



UNIVERSIDADE CATÓLICA PORTUGUESA

A Discriminação Algorítmica na era do Constitucionalismo Digital

Diana Camões Moreira

Mestrado em Direito

Faculdade de Direito | Escola do

Porto

2024



UNIVERSIDADE CATÓLICA PORTUGUESA

A Discriminação Algorítmica na era do Constitucionalismo Digital

Diana Camões Moreira

Orientador: Catarina Santos Botelho

Mestrado em Direito

Faculdade de Direito | Escola do

Porto

2024

À minha mãe,

Agradecimentos

Em abril de 2023, numa ida à Bertrand de Espinho, na rua 19, a minha mãe ofereceu-me o livro *A era da inteligência artificial*. Foi aí que se fez luz sobre o tema da dissertação: a discriminação algorítmica, concretizando-se o meu desejo de conciliar as duas áreas do direito que mais me apaixonam: o Direito Digital e o Direito Constitucional.

Agradeço à minha mãe, a minha principal referência de vida, por todo o apoio, dedicação e carinho. Estarei eternamente grata por tudo aquilo que fez por mim, por me dar a motivação mesmo nos dias mais cinzentos e por me fazer acreditar e lutar pelos meus objetivos. Todos os meus sonhos e, em particular, todas as minhas conquistas do meu percurso académico seriam inalcançáveis sem a sua presença.

À minha avó Fernanda, cuja memória guardarei para sempre na minha alma, por ter sido um exemplo de força e resiliência.

À minha orientadora, Sra. Professora Doutora Catarina Santos Botelho, cuja inestimável orientação foi essencial para a elaboração da minha Dissertação. Agradeço toda a partilha de conhecimento e por me demonstrar que ainda é possível acreditar na academia.

Resumo

Num mundo cada vez mais digital, os algoritmos são a sua estrela em ascensão. A discriminação algorítmica coloca-nos desafios de difícil resolução, implicando uma análise crítica do enquadramento legal existente. Assim, a presente Dissertação procede à reflexão sobre as diferentes situações em que a discriminação algorítmica pode surgir, avaliando se fará ou não sentido aplicar o direito antidiscriminação. De seguida, destacamos o papel que o Regulamento da Inteligência Artificial terá na regulamentação desta matéria, fazendo a articulação necessária com o Regulamento Geral da Proteção de Dados. Finalmente, apontamos as soluções que, em nosso entendimento, poderão dar uma resposta efetiva à discriminação algorítmica nesta era do constitucionalismo digital.

Palavras-chave

Discriminação Algorítmica; Algoritmo; Constitucionalismo Digital; Direitos Fundamentais, Inteligência Artificial.

Abstract

In a more digitalized world, algorithms are its rising star. Algorithmic bias poses difficult challenges, implying a critical analysis to the existent legal framework. Hence, the following Dissertation proceeds to the reflection about the different situations where algorithm bias can arise, evaluating if it will make sense to apply an antidiscrimination law. Then, we highlight the role that the Artificial Intelligence Act will have in the regulation of this matter, by doing the necessary articulation with the General Data Protection Regulation. Finally, we point out solutions that, in accordance to our understanding, will give an effective answer to the algorithmic bias in this new era of digital constitutionalism.

Key Words

Algorithmic Bias; Algorithm; Digital Constitutionalism; Fundamental Rights; Artificial Intelligence.

Índice

Agradecimentos	5
Resumo	6
Palavras-chave	6
Abstract.....	6
Key Words.....	6
Advertência.....	9
Lista de Siglas e Abreviaturas	10
Parte I.....	11
1. Introdução.....	11
2. Os Algoritmos e a ML	13
3. A Discriminação Algorítmica.....	17
3.1 Dados correlacionados.....	17
3.2 Dados que refletem discriminação social	19
3.3. Discriminação resultante da programação do algoritmo	21
3.4 A natureza da Discriminação algorítmica	22
4. Enquadramento no âmbito do direito antidiscriminação.....	23
4.1 Problemas resultantes da aplicação do direito antidiscriminação	25
Parte II	30
5. A discriminação algorítmica na era do constitucionalismo digital.....	30
6. A Regulação como mecanismo preventivo para combater a discriminação algorítmica	31
6.1 Regulamento da IA.....	31
6.2 O RGPD.....	36
a) A consideração das inferências.....	37
b) Uma nova exceção para o tratamento de dados sensíveis	38
c) Clarificar o nível de intervenção humana.....	40

7. Adaptação na era do constitucionalismo digital – a criação de novos direitos?.....	42
8. Conclusão	43
Bibliografia.....	46
Legislação.....	54
Jurisprudência.....	55

Advertência

Por forma a utilizar menos caracteres, nas notas de rodapé é usado o sistema autor, data, página. A informação bibliográfica completa está disponível no final da Dissertação.

Lista de Siglas e Abreviaturas

al.	Alínea
als.	Alíneas
art.	Artigo
CDFUE	Carta dos Direitos Fundamentais da União Europeia
CEDH	Convenção Europeia dos Direitos Humanos
consult.	Consultado
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
CRP	Constituição da República Portuguesa
CPDHED	Carta Portuguesa de Direitos Humanos na Era Digital
Ed.	Editor
Eds.	Editores
EDPB	European Data Protection Board
GPAI	General Purpose Artificial Intelligence Systems
GT29	Grupo de Trabalho de Proteção de Dados do Artigo 29
IA	Inteligência Artificial
ML	Machine Learning
n.º	Número
p.	Página
pp.	Páginas
RGPD	Regulamento Geral da Proteção de Dados
ss.	Seguintes
TFUE	Tratado sobre o Funcionamento da União Europeia
TJ	Tribunal de Justiça
TJUE	Tribunal de Justiça da União Europeia
TUE	Tratado da União Europeia

Parte I

1. Introdução

Num mundo cada vez mais digital, a IA é a sua estrela em ascensão. O seu progressivo avanço parece inevitável, mas o seu destino não o é¹. Vivemos numa sociedade algorítmica², tendo os processos manuais realizados por seres humanos sido, continuamente, substituídos por decisões automatizadas e práticas algorítmicas³. Já em 1950, ALAN TURING, perçecionado como o “pai” da IA, propôs que o foco passasse a residir não na inteligência da máquina, mas sim na sua manifestação, a qual apelidou de “jogo da imitação”⁴. Desde então, assistiu-se a verdadeira revolução científica e, sobretudo, tecnológica.

Tal não é isento de desafios, já que esta nova forma de “algocracia”⁵ molda práticas sociais, humanas e governativas, sendo inegável o seu impacto no mundo como o conhecemos⁶. Consequentemente, cria-se um paradigma: ao passo que a IA permite incrementar a eficiência dos serviços⁷, poderá contribuir igualmente para colocar em causa direitos fundamentais⁸ e o sistema democrático tal como o conhecemos⁹. Na esteira de HELGA NOWOTNY, “confiamos na IA ao mesmo tempo que dela desconfiamos”¹⁰. Esta ambivalência de confiança e desconfiança é essencial para nos fazer refletir criticamente sobre os problemas decorrentes da sua utilização.

Surge-nos, assim, a problemática da nossa Dissertação: a discriminação algorítmica. À semelhança das decisões humanas, os algoritmos também discriminam¹¹, condenando-nos “a repetir o passado do qual queríamos fugir, ao replicar os preconceitos que definem muitas

¹ H. A. KISSINGER / ERIC SCHMIDT / DANIEL HUTTENLOCHER (2021), p. 21

² GIOVANNI DE GREGORIO (2022), p. 6. ORESTE POLLICINO / GIOVANNI DE GREGORIO (2022), p. 4.

³ Tal tem ocorrido em sucessivos setores: banca, sistema judicial, mundo laboral e política, por exemplo.

⁴ ALAN TURING (1950), p. 433. Deste modo, o objetivo passaria por aferir se a máquina conseguia, ou não, atuar de tal forma que não fosse possível distinguir o seu comportamento dos seres humanos, caso em que seria considerada inteligente. Permitiu-se, como postulam HENRY A. KISSINGER / ERIC SCHMIDT / DANIEL HUTTENLOCHER (2021), p. 59-60, marginalizar “séculos de debates sobre a natureza da inteligência.”

⁵ PAULO NUNO VICENTE (2022), pp. 52-55,

⁶ JULIANA CASTANEDA *et. al* (2022), p. 4. IGNACIO N. COFONE (2019), p. 1391

⁷ Na esteira de CATARINA SANTOS BOTELHO (2017), p. 52, “o mundo e o espaço digital surgem então como um armazenamento contínuo e inesgotável de dados.”

⁸ Sendo de salientar, à cabeça, o direito à proteção de dados. GIOVANNI DE GREGORIO (2022), p. 230. A IA afeta praticamente todos os direitos fundamentais, como recorda CHARLINE DAELMAN (2021), p. 124. EMILE LOZA SILES (2021), p. 518 evidencia que será impossível impedir o curso da IA, agora que esta se encontra difundida por todo o lado.

⁹ CATARINA SANTOS BOTELHO (2024), p. 2.

¹⁰ (tradução nossa) HELGA NOWOTNY (2021), p. 21.

¹¹ Como assinala IFEOMA AJUNWA (2020), p. 1673, a discriminação algorítmica tem a desvantagem de reproduzir significativamente as ocorrências discriminatórias.

vezes as pessoas que os programam”¹². À primeira vista, estes casos não corresponderão à típica discriminação humana (a qual será muitas vezes intencional). Ao invés, eles serão difíceis de detetar e implícitos¹³, mantendo uma aparência de objetividade¹⁴. Tendo em consideração que “um algoritmo não deve ser visto como um oráculo que tudo sabe”¹⁵, devemos ter a capacidade de refletir criticamente sobre o seu impacto no quotidiano.

Começaremos por discorrer sobre o conceito de algoritmo, para seguidamente analisar as diferentes situações em que esta discriminação pode surgir, aferindo qual é, afinal, a sua natureza. Como denota MICHAEL SELMI, um ponto importante de reflexão estará em saber se estes algoritmos podem, ou não, produzir resultados mais nefastos do que os sistemas “tradicionais”¹⁶. Assim, num momento em que a máquina consegue processar informação a uma escala nunca antes vista (e bastante superior à capacidade humana), quais deverão ser os limites estabelecidos¹⁷? Adicionalmente, o núcleo essencial do nosso estudo residirá em aferir o tipo de respostas que podem ser apresentadas, avaliando igualmente se o *law in books* está a acompanhar devidamente o *law in action*, isto é, se as soluções legais criadas consubstanciam um acompanhamento coerente dos avanços tecnológicos.

¹² TERESA COELHO MOREIRA (2021), p. 90.

¹³ SOLON BAROCAS / ANDREW D. SELBST (2016), p. 673.

¹⁴ SANDRA G. MAYSON (2019), p. 2221.

¹⁵ MILENA DA SILVA ROUXINOL (2021), p. 236.

¹⁶ MICHAEL SELMI (2020), p. 410.

¹⁷ LUCAS CORTIZO (2021), p. 99 relembra que a IA construiu um “ideal de apreensão” já que pela primeira vez existe uma tecnologia que pode a breve trecho ultrapassar a inteligência humana.

2. Os Algoritmos e a ML

Os algoritmos, devido à sua complexidade e obscuridade, conjuram certos pensamentos de “magia das trevas”¹⁸, já que “são muito complexos e desafiam a compreensão humana”¹⁹. Estes podem ser definidos como uma “descrição formalizada e abstrata de procedimentos computacionais”²⁰. Por outro lado, seguindo uma perspectiva de engenharia, os algoritmos podem ser entendidos como “um sistema de dois componentes, o componente lógico e o componente de controlo”²¹. A evolução tecnológica tem levado a que estes deixem de promover exclusivamente meros cálculos computacionais, “enveredando em searas artísticas, seja para auxílio dos indivíduos da área, seja como entidades inteligentes e criativas”²². Vejamos.

Podemos distinguir duas situações distintas: (i) os algoritmos tradicionais (*knowledge-based algorithms*) e (ii) algoritmos de aprendizagem automática (*ML algorithms*)²³. Os primeiros apresentam um carácter determinístico²⁴, já que, com base nos *inputs*²⁵ recebidos, há a produção do resultado específico e expectável à saída (*outputs*)²⁶. Assim, os dados entram no computador, “o algoritmo faz o que quer com eles, e o resultado sai”²⁷. Os algoritmos tradicionais assentam numa perspectiva matemática²⁸, tornando possível a explicação de como determinado processo é gerado²⁹, permitindo igualmente fazer uma distinção relevante: ao passo que todos “os empreendimentos em torno da IA implicam sempre algoritmos, nem todos os algoritmos recorrem necessariamente a IA”³⁰.

A evolução tecnológica fulminante fez surgir os algoritmos de aprendizagem automática. Ao contrário dos primeiros que requerem programação manual, os algoritmos

¹⁸ DEVEN R. DESAI / JOSHUA A. KROLL (2017), p. 23. Metaforicamente, DIOGO QUEIROZ DE ANDRADE (2023), p. 7 advoga que o algoritmo é como uma receita: “tem a listagem dos ingredientes, tem uma ordem e tem as instruções de utilização para se seguir um objetivo pré-determinado.”

¹⁹ CATARINA DOS SANTOS BOTELHO (2024), p. 4 (tradução nossa).

²⁰ PAUL DOURISH (2016), p. 3, FREDERIK ZUIDERVEEN BORGESIU (2018), p. 11

²¹ JULIANA CASTANEDA *et. al* (2022), p. 3. (tradução nossa).

²² LUCA CORTIZO (2021), p. 102.

²³ IGNACIO N. COFONE (2019), p. 1394, HENRY A. KISSINGER / ERIC SCHMIDT / DANIEL HUTTENLOCHER (2021), pp. 61-62, PAULO NUNO VICENTE (2023), p. 24.

²⁴ Daí que sejam também conhecidos por algoritmos determinísticos. *Vide* PAULO NUNO VICENTE (2023), p. 24.

²⁵ Os *inputs* correspondem aos dados fornecidos à entrada do algoritmo.

²⁶ Os *outputs* traduzem os dados à saída, isto é, o resultado final produzido pelo algoritmo.

²⁷ PEDRO DOMINGOS (2017), p. 30

²⁸ *Vide* NIZAN GESLEVICH PACKIN / YAFIT LEV-ARETZ (2020), p. 88-89.

²⁹ Tal como é assinalado por PAULO NUNO VICENTE (2023), pp. 23-24.

³⁰ PAULO NUNO VICENTE (2023), p. 25.

de aprendizagem automática “aprendem” e melhoram a sua prestação através da exposição a um vasto conjunto de dados³¹. *In casu*, há uma inversão da situação: os dados e os resultados desejados são fornecidos, saindo o algoritmo que “transforma o primeiro no segundo”³². Permitem, por isso, asseverar “a qualidade dos resultados e proporcionar meios para os melhorar”³³.

Consequentemente, estes dispõem de uma capacidade fenomenal para fazer previsões, classificações e identificar padrões³⁴ para criar novos dados de *input*. Deste modo, o “trabalho sujo”³⁵ é realizado pelos algoritmos de aprendizagem automática, sendo estes que identificam os diversos padrões, bem como a dependência na relação entre os diversos dados³⁶. Pensemos no exemplo referido por DAVID S. RUBENSTEIN: se o objetivo for o de detetar e diagnosticar doenças (como o cancro), poder-se-á “alimentá-lo” com várias imagens de tumores malignos e benignos, desenvolver como poderá proceder à distinção entre ambas com base em correlações estabelecidas, possibilitando-se a criação de um algoritmo que permita aprender a fazer previsões de diagnósticos clínicos³⁷.

Devemos fazer uma distinção entre aprendizagem supervisionada, aprendizagem não supervisionada e a aprendizagem por reforço³⁸.

A aprendizagem supervisionada é aquela onde o programador assume um papel de “programador pessoal”³⁹, pois como o próprio nome indica, é aqui que eles requerem mais atenção⁴⁰. Tal deve-se ao facto de serem fornecidos ao algoritmo um conjunto diversificado de dados, os quais contendo exemplos de *inputs*, são “rotulados” segundo o *output* ou resultado pretendido⁴¹. *In casu*, eles podem constituir algoritmos de classificação (o que é muito frequente para reconhecimento de voz, spam ou imagem)⁴², podendo apresentar duas

³¹ DAVID S. RUBENSTEIN (2021), p. 759.

