

O algoritmo no tecido empresarial: Das vicissitudes de contratação ao despedimento de trabalhadores – Uma arma nociva que necessita de ser desarmada?

FREDERICO VENTURA SEQUEIRA*

Resumo: A proliferação da inteligência artificial em contextos de tomada de decisão é saudada como uma bala de prata prometendo substituir a subjetividade humana por decisões objetivas e infalíveis, encobrendo males maiores que espelham a permissibilidade de decisões algorítmicas discriminatórias e arbitrárias, tomando a proporção de verdadeiras *WMDs*. Tendo presente o exponencial crescimento da adoção de algoritmos de *ML* nas relações laborais, propomo-nos exibir uma imagem mais nítida dos potenciais e efetivos malefícios juntamente com interrogações que carecem de respostas claras fruto do advento e da paulatina concretização da quarta Revolução Industrial, tentando desarmar algo que à primeira vista aparenta tratar-se de uma arma nociva que atua silenciosamente, procurando balizá-la e instruí-la de forma a (re)equilibrar os pratos da balança e conferir uma efetiva tutela aos candidatos a emprego e aos trabalhadores na iminência de mais um (grande) desafio que parece desafiar as condições da sua contratação, prestação do seu labor e cessação da sua atividade profissional, quer no que respeita ao trabalho tradicional, quer quanto às novas formas de o prestar. Para tal, defendemos uma reforma do uso de algoritmos de *ML* assente na sua regulação ético-jurídica, num processo completo e estruturado de auditoria, num efetivo direito à intervenção humana e num acesso informado aos critérios subjacentes à *ratio*

* Licenciado em Direito pela Faculdade de Direito da Universidade de Lisboa. Na mesma Instituição de ensino, frequenta o Mestrado em Direito e Ciência Jurídica, na especialidade de Direito Laboral, encontrando