativo da opacidade decisória referida. Tirando o investimento que pode representar a aquisição ou utilização do sistema algorítmico, o decisor humano tem, em abstrato, os mesmos incentivos em manter a confidencialidade do seu processo decisório (é velho – e pré-algorítmico – o axioma de acordo com o qual “o segredo é a alma do negócio”). Em rigor, nos regimes que protegem o segredo comercial e a propriedade intelectual, não há diferenças relevantes na proteção conferida ao sistema algorítmico ou ao conhecimento humano não vertido em linguagem computacional.

<sup>149</sup> Cf. TAL Z. ZARSKY, “Transparent Predictions”, *University of Illinois Law Review*, 4, 2013 (1503-1570), p. 1519.

<sup>150</sup> Cf. BRENT DANIEL MITTELSTADT *et al.*, “The ethics of algorithms: Mapping the debate”, *Big Data & Society*, 3/2, 2016, disponível em <<https://journals.sagepub.com/doi/epdf/10.1177/2053951716679679>> [consultado a 20.07.2022].

47. Iniciou-se, atrás, um processo de desconstrução desta ideia – deste preconceito algorítmico –, de acordo com a qual existe um “fosso de transparência” no percurso e no discurso decisório entre decisões humanas e decisões algorítmicas. Mesmo que não se conteste que a opacidade de um sistema algorítmico seja fruto da sua complexidade técnica e da sua imprevisibilidade operacional e das dificuldades em traduzir a linguagem computacional para linguagem natural, consideramos serem desprovidos de fundamento os entendimentos que defendem, numa comparação com as decisões humanas, um maior grau de impenetrabilidade das decisões algorítmicas – pelo menos, quando apresentados em moldes generalizáveis<sup>151</sup>.

Concluiu-se já que a opacidade é uma propriedade tendencial e não exclusiva das decisões algorítmicas e que o processo decisório humano é naturalmente mais opaco – seja externamente, seja frequentemente para o próprio decisor – do que o processo decisório algorítmico.

Neste sentido, depõem, à cabeça, duas razões.

A primeira é a de que o processo decisório humano, ainda que, em abstrato, se possa decompor analiticamente num curso encadeado de fases à luz das clássicas teorias da decisão<sup>152</sup>, é marcado pela influência de um conjunto de fatores cujo peso na formação de uma decisão é dificilmente aferível e dificilmente explicável: as decisões humanas *também* são tomadas com influência das emoções, da intuição, de preconceitos, de erros de raciocínio e défices de racionalidade, de cansaço, de *stress*, de incentivos evidentes, de *nudges*, ou de outras formas de motivação<sup>153</sup>. Acresce que qualquer decisão é sempre tomada em contexto de incerteza<sup>154</sup>. Quando as pessoas tomam decisões sob incerteza, entram em jogo, muitas vezes de forma inconsciente, métodos heurísticos que facilitam o processo decisório, pese embora também o tornem mais permeável a vieses cognitivos. Esses métodos baseiam-se em padrões de pensamento assentes em dados e experiências prévios (que podem, não obstante, referir-se a experiências “erradas”, assentes em perceções ou correlações distorcidas), ou mesmo comprovados pelas regras da experiência, funcionando esses

---

<sup>151</sup> Cf. AZIZ Z. HUQ, “A Right to a Human Decision”, cit., pp. 643 ss.

<sup>152</sup> Cf. DAVID NINK, *Justiz und Algorithmen*, cit., pp. 29 ss.

<sup>153</sup> Cf. DAVID NINK, *Justiz und Algorithmen*, cit., p. 28.

<sup>154</sup> Cf. AMOS TVERSKY / DANIEL KAHNEMAN, “Judgment under Uncertainty: Heuristics and Biases Biases in judgments reveal some heuristics of thinking under uncertainty”, *Science*, 185, 1974 (1124-1131). Entre nós, o tema é desenvolvido por MARGARIDA LIMA REGO, “Decisões em ambiente de incerteza: probabilidade e convicção na formação das decisões judiciais”, cit. Segundo a Autora, “quando se está na posse de toda a informação necessária, se é que isso alguma vez sucede, a decisão não representa um desafio. «A teoria da decisão é a teoria de decidir o que fazer quando não se sabe o que irá acontecer.»” (p. 126).

métodos como atalhos automáticos dos processos cognitivos e volitivos humanos<sup>155</sup>.