³² PEDRO DOMINGOS (2017), p. 30.

³³ HENRY A. KISSINGER / ERIC SCHMIDT / DANIEL HUTTENLOCHER (2021), p. 67.

³⁴ PAULO NUNO VICENTE (2023), p. 28. IGNACIO N. COFONE (2019), p. 1395, DAVID S. RUBENSTEIN (2021), p. 759-760 e FILIPA URBANO CALVÃO (2023), p. 79.

³⁵ Expressão usada por PAULO NUNO VICENTE (2023), p. 28.

³⁶ *Ibid*, p. 28.

³⁷ DAVID S. RUBENSTEIN (2021), p. 760

³⁸ Sobre o tema na doutrina, *vide* igualmente LILIAN EDWARDS / MICHAEL VEALE (2017), pp. 25-27, IGNACIO N. COFONE (2019), pp. 1395-1936, HENRY A. KISSINGER / ERICH SCHMIDT / DANIEL HUTTENLOCHER (2021), pp. 69-72.

³⁹ PAULO NUNO VICENTE (2023), p. 28.

⁴⁰ DAVID LEHR / PAUL OHM (2017), p. 672.

⁴¹ Assim, HENRY A. KISSINGER / ERICH SCHMIDT / DANIEL HUTTENLOCHER (2021), p. 69. *Vide* igualmente PAULO NUNO VICENTE (2023), p. 28 e DAVID LEHR / PAUL OHM (2017), pp. 672-673.

⁴² DAVID LEHR / PAUL OHM (2017), p. 673.

ou mais variáveis. Recordemos o exemplo referido por PAULO NUNO VICENTE⁴³. Se o objetivo de um algoritmo for o de identificar que imagens correspondem a maçãs, então, o seu treino seria composto por um conjunto diversificado de maçãs (com várias cores) – as quais corresponderiam a uma etiqueta (*label*) “positivo” – e, adicionalmente, imagens de outras frutas (manga, morangos, abacaxi, por exemplo), as quais seriam associadas à etiqueta “negativo”⁴⁴. Por sua vez, os algoritmos de regressão correspondem a modelos de aprendizagem supervisionada, que se caracterizam pela sua forma fácil de apreensão. Em termos simples, nos algoritmos de regressão, há a representação de um modelo matemático com a capacidade para estimar os coeficientes de equação linear que melhor preveem o valor da variável dependente.

Por sua vez, a aprendizagem não supervisionada, como o próprio nome indica, é aquela que “prescinde” dos especialistas e outros peritos humanos⁴⁵, os quais se limitam a inserir etiquetas e as classificações nos dados de entrada⁴⁶. Assim, estes “agrupam os sujeitos em função da semelhança dos valores dos seus dados de entrada”⁴⁷, daí que também possa ser conhecida por “*knowledge discovery*”⁴⁸. Como tal, a aprendizagem não supervisionada será mais útil nas situações em que os investigadores disponham de meras coleções de dados não articuladas entre si⁴⁹, encetando ao algoritmo estabelecer padrões que poderiam escapar ao olho humano, almejando-se um resultado mais profícuo à situação em concreto. Permite-se, mediante o objetivo pretendido, diversificar um conjunto de situações: análise de agrupamento (“*clustering*”⁵⁰) ou a redução da dimensionalidade (“*dimensionality reduction*”)⁵¹. Há situações de aprendizagem semiautomática, as quais conjugam características de ambos os tipos de aprendizagem agora evidenciados⁵².

Por último, devemos referenciar a aprendizagem por reforço, através da qual “o algoritmo aprende com base nos erros e nos sucessos obtidos”⁵³. Deste modo, permite-se

⁴³ PAULO NUNO VICENTE (2023), p. 29.

⁴⁴ Ibid, p. 29.

⁴⁵ O que será equivalente a aprender sem o professor, como denota PEDRO DOMINGOS (2017), p.230.

⁴⁶ PAULO NUNO VICENTE (2023), p. 29.

⁴⁷ DAVID LEHR / PAUL OHM (2017), p. 676 (tradução nossa)

⁴⁸ PAUL MURPHY (2012), p. 9.

⁴⁹ HENRY A. KISSINGER / ERIC SCHMIDT / DANIEL HUTTENLOCHER (2021), p. 70.

⁵⁰ Conjunto de técnicas através da qual se promove o agrupamento automático de dados com semelhança. Muito usado na biologia, astronomia ou e-commerce, como salienta PAUL MURPHY (2012), p. 11.

⁵¹ PAUL MURPHY (2012), p. 11 afirma que esta é a técnica através da qual se procede à redução da “dimensionalidade ao projetar os dados para um subespaço de dimensão inferior de maneira a capturar a essência dos dados.” (tradução nossa).

⁵² REMBRANDT DEVILLÉ / NICO SERGEYSSELS / CATHERINE MIDDAG (2021), p. 6.

⁵³ PAULO NUNO VICENTE (2023), p. 30.

que este faça uma exploração por sua conta, encontrem recompensas e descubram como voltar a obtê-las no futuro⁵⁴. O objetivo passará por adotar uma estratégia de forma a maximizar-se a recompensa geral⁵⁵, o que faz com que a IA deixe de ser “um mero observador passivo” para atuar como um “agente num ambiente simplificado e controlado, observando e registando reações e ações”⁵⁶. PEDRO DOMINGOS, de um ponto de vista metafórico, advoga que tal é semelhante aos bebés que gatinham e colocam coisas na boca, aprendendo sucessivamente maneiras de obter as desejadas recompensas⁵⁷. Como tal, os algoritmos são treinados numa lógica de tentativa e erro, sendo expostos a novas situações, assentando num sistema de recompensas e penalizações⁵⁸: sendo o resultado desejável obtido, o algoritmo será premiado. Sempre que não consiga obter o resultado desejável, o algoritmo será desencorajado⁵⁹. Desta forma, o algoritmo “aprende” com o feedback recebido, sendo necessária menor intervenção humana.

Uma breve menção deve igualmente ser feita à *deep learning* (ou aprendizagem profunda), técnica muito recente, através da qual os dados são recolhidos, aprendendo-se e aplicando-se posteriormente a conjuntos maiores de dados para determinar-se ou prever-se algo sobre a realidade⁶⁰. O seu objetivo, como evidencia PAULO NUNO VICENTE, consiste em “imitar o cérebro humano”⁶¹, já que ao contrário do que se sucede na ML, aqui o algoritmo tem a própria capacidade para fazer a sua própria seleção dos dados, sendo muito mais difícil para os cientistas conseguir explicar como determinado resultado foi atingindo⁶². Caracteriza-se, enquanto subcategoria de ML, pela sua maior opacidade e “blindagem”, incorporando em si “a potencialidade de uma disrupção social”⁶³. Consequentemente, contrariamente à ML, torna-se ainda mais difícil (senão mesmo impossível) controlar o processo decisório⁶⁴.

A *deep learning* está em todo o lado, consubstanciando um ponto sem retorno. O seu exemplo mais paradigmático pode ser encontrado na IA Generativa, a qual se baseia em

⁵⁴ PEDRO DOMINGOS (2017), p. 246.

⁵⁵ MATT HERVEY / MATTHEW LAVY (2021), p. 25.

⁵⁶ HENRY A. KISSINGER / ERIC SCHMIDT / DANIEL HUTTENLOCHER (2021), p. 71.

⁵⁷ PEDRO DOMINGOS (2017), p. 246.

⁵⁸ PAULO NUNO VICENTE (2023), p. 30

⁵⁹ *Ibid*, p. 30.

⁶⁰ IGNACIO N. COFONE (2019), p. 1395.

⁶¹ PAULO NUNO VICENTE (2023), p. 32.

⁶² PAULO NUNO VICENTE (2023), p. 35.

⁶³ *Ibid*, p. 36.

⁶⁴ FILIPA URBANO CALVÃO (2023), p. 80.

modelos de *deep learning* e possui a capacidade para criar novos conteúdos de imagem, texto, vídeo, imagens e áudio. Esta poderá, igualmente, ser usada com um intuito malicioso, algo que é sobretudo visível no *deepfake*⁶⁵.

3. A Discriminação Algorítmica

A Discriminação Algorítmica é uma realidade patente no nosso quotidiano. Na esteira de GIOVANNI DE GREGORIO, podemos denotar que a opacidade inerente aos algoritmos constitui o maior desafio à dimensão constitucional da *Big Data*⁶⁶. Há doutrina que considera que o Direito não tem estado à altura de regular algoritmos que produzem resultados discriminatórios. SONIA K. KATYAL considera que pouco tem sido feito e que o atual enquadramento constitucional não é adequado a regular este problema⁶⁷, enquanto outros autores postulam que a doutrina jurídica não dispõe de preparação suficiente para enfrentar os desafios colocados pelos algoritmos⁶⁸.

3.1 Dados correlacionados

Os algoritmos necessitam de dados⁶⁹. Tal como é afirmado por IGNACIO N. COFONE, “um algoritmo só pode ser tão bom quanto os dados que lhe são fornecidos”⁷⁰. Assim, existirá um problema caso os dados de entrada (*inputs*) sejam, eles próprios, enviesados e discriminatórios. Gera-se, assim, um “efeito bola de neve” ou, como denomina alguma doutrina, a “profecia autorrealizável”⁷¹. Estamos diante dados correlacionados (dados proxy), assentando estes dados em “estereótipos e generalizações relativamente a determinados grupos”⁷².

Se os dados de entrada estão incorretos ou incompletos quanto a uma categoria da população ou se, adicionalmente, mantêm um padrão de discriminação, então, o resultado,

⁶⁵ O *deepfake* tem-se caracterizado pela utilização adulterada de imagens, vídeos e áudios criados a partir de ferramentas de IA Generativa, com o intuito de espalhar desinformação, cometer crimes de burla e, inclusive, criar conteúdo pornográfico.

⁶⁶ GIOVANNI DE GREGORIO (2022), p. 229.

⁶⁷ SONIA K. KATYAL (2019), p. 100.

⁶⁸ TALIA B. GILLIS / JANN L. SPIESS (2019), p. 460.

⁶⁹ Na esteira de STEPHANIE BORSTEIN (2018), p. 570, podemos afirmar que “qualquer melhoria na tomada de decisões tradicionais dependerá dos dados que estão a ser utilizados e da forma como são utilizados” (tradução nossa).

⁷⁰ (tradução nossa) IGNACIO N. COFONE (2019), p. 1402. Igualmente, TERESA COELHO MOREIRA (2021), p. 96.

⁷¹ MARIANA ALVES TEIXEIRA, (2023), p. 37.

⁷² *Ibid*, p. 36.

a posteriori, só poderá ser discriminatório. Tal pode acontecer por as bases de dados (individuais ou de empresas) conterem erros e serem pouco representativas de certos grupos sociais⁷³.

Esses acontecimentos afetarão, mais intensivamente, os setores da população *a priori* mais desfavorecidos. Como explicam SOLON BAROCAS E ANDREW D. SELBST, certos grupos sociais, devido a situações antecedentes, poder-se-ão encontrar menos envolvidos na economia formal, bem como nas suas atividades produtivas, ter menos acesso à tecnologia ou serem considerados consumidores “menos atrativos” para observação⁷⁴.

Deste modo, a distorção estatística, mesmo sem qualquer tipo de intenção dolosa⁷⁵, será sempre problemática. Iremos partir do exemplo apresentado por alguns autores. Se um algoritmo for treinado e criado com o objetivo de auxiliar a polícia para parar e revistar os peões, caso os *inputs* apresentem uma incidência excessiva da possibilidade de crime em certos grupos sociais⁷⁶, então, o algoritmo poderá orientar a polícia para fiscalizar membros de determinados grupos sociais numa escala maior e desproporcional face a outros⁷⁷. Também em situações de reconhecimento facial, os algoritmos têm sido treinados com um conjunto de dados que contêm uma reduzida representação de pessoas negras, o que mina o rigor de identificação que se visa almejar⁷⁸. Acima de tudo, devemos ter presente que o algoritmo não poderá produzir um resultado adequado, caso tenha sido treinado com dados enviesados que estão errados ou traduzem uma má representação de determinados grupos sociais⁷⁹.

A nível laboral, o uso de algoritmos para contratação de trabalhadores resulta, em muitas situações, em discriminação⁸⁰. A Amazon pretendia usar um algoritmo destinado a aferir os candidatos mais ajustados ao perfil de contratação pretendido. Até aqui, não se colocaria nenhum problema, até porque cada vez mais (inclusive no setor da advocacia) a triagem inicial é efetuada pelo algoritmo. Todavia, os dados introduzidos tinham em conta

⁷³ SOLON BAROCAS E ANDREW D. SELBST (2016), p. 684. IGNACIO N. COFONE (2019), pp. 1402-1043.

⁷⁴ SOLON BAROCAS E ANDREW D. SELBST (2016), p. 685.

⁷⁵ Joshua A. Kroll *et. al.* (2017), p. 681.

⁷⁶ Os quais já são alvo de perseguição por parte das forças policiais por motivos raciais. Basta tomar em consideração o que constantemente acontece nos Estados Unidos da América, onde latinos e negros, por tendência, tendem a ser vítimas de atuações desproporcionais das forças policiais.

⁷⁷ JOSHUA A. KROLL *et. al.* (2017), p. 681.

⁷⁸ HENRY A. KISSINGER / ERICH SCHIMDT / DANIEL HUTTENLOCHER (2021), p. 83.

⁷⁹ LUCA CORTIZO (2021), p. 111.

⁸⁰ Sobre o tema no âmbito laboral, *vide* LUÍS FELIPE MIRANDA RAMOS (2020), pp. 210-213 e MARIANA ALVES TEIXEIRA (2023), pp. 41 e ss.

os trabalhadores da empresa nos dez anos anteriores, tomando-se em consideração dados dos trabalhadores bem ou malsucedidos na Amazon. Sucede que, nesse período de tempo, a maioria dos trabalhadores tinham sido homens. Consequentemente, o algoritmo começou a associar aptidões profissionais ao gênero masculino (direta ou indiretamente)⁸¹, levando a que as candidatas femininas fossem prejudicadas e menos contratadas.

3.2 Dados que refletem discriminação social⁸²

Nesta situação, os dados são representativos da sociedade, mas ainda assim estes são suscetíveis de produzir um resultado discriminatório refletindo determinadas tendências sociais⁸³. Naquilo que a doutrina apelida de “*garbage in, garbage out*”⁸⁴, como denota MILENA ROUXINOL, “se o *input* for, direta ou indiretamente, deliberada ou inconscientemente, viciado por alguma espécie de preconceito, então isso será refletido pelo *output*”⁸⁵. Nesta circunstância, os dados até podem ser representativos da população, mas devido a questões sociais (muitas vezes inócuas), o resultado que daí advém é igualmente discriminatório⁸⁶. Tal é muito evidente no acesso ao emprego⁸⁷. No exemplo já referido da Amazon, o algoritmo conseguia estabelecer o gênero do candidato através de *proxies*, com o uso das palavras “executei” e “apanhei”, mais utilizadas por homens⁸⁸.

Adicionalmente, fatores relacionados, por exemplo, com o gênero ou raça, poderão ser usadas em desfavor de determinada pessoa. Um exemplo apontado é o *Linked-in*. Se determinado empregador estiver à procura de candidatos para uma vaga, o mais comum será

⁸¹ MILENA ROUXINOL (2021), p. 247-248 reflete ainda sobre a ideia de saber se este algoritmo não será menos nocivo do que uma intervenção humana.

⁸² A qual também pode ser apelidada de “*pre-existing bias*”. Assim, AHMET BILAL AYTEKIN (2023), p. 5.

⁸³ IGNACIO N. COFONE (2019), pp. 1404-1405. Saliente-se que a discriminação resultante de valores sociais não é nova. Relembremos o exemplo dos EUA, onde ainda existe uma discriminação racial que se tem vindo a acumular ao longo dos anos. MAYA C. JACKSON (2021), pp. 306-307 dá o exemplo da Constituição Norte-Americana, a qual no século XVIII indiretamente considerava indivíduos de raça negra como sendo inferiores. Rapidamente, com a proibição da escravatura, foram criados os “*Black codes*”, os quais visavam estabelecer uma alternativa para impedir a sua progressão na sociedade, fosse no acesso ao emprego ou aquisição de propriedade.

⁸⁴ SOLON BAROCAS / ANDREW D. SELBST (2016), p. 683, MILENA ROUXINOL (2021), p. 654, MICHAEL SELMI (2021), p. 616.

⁸⁵ MILENA ROUXINOL, (2021), p. 252.

⁸⁶ Ibid, pp. 252-253 e KIMBERLY A. HOUSER (2019), p. 334.

⁸⁷ Incidindo especificamente sobre o tópico, vide MARIANA ALVES TEIXEIRA (2023), pp. 17-118.

⁸⁸ NATALIA RAMÍREZ-BUSTAMANTE / ANDRÉS PÁEZ (2023), p. 8.

o algoritmo do *Linked-in* apresentar resultados que traduzam, igualmente, preferências da entidade empregadora as quais, em si mesma, poderão ser discriminatórias⁸⁹.

Num universo de todos os utilizadores na plataforma que se possam adequar ao perfil pretendido, a tendência será para que lhe sejam oferecidas recomendações com base em contratações anteriores, as quais podem ter um carácter discriminatório. Caso, por exemplo, se demonstre uma menor preferência por candidatos de um determinado género ou raça, então, o algoritmo continuará a replicar aquelas que são as tendências, não obstante discriminatórias, da entidade empregadora⁹⁰. Assim, como afirma MILENA ROUXINOL, “o preconceito perpetua-se”⁹¹.

Outro exemplo paradigmático é o COMPAS, algoritmo usado por tribunais norte-americanos que permite aferir a probabilidade de reincidência na prática de futuros crimes durante a liberdade condicional. Ora, veio a demonstrar-se uma enorme disparidade racial durante a sua utilização, já que condenados de raça negra eram considerados como tendo um maior risco de reincidência face à realidade (classificados de alto risco) e, contrariamente, condenados brancos eram avaliados mais positivamente (baixo risco). Assim, os indivíduos de raça negra teriam 45% mais probabilidade de receber classificações de risco mais elevado do que condenados brancos⁹². Mesmo não usando a raça como *input*, ainda assim os *proxies* conduziram a resultados discriminatórios, devido a desigualdades sociais enraizadas⁹³, na medida em que, nos dados usados, os indivíduos de raça negra foram estatisticamente sujeitos a mais detenções do que indivíduos brancos⁹⁴, o que posteriormente levou a uma associação errada entre detenção e reincidência⁹⁵.

Como tal, e não obstante os dados serem representativos, ainda assim poderão manifestar desequilíbrios sociais, levando a um resultado também ele discriminatório. Podemos concluir, como afirma MICHAEL SELMI, que “os algoritmos são treinados com dados, não são fórmulas mágicas, e se os dados estiverem enviesados de alguma forma, então os resultados também estarão”⁹⁶.

⁸⁹ SOLON BAROCAS / ANDREW D. SELBST (2016), p. 683, MILENA ROUXINOL (2021), pp. 254-255.

⁹⁰ MILENA ROUXINOL (2021), p. 255 e SOLON BAROCAS / ANDREW D. SELBST (2016), p. 683.

⁹¹ MILENA ROUXINOL (2021), p. 255.

⁹² CATARINA SANTOS BOTELHO (2024), pp. 5-6. JEFF LARSON *et. al* (2016).

⁹³ CATARINA SANTOS BOTELHO (2024), p. 6.

⁹⁴ JEFF LARSON *et. al* (2016). IGNACION N. COFONE (2019), p. 1406.

⁹⁵ *Ibid*, p. 1422.

⁹⁶ MICHAEL SELMI (2021), p. 625 (tradução nossa).

3.3. Discriminação resultante da programação do algoritmo⁹⁷

Por último, iremos analisar a discriminação que provém não dos dados nela introduzidos, mas sim da programação do algoritmo. Temos de ter em consideração que os algoritmos são, no final do dia, uma criação humana. Assim sendo, a forma de programação poderá, em determinadas situações, traduzir discriminações humana, sendo isso codificado e, até, ampliado no algoritmo⁹⁸. Nessa medida, SOLON BAROCAS e ANDREW D. SELBST identificam que tal poderá ocorrer na definição da *target variable* e das *class labels*⁹⁹. Ao passo que a *target variable* define aquilo que se está à procura, as *class labels* dividem todos os valores possíveis da *target variable* em categorias mutuamente exclusivas¹⁰⁰.

Por conseguinte, a definição da *target variable* nem sempre é de fácil execução, pois os programadores devem traduzir um determinado problema numa questão que seja colocada em termos formais para que o computador a possa entender¹⁰¹. Tal não terá problemas, como salientam os autores, quando os conceitos adotados não geram querela¹⁰². Não obstante, quando se pretende que o algoritmo selecione um candidato para um emprego¹⁰³ ou decida a capacidade de determinada pessoa para cumprir o pagamento de um crédito habitacional/consumo, a resposta será menos linear.

Nessas circunstâncias, haverá que definir os critérios para se determinar o que poderá vir a ser um bom trabalhador ou um cumpridor de crédito. Nessa definição de critérios entrará, sem dúvida, alguma subjetividade, o que, concomitantemente, poderá gerar discriminação. MILENA ROUXINOL, no âmbito laboral, evidencia que o empregador/programador poderá treinar o algoritmo a optar por selecionar candidatos apenas de uma determinada zona, excluindo outras associadas a outros grupos sociais¹⁰⁴ ou estabelecer a antiguidade em empregos anteriores¹⁰⁵. Deste modo, a seleção das variáveis

⁹⁷ A qual pode também ser apelidada de discriminação consequencial – CATARINA SANTOS BOTELHO (2024), p. 5.

⁹⁸ IGNACIO N. COFONE (2021), p. 1401.

⁹⁹ SOLON BAROCAS / ANDREW D. SELBST (2016), p. 677.

¹⁰⁰ Ibid, p. 678. *Vide*, igualmente, KIMBERLY A. HOUSER (2019), p. 338

¹⁰¹ SOLON BAROCAS / ANDREW D. SELBST (2016), p. 677.

¹⁰² Os autores dão o exemplo da deteção de e-mails *spam* ou fraude, já que tal não é controverso, cabendo posteriormente ao algoritmo classificar os respetivos e-mails. Bastará aos programadores recorrer a categorias simples e pré-existentes para definir as *class labels*, como denotam SOLON BAROCAS / ANDREW D. SELBST (2016), pp. 678-679.

¹⁰³ MILENA ROUXINOL (2021), p. 251.

¹⁰⁴ MILENA ROUXINOL (2021), pp. 251-252.

¹⁰⁵ Ibid, pp. 251-252.

poderá, *a posteriori*, ser preditiva da pertença a determinado grupo¹⁰⁶, promovendo resultados discriminatórios.

3.4 A natureza da Discriminação algorítmica

Para alguns autores, a discriminação algorítmica apresenta um caráter endógeno, isto é, a discriminação será inerente ao próprio algorítmico¹⁰⁷. Assim, a solução para este problema passaria por assegurar mais transparência e controlo no momento da sua criação¹⁰⁸. Porém, a transparência *per se* não resolve, antes levantando outras questões. Desde logo, mesmo que se vise asseverar mais transparência, ainda assim o procedimento do algoritmo será inteligível para todos aqueles que não sejam da área¹⁰⁹, sendo muito mais difícil de assegurar transparência comparativamente aos processos de decisão humana¹¹⁰. Ademais, a transparência nem sempre é desejável, considerando a necessidade de ponderar outros interesses em confronto, como os segredos de negócio atinentes à criação do algoritmo¹¹¹. Mais, concordamos com os autores que chamam à atenção para o facto de a transparência ser, em si mesma, uma ilusão semelhante ao consentimento no contexto das plataformas digitais (“*transparency fallacy*”)¹¹², o qual se encontra banalizado dando uma falsa sensação de segurança¹¹³. A transparência não só é difícil de definir, como funciona como um curativo rápido demasiado simplista.

Tendo em conta as situações mais comuns em que a discriminação algorítmica ocorre, na maior parte das situações o problema está nos dados de *input* e, mesmo quando resulte da programação humana, o algoritmo *per se* não é discriminatório, mas depende de um conjunto externo de fatores que influenciarão o resultado final.

Por este motivo, concordamos com a doutrina que propugna pela natureza exógena da discriminação algorítmica¹¹⁴, já que o problema não reside na “*black box*” propriamente dita, mas dos fatores sociais, políticos e históricos no qual é criado¹¹⁵. Na esteira de ANUPAM

¹⁰⁶ ROBERT BARTLETT *et. al* (2021), p. 686

¹⁰⁷ Sobre o tópico, CATARINA SANTOS BOTELHO (2023), pp. 31-32.

¹⁰⁸ *Vide*, a este propósito, DONGHEE SHIN E YONG JIN PARK (2019), pp. 277-284.

¹⁰⁹ JOSHUA A. KROLL *et. al* (2017), p. 638.

¹¹⁰ JOSHUA A. KROLL *et. al* (2017), p. 638.

¹¹¹ *Ibid*, p. 639, SONIA K. KAYTAL (2019), p. 138 e CATARINA SANTOS BOTELHO (2023), p. 32.

¹¹² LILIAN EDWARDS / MICHAEL VAELE (2017), p. 65 e ss.

¹¹³ CATARINA SANTOS BOTELHO (2023), p. 32.

¹¹⁴ CATARINA SANTOS BOTELHO (2024), p. 8, ANUPAM CHANDER (2017), p. 1025

¹¹⁵ CATARINA SANTOS BOTELHO (2024), p. 8.

CHANDER, “o problema não é a Caixa Negra, a qual é frequentemente mais neutra do que o decisor humano que substitui, mas o mundo real em que funciona”¹¹⁶. Deste modo, mais do que olhar à raiz do problema (os algoritmos), importa refletir sobre os mecanismos adequados para tutelar esta questão, já que a sua discriminação é o resultado dos fatores que lhe são externos. Consideramos mais relevantes para este efeito, no ordenamento jurídico europeu, o RGPD e o Regulamento IA.

4. Enquadramento no âmbito do direito antidiscriminação

Apresentadas as principais situações em que pode ocorrer a discriminação algorítmica, iremos por agora analisar o seu enquadramento no âmbito do direito antidiscriminação. Afinal, não podemos olvidar que “o imperativo antidiscriminatório cobra a sua utilidade na medida em que enaltece o direito à diferença, isto é, o direito a não ser alvo de um tratamento diferenciado em razão de características pessoais”¹¹⁷. Tal remete-nos, em primeira linha, para o princípio da igualdade¹¹⁸, princípio que se encontra positivado no art. 13.º da CRP¹¹⁹. Adicionalmente, o art. 14.º da CEDH consagra a proibição da discriminação, ao passo que a CDFUE¹²⁰ preceitua o princípio da igualdade perante a lei (art. 20.º) e, ainda, a proibição de discriminação (art. 21.º). Adicionalmente, no âmbito do Direito da UE (originário), tal referência poderia ser sempre encontrada também nos tratados, nomeadamente no art. 3.º, n.º 3, 2.º parágrafo, do TUE (“a União combate a exclusão social e as discriminações”) e art. 10.º do TFUE¹²¹. De um ponto de vista do direito derivado da UE, o tratamento dado pelo legislador tem sido denso e reforçado nestas

¹¹⁶ (tradução nossa) ANUPAM CHANDER (2017), p. 1025.

¹¹⁷ MILENA ROUXINOL (2021), p. 243.

¹¹⁸ O seu sentido imediato, na esteira de GOMES CANOTILHO / VITAL MOREIRA (2007), p. 337 consiste “na proclamação da idêntica validade cívica de todos os cidadãos, independentemente da sua inserção económica, social, cultural e política.”

¹¹⁹ Trata-se, assim, de um dos princípios estruturantes do sistema constitucional. GOMES CANOTILHO / VITAL MOREIRA (2007), p. 336, JORGE MIRANDA / RUI MEDEIROS (2017), p. 164.

¹²⁰ A CDFUE foi proclamada a 7 de dezembro de 2000 e formalmente alterada a 12 de dezembro de 2007. Inicialmente, a CDFUE não dispunha de força vinculativa, o que se ficou a dever por duas razões, como salienta CATARINA SANTOS BOTELHO (2013), p. 322: reforçar o processo de integração e acalmar os ânimos sentidos pela forte oposição exercida por países como o Reino Unido e a Dinamarca. Sobre a natureza da CDFUE, tópico que aqui não nos apraz desenvolver, remetemos para SOFIA PAIS (2018), pp. 156 e CATARINA SANTOS BOTELHO (2013), pp. 323-326.

¹²¹ Este princípio basilar do sistema encontra, assim, consagração em todos os níveis de fontes do direito da união, seja ele direito originário ou derivado. *Vide* AHMET BILAL AYTEKIN (2023), p. 6.

matérias, podendo salientar-se a título de exemplo um conjunto diversificado de atos legislativos: Diretiva (UE) 2000/43/CE¹²², Diretiva (UE) 2006/54/CE¹²³ ou a Diretiva (UE) 2000/78/CE¹²⁴.

Releva, para o âmbito do nosso estudo, distinguir entre discriminação direta e indireta. A discriminação direta refere-se às situações em que o ato discriminatório ocorre pelo facto de o indivíduo pertencer a um determinado grupo ou deter certas características protegidas (origem étnica, género, religião, idade, entre outros), sendo objeto de um tratamento menos favorável¹²⁵. Estamos, como refere MARIANA CANOTILHO, diante situações que se caracterizam “no facto de determinada medida se fundar diretamente e sem justificação num critério interdito pela ordem jurídica”¹²⁶. A possibilidade de justificação dependerá, em grande medida, da existência de uma disposição desse tipo¹²⁷.

Nos EUA, adota-se a formulação de “*disparate treatment*”, prevista no capítulo VII do *Civil Rights Act of 1964*, referindo-se às situações em que um empregador trata o trabalhador de modo diferente por determinadas características protegidas (raça, género, religião ou nacionalidade de origem)¹²⁸. O *Supreme Court of Justice* tem exigido a demonstração de um motivo ou intenção discriminatória¹²⁹.

¹²² Diretiva (UE) 2000/43/CE do Conselho, de 29 de junho de 2000, que aplica o princípio da igualdade de tratamento entre pessoas, sem distinção de origem racial ou étnica.

¹²³ Diretiva (UE) 2006/54/CE do Conselho, de 5 de julho de 2006, relativa à aplicação do princípio da igualdade de oportunidades e igualdade de tratamento entre homens e mulheres em domínios ligados ao emprego e à atividade profissional.

¹²⁴ Diretiva (UE) 2000/78/CE do Conselho, de 27 de novembro de 2000, que estabelece um quadro geral de igualdade de tratamento no emprego e na atividade profissional.

¹²⁵ CATARINA SANTOS BOTELHO (2024), p. 3. MILENA ROUXINOL (2021), p. 245, SANDRA WATCHER / BRENT MITTELSTADT / CHRIS RUSSELL (2021), p. 7, AHMET BILAL AYTEKIN (2023), p. 7, FREDERIK J. ZUIDERVEEN BORGESIU (2020), p. 1576. Nas palavras de BRUNO MESTRE (2015), p. 393 “a discriminação direta é encarada de uma forma assaz negativa pelo Direito antidiscriminação.”

¹²⁶ MARIANA CANOTILHO (2011), p. 107.

¹²⁷ A título de exemplo, o art. 4.º, n.º 5, da Diretiva (UE) 2004/113/CE não exclui em absoluto diferenças de tratamento, conquanto o fornecimento de bens e a prestação de serviços exclusivamente ou prioritariamente aos membros de um dos sexos for justificado por um objetivo legítimo e os meios para atingir esse objetivo forem adequados e necessários. Trata-se de uma das situações em que a discriminação direta poderá ser justificada. *Vide* AHMET BILAL AYTEKIN (2023), p. 8 e, na doutrina portuguesa, BRUNO MESTRE (2015), pp. 393-395.