A segunda razão para a opacidade natural das decisões humanas prende-se com a tendencial falta de pressão normativa, no quadro do direito civil, no sentido da exposição e no sentido da sindicabilidade do percurso decisório humano<sup>156</sup>. Só assim não é em domínios muito circunscritos, marcados pela necessidade de tutela dos interesses de uma parte tida como negocialmente mais fraca:

Novamente no contexto da concessão de crédito bancário: na avaliação da solvabilidade do consumidor bancário, aos mutuantes é exigido, a título prévio, que elaborem e implementem um documento interno que descreva o método de avaliação da solvabilidade, definindo os elementos informativos e os procedimentos em que a mesma deve fundar-se (artigo 16.º, n.º 6, do Decreto-Lei n.º 74-A/2017, de 23 de junho de 2017). Esses elementos encontram-se elencados no artigo 5.º do Aviso do Banco de Portugal n.º 4/2017. Se o pedido de crédito for rejeitado com fundamento nos dados constantes das bases de dados de responsabilidades de créditos, ou da lista pública de execuções, ou de outras bases de dados consideradas úteis para a avaliação da solvabilidade dos consumidores, o mutuante deve informar o consumidor imediata, gratuita e justificadamente desse facto e dos elementos constantes da respetiva base de dados, salvo se a prestação destas informações for proibida por lei ou for contrária a objetivos de ordem pública ou de segurança pública (artigo 16.º, n.º 3, do Decreto-Lei n.º 74-A/2017, de 23 de junho de 2017).

É patente o cariz restritivo que subjaz à cominação deste dever de apresentar uma explicação (e não necessariamente uma justificação) para a decisão de recusa de concessão de crédito: deste regime não resulta a cominação desse dever a propósito de qualquer decisão de recusa, fundada na falta ou no preenchimento insuficiente dos elementos em que se funda a avaliação da solvabilidade. O dever apenas é cominado quando o pedido é rejeitado com fundamento em elementos consultados em bases de dados (bases de dados de responsabilidades de créditos, ou da lista pública de

---

<sup>155</sup> Cf. DAVID NINK, *Justiz und Algorithmen*, cit., pp. 35 ss.

<sup>156</sup> Esta ausência de uma imposição na gama de fontes foi apontada em “Aspetos jurídicos da análise de dados na Internet (*big data analytics*) nos setores bancário e financeiro: proteção de dados pessoais e deveres de informação”, cit., p. 202. Em particular, deve concretizar-se a falta de uma pressão normativa, no direito civil, para que o sujeito siga um determinado processo decisório, ou para que tome uma decisão com um certo conteúdo. Não se trata de indiferença do ordenamento, mas antes uma reação de fundo, ancorada no princípio da autodeterminação, que rejeita a ideia de um controlo do processo de tomada de decisão pelos sujeitos de direito privado. Como regra geral, estes podem decidir segundo o processo decisório que quiserem e no sentido que bem entenderem, porque têm, à partida, o poder de dispor sobre o conteúdo da sua própria esfera jurídica e conformar a sua atividade em função daquilo que sejam as suas preferências subjetivas (ainda que não racionalmente justificadas à luz da teoria da decisão).

execuções, ou de outras bases de dados consideradas úteis para a avaliação da solvabilidade dos consumidores).

Esta falta de pressão normativa, acompanhada das dificuldades naturais em decompor mentalmente, de forma criteriosa, um processo decisório humano, leva a que o decisor humano não esteja treinado a problematizar e a desconstruir analiticamente o *iter* da sua decisão e conduz a dificuldades óbvias de deteção e de prova de vieses cognitivos do decisor humano que distorçam esse processo<sup>157</sup>.

48. Por isto se afirmou anteriormente (3.3.) que as decisões humanas são potencialmente mais discriminatórias do que as decisões algorítmicas e que a falta de evidências nesse sentido explica-se intuitivamente: essa discriminação não é revelada pelas dificuldades, ou mesmo impossibilidade, de acesso à fórmula mental utilizada pelo decisor humano, diversamente do que sucede com as decisões algorítmicas, em que – com mais ou menos dificuldades de compreensão para o intérprete, e com mais ou menos onerosidade para o sujeito a quem incumba uma explicação – a fórmula algorítmica utilizada, comportando e registado a matriz dessa discriminação, exporá os elementos discriminatórios que são relevados, facilitando a prova dessa discriminação.