¹²⁸ CATARINA SANTOS BOTELHO (2024), p. 3, STEPAHNIE BORSTEIN (2018), p. 558, SOLON BAROCAS / ANDREW D. SELBST (2016), pp. 696 e ss.

¹²⁹ SANDRA FREEDMAN (2022), p. 259. *Vide* também JERAMIAS ADAMS-PRASSL / REUBEN BINNS / AISLINN KELLY-LYTH (2023), p. 148. Tal foi pela primeira vez articulado no caso *McDonnell Douglas Corp. v. Green*, em 1973, onde o *Supreme Court of Justice* defendeu não ser legal a conduta de uma entidade empregadora que não contrata alguém por motivos relacionados com raça, cor, género, religião ou nacionalidade de origem. Para mais desenvolvimentos, *cfr.* STEPAHNIE BORSTEIN (2018), pp. 558-559.

Por sua vez, a discriminação indireta poderá ocorrer sempre que haja a adoção de uma prática, critério ou disposição aparentemente neutra no plano formal, mas que, de um ponto de vista material, tem um impacto desproporcional em indivíduos de uma determinada categoria¹³⁰. O seu foco encontra-se mais vocacionado para grupos do que indivíduos, sendo possível justificar a prática caso seja necessária e proporcional ao interesse a seguir¹³¹. Tal encontrará um paralelismo com a doutrina do *disparate impact*, a qual tem uma afinidade com a discriminação indireta, referindo-se a todas as situações em que há uma prática formalmente neutra, mas suscetível de gerar um impacto díspar em relação a uma classe protegida, sendo o foco dirigido para o efeito adverso da medida¹³².

4.1 Problemas resultantes da aplicação do direito antidiscriminação

Poderemos, afinal, encontrar no direito antidiscriminação uma resposta adequada a tutelar as situações de discriminação algorítmica? Vejamos.

Em primeiro lugar, devemos começar por assinalar que a discriminação algorítmica é um problema recente, pelo que o atual estado de arte carecerá necessariamente de uma adaptação. Ao contrário das situações discriminatórias normais, o *output* de um algoritmo, em princípio, não terá um caráter individualizado, já que visa ser aplicável a uma maior escala¹³³, sendo suscetível de ser replicado em diversas situações. A discriminação direta, além do seu pendor individualizado¹³⁴, tem sido encarada como de difícil aplicação¹³⁵. Desde logo, teremos de conseguir identificar que o *output* discriminatório do algoritmo se ficou a dever ao facto de pertencer a um grupo protegido. Ora, tal é difícil pois ao contrário das decisões humanas individualizadas, o algoritmo tem em mente um âmbito de aplicação muito mais lato. A discriminação direta por si já é difícil de provar, mais ainda atendendo à complexidade de uma decisão algorítmica¹³⁶.

¹³⁰ Ibid, p. 107. Igualmente, MILENA ROUXINOL (2021), p. 245, FREDERIK J. ZUIDERVEEN BORGESIU (2020), p. 1577, CATARINA SANTOS BOTELHO (2024), p. 3, SANDRA WATCHER / BRENT MITTELSTADT / CHRIS RUSSELL (2021), p. 7.

¹³¹ Algo que resulta diretamente da noção de discriminação indireta das diretivas. Mencione-se, a título exemplificativo, o art. 2.º, n.º 2, al. b), da Diretiva (UE) 2000/43/CE do Conselho de 29 de junho de 2000.

¹³² SANDRA FREDMAN (2022), p. 259, ANDREW D. SELBST (2017), p. 161.

¹³³ AHMET BILAL AYKETIN (2023), p. 11.

¹³⁴ PHILIPP HACKER (2018), p. 1151.

¹³⁵ A título de exemplo, *vide* MILENA ROUXINOL (2021), p. 258, FREDERIK J. ZUIDERVEEN BORGESIU (2020), p. 1576, PHILIPP HACKER (2018), pp. 1151 e ss. e RAPHAËLE XENIDIS / LINDA SENDEN (2020), p. 29.

¹³⁶ IGNACIO N. COFONE (2019), p. 1431.

Não basta, como corretamente postula AIBA SORIANO ARNANZ, que o algoritmo discrimine explicitamente com base em características de um grupo protegido, devendo isso ser acessível ao tribunal¹³⁷. SOLON BAROCAS e ANDREW D. SELBST, embora focando na teoria do *disparate treatment*, consideram que tal poderá suceder-se se houver a decisão de aplicar um modelo preditivo que de antemão já se sabe que irá tratar membros de um grupo com uma dada característica de forma diferente, devido aos *inputs* nele colocados¹³⁸, o que pode ser feito aquando da criação do algoritmo. Ficarão, por isso, de fora as situações mais comuns relacionadas com a introdução de dados não representativos ou que reflitam discriminações sociais antecedentes¹³⁹.

Ainda assim, teria de ser identificado no caso que a discriminação teve por base uma das características protegidas previstas nas diretivas (raça, género ou religião, por exemplo). As diretivas assentam, por isso, num modelo fechado¹⁴⁰, pelo em muitos casos não são incluídos como *inputs* estas características, mas outras que fazem cair fora deste escopo. Parece-nos, na linha daquilo que tem seguido a doutrina maioritária, que é muito difícil equacionar situações de discriminação direta e que os problemas referidos dificilmente podem ser ultrapassados¹⁴¹.

Importa tecer breves comentários quanto à discriminação indireta. Esta tem sido apontada, dentro do direito antidiscriminação, como aquela que de algum modo se poderá melhor adequar às situações de discriminação algorítmica¹⁴². Se o algoritmo tiver como dado de *input* a exclusão de candidatos no acesso ao emprego de determinadas zonas (mediante o código postal), esta opção (ainda que aparentemente neutra) afetará indubitavelmente de forma desequilibrada os membros de determinadas categorias sociais¹⁴³, especialmente se

¹³⁷ AIBA SORIANO ARNANZ (2022), p. 153.

¹³⁸ SOLON BAROCAS / ANDREW D. SELBST (2016), p. 699.

¹³⁹ Denominadas por Philipp Hacker (2018), p. 1152 de “*accidental discrimination*.”

¹⁴⁰ Constituem uma lista fechada de motivos (“*closed list of grounds*”). Sobre o tópico, JANNEKE GERARDS / Frederik Zuiderveen Borgesius (2022), p. 24.

¹⁴¹ Assim, MILENA ROUXINOL (2021), p. 258, SOLON BAROCAS / ANDREW D. SELBST (2016), pp. 696-701. STEPHANIE BORSTEIN (2018), p. 535, AHMET BILAL AYTEKIN (2023), p. 14. Defendendo a aplicação do regime da discriminação direta temos JEREMIAS ADAMS-PRASSL / REUBEN BINNS / AISLINN KELLY-LYTH (2023), p. 158. Os autores dão como exemplo o caso das decisões automatizadas no crédito ao consumo, nomeadamente quando a entidade bancária, para aferir a solvabilidade do consumidor, tenha o seu algoritmo a identificar uma correlação entre o estado civil (casamento) e a capacidade de pagamento. Reiteram, ainda, que se for excluído do conceito o casamento civil tal poderá levar a que casais homossexuais tenham classificações mais baixas. Não nos parece, porém, que este seja o melhor exemplo para concluir que há a discriminação com base numa categoria desprotegida. Ao invés, parece-nos inserir-se na discriminação indireta.

¹⁴² Vide PHILIPP HACKER (2018) p. 1153, MILENA ROUXINOL (2021), p. 258, a título de exemplo.

¹⁴³ MILENA ROUXINOL (2021), p. 258.

determinadas zonas forem habitadas predominantemente por certos grupos. Alguns autores salientam que uma das vantagens de recorrer a este mecanismo prende-se com o facto de ser mais vocacionada para modificar a injustiça sistemática e promover uma mudança social¹⁴⁴.

No entanto, coloca-se uma questão no plano prático que nos parece da maior importância: como fazer prova desta situação? Não podemos olvidar que os algoritmos de aprendizagem automática constituem verdadeira impenetrável “*black box*”¹⁴⁵. Não só porque a sua compreensão implicará conhecimentos técnicos que a maioria não tem, assim como inerente à criação do algorítmico estão associados segredos de negócio que não podem, sem mais, ser revelados¹⁴⁶. Adicionalmente, PHILIPP HACKER lembra que várias técnicas de aprendizagem automática¹⁴⁷ não permitem a recuperação dos fatores relevantes para gerar o *output*¹⁴⁸.

Colocando estas dificuldades de parte, os *inputs* colocados nos algoritmos muitas vezes não têm por base categorias de grupo protegidos (ou pelo menos diretamente), apresentando um carácter neutral¹⁴⁹, mas ainda assim serão suscetíveis de produzir resultados desequilibrados¹⁵⁰. Vimos como a existência de uma base de dados não representativa poderá, indubitavelmente, afetar mais determinados membros de grupos sociais: aí, em bom rigor, a escolha do algoritmo não promove discriminação por assentar numa característica protegida, mas sim pela falta sua de representatividade. A chave residirá em conseguir provar e, para isso, torna-se necessário apreender o modo de funcionamento do algoritmo¹⁵¹.

Sendo possível estabelecer um caso de discriminação indireta, o ónus da prova recairá sobre a parte demandada¹⁵², a qual deverá invocar um motivo objetivo, razoável, e

¹⁴⁴ SANDRA WATCHER / BRENT MITTELSTADT / CHRIS RUSSELL (2021), p. 7. Comparativamente à discriminação direta que estará mais focada em cada situação em específico.

¹⁴⁵ Em especial sobre o tema SANDRA WATCHER / BRENT MITTELSTADT / CHRIS RUSSELL (2018), pp. 842-887.

¹⁴⁶ Como denota VÍTOR PALMELA FIDALGO (2021), p. 163, a generalidade das pessoas (padrão do consumidor médio) não terá conhecimentos suficientes de programação para entender a *ratio* que subjaz à decisão automatizada.

¹⁴⁷ Como *deep neural networks*.

¹⁴⁸ PHILIPP HACKER (2018), p. 1153.

¹⁴⁹ RAPHAËLE XENIDIS/ LINDA SENDEN (2020), p. 20.

¹⁵⁰ MILENA ROUXINOL (2021), p. 258.

¹⁵¹ RAPHAËLE XENIDIS/ LINDA SENDEN (2020), p. 21.

¹⁵² Vide AIBA SORIANO ARNANZ (2022), p. 159.

que passe o teste da proporcionalidade¹⁵³, não se aplicando a proibição¹⁵⁴. Seja como for, o direito antidiscriminação europeu poderá não se encontrar ainda vocacionado para estas questões. Desde logo, o âmbito de aplicação material das diretivas é restrito a três grupos distintos: emprego, acesso a bens ou serviços e sistema de segurança social, sendo que a discriminação algorítmica vai muito para além disso, o que é um entrave¹⁵⁵. Sempre se poderá convocar a CDFUE ou o TFUE. Destarte, tendo em consideração que a maior parte destas situações surgem em relações entre particulares, tal implicará a aceitação da eficácia horizontal. Se a questão não é controversa para o TFUE¹⁵⁶, para a CDFUE é menos líquida, embora aceitemos – por uma questão de coerência da tutela dos direitos fundamentais – a possibilidade da sua invocação na relação entre particulares¹⁵⁷.

Consideramos, por isso, que o direito antidiscriminação, no seu estado atual, não permite dar resposta a este tipo de situações. Ainda assim, reconhecemos mérito à solução apresentada por JANNEKE GERARDS e FREDERIK ZUIDERVEEN. Os autores defendem a adoção de um sistema que combine uma lista semifechada de motivos (isto é, embora à partida possa fazer uma enumeração taxativa das situações em que há discriminação, um olhar mais próximo revela que a lista é aberta)¹⁵⁸. Na sua ótica, tal seria suscetível de reforçar o valor simbólico do direito antidiscriminação¹⁵⁹, facilitar a demonstração em tribunal, sendo que o julgador não teria de subsumir a situação de discriminação algorítmica a uma das concretas provisões, como também permitiria a construção jurisprudencial de novos motivos de

¹⁵³ A medida deverá ser: adequada (as medidas restritivas de direitos, liberdades e garantias devem revelar-se como um meio para a prossecução dos fins pretendidos), necessária (as medidas têm de ser exigidas para alcançar os fins pretendidos, não se dispondo de meios menos restritivos para almejar o mesmo objetivo) e proporcional *stricto sensu* (não poderão adotar-se medidas excessivas, desproporcionadas para alcançar os fins pretendidos).

¹⁵⁴ FREDERIK J. ZUIDERVEEN BORGESIU (2020), p. 1577.

¹⁵⁵ AHMET BILAL AYTEKIN (2023), p. 14. Ficam de fora, como salienta AIBA SORIANO ARNANZ (2022), p. 144, situações em que a discriminação tenha por base a origem social, propriedade ou outro fator que possa indicar a condição socioeconómica.

¹⁵⁶ Em *Defrenne c. Sabena*, C-43/75, 8 de abril de 1976, parágrafos 23 e ss., o TJ reconheceu a possibilidade de ser invocado o art. 157.º do TFUE numa relação entre particulares, conferindo eficácia horizontal ao TFUE.

¹⁵⁷ A defesa da eficácia vertical da CDFUE poderá assentar no elemento literal do art. 51.º, segundo o qual o diploma destina-se às instituições, órgãos e organismos da EU. Isto teria a vantagem de facilitar a sua proclamação. Destarte, na linha do que assinala SOFIA PAIS (2018), pp. 161-162, consideramos que não faria sentido dar um tratamento diferente ao mesmo direito (basta pensar no art. 157.º do TFUE e 23.º da CDFUE o qual tem um teor idêntico) somente por se encontrar num diploma diverso. Ademais, a distinção entre relações jurídicas particulares e públicas tem-se esvanecido e, mais do que nunca, assegurar a proteção dos direitos fundamentais nestas situações é essencial. No caso *Egenberger*, C-414/16, 17 de abril de 2018, o TJ, numa questão prejudicial relacionada com a discriminação em função da religião no acesso ao emprego, reconheceu eficácia horizontal às disposições da CDFUE, mormente ao art. 21.º, n.º 1.

¹⁵⁸ JANNEKE GERARDS / FREDERIK ZUIDERVEEN (2022), p. 25. Sobre o tema, *cfr* CATARINA SANTOS BOTELHO (2024), p. 4.

¹⁵⁹ JANNEKE GERARDS / FREDERIK ZUIDERVEEN (2022), p. 51.

proteção mais adaptados a estes casos. Defendem, ainda, a adoção de uma lista semifechada de exceções¹⁶⁰, pois isso possibilitaria que, para todas as formas de discriminação resultantes da IA, os detentores do sistema fossem capazes de demonstrar motivos ponderosos da sua utilização.

Trata-se de uma perspectiva que tem o mérito inegável de chamar à colação para a necessidade de o direito antidiscriminação atual não ser adequado para regular estas situações. Não obstante, na era do constitucionalismo digital, urge refletir sobre mecanismos mais aptos a regular este problema.

¹⁶⁰ Ibid, p. 53.

Parte II

5. A discriminação algorítmica na era do constitucionalismo digital

O constitucionalismo digital surge da necessidade de se adaptar os valores constitucionais aos desafios e necessidades da sociedade digital¹⁶¹. Assim, o maior desafio passa por criar um sistema capaz de salvaguardar os instrumentos normativos necessários aptos a permitir a evolução da IA, mas também capazes de defender os direitos fundamentais dos cidadãos e os valores democráticos¹⁶².

Embora, após as revoluções oitocentistas as constituições tenham sido criadas com o objetivo de limitar o poder estatal e proteger os indivíduos de exercícios de poder arbitrários, a sociedade algorítmica mudou esse paradigma. Como afirmam ORESTE POLLICINO e GIOVANNI DE GREGORIO:

*O poder público já não é a única fonte de preocupação relativamente à proteção dos direitos fundamentais e da democracia, onde as fronteiras jurisdicionais estão em mutação, e onde as doutrinas e os procedimentos desenvolvidos na era pré-cibernética não captam as violações dos direitos num período relevante de tempo*¹⁶³.