J. Kleinberg, J. Ludwig, S. Mullainathany e Cass R. Sunstein apresentam uma constelação de testes para justificar esta conclusão<sup>158</sup>.

(1) Uma empresa pretende contratar vendedores. Para tal, elege o seguinte critério de contratação: favorecer a contratação de funcionários que demonstrem maior aptidão para maximizar as vendas. Com isto, acaba por dar preferência a candidatos caucasianos por um motivo: os clientes da empresa são mais propensos a fazer compras a vendedores caucasianos.

(2) Uma empresa pretende contratar gestores. Em igualdade de circunstâncias, a empresa procura gestores que se mantenham por mais tempo na empresa. De acordo com a experiência da empresa, as mulheres apresentam, face aos homens, uma maior probabilidade de cessar a sua relação laboral após um curto período (menos de cinco anos). Por isso, a empresa dá preferência a candidatos do sexo masculino.

(3) Uma entidade governamental pretende contratar orçamentistas para uma posição de júnior. Dá preferência a candidatos que tenham estudado nas faculdades e universidades mais prestigiadas do país. Esta preferência tem um efeito adverso desproporcional em candidatos afro-americanos.

---

<sup>157</sup> Cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, cit., p. 114.

<sup>158</sup> Cf. JON KLEINBERG / JENS LUDWIG / SENDHIL MULLAINATHANY / CASS R. SUNSTEIN, “Discrimination in the Age of Algorithms”, cit., pp. 148 ss.

(4) Uma empresa pretende contratar seguranças. Com base na sua experiência, considera que os seguranças com mais de trinta anos de experiência são menos eficientes e cuidadosos na realização das suas funções. Por conseguinte, dá preferência aos candidatos que tenham tido menos de trinta anos de experiência. O resultado é um impacto desfavorável na contratação de pessoas mais velhas.

Em todos os quatro casos, a análise jurídica é razoavelmente intuitiva: estes casos apenas terão relevância jurídica na medida em que, no sistema de fontes, se encontre norma ou princípio que vede algum destes tipos de procedimentos decisórios ou resultados. A análise jurídica reduz-se a uma ideia simples (ainda que a uma tarefa complexa, que é a de encontrar a fonte da proibição da discriminação), mas o litígio pode não o ser. Admitamos que das fontes se extrai ato de conteúdo paramétrico que veda processos decisórios que adotem critérios discriminatórios.

Estando em causa decisões humanas, a prova pode ser difícil ou impossível: como se prova que a empresa discriminou os sujeitos excluídos da contratação, quando os critérios que subjazem às decisões de contratar – mesmo que revelados abertamente – são neutros e, de uma perspectiva económica, absolutamente racionais?

Estando em causa decisões algorítmicas, os resultados seriam, à partida, os mesmos. Mas o grau de dificuldade da prova da utilização de critérios discriminatórios estaria significativamente mitigado. Visto à lupa, o algoritmo revelaria não só os critérios que subjazem às decisões de contratar (neutros e absolutamente racionais), mas também os subcritérios – esses sim, discriminatórios (em função do género, raça e idade) e que estavam implícitos na decisão humana anterior – embutidos no algoritmo pelo seu criador ou através de processos de *machine learning*. A auditabilidade dos algoritmos consente, portanto, com maior ou menor grau de dificuldade de compreensão do intérprete-aplicador, a revelação de subcritérios que a mente humana, à partida, encobriria.

49. Perante isto, fica em crise a coerência valorativa dos entendimentos que sustentam, com veemência, um dever de explicação da decisão algorítmica sem, de igual forma, estender – ou reconhecer – essa exigência às decisões humanas. *A fortiori*, pela opacidade natural das decisões humanas, menos debelável do que a opacidade das decisões algorítmicas, a defesa dessa explicação até encontraria justificação mais enérgica no domínio dos processos decisórios humanos. Esta ideia não é derrubada pelo entendimento de que as opacidades do processo decisório humano são atenuadas através de mecanismos menos complexos do que os que necessários à mitigação das opacidades algorítmica. De acordo com certa doutrina<sup>159</sup>, ao contrário dos algoritmos, o decisor humano pode ser questionado diretamente pelo destinatário da decisão e apresentar razões para aquela decisão. Ora, este questionamento não assegura que o

---

<sup>159</sup> Cf. REBECCA CROOTOF / MARGOT E. KAMINSKI / WILLIAM NICHOLSON PRICE II, “Humans in the Loop”, cit., p. 30.

decisor humano responda; se responder, não garante que o decisor humano responda de forma completa, verdadeira, transparente e sem racionalizações ou fundamentações arquitetadas *post hoc*. Os processos decisórios humanos não são, em nenhum cenário, controláveis sem a colaboração voluntária da pessoa. Por outras palavras: a mente humana não é auditável. Mas os algoritmos são.