Razão assiste a GIOVANNI DE GREGORIO, quando sustenta que a próxima fase do constitucionalismo digital passará, necessariamente, por reagir às ameaças à dignidade da pessoa humana na sociedade algorítmica¹⁶⁴. No nosso entendimento, isso pode ser feito de duas formas: (i) através da a regulação normativa da IA como mecanismo preventivo e (ii) mediante o reforço dos direitos fundamentais dos cidadãos na era digital, adaptando-se às particularidades da IA. Isto é o que iremos explorar na Parte II da nossa dissertação.

¹⁶¹ GIOVANNI DE GREGORIO (2021), pp. 41-70 identifica quatro fases distintas para a evolução do constitucionalismo digital: (i) o liberalismo digital, (ii) o ativismo judicial, (iii) o constitucionalismo digital e (iv) fortalecimento dos direitos.

¹⁶² EDOARDO CELESTE (2023), p. 215

¹⁶³ (tradução nossa) ORESTE POLLICINO / GIOVANNI DE GREGORIO (2022), p. 6.

¹⁶⁴ GIOVANNI DE GREGORIO (2021), p. 70.

6. A Regulação como mecanismo preventivo para combater a discriminação algorítmica

6.1 Regulamento da IA

A Proposta de Regulamento da IA (*AI Act*)¹⁶⁵ constitui o primeiro instrumento normativo de regulamentação geral e uniforme para a IA, o que por si só já é um marco histórico na luta pela maior humanização do processo. Após um debate intenso e difícil, chegou-se a um consenso político, tendo sido adotada a Proposta Final do Regulamento.

Cumprir notar que a discriminação resultante de sistemas de IA é uma preocupação real, motivo pelo qual no documento podem ser encontradas menções e medidas para mitigar o seu impacto¹⁶⁶.

No que diz respeito ao seu âmbito subjetivo, este aplicar-se-á (i) aos fornecedores que coloquem no mercado ou em serviço sistemas de IA e modelos de IA de finalidade geral no território da União, independentemente de estarem estabelecidos ou localizados na União ou em país terceiro, (ii) aos responsáveis pela implementação de sistemas de IA que tenham o seu local de estabelecimento ou que estejam localizados na União, (iii) aos fornecedores e responsáveis pela implementação de sistemas de IA que tenham o seu local de estabelecimento ou estejam localizados num país terceiro, se o resultado produzido pelo sistema for utilizado na União¹⁶⁷, (iv) aos importadores e distribuidores de sistemas de IA¹⁶⁸, (v) aos fabricantes de produtos que coloquem no mercado ou coloquem em serviço um sistemas de IA¹⁶⁹, (vi) mandatários dos fornecedores que não estejam estabelecidos na União¹⁷⁰, (vii) e às pessoas afetadas localizadas na união¹⁷¹.

A proposta adota uma sistematização baseada nos diversos níveis de risco¹⁷², variando as obrigações daí decorrentes consoante o tipo de sistema em causa.

¹⁶⁵ À data em que redigimos a nossa Dissertação, o Regulamento ainda não foi publicado. Iremos, por isso, basear a nossa análise na Proposta Final de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (*AI Act*).

¹⁶⁶ Considerandos 45, 48, 56, 58, 60.

¹⁶⁷ Art. 2.º, n.º 1, als a), b) e c).

¹⁶⁸ Art. 2.º, n.º 1, al. d).

¹⁶⁹ Art. 2.º, n.º 1, al. e).

¹⁷⁰ Art. 2.º, n.º 1, al. f).

¹⁷¹ Art. 2.º, n.º 1, al. g).

¹⁷² Vide MAFALDA MIRANDA BARBOSA (2023), p. 48, JOSÉ RICARDO MARCONDES RAMOS (2023), pp. 34 e ss e JOSÉ JOÃO ABRANTES / ISABEL VALENTE DIAS (2023), p. 48.

A adoção deste critério apresenta a inegável vantagem de permitir adaptar o tipo de regulação ao sistema de IA em causa, tomando em consideração as suas funcionalidades, os dados fornecidos e o destino do sistema¹⁷³, embora reconheçamos o risco de rápida desatualização¹⁷⁴.

Desde logo, são estabelecidas as práticas absolutamente proibidas de IA, por serem aqueles cuja utilização colocaria em causa os direitos fundamentais, devido ao seu carácter malicioso, manipulador e perigoso¹⁷⁵. A título de exemplo, proíbe-se a colocação no mercado, a entrada em serviço ou utilização de sistemas de IA que empreguem técnicas subliminares que contornem a consciência de uma pessoa, ou técnicas manifestamente manipuladoras ou enganadoras, com o objetivo de distorcer o comportamento de uma pessoa ou grupo de pessoas¹⁷⁶.

São igualmente interditos os sistemas de IA que exploram vulnerabilidades de uma pessoa ou grupo específico devido à sua idade, incapacidade e situação socioeconómica específica, com o objetivo ou efeito de distorcer substancialmente o comportamento dessa pessoa ou de uma pessoa pertencente a esse grupo, de uma forma que cause ou seja razoavelmente suscetível de causar danos significativos¹⁷⁷. Serão igualmente proibidos aqueles que visam a avaliação ou de classificação social de pessoas singulares ou de grupos de pessoas durante um certo período com base no seu comportamento social ou em características de personalidade ou pessoais, conhecidas, inferidas ou previsíveis, em que a classificação social conduza a uma das seguintes situações ou a ambas.¹⁷⁸

Igualmente, não serão admissíveis os sistemas de categorização biométrica que categorizem individualmente pessoas singulares com base nos seus dados biométricos para deduzir ou inferir a sua raça, opiniões políticas, filiação sindical, convicções religiosas/filosóficas, vida sexual ou orientação sexual¹⁷⁹. Uma das razões que levou à sua proibição foi, precisamente, o facto de lesar a dignidade da pessoa humana e o princípio da igualdade¹⁸⁰.

¹⁷³ MARTA BOURA (2023), p. 116.

¹⁷⁴ MAFALDA MIRANDA BARBOSA (2023), p. 51 e DIOGO QUEIROZ DE ANDRADE (2023), p. 42.

¹⁷⁵ MARTA BOURA (2023), p. 116.

¹⁷⁶ Art. 5.º, n.º 1, al. a).

¹⁷⁷ Art. 5.º, n.º 1, al. b).

¹⁷⁸ Art. 5.º, n.º 1, al. c).

¹⁷⁹ Art. 5.º, n.º 1, al. g).

¹⁸⁰ Considerando 17. JOSÉ RICARDO MARCONDES RAMOS (2023), p. 37.

O Regulamento de IA regula os modelos de IA de finalidade geral, isto é, os modelos de IA, incluindo-se todos aqueles treinados com uma grande quantidade de dados em escala, que se caracterizem pela generalidade significativa e sejam aptos para executar um conjunto de tarefas distintas, independentemente do modo como são colocados no mercado da UE¹⁸¹.

Estes modelos serão considerados como tendo risco sistémico, caso apresentem capacidades de grande impacto ou exista uma decisão da Comissão Europeia nesse sentido¹⁸².

Consideramos relevante para o nosso estudo a sistematização que é dada aos sistemas de risco elevado¹⁸³. Considerar-se-ão sistemas de risco elevado aqueles que se destinem a ser utilizados como componente de segurança de um produto ou o sistema de IA e, sejam eles próprios, produtos abrangidos pela legislação de harmonização da UE mencionada no anexo I, devendo este ser sujeito a uma avaliação de conformidade por terceiros com vista à colocação no mercado ou em serviço¹⁸⁴.

Adicionalmente, os sistemas de IA referidos no anexo III serão considerados de risco elevado, caso representem um risco considerável para a saúde, a segurança ou os direitos fundamentais de pessoas singulares¹⁸⁵ e influenciem de forma significativa o resultado da tomada de decisões¹⁸⁶.

À luz do Anexo III, podemos constatar que grande parte dos exemplos que mencionamos na nossa Dissertação de discriminação algorítmica consubstanciarão sistemas de alto risco, à luz do Regulamento IA. A título de exemplo, estão aqui incluídos os sistemas de IA concebidos para serem usados no âmbito laboral para recrutamento, avaliação de candidatos, entrevistas ou testes¹⁸⁷ (basta pensar no caso da Amazon), sistemas concebidos para serem utilizados por autoridades policiais, ou em seu nome, ou por agências, serviços ou organismos da União em apoio das autoridades policiais (COMPAS), sistemas

¹⁸¹ Art. 3.º, n.º 63.

¹⁸² Art. 51.º, n.º 1, als. a) e b). Nos termos do art. 51, n.º 2, do Regulamento IA, presumir-se-á que o sistema tem capacidades de grande impacto, sempre que a quantidade acumulada de computação utilizada para o seu treino, medida em operações de vírgula flutuante, seja superior a 10^{25} .

¹⁸³ Cujas explicações são feitas no Título III.

¹⁸⁴ Art. 6.º, n.º 1, als. a) e b).

¹⁸⁵ Art. 6.º, n.º 2 e art. 6.º, n.º 3.

¹⁸⁶ Nomeadamente se estiver prevista uma das situações do art. 6.º, n.º 3.

¹⁸⁷ Encontrando-se aqui incluídos os sistemas usados para recrutamento ou seleção de pessoas, filtrar listas de candidatos e avaliá-los e, ainda, sistemas que sejam usados para adotar decisões no âmbito laboral (a título de exemplo, os despedimentos ou promoções).

concebidos para aceder a serviços privados ou serviços públicos essenciais e respetivos benefícios¹⁸⁸, entre outros.

Estes sistemas encontram-se adstritos ao cumprimento de requisitos obrigatórios, os quais abrangem o processo de programação e utilização do algoritmo.

Desde logo, prevê-se a criação de um sistema de gestão de riscos¹⁸⁹. Este deverá ser criado, implementado, documentado e mantido durante todo o ciclo de vida do sistema de IA, podendo ser este integrado nos sistemas já existentes. Trata-se, assim, de um processo interativo contínuo e planeado, requerendo-se a sua revisão e atualização regularmente, de maneira a identificar e analisar os diversos riscos¹⁹⁰.

Este deverá compreender uma fase de eliminação dos riscos (quando possível) ou a sua atenuação¹⁹¹, alcançando-se um equilíbrio adequado na aplicação das medidas destinadas a cumprir este objetivo¹⁹². Esta obrigação parece-nos da maior relevância, já que produzindo o algoritmo um resultado discriminatório, deverá ser salvaguardado um cuidadoso acompanhamento do ciclo de vida da IA e, quando possível, eliminação ou mitigação da situação discriminatória.

Igualmente relevante é a obrigação de dados e governação de dados¹⁹³, prevendo-se que os dados de *input* deverão ser desenvolvidos com base em conjuntos de dados de treino, validação e teste que cumpram os requisitos de qualidade¹⁹⁴. Os dados que alimentam o algoritmo são uma das razões que levam à discriminação, seja por não serem representativos ou por traduzirem eles próprios discriminações sociais. Este dever pressupõe que sejam realizados exames aos dados para detetar possíveis enviesamentos¹⁹⁵, especialmente quando estes possam influenciar os dados de entrada para operações futuras, bem como a adoção de medidas apropriadas para detetar, prevenir e atenuá-los¹⁹⁶.

A proposta prevê a possibilidade de, excepcionalmente, serem tratados categorias especiais de dados nos termos do art. 9.º, n.º 1 do RGPD¹⁹⁷ sempre que tal seja estritamente

¹⁸⁸ Dentro dos quais se incluem os sistemas usados pelas autoridades públicas para aferir a elegibilidade de determinada pessoa para gozar de benefícios/direitos sociais, os sistemas para avaliar a capacidade de endividamento ou estabelecer a classificação de crédito, entre outros.

¹⁸⁹ Art. 9.º. Igualmente sobre o tema, MICHAEL VEALE / FREDERIK ZUIDERVEEN (2021), p. 103.

¹⁹⁰ O art. 9.º, n.º 2 elenca os respetivos passos que devem ser adotados.

¹⁹¹ Art. 9.º, n.º 5, al. a).

¹⁹² Art. 9.º, n.º 4.

¹⁹³ Art. 10.º.

¹⁹⁴ Previstos no Art. 10.º, n.º 2.

¹⁹⁵ Art. 10.º, n.º 2, al. f).

¹⁹⁶ Art. 10.º, n.º 2, al. g).

¹⁹⁷ Assim como do art. 10º da Diretiva (UE) 2016/680 e 10.º, n.º 1, do Regulamento (UE) 2018/1725.

necessário para assegurar a detecção e a correção de enviesamentos negativos¹⁹⁸. Note-se que esta possibilidade não é criada arbitrariamente, já que deverá ser demonstrado que: (i) a detecção e correção da discriminação não pode ser almejada de outra forma, (ii) são consagradas limitações técnicas à reutilização desses dados pessoais, criando-se as respetivas medidas de segurança (incluindo a pseudonimização), (iii) os dados pessoais não são transmitidos ou cedidos por entidades terceiras, devendo igualmente ser apagados após a sua utilização e (iv) deverá ser apresentada uma justificação sobre a necessidade deste mecanismo para detetar, corrigir e eliminar a discriminação, não podendo ser almejado através de outro modo¹⁹⁹.

Em nosso entendimento, esta é uma boa solução. Não se seguiu (e muito bem) o entendimento do “*fairness through blindness*”, de acordo com a qual a melhor maneira para mitigar a discriminação algorítmica passaria por eliminar do algoritmo características protegidas (raça, género, sexo, etc). Tal possibilidade é ingénua²⁰⁰ e pouco efetiva, já que os algoritmos possuem uma capacidade inesgotável para identificação de padrões e algumas variáveis poderão, elas próprias, revelar atributos sensíveis²⁰¹, agravando-se a possibilidade de discriminação²⁰².

Aqui chegados, podemos apontar dois caminhos. Em primeiro lugar, a via do *group fairness*, avaliando-se a justiça do *output* do algoritmo mediante a comparação entre dois grupos distintos e asseverando-se se o resultado é proporcional à representação desse grupo na população geral ou, ainda, fazendo-se uma comparação entre os falsos positivos e os falsos negativos.

Poderíamos, igualmente, adotar uma estratégia de *individual fairness* (as equidades individuais medem as disparidades de tratamento a nível individual para indivíduos com características semelhantes²⁰³). Nesta última situação, como bem realça ALICE XIANG, não há garantias que o decisor use essas características protegidas de uma forma que permita almejar resultados mais justos²⁰⁴. Cremos que, do ponto de vista legal, importa mais a

¹⁹⁸ Art. 10.º, n.º 3.

¹⁹⁹ Art. 10.º, n.º 5.

²⁰⁰ JOSHUA A. KROLL *et. al* (2017), p. 685.

²⁰¹ *Ibid*, pp. 685 e ss, JASON R. BENT (2020), pp. 814-816 e SOLON BAROCAS / ANDREW D. SELBST (2016), p. 691 e MICHAEL VEALE / REUBEN BINNS (2017), p. 4 e MARIANA ALVES TEIXEIRA (2023), pp. 96 e ss.

²⁰² CATARINA SANTOS BOTELHO (2024), p. 11.

²⁰³ *Cfr.* JASON R. BENT (2020), p. 819 e ALICE XIANG (2021), p. 716 e ss.

²⁰⁴ *Ibid*, p. 723.

consagração desta possibilidade do que o método propriamente dito, pois isso levar-nos-á para o campo da programação do algoritmo.

O essencial será assegurar que, a par desta utilização, sejam salvaguardadas as medidas técnicas e organizativas necessárias para mitigar qualquer risco que daqui possa advir ao nível da segurança do tratamento dos dados. Particularmente relevante, na nossa ótica, é a obrigatoriedade da Avaliação de Impacto ao nível dos Direitos Fundamentais²⁰⁵, antes do sistema de IA entrar em vigor, à semelhança do que já se prevê no RGPD²⁰⁶. A sua aplicação permitirá mitigar os riscos *ex ante* dos diversos sistemas, detalhando os perigos que daí possam advir, sobretudo para a proteção dos direitos fundamentais²⁰⁷. Além disso, reforçam-se as obrigações de transparência²⁰⁸, visando-se com isso assegurar que todos os meios técnicos disponíveis são utilizados para assegurar o cumprimento das obrigações que incumbem ao fornecedor e responsável pela implantação.

A UE, uma vez mais, assume-se como pioneira ao criar um modelo de regulamentação para a IA. Permite-se, com isto, mitigar o vazio jurídico existente. Cremos que o primeiro passo para mitigar a discriminação algorítmica passará pela regulação da IA.

6.2 O RGPD

O RGPD²⁰⁹, tendo como objetivo primordial a livre circulação dos dados pessoais e a proteção dos direitos fundamentais²¹⁰, não tutela diretamente o problema da discriminação algorítmica, existindo uma singela menção no Considerando 71. Ainda assim, o seu papel não deve ser menosprezado na temática em questão²¹¹.

Isto dito, as obrigações a que se encontram adstritos os responsáveis pelo tratamento/subcontratantes são significativas, devendo a sua atuação pautar-se sempre pelos princípios basilares do RGPD, mormente: licitude, lealdade e transparência, limitação das finalidades, minimização dos dados, exatidão, limitação da conservação, integridade e confidencialidade e responsabilidade²¹².

²⁰⁵ Art. 27.º da Proposta de Regulamento IA.

²⁰⁶ *Vide* art. 35.º do RGPD. Uma das situações de obrigatoriedade é precisamente quando estejam em causa tratamentos em larga escala os dados pessoais previstos no artigo 9.º ou no artigo 10.º do RGPD.

²⁰⁷ Art. 27.º, n.º 1, als. (a)-(f) da Proposta de Regulamento IA.

²⁰⁸ Art. 13.º da Proposta de Regulamento IA.

²⁰⁹ Regulamento (UE) n.º 679/2016, de 27 de abril. Substituiu a Diretiva 95/46/CE.

²¹⁰ Art. 1.º, n.º 2 e 3 do RGPD.

²¹¹ PHILIPP HACKER (2018), p. 1172

²¹² Art. 5.º do RGPD.

Com este desiderato, o art. 22.º aplica-se às decisões individuais automatizadas, incluindo definição de perfis. O preceito deverá ser configurado, em nosso entendimento, como uma proibição. Bem sabemos que há quem na doutrina defenda existir um verdadeiro direito do titular dos dados, atendendo à letra da norma e à sua inserção sistemática no capítulo III²¹³. Não consideramos esses argumentos procedentes, já que não fará sentido o titular dos dados opor-se *ex ante* à decisão automatizada para depois objetar *ex post* a essa mesma decisão²¹⁴. Só faz sentido haver exceções no n.º 2 se a regra geral for a de uma proibição, como esta é a única maneira de proteger *ab initio* o titular dos dados dos efeitos deste tipo de tratamento²¹⁵, não ficando o seu conteúdo dependente do exercício do direito²¹⁶. Esta é a posição adotada pelo GT29²¹⁷, e, mais recentemente, pelo TJ no caso *Shufa*²¹⁸.

De seguida, apresentaremos três caminhos que, no âmbito do RGPD, poderão fortalecer a mitigação da discriminação algorítmica.

a) A consideração das inferências

O RGPD apenas poderá ser aplicável ao tratamento de dados pessoais. O conceito de dado pessoal²¹⁹ pressupõe que esteja em causa uma informação relativa a uma pessoa singular identificada ou identificável²²⁰. Tendo em consideração o contexto da decisão algorítmica, em muitas situações não será possível afirmar que os dados são identificados ou identificáveis em relação a alguém em concreto, o que, seguindo uma interpretação estritamente literal, levar-nos-ia a excluir os dados inferidos (dados que são calculados a partir de outros dados).

²¹³ Assim, GRAÇA CANTO MONIZ (2023), p. 199, LEE A. BYGRAVE (2020), p. 531, FILIPA URBANO CALVÃO (2023), p. 82 e ANA ALVES LEAL (2017), p. 152.

²¹⁴ DIOGO MORGADO REBELO (2023), pp. 284-288.

²¹⁵ INÊS DA SILVA COSTA (2021), p. 61, FLORENT THOUVENIN / ALFRED FRUH / SIMON HENSELER (2022), pp. 193-198.

²¹⁶ MARGOT E. KAMINSKI (2019), p. 197.

²¹⁷ GT29 (2017), pp. 25 e ss.

²¹⁸ *Schufa Holding AG* (C-634/21), par. 52.

²¹⁹ Art. 4.º, n.º 1, do RGPD.

²²⁰ Para maior aprofundamento do conceito de dado pessoal, *vide* A. BARRETO MENEZES CORDEIRO (2018), pp. 297 e GT29 (2007), pp. 1-28. Como afirma FILIPA URBANO CALVÃO (2018), p. 15, “tudo hoje depende das tecnologias de comunicação e, nessa medida, mesmo o que fazemos no mundo físico tem expressão no mundo virtual.”

Estes têm, por isso, uma identidade “fluida”²²¹. Nada impede que, num momento ulterior, essas inferências venham permitir identificar um indivíduo²²². Nesta matéria, se há quem na doutrina evidencie que um conceito amplo de dados pessoais faz com que o RGPD se torne na “lei do tudo”²²³, o GT29 reconhece que os dados inferidos constituem também eles dados pessoais²²⁴, não obstante a jurisprudência restrita do TJ em *YS, M, S*²²⁵ (onde se concluiu que uma análise jurídica que figure numa minuta não constitui em si mesma um dado pessoal)²²⁶. Todavia, em *Novak* o TJ afirmou exatamente o contrário, embora tenha uma limitação ao exercício de direitos nestas situações²²⁷. Mesmo que o GT29 reconheça que os dados inferidos são dados pessoais, as suas *guidelines* são meramente orientadoras, contrariamente à jurisprudência do TJ, o que faz com que a situação seja ambígua, confusa²²⁸ e limitada²²⁹. Por isso, poderá o RGPD ficar aquém caso se excluam os dados inferidos do conceito de dado pessoal, desprotegendo-se inadvertidamente o titular dos dados.

b) Uma nova exceção para o tratamento de dados sensíveis

Conforme tivemos a oportunidade de desenvolver, a proposta de regulamento IA vai no sentido de permitir a utilização de dados sensíveis para mitigar a discriminação algorítmica.

O RGPD, no art. 9.º, proíbe o tratamento de categorias especiais de dados pessoais, salvo se verificados uma das exceções do art. 9.º, n.º 2. Na nossa ótica, consideramos mais acertado defender, na linha do EDPB, a necessidade de se demonstrar a observância de um dos fundamentos do 9.º, n.º 2, bem como do art. 6.º²³⁰, até porque a solução contrária não se compreenderia tendo em conta o reforço da proteção do titular dos dados nestas situações, não se seguindo a posição daqueles que defendem a natureza especial do regime face ao geral²³¹.

²²¹ MARIANA ALVES TEIXEIRA (2023), p. 97.

²²² SANDRA WATCHER / BRENT MITTELSTADT (2019), pp. 538 e ss e ALLAN E. HOLDER (2020), p. 1343.

²²³ LEE A. BYGRAVE / LUCA TOSONI (2020), p. 114.

²²⁴ GT29 (2018), p. 8.

²²⁵ *YS, M, S* (C-141/12), 17 de julho de 2014.

²²⁶ *Ibid*, par. 39.

²²⁷ *Novak*, C-434/16, 20 de dezembro de 2017.

²²⁸ ANNA E. HOLDER (2020), p. 1364.

²²⁹ SANDRA WATCHER / BRENT MITTELSTADT (2019), p. 537.

²³⁰ EDPB (2020), p. 17. LUDMILA GEORGIEVA / CHRISTOPHER KUNER (2020), p. 376.

²³¹ A. BARRETO MENEZES CORDEIRO (2020), p. 238.

Consideramos que seria interessante equacionar a possibilidade de alargar o leque de categorias de dados especiais sensíveis, nomeadamente no que à naturalidade diz respeito. Este dado (em si mesmo) pode parecer insignificante – pelo menos para os nacionais, mas relativamente às pessoas oriundas de outro país, tal poderá ser fonte de problema e levar ao preconceito.

Além disso, o que aqui propomos passa pela criação de uma nova exceção no art. 9.º, n.º 2 do RGPD que permita o tratamento destes dados para mitigar a discriminação existente no algoritmo, em linha com a regulamentação da IA. Tal, na nossa ótica, apresenta inegáveis vantagens. Desde logo, permite adotar uma postura ativa no sentido de mitigar a discriminação algorítmica, já que dificilmente consideramos adequada *in casu* os fundamentos do preceito²³². Seria contraditório o Regulamento da IA prever esta possibilidade e o RGPD negá-la, sendo que o uso de dados adquire um carácter simbólico²³³.

Sabemos que esta solução não é consensual e que há quem diga que tal possibilidade poderá contribuir para usar esta categoria de dados para outros fins e, pior, incrementar a possibilidade de discriminação²³⁴.

Embora compreendendo as preocupações, cremos que o tratamento de dados sensíveis poderá, em última instância, resultar sempre em maior discriminação. Caso fosse a intenção do legislador europeu evitar o seu tratamento, ter-se-ia procedido à sua proibição total. Aliás, no contexto das decisões automatizadas, o art 22.º, n.º 4, do RGPD prevê que nos casos do art. 9.º, n.º 2, do RGPD as decisões automatizadas poderão basear-se nestes dados, conquanto sejam salvaguardados os direitos, liberdades e interesses do titular.

A introdução de um fundamento para combate à discriminação algorítmica terá, claro está, de se limitar ao estritamente necessário²³⁵ e não dispensaria a adoção reforçada de mecanismos de proteção (anonimização, pseudonimização, encriptação, por exemplo). Claro que isto dependerá da evolução da ciência, pressupondo-se que os sistemas que venham a ser desenvolvidos possam realizar auditorias eficazes ao sistema de IA, de maneira a que seja possível mitigar a discriminação. Na nossa visão, a adotar-se tal exceção, seria essencial limitá-la a determinadas entidades (quando estejam em causa os sistemas de alto risco, à semelhança do Regulamento IA, para aí se remetendo). A par disto, sempre se poderia

²³² MARVIN VAN BEKKUM / FREDERIK ZUIDERVEEN BORGESIU (2022), p. 9.

²³³ Ibid, p. 9.

²³⁴ AGATHE BALAYN / SEDA GÜRSES (2021), pp. 94-95.

²³⁵ Algo a ser demonstrado pelo responsável pelo tratamento.

equacionar a possibilidade de dar a uma entidade terceira, independente e qualificada, a auditoria destes sistemas, embora esta seja uma medida que só após a consagração da exceção fará sentido equacionar.

c) Clarificar o nível de intervenção humana

A decisão algorítmica, como assinala CATARINA SANTOS BOTELHO, não é melhor nem pior do que a decisão humana: é apenas diferente²³⁶. Temos ciente que, em algum momento, a intervenção humana será necessária. Nesta matéria, e considerando a evolução realizada pela proposta de regulamento de IA, o RGPD continua um imbróglio no que à intervenção humana diz respeito, o que urge mudar. BRYCE GOODMAN e SETH FLAXMAN advogam que as incoerências legais sentidas demonstram “a necessidade de [adotar] explicações inteligíveis para o ser humano sobre a tomada de decisões algorítmicas”²³⁷.

Ex ante, começamos logo por considerar que, não obstante o art. 22.º do RGPD falar “em decisão exclusivamente com base no tratamento automatizado”, tal deverá ser interpretado de modo restritivo. À primeira vista, poder-se-ia advogar que qualquer tipo de intervenção humana (mesmo que formal) afastaria a aplicação do preceito²³⁸, algo que consideraríamos excessivo e permitiria os responsáveis pelo tratamento encontrar um escudo contra a proibição. Ao invés, tal como defendido por BENEDIKT BUCHNER²³⁹, para ser afastada a sua aplicação será necessária uma intervenção material/substancial (isto é, que a intervenção humana tenha uma influência efetiva no processo da decisão, não se limitando a confirmar aquilo que foi decidido pelo algoritmo). Esta é também a posição do GT29, ao defender-se que para o afastamento do 22.º será necessária uma intervenção humana substancial²⁴⁰.

Isto leva-nos inevitavelmente à querela do direito à explicação. Defendido pela primeira vez na doutrina por BRYCE GOODMAN e SETH FLAXMAN, ambos propugnaram pela existência de um direito à explicação (baseando-se no considerando 71) defendendo uma aplicação adequada de como os dados recolhidos articulam-se com as previsões

²³⁶ CATARINA SANTOS BOTELHO (2024), p. 18.

²³⁷ (tradução nossa) BRYCE GOODMAN / SETH FLAXMAN (2017), p. 55.

²³⁸ SANDRA WATCHER / BRENT MITTELSTADT / LUCIANO FLORIDI (2017), p. 92.

²³⁹ BENEDIKT BUCHNER (2018), p. 512.

²⁴⁰ GT29 (2017), p. 10.

realizadas²⁴¹. As vozes mais críticas invocam a sua inexecuibilidade e ausência de base legal suficiente para a sua existência, podendo os titulares dos dados, se assim o entenderem, contestar a decisão²⁴². Ao passo que autores como ANDREW SELBST E JULIA POWLES defendem que da interpretação dos arts. 13.º, n.º 2, al. f), 14.º, n.º 2, al. g) e 15.º al. h) resulta a existência de um direito à explicação *ex post* da decisão, devendo esta ser acompanhada de um mínimo de funcionalidade da informação²⁴³.

O direito à explicação *ex post* configura em si uma dificuldade praticamente insanável: o titular dos dados afetado pela decisão algorítmica não terá o conhecimento suficiente para apreender o funcionamento dos sistemas de ML (será necessário adotar o padrão do Homem Médio?), como também não se encontra claro o alcance de informação que deve ser fornecido. Bastará dizer que houve uma decisão automatizada e o seu resultado? Estará adstrito a explicar o funcionamento do sistema? Parece-nos duvidoso. Perante estas dificuldades, parece-nos correta a posição de FILIPA URBANO CALVÃO, segundo as ferramentas do RGPD terão um efeito inútil caso não sejam acompanhadas de uma inversão do ónus da prova, cabendo às diferentes organizações que utilizam o algoritmo encontrar formas de o explicar²⁴⁴.

Além disso, o art 22.º, n.º 3, do RGPD prevê o direito a obter uma intervenção humana, mas não esclarece o alcance desta intervenção. Na linha de GUILLERMO LAZCOZ E PAUL DE HERT, a intervenção funciona como um pré-requisito para o exercício do direito de contestação²⁴⁵. Com esta intervenção humana, quem aí intervenha deverá explicar aquela que foi a decisão, realizando uma avaliação exaustiva dos dados que sejam pertinentes²⁴⁶. Logicamente, sempre se poderá dizer que isto é um passo importante para o titular dos dados poder estar ciente daquela que foi a decisão, mas será que este é um passo essencial para este contestar a decisão? O RGPD não esclarece, como também não esclarece um fator adicional: o titular dos dados irá contestar a decisão contra quem? Poderá exigir que a sua decisão seja num segundo passo tomada por um ser humano? Poderá exigir a mudança de algoritmo decisório?

²⁴¹ BRYCE GOODMAN / SETH FLAXMAN (2017), p. 55.

²⁴² SANDRA WATCHER / BRENT MITTELSTADT / LUCIANO FLORIDI (2017), pp. 80-81.

²⁴³ ANDREW SELBST / JULIA POWLES (2017), p. 236. Igualmente defendendo GIOVANNI DE GREGORIO / ORESTE POLLICINO (2022), p. 21.

²⁴⁴ FILIPA URBANO CALVÃO (2023), p. 85.

²⁴⁵ GUILLERMO LAZCOZ / PAUL DE HERT (2022), p. 20.

²⁴⁶ GT29 (2018), p. 30.