#### 4.3. Processo formativo e declaração negocial: a explicação como descrição e a tendencial irrelevância da motivação

50. A declaração negocial, como referido (4.1.), é o produto de uma decisão<sup>160</sup>. O processo formativo da declaração negocial abrange, assim, o processo decisório (baseado, ou não, em algoritmos) que conduziu à decisão de emissão de uma declaração negocial. Por outras palavras, a contemplação analítica da declaração negocial permite nela distinguir o processo formativo, o texto do ato e a sua eficácia jurídica.

Vincar a distinção entre o processo formativo e a declaração negocial – pese embora se reconheça a sua evidência – tem a utilidade de explicitar duas ideias.

51. A primeira ideia é a de que as descrições sobre o processo formativo não se confundem com o ato comunicativo e performativo que é a declaração negocial. Essas descrições – que são um metatexto – tanto poderão dirigir-se à *explicação* como à *justificação* da declaração negocial<sup>161</sup>. O texto da declaração negocial (que é tanto

<sup>160</sup> É conhecida a expressão segundo a qual “a decisão corresponde ao fim do pensamento e ao início da ação” (“decision implies the end of deliberation and the beginning of action”) – comumente atribuída a William Starbuck, professor emérito da University of Oregon’s Charles H. Lundquist College of Business. Note-se ainda que *nada decidir* pode corresponder a uma decisão (negativa) (*v.g.*, em fase pré-contratual, a seguradora, munida com a informação disponibilizada pelo cliente e com a informação obtida via *big data*, *decide não fazer* qualquer proposta de contrato de seguro ao cliente).

<sup>161</sup> Recorremos, sem problematizar, à distinção clássica, desenvolvida na literatura da filosofia da ação, entre explicação e justificação, ou entre razões explicativas e razões justificativas. Com mais ou menos *nuances*, trata-se de distinção largamente acolhida pelos filósofos contemporâneos – cf., entre outros, Joseph Raz, *Practical Reasoning and Norms*, London, Hutchinson & Co., Oxford University Press (1990, reprinted); Jonathan Dancy, “Why There Is Really No Such Thing as the Theory of Motivation”, *Proceedings of the Aristotelian Society*, 95, 1995 (1-18); do mesmo Autor, *Practical Reality*, Oxford, Clarendon Press, 2000; para uma visão histórica e completa destes temas, cf. Maria Alvarez, “Reasons for Action: Justification, Motivation, Explanation”, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), disponível em <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/> [consultado a 20.07.2022]. Muito depuradamente, são razões explicativas aquelas que se identificam com as circunstâncias envolvidas na ação e com os motivos do agente, e que são constituídas por estados mentais ou subjetivos que são antecedentes causais de certas ações (paradigmáticamente: as crenças ou os desejos que baseiam uma certa decisão); são razões justificativas (ou normativas) para a ação aquelas que assentam em determinadas normas, princípios ou códigos que

o objeto como o resultado final da interpretação) não é tudo o que esta representa: o texto é o núcleo expressivo onde se cristaliza a ação comunicativa negocial concreta e que prefigura os efeitos da declaração negocial, diz o que se *põe-a-valer*<sup>162</sup>, mas não reproduz toda a atividade desencadeada *ex ante*, não diz nem *como* nem *porquê*.

Excetuadas, para um certo entendimento, as descrições que possam ser vistas como “manifestações de vontade” jurígenas (dados os problemas substantivos, de intenso debate doutrinário, da oposição entre *declaração* e *vontade*<sup>163</sup>), a descrição de elementos do processo formativo não conforma a declaração negocial. Os enunciados assertivos que podem conter-se numa explicação da declaração negocial não integram, à partida, essa declaração.

Pense-se nas asserções que sejam veiculadas pelo decisor-declarante sobre um estado subjetivo no momento da tomada da decisão negocial, sobre a descrição de pessoas ou de objetos, sobre motivações e finalidades primárias e secundárias, sobre fatores determinantes para a tomada daquela decisão, ou mesmo sobre cenários hipotéticos ponderados. Todas estas asserções são candidatas a integrar uma explicação da declaração negocial (e do seu processo formativo), mas *não a formam*. Seja esta, ou não, uma declaração negocial baseada num processo decisório algorítmico.