Estas são algumas das questões que permanecem em aberto e que precisariam de ser clarificadas. Não podendo descurar a importância que a atuação humana continua a ter (em todo o processo de tratamento e não somente com a decisão automatizada), as obrigações gerais previstas para os responsáveis pelo tratamento existem, mas no que diz respeito as decisões automatizadas (que é o relevante para o nosso objeto de estudo) conviria dar mais certeza e criar um regime mais claro. Se o RGPD deu um passo importante, parece-nos que ainda não consubstancia um mecanismo totalmente eficaz para mitigar a discriminação algorítmica.

7. Adaptação na era do constitucionalismo digital – a criação de novos direitos?

A eliminação da discriminação algorítmica é uma utopia distante. Como vimos, regular, proteger os direitos fundamentais e responsabilizar as entidades é primordial. Todavia, não podendo ser eliminado o risco, nesta era do constitucionalismo digital, a proteção dos direitos fundamentais necessita de acompanhar as constantes mutações, Nesta medida, GIOVANNI DE GREGORIO sustenta que estamos a entrar na quarta fase de constitucionalismo digital, caracterizada pelo fortalecimento dos diversos direitos²⁴⁷.

A este propósito, GIOVANNI DE GREGORIO e ORESTE POLLICINO propõem um conjunto de direitos individuais que nos parecem relevantes, tais como o direito de fácil acesso (*right to accessibility* - o qual visaria o direito de fácil acesso e interação com o algoritmo) e o direito a uma tradução da linguagem tecnológica para linguagem humana (que implicaria o uso de linguagem clara e entendível da decisão algorítmica)²⁴⁸. Este último parece-nos da maior relevância, mas claro que teria de ser acompanhado dos devidos cuidados procedimentais, sob pena de se tornar noutra direito oco e sem concretização prática. Tal poderia ainda diminuir as dificuldades que acentuamos propósito da explicabilidade que deve ser fornecida no âmbito do RGPD.

De um ponto de vista multinível, encontramos tutelado o direito da proteção de dados e o direito à reserva sobre a intimidade da vida privada (consagrado na CDFUE e no TFUE). Em nosso entendimento, e para uma tutela mais adequada dos casos de discriminação algorítmica, faria sentido (numa lógica de acompanhamento da realidade tecnológica) criar

²⁴⁷ GIOVANNI DE GREGORIO (2021), p. 67.

²⁴⁸ ORESTE POLLICINO / GIOVANNI DE GREGORIO (2022), p. 21.

um “direito fundamental à proteção das decisões algorítmicas”²⁴⁹, o qual permitiria criar uma nova geração de direitos previstos no plano constitucional²⁵⁰. Defendemos isto com o intuito de reforçar a tutela dos direitos fundamentais, equacionando um direito o mais amplo possível para permitir a sua adaptabilidade às mudanças tecnológicas. Com efeito, no plano do direito constitucional, as atuações não podem ser arbitrárias e o foco deverá estar sempre na garantia dos direitos, liberdades e garantias essenciais do estado democrático.

Adicionalmente, há quem tenha proposto a criação de uma entidade independente e fiscalizadora das decisões algorítmicas (adotando um papel semelhante à CNPD no direito da proteção de dados)²⁵¹. Ora, o Regulamento de IA deu um passo em frente ao prever a criação de uma entidade de fiscalização, algo que faria todo o sentido replicar nos diversos Estados-Membros, criando-se uma instância mais próxima e especializada de reação das decisões algorítmicas discriminatórias. Uma coisa é certa: independentemente das diferentes medidas, nunca nos poderemos esquecer que no final do dia importará sempre criar os meios de tutela mais eficazes para que os cidadãos possam exercer os seus direitos. Na nossa perspetiva, o primeiro passo passará por criar um direito fundamental à proteção das decisões algorítmicas.

8. Conclusão

Em suma, a discriminação algorítmica constitui um dos principais desafios nesta nova era de constitucionalismo digital, levantando questões que outrora pareciam inexistentes. Mesmo perante um “véu da objetividade” vimos como pode o algoritmo discriminar nas mais diversas situações do quotidiano, impactando a justiça, o mundo laboral, o crédito ao consumo e a educação.

O principal problema reside nos dados (seja quando estes não são representativos da população encontrando-se enviesados ou quando eles próprios traduzem discriminação social pré-existente) e na programação do algoritmo (caso em que são os próprios

²⁴⁹ A CPDHED prevê, no art. 9.º, que a utilização da IA deve ser orientada pelo respeito dos direitos fundamentais, prevenindo aspetos gerais de orientação.

²⁵⁰ MARGOT E. KAMINSKI / JENNIFER M. URBAN (2021), pp. 2032-2035 defendem a adoção de um direito a contestar as decisões de IA, o qual, ao contrário do RGPD, assevera a dimensão procedimental necessária.

²⁵¹ Sobre o tema CATARINA SANTOS BOTELHO (2024), p. 14 e LILIAN EDWARDS / MICHAEL VEALE (2017), p. 83

programadores que, consciente ou inconscientemente, colocam no algoritmo concepções que poderão ser discriminatórias).

Perante este desafio que convoca a tecnologia, o direito não pode ficar indiferente e, muito menos, continuar a aplicar institutos que não foram pensados para esta questão. Por este motivo, consideramos desadequada a utilização do Direito Antidiscriminação para tutelar este problema. Desde logo, no que respeita à discriminação direta, será sempre muito difícil demonstrar que o algoritmo (perante os exemplos que evidenciamos) discriminou devido a uma característica protegida específica, como o seu carácter individualizado não se coaduna com a potencialidade destas decisões automatizadas.

Ademais, a discriminação indireta, assentando numa lógica grupal, poderia à primeira vista ser como uma melhor solução, mas não o é. Não só pela dificuldade probatória evidente, como pela neutralidade dos algoritmos.

Por essa razão, sendo a natureza da discriminação algorítmica exógena, cremos que o melhor caminho passa por melhorar a regulação nesta matéria, alicerçando com uma defesa ativa dos direitos fundamentais.

Sabemos que o legislador deve adaptar-se às mutações tecnológicas, motivo pelo qual o Regulamento da IA será um mecanismo essencial, sobretudo para os sistemas de alto risco. Não obstante dificuldades que se possam apontar noutros domínios, o regulamento aborda diretamente a discriminação algorítmica, criando novas obrigações para os fornecedores e utilizadores dos sistemas de IA, reforçando o acompanhamento no seu processo de criação e decisão. Além disso, assenta numa estratégia de defesa dos direitos fundamentais, algo essencial tendo em conta a temática em questão.

Um ponto relevante e que, em nosso entendimento, deverá ser articulado com o RGPD prende-se com a possibilidade de tratamento de dados sensíveis, conquanto observando-se o princípio da necessidade, de maneira a detetar as situações discriminatórias no algoritmo e, com isso, tentar mitigá-las.

Claro que, para uma maior harmonia no ordenamento jurídico, faria todo o sentido proceder à alteração do RGPD, para que o art. 9.º, n.º 2 introduzisse uma nova exceção para estas situações, contrariando-se a tese da *fairness through blindness*. Afinal, não nos parece que o caminho passe pela ocultação das características protegidas, tendo em conta que a discriminação pode ocorrer na mesma através dos dados correlacionados.

Bem sabemos que o RGPD não foi pensado diretamente para esta finalidade e que algumas das suas disposições podem *à priori* parecer incoerentes. Por essa razão, foram apontadas certas soluções que poderiam ser idóneas a este fim: não só com a consideração dos dados sensíveis no processo de tratamento de dados para efeitos de detetar e eliminar a discriminação algorítmica, bem como com a consideração das inferências no conceito de dado pessoal e a clarificação do nível de intervenção humana.

Doravante, além de regular, dever-se-á assegurar que o processo de decisão automatizada não fica desprovido de intervenção humana e da necessária *accountability*. Aliás, não podemos olvidar que com o Regulamento da IA poderá haver situações em que exista uma cumulação de responsabilidades (quer ao abrigo desse regulamento, quer ao abrigo do RGPD).

Além do mais, consideramos que a tendência regulatória atual deverá ser acompanhada de maior tutela dos direitos fundamentais dos cidadãos. Nesta nova era de constitucionalismo digital, onde os algoritmos desempenham um papel certo no futuro decisório, será primordial acautelar direitos que permitam uma reação efetiva e adequada. Não acreditamos em utopias, nem que a discriminação algorítmica possa ser eliminada *ab initio*.

O segredo reside em adotar as melhores práticas para mitigar este risco e, no caso da sua ocorrência, criar os melhores mecanismos de reação. Por essa razão, advogamos que faria sentido criar no plano constitucional o direito fundamental à proteção das decisões algorítmicas, adaptando-se às mutações tecnológicas recentes e reforçando-se o plano fundamental de proteção. Tudo isto, como defendemos, poderia ser supervisionado por uma entidade externa de controlo especializada, asseverando no plano prático o respeito pelos direitos fundamentais no processo de decisão algorítmica.

O futuro da sociedade algorítmica é incerto e as respostas dadas pelo ser humano à evolução da IA serão essenciais. No plano da discriminação algorítmica, será crucial assegurar sempre em última linha os direitos individuais. Esperemos que esta Dissertação tenha contribuído para a evolução científica na área e que, doravante, seja possível a continuidade da investigação na discriminação algorítmica na nova era digital.

Bibliografia

- ABRANTES, José João e Isabel Valente Dias (2023) - “Regulação do algoritmo nas relações laborais”, *Revista Questões Laborais*, Ano XXX, n.º 63, Jul/Dez 2023, pp. 37-50.
- AJUNWA, Ifeoma (2020) - “The paradox of automation as anti-bias intervention”, *Cardozo Law Review*, Volume 41, pp. 1671-1742.
- ANDRADE, Diogo Queiroz de (2023) - *Algoritmos – uma revolução em curso?*, Lisboa, Fundação Francisco Manuel dos Santos.
- ARNAZ, Aiba Soriano (2022) - “Discriminación algorítmica: garantías y protección jurídica”, in Marcelo Bauzá Reilly (Coord.) *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, n.º 48, Navarra, Thomson Reuters Aranzadi, pp. 139-169.
- AYTEKIN, Ahmet Bilal (2023) - “Algorithmic Bias in the context of European Union Anti-Discrimination Directives”, *EWAf’23: European Workshop on Algorithmic Fairness*, <https://ceur-ws.org/Vol-3442/paper-41.pdf>, consult. em 02/Out/2023.
- BALAYN, Agathe E SEDA GÜRSES (2021) - *Beyond Debiasing. Regulating IA and its inequalities*, European Digital Rights, https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf, consult. em 18/Dec/2023.
- BARBOSA, Mafalda Miranda (2023) - “Inteligência artificial e proteção de dados”, in Ana Miranda Rodrigues e Susana Aires de Sousa, *I Congresso de Inteligência Artificial e Direito*, Coimbra, Almedina, pp. 37-57.
- BAROCAS, Solon e Andrew D. Selbst (2016) - “Big Data’s Direct Impact”, *California Law Review*, Volume 104, n.º 3, pp. 671-732.
- BARTLETT, Robert *et. al* (2021) - “Algorithmic Discrimination and Input Accountability under the civil rights act”, *Berkeley Technology Law Journal*, Volume 86, pp. 675-736.
- BENT, Jason R. (2020) - “Is Algorithmic Affirmative Action Legal?”, *The Georgetown Law Journal*, Volume 108, pp. 803-853.
- BORGESIUUS, Frederik J. Zuiderveen (2020) - “Strengthening legal protection against discrimination by algorithms and artificial intelligence”, *The International Journal of Human Rights*, Volume 24, pp. 1572-1593.
- BORSTEIN, Stephanie (2018) - “Antidiscriminatory Algorithms”, *Alabama Law Review*, Volume 70, Issue 2, pp. 521-572.
- BOTELHO, Catarina Santos (2013) - “A recepção da Carta dos Direitos Fundamentais da União Europeia na ordem jurídico-constitucional portuguesa: uma dinâmica *pro unione* ou *pro*

- constitutione?*”, in ANTÓNIO PINTO PEREIRA *et. al.* (ed.), *Liber Amicorum em homenagem ao Professor Doutor João Mota de Campos*, Coimbra, Coimbra Editora, pp. 315-359.
- BOTELHO, Catarina Santos (2017) - “Novo ou Velho Direito? – O Direito ao esquecimento e o princípio da proporcionalidade no constitucionalismo global”, *AB Instantia*, Ano V, n.º 7, pp. 49-71.
- BOTELHO, Catarina Santos (2023) - “Algoritmos Discriminatórios”, in Ana Miranda Rodrigues e Susana Aires de Sousa (Ed.) *I Congresso Inteligência Artificial e Direito*, Coimbra, Almedina, pp. 27-35.
- BOTELHO, Catarina Santos (2024) - “The end of the deception? – Counteracting algorithmic bias in the digital age”, in Giovanni De Gregorio, Oreste Pollicino and Peggy Valcke (eds) *The Oxford handbook of Digital Constitutionalism*, Oxford University Press, forthcoming, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4659451, consult. em 27/Jul/2023.
- BOURA, Marta (2023) - “Inteligência Artificial. Quadro jurídico e reflexões sobre a Proposta de Regulamento de Inteligência Artificial”, *Revista Eletrónica de Direito*, n.º 3, volume 32, pp. 99-123.
- BUCHNER, Benedikt (2018) - “Art. 22. Automatisierte Entscheidungen im Einzelfall einschließlic Profiling”, in Jürgen Kühling e Benedikt Buchner (ed.), *Datenschutz Grundverordnung/BDSG Kommentar*, 2.^a Auflage, C.H. BECK, pp. 508-520.
- BUSTAMANTE, Natalia Ramirez e Andrés PÁEZ (2023), “Análisis jurídico de la discriminación algorítmica en los procesos de selección laboral”, [https://www.researchgate.net/publication/343141239 Analisis juridico de la discriminacion algoritmica en los procesos de seleccion laboral](https://www.researchgate.net/publication/343141239_Analisis_juridico_de_la_discriminacion_algoritmica_en_los_procesos_de_seleccion_laboral), consult. em 28/Set/2023.
- BYGRAVE, Lee A. (2020) - “Article 22.º Automated Individual Decision-Making including profiling”, in Christopher Kuner *et. al* (ed.) *The EU General Data Protection Regulation (GDPR) A Commentary*, Oxford, Oxford University Press, pp 523-542.
- BYGRAVE, Lee A. e Luca TOSONI (2020) - “Article 4(1). Personal Data”, in Christopher Kuner *et. al* (ed.) *The EU General Data Protection Regulation (GDPR)*, Oxford, Oxford University Press, pp. 103-115.
- CALVÃO, Filipa Urbano (2018) - *Direito da Proteção de Dados Pessoais – Relatório sobre o programa, os conteúdos e os métodos de ensino da disciplina*, Porto, Universidade Católica Editora.