Só assim não será quando estas descrições – estes enunciados assertivos – forem acolhidos no enunciado negocial (por exemplo, quando o enunciado da declaração inclua, explícita ou implicitamente, algum destes elementos assertivos que confira ao conjunto natureza jurídico-negocial). De outra forma, estas meras descrições não relevam enquanto declaração negocial. Mais discutível é se têm importância, em sede interpretativa, no apuramento do sentido da declaração negocial<sup>164</sup>.

---

impõem, permitem ou proíbem ações e que, desse prisma normativo, permitem valor a ação como *correta* ou *incorreta*, *boa* ou *má*. Apenas estas são razões suscetíveis de uma análise de tipo lógico. Num exemplo cristalino apresentado por Manuel Atienza, *Las Razones del Derecho. Teorías de la Argumentación Jurídica*, 2.ª ed., México, Universidad Nacional Autónoma de México, 2005, p. 5 “decir que el juez tomó [la] decisión debido a sus firmes creencias religiosas significa enunciar una razón explicativa; decir que la decisión del juez se basó en determinada interpretación del artículo 15 de la Constitución significa enunciar una razón justificatoria. Los órganos jurisdiccionales o administrativos no tienen, por lo general, que explicar sus decisiones, sino justificarlas”.

<sup>162</sup> Para usar a conhecida expressão (“*in-Geltung-setzen*”) de Karl Larenz, usada a propósito da sua “teoria da validade” (*Geltungstheorie*) da declaração (cf. *Die Methode der Auslegung des Rechtsgeschäfts*, Frankfurt a. M., Berlin, Metzner, 1930, pp. 43 ss.).

<sup>163</sup> Entre nós, é em Paulo Mota Pinto, *Declaração tácita e comportamento concludente no negócio jurídico*, cit., pp. 553 ss. que encontramos uma súmula e um tratamento aturado da (ir)relevância das manifestações de vontade no conceito de «declaração» adotado no nosso Código Civil.

<sup>164</sup> Pode discutir-se se a explicação da declaração negocial pelo declarante poderá valer como “interpretação autêntica” desta. Numa visão restritiva, sobre os problemas deste tipo de interpretação, cf. Maria Raquel Rei, *Da interpretação da declaração negocial no Direito civil português*, dissertação de doutoramento

A confirmação desta conclusão chega-nos, desde logo, pelo regime do erro sobre os motivos (artigo 252.º, n.º 1, do Código Civil): é irrelevante o erro sobre os motivos como fundamento de anulação do negócio porque “os motivos não integram o conteúdo do negócio”<sup>165</sup>. O erro sobre os motivos, também perspetivado como um “erro no processo decisório”<sup>166</sup> (um erro sobre a existência, subsistência ou verificação de uma circunstância presente ou atual que era determinante para a decisão de emitir aquela declaração negocial<sup>167</sup>), apenas será relevante quando as partes reconhecerem por acordo (ainda que tácito) a essencialidade do motivo. Nestes casos, a relevância do erro não resulta da representação subjetiva de uma das partes (como sucede no erro sobre a pessoa ou no erro sobre o objeto), mas antes da *recepção* desses motivos no negócio, através da concordância de ambas as partes, prévia à formação do negócio, sobre a importância, para esse negócio, dos motivos em jogo<sup>168</sup>. Neste condicionamento consensualizado da validade do negócio à verificação de um determinado motivo, não está, pois, em causa, conferir relevância à vontade interna do declarante, mas antes conferir relevância ao acordo<sup>169</sup>. Este regime do erro sobre

---

(polic.), Lisboa, FDUL, 2010, pp. 429 ss. Seja qual for a tese acolhida, haverá que considerar o declarante-explicador numa posição exterior relativamente à sua declaração negocial, uma vez identificado o estabilizado o *quid* interpretando (em exemplo dado pela Autora, “[s]e o declarante, aquando da emissão da declaração negocial, afirma que a expressão “os meus códigos” inclui “as minhas compilações de leis e ainda a minha coleção de *Colectâneas de Jurisprudência*”, este esclarecimento não é – ainda – interpretação da declaração negocial. Trata-se de um trecho da própria declaração, em que o seu autor explicita o sentido da afirmação anterior (ou seguinte). Estas explicitações são inteiramente lícitas. São, inclusivamente, muito naturais. O declarante é livre de se exprimir mais sucinta ou mais explicadamente” (pp. 429-430). Fora destes casos em que a explicação forma a declaração negocial, julgamos que a interpretação feita pelo declarante ou mesmo pelo declaratário, *qua tale*, não releva à luz do regime primariamente aplicável: o critério interpretativo gizado no artigo 236.º, n.º 1, do Código Civil, não acolhe o juízo que algum destes possa realizar e expressar sobre o sentido da declaração.