- CALVÃO, Filipa Urbano (2023) - "IA y el algoritmo discriminatorio: ¿Cómo promover la igualdad?", AA.VV. *Desafíos éticos, jurídicos y tecnológicos del avance digital*, Iustel, pp. 75-88.
- CANOTILHO, Gomes e Vital MOREIRA (2007) - *Constituição da República Portuguesa Anotada*- Volume 1.º a 107.º, 4.ª Edição Revista, Coimbra, Coimbra Editora.
- CANOTILHO, Mariana (2011) - “Brevísimos apontamentos sobre a não discriminação no Direito da União Europeia”, *Revista Julgar*, n.º 14, pp. 101-111.
- CARLSON, Alyssa M. (2017) - “The Need for Transparency in the Predictive Sentencing Algorithms”, *Iowa Law Review*, Volume 105, pp. 303-329.
- CASTANEDA, Juliana *et. al* (2022) - “Dealing with Gender Bias Issues in Data-Algorithmic Processes: a Social-Statistical Perspective”, *in Algorithms*, 15, 303, pp. 1-16.
- CELESTE, Edoardo (2023) - *Digital Constitutionalism – The Role of Internet Bill of Rights*, New York, Routledge.
- CHANDER, Anupam (2017) - “The Racist Algorithm?”, *Michigan Law Review*, Volume 115, Issue 6, pp. 1023-1045.
- COFONE, Ignacio N. (2019) - “Algorithmic Discrimination Is an Information Problem”, *Hastings Law Journal*, Volume 70, Issue 6, pp. 1389-1444.
- CORDEIRO, A. Barreto Menezes (2018) - “Dados Pessoais: conceito, extensão e limites”, *Revista de Direito Civil*, n.º 2, pp. 297-321.
- CORDEIRO, A. Barreto Menezes (2020) - *Direito da Proteção de Dados à luz do RGPD e da Lei n.º 58/2019*, Reimpressão, Coimbra, Almedina.
- CORTIZO, Lucas (2021) - “Dados e Inteligência Artificial: os efeitos jurídicos da discriminação algorítmica”, *Anuário da Proteção de Dados 2021*, pp. 95-120.
- COSTA, Inês da Silva (2021) - “A Proteção da pessoa na era dos big data: a opacidade do algoritmo e as decisões automatizadas”, *Revista Eletrónica de Direito*, volume 24, n.º 1, 2021, pp. 34-82.
- DAELMAN, Charline (2021) - “AI Through a Human Rights Lens. The Role of Human Rights in Fulfilling AI’s Potential”, *Artificial Intelligence and the Law*, Jan de Bruyne e Cedric Vanleenhove (ed.), Cambridge, Intersentia, pp. 123-149.
- DESAI, Deven R. e Joshua, A. KROLL (2017) - “Trust but verify: a Guide to Algorithms and the Law”, *Harvard Journal of Law & Technology*, Volume 31, number 1, pp. 1-64.

- DEVILLÉ, Rembrandt, Nico SERGEYSSELS e Catherine MIDDAG (2021) - “Basic Concepts of AI for Legal Scholars”, in JAN DE BRUYNE AND CEDRIC VANLEENHOVE (eds.) *Artificial Intelligence and the Law*, Cambridge, Intersentia, pp. 1-22.
- DOMINGOS, Pedro (2017) - *A Revolução do algoritmo mestre – como a aprendizagem automática está a mudar o mundo*, 11.^a Edição, Barcarena, Manuscrito.
- DOURISH, Paul (2016) - “Algorithms and their others: Algorithmic culture in context”, *Big Data & Society*, Volume 3, Issue 2, pp. 1-11, <https://journals.sagepub.com/doi/10.1177/2053951716665128>, consult. em 15/Set/2023.
- EDWARDS, Lilian e Michael VEALE (2017) - “Slave to the Algorithm? Why a 'Right to An Explanation' is probably not the remedy you are looking for”, *Duke Law and Technology Law Review*, Volume 16, Issue 1, pp. 18-84.
- FIDALGO, Vítor Palmela (2021) - “Artigo 13.º Informações a facultar quando os dados pessoais são recolhidos junto do titular”, in A. Barreto Menezes Cordeiro (Coord.) *Comentário ao Regulamento Geral de Proteção de Dados e à Lei n.º 58/2019*, Coimbra, Almedina, pp. 156-164.
- FREEDMAN, Sandra (2022) - *Discrimination Law*, Oxford, Oxford University Press.
- GEORGIEVA, Ludmila e Christopher KUNER (2020) - “Article 9. Processing of special categories of personal data”, in Christopher Kuner (ed.) *The EU General Data Protection Regulation (GDPR)*, Oxford, Oxford University Press, pp. 365-384.
- GERARDS, JANNEKE e Frederik Zuiderveen BORGESIU (2022) - “Protected Grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence”, *Colorado Technology Law Journal*, Volume 20/1, pp. 1-56.
- GILLIS, Talia B. e Jann L. SPIESS (2019) - “Big Data and Discrimination”, *The University of Chicago Law Review*, Volume 86, pp. 459-487.
- GOODMAN, Bryce e Seth FLAXMAN (2017) - “European Union Regulations on Algorithmic Decision Making and a Right to explanation”, *AI Magazine*, pp. 50-57.
- GREGORIO, Giovanni de (2021) - “The rise of digital constitutionalism in the European Union”, *International Journal of Constitutional Law*, Volume 19, Issue 1, pp. 41–70.
- GREGORIO, Giovanni de (2022) - *Digital Constitutionalism in Europe – Reframing Rights and Powers in the Algorithmic Society*, Cambridge, Cambridge University Press.

- HACKER, Philipp (2018) - “Teaching fairness to Artificial Intelligence: existing and novel strategies against algorithmic discrimination in EU Law”, *Common Market Law Review*, Volume 55, pp. 1143-1186.
- HERVEY, Matt e Matthew LAVY (2021) - *The Law of Artificial Intelligence*, London, Sweet & Maxwell.
- HOLDER, Allan E. (2020) - “What we don’t know they know: what to do about inferences in European and California data protection law”, *Berkeley Technology Law Journal*, Volume 35, pp. 1331-1364.
- HOUSER, Kimberly A. (2019) - “Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making”, *Stanford Technology Law Review*, n.º 22, pp. 290-354.
- HUQ, Aziz Z. (2020) - “A right to a human decision”, *Virginia Law Review*, Volume 106, pp. 611-688.
- JACKSON, Maya C. (2021) - “Artificial Intelligence & Algorithmic Bias: The Issues with Technology Reflecting History & Humans”, *Journal of Business & Technology Law*, Volume 16, Issue 2, pp. 299-316.
- KAMINSKI, Margot E. (2019) - “The Right to Explanation, Explained”, *Berkeley Technology Law Journal*, Volume 34, pp. 189-218.
- KAMINSKI, Margot E. e Jennifer URBAN (2021) - “The right to contest AI”, *Columbia Law Review*, Volume 121, n.º 7, pp. 1957-2048.
- KAYTAL, Sonia K. (2019) - “Private accountability in the age of Artificial Intelligence”, *U.C.L.A Law Review*, Volume 66, Issue 54, pp. 54-141.
- KISSINGER, Henry A., Eric SCHMIDT e Daniel HUTTENLOCHER (2021) - *A Era da Inteligência Artificial e o nosso futuro humano*, Alfragide, D. Quixote.
- KROLL, Joshua A. et al. (2017) - “Accountable Algorithms”, *University of Pennsylvania Law Review*, Volume 165, pp. 633-705.
- KUŚMIERCZYK, Małgorzata (2022) - “Algorithmic bias in the light of the GDPR and the proposed AI Act”, Maciej Olejnik (ed.), *(In)equality. Faces of modern Europe*, (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4117936, consult. em 02/Dez/2023.

- LARSON, Jeff *et. al* (2016) - “How we analyzed the COMPAS Recidivism Algorithm”, *ProPublica*, <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> consult. em 30/Sep/2023.
- LAZCOZ, Guillermo e Paul de HERT (2022) - “Humans in the GDPR and AIA Governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities”, *Brussels Privacy Hub*, Volume 8, n.º 32, <https://brusselsprivacyhub.eu/publications/BPH-Working-Paper-VOL8-N32.pdf>, consult. em 29/Dez/2023.
- LEAL, Ana Alves (2017) - “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 A. Menezes Cordeiro *et. al* (coord.) *FinTech: Desafios da Tecnologia Financeira*, 2.ª edição, Almedina, pp. 89-220.
- LEHR, David e Paul OHM (2017) - “Playing with Data: What legal scholars should learn about machine learning”, *UC David Law Review*, Volume 51, pp. 653-717.
- MAYSON, Sandra G. (2019) - “Bias In, Bias Out”, *The Yale Law Journal*, Volume 128, n.º 8, pp. 2122-2473.
- MESTRE, Bruno (2015) - “Sobre o conceito de discriminação – uma perspetiva contextual e comparada”, AA.VV. *Estudos dedicados ao Professor Bernardo da Gama Lobo Xavier*, Volume I (Especial), pp. 377-410.
- MIRANDA, Jorge e Rui MEDEIROS (2017) - “Artigo 13.º Princípio da Igualdade”, Jorge Miranda e Rui Medeiros (Ed.) *Constituição da República Portuguesa Anotada Volume I*, Lisboa, Universidade Católica Editora, pp. 159-179.
- MONIZ, Graça Canto (2023) - *Manual de Introdução à Proteção de Dados Pessoais*, Coimbra, Almedina.
- MOREIRA, Teresa Coelho (2021) - “A Discriminação Algorítmica”, *Revista Questões Laborais*, n.º 58, pp. 85-103.
- MURPHY, Paul (2012) - *Machine Learning – A Probabilistic Perspective*, Massachusetts Institute of Technology, http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf, consult. em 31/Out/2023.
- NOWOTNY, Helga (2021) - *In AI We Trust – Power, Illusion and Control of predictive algorithms*, Cambridge, Policy.

- PACKIN, Nizan Geslevich e Lev Yafit ARETZ (2020) - “Learning Algorithms and discrimination”, in Woodrow Barfield e Ugo Pagallo (Ed.) *Research Handbook on the Law of Artificial Intelligence*, Cheltenham, Edward Elgar Publishing, pp. 87-113.
- PAIS, Sofia (2018) - *Estudos de Direito da União Europeia*, 4.^a edição, Coimbra, Almedina, 2018.
- POLLICINO, Oreste e Giovanni de GREGORIO (2022) - “Constitutional Law in the Algorithmic Society”, in Hans-W. Micklitz *et. al*, (Ed.) *Constitutional Challenges in the Algorithmic Society*, Cambridge, Cambridge University Press, pp. 3-24.
- PRASSL, JERAMIAS ADAMS, Reuben BINNS E Aislinn Kelly LYTH (2023), “Directly discriminatory algorithms”, *The Modern Law Review*, Volume 186, Issue 1, pp. 144-175.
- RAMOS, José Ricardo Marcondes (2023) - “Supervisão, classificação e certificação dos sistemas de IA na Proposta de Regulamento sobre Inteligência Artificial”, in Susana Aires de Sousa (Coord.) *Direito em Mudança – A Proposta de Regulamento Europeu sobre Inteligência Artificial Algumas Questões*, Coimbra, Coimbra Editora, pp. 31-63.
- RAMOS, LUÍS FELIPE MIRANDA (2020) - “Os riscos invisíveis de discriminação na utilização de algoritmos no processo de contratação de pessoal”, *Revista Questões Laborais*, n.º 56, pp. 203-217.
- REBELO, Diogo Morgado (2023) - *Inteligência Artificial e Scoring no crédito ao consumo*, Coimbra, Almedina.
- ROUXINOL, Milena da Silva (2021) - “Os algoritmos – Licença para discriminar? (Um segundo olhar sobre a seleção de candidatos a trabalhadores através de técnicas de inteligência artificial)”, *Revista do CEJ*, I, pp. 235-268.
- RUBENSTEIN, David. S. (2021) - “Acquiring Ethical AI”, *Florida Law Review*, Volume 73, n.º 4, pp. 747-819.
- SELBST, Andrew D. (2017) - “Disparate Impact in Big Policing”, *Georgia Law Review*, Volume 52, pp. 109-195.
- SELBST, Andrew D. (2020) - “Negligence and AI’s Human Users”, *Boston University Law Review*, Volume 100, p. 1315-1376.
- SELBST, Andrew D. e Julia POWLES (2017) - “Meaningful information and the right to explanation”, *International Data Privacy Law*, Volume 7, no. 4, pp. 233-242.
- SELMI, Michael (2020) - “Algorithms, Discrimination and the Law”, *Ohio State Law Journal*, Volume 82, Issue 4, pp. 406-405.

- SHIN, Donghee e Yong Jin PARK (2019) - “Role of fairness, accountability, and transparency in algorithmic affordance”, *Computers in Human Behavior*, Volume 98, pp. 277-284.
- SILES, Emile Loza (2022) - “Artificial Intelligence Bias and Discrimination: Will we pull the arc of the moral universe towards justice”, *Journal of International and Comparative Law*, Volume 8, No. 2, pp. 513-544.
- TEIXEIRA, Mariana Alves (2023) - *Da (des)proteção do Candidato a Emprego em Face da Inteligência Artificial*, Coimbra, Almedina.
- THOUVENIN, Florent, Alfred FRUH e Simon HENSELER (2022) - “Article 22 GDPR on Automated Individual Decision-Making: Prohibition or Data Subject Right”, *European Data Protection Law Review*, Volume 183, pp. 183-198.
- TURING, Alan (1950) - “*Computing Machinery and Intelligence*”, *Mind a Quarterly Review of Psychology and Philosophy*, Volume 59, N. ° 236, pp. 433-460.
- VEALE, Michael e Reuben BINNS (2017) - “Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data”, *Big Data & Society*, July–December, pp. 1–17, <https://doi.org/10.1177/2053951717743530>, consult. em 06/Dec/2023.
- VEALE, Michael e Frederik ZUIDERVEEN (2021) - “Demystifying the Draft EU Artificial Intelligence Act”, *Computer Law Review International*, 4/2021, pp. 97-112.
- VICENTE, Paulo Nuno (2022) - *Os Algoritmos e Nós*, Lisboa, Fundação Francisco Manuel dos Santos.
- XENIDIS, Raphaële e Linda SENDEN (2020) - “EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic decision”, in Ulf Bernitz *et. al* (eds.), *General Principles of EU law and the EU Digital Order*, Kluwer Law, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529524, consult. em 14/Out/2023.
- WATCHER, Sandra e Brent MITTELSTADT (2019) - “A Right to Reasonable inferences: re-thinking data protection law in the age of big data and AI”, *Columbia Business Law Review*, n.º 2, pp. 494-620.
- WATCHER, Sandra, Brent MITTELSTADT e Chris RUSSEL (2017) - “Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation”, *International Data Privacy Law*, Volume 7, n.º 2, pp. 76-99.
- WATCHER, Sandra, Brent MITTELSTADT e Chris RUSSEL (2018) - “Counterfactual explanations without opening the Black Box: Automated Decisions and the GDPR”, *Harvard Journal of Law & Technology* Volume 31, Number 2, pp. 841-887.

WATCHER, Sandra, Brent MITTELSTADT e Chris RUSSEL (2021) - “Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI”, *Computer Law & Security Law Review*, Volume 41, pp. 1-31, <https://doi.org/10.1016/j.clsr.2021.105567>, consult. em 05/Oct/2023.

XIANG, Alice (2021) - “Reconciling Legal and Technical Approaches to Algorithmic Bias”, *Tennessee Law Review*, Volume 88, Issue 3, pp. 649-724.

Legislação

Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE.

Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial, 2021/0106.

Diretiva (UE) 2000/43/CE do Conselho, de 29 de junho de 2000, que aplica o princípio da igualdade de tratamento entre pessoas, sem distinção de origem racial ou étnica.

Diretiva (UE) 2000/78/CE do Conselho, de 27 de novembro de 2000, que estabelece um quadro geral de igualdade de tratamento no emprego e na atividade profissional.

Diretiva (UE) 2006/54/CE do Conselho, de 5 de julho de 2006, relativa à aplicação do princípio da igualdade de oportunidades e igualdade de tratamento entre homens e mulheres em domínios ligados ao emprego e à atividade profissional.

Outros documentos

European Data Protection Board (2020) - *Diretrizes 3/2019 sobre tratamento de dados pessoais através de dispositivos de vídeo*, Versão 2.0, adotado em 29 de janeiro de 2020.

Grupo de Trabalho de Proteção de Dados do Artigo 29 (2007) - *Parecer 4/2007 sobre o conceito de dados pessoais*, adotado em 20 de junho.

Grupo de Trabalho de Proteção de Dados do Artigo 29 (2017) - *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, adotado em 3 de outubro de 2017.

Grupo de Trabalho de Proteção de Dados do Artigo 29 (2018) - *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*, adotado em 3 de outubro de 2017.

Juriprudência

McDonnell Douglas Corp. v. Green, 411 U.S. 792.

Tribunal de Justiça, *Defrenne c. Sabena*, C-43/75, 8 de abril de 1976.

Tribunal de Justiça, *YS, M, S*, C-141/12, 17 de julho de 2014.

Tribunal de Justiça, *Egenberger*, C-414/16, 17 de abril de 2018.

Tribunal de Justiça, *Schufa Holding AG* (C-634/21), 7 de dezembro de 2023.