<sup>165</sup> Cf. A. VON TUHR, *Der allgemeine Teil des deutschen bürgerlichen Rechts*, II, Duncker & Humblot, reimp. 2013, § 67, Rn. 578. Esta irrelevância encontra justificação no facto de os motivos corresponderem a representações psicológicas, o que é dificilmente apreensível pela contraparte, mesmo que sejam exteriorizadas. Defende-se, por isso, que são razões de segurança e certeza jurídicas que excluem, para estes efeitos, a atendibilidade dos motivos do perímetro da declaração negocial (e do negócio) (cf. PAULO MOTA PINTO, *Declaração tácita e comportamento concludente do negócio jurídico*, cit. p. 355), cabendo a cada sujeito suportar o risco atinente à sua formação de vontade, sob pena de se introduzir uma incerteza insuportável em quase todas as transações (Karl Larenz / Manfred Wolf, *Allgemeiner Teil des Bürgerlichen Rechts*, 9. Auflage, München, C-H.Beck, 2004, § 36, Rn. 2).

<sup>166</sup> Ac. STJ, 15-05-2012, Proc. 5223/05.3TBOER.L1.S1, disponível em [www.dgsi.pt](http://www.dgsi.pt).

<sup>167</sup> Ac. STJ, 18-06-2015, Proc. 3200/04.0TVLSB.L2.S1, disponível em [www.dgsi.pt](http://www.dgsi.pt).

<sup>168</sup> Cf. MANUEL DE ANDRADE, *Teoria geral da relação jurídica*, II, Coimbra, Almedina, 1966, pp. 341 ss.

<sup>169</sup> Cf. PAULO MOTA PINTO, *Declaração tácita e comportamento concludente do negócio jurídico*, cit., p. 360.



os motivos fornece-nos pistas no sentido da irrelevância, em geral, das descrições e das perturbações do processo decisório negocial e conduz-nos inelutavelmente à conclusão de que este tipo de asserções não integram o texto da declaração negocial.

Em suma: a explicação da declaração negocial não se confunde com esta, e os dados positivos do nosso direito civil indiciam a irrelevância dessa explicação no plano da eficácia negocial.

52. A segunda ideia, alinhada com a primeira, é a de que, no direito civil, apesar da atenção dada aos vícios que atinjam a formação da vontade do declarante, esse processo formativo não é, *regra geral*, normativamente condicionado<sup>170</sup>.

Diversamente do que sucede com o decisor administrativo ou com o decisor judicial, o decisor-declarante de direito privado pode decidir segundo o processo decisório que quiser, em função daquilo que sejam as suas preferências subjetivas (ainda que irracionais, à luz das mais diversas teorias da decisão), e sem que a apresentação de uma explicação ou de uma justificação da sua decisão seja condição da eficácia da mesma como declaração negocial<sup>171</sup>.

Regra geral, não está o decisor-declarante condicionado nem no processo empírico, heurístico ou psicológico de produção da decisão, nem obrigado a apresentar os fundamentos em que assenta essa decisão. À liberdade na condução do processo decisório e na conformação do conteúdo da decisão *como regra* preside a autonomia privada do decisor. À ausência de um dever geral de explicação ou de justificação das declarações negociais preside uma ideia de inutilidade: não estando em jogo qualquer ato ou resultado antijurídico, de nada serve, para o respetivo destinatário, a explicação ou justificação, quando, mesmo que o processo decisório ou a decisão fossem sindicáveis, ao decisor-declarante não seria imposto um determinado processo decisório, nem estaria obrigado a emitir uma declaração negocial com determinado conteúdo (mesmo que esta fosse considerada inválida). Isto vale tanto para os casos em que a explicação é possível, como, por maioria de

---

<sup>170</sup> Não se ignora, porém, a tendência de juridificar teorias económicas sobre a racionalidade do processo decisório e torná-las modelos de decisão, com critérios jurídicos operativos, *maxime*, no domínio do direito privado – e mesmo na construção, aí, de novos modelos de informação. Destaca-se, neste contexto, o estudo de HYPERLINK “[https://www.mohr.de/suche/suchergebnisse?tx\\_sgpublisher\\_pi1%5BauthorUid%5D=7725&cHash=a82c82ddc36041669aa002e819ca093d](https://www.mohr.de/suche/suchergebnisse?tx_sgpublisher_pi1%5BauthorUid%5D=7725&cHash=a82c82ddc36041669aa002e819ca093d)” Philipp Hacker, *Verhaltensökonomik und Normativität, Die Grenzen des Informationsmodells im Privatrecht und seine Alternativen*, Tübingen, Mohr Siebeck, 2017.

<sup>171</sup> Apontando as dificuldades em encontrar normas de direito privado sobre os motivos ou sobre a motivação do declarante, cf. ALESSIO D. SCANO, *La motivazione delle decisioni nelle società di capitali*, Milano, Giuffrè, pp. 11 ss.

razão, para todos aqueles casos referentes a explicações difíceis (excessivamente onerosas), a explicações impossíveis e a explicações inúteis.

Como já desenvolvemos em estudo anterior<sup>172</sup>, esta ideia surge, contudo, enfraquecida se se considerar, em regimes especiais, a existência de pontuais normas específicas que regulam processos decisórios, seja através da fixação de *regras procedimentais*<sup>173</sup>, seja através do *fornecimento de critérios decisórios e do seu peso relativo*<sup>174</sup>, seja através da *determinação da relevância ou da irrelevância de certa informação* na tomada de decisão<sup>175</sup>, seja através da *imposição de uma explicação ou justificação* da decisão<sup>176</sup>. Estas normas estão geralmente orientadas à proteção e garantia de um “núcleo intangível” do conteúdo da decisão, considerando-se violadas mesmo que não se verifique o resultado desvantajoso a que atendia o *perigo* para cuja evitação estava teleologicamente orientada a norma violada (um exemplo óbvio é o da nulidade dos negócios celebrados em conflito de interesses proibido, independentemente do conteúdo desse negócio – artigo 397.º, n.º 2, do Código das Sociedades Comerciais).

A regulação descrita sinaliza a independência do processo decisório face à própria decisão, e atribui-lhe uma relevância própria, por se poder identificar um conjunto de situações jurídicas passivas de *dimensão procedimental* cominadas ao decisor. Deixa também o caminho aberto à identificação dessas situações fora dos ditames da lei estrita.

---

<sup>172</sup> Cf. ANA ALVES LEAL, “Aspetos jurídicos da análise de dados na internet”, cit., pp. 200 ss.

<sup>173</sup> Pense-se na *business judgment rule* (artigo 72.º, n.º 2, do Código das Sociedades Comerciais) e na relevância do conjunto de prescrições relativas ao processo de decisão dos administradores de sociedades comerciais.

<sup>174</sup> Por exemplo: os administradores das sociedades comerciais devem, na formação das suas decisões, atender “aos interesses de longo prazo dos sócios e ponderando os interesses dos outros sujeitos relevantes para a sustentabilidade da sociedade, tais como os seus trabalhadores, clientes e credores” (artigo 64.º, n.º 1, alínea b), do Código das Sociedades Comerciais).

<sup>175</sup> Por exemplo: no sentido da relevância de certa informação no cumprimento do dever de adequação dos intermediários financeiros, determina-se que estes devem “basear-se nas informações prestadas pelos seus clientes ou clientes potenciais, salvo se tiver[em] conhecimento, ou deves[em] ter conhecimento, de que as informações estão manifestamente desatualizadas ou são inexatas ou incompletas” (artigo 55.º, n.º 3, do Regulamento Delegado (UE) 2017/565 da Comissão, de 25 de abril de 2016).

<sup>176</sup> Em exemplo já abordado, determina o artigo 16.º, n.º 3, do Decreto-Lei n.º 74-A/2017, de 23 de junho, que “se o pedido de crédito for rejeitado com fundamento nos dados constantes das bases de dados de responsabilidades de créditos, ou da lista pública de execuções, ou de outras bases de dados consideradas úteis para a avaliação da solvabilidade dos consumidores, o mutuante deve informar o consumidor imediata, gratuita e justificadamente desse facto”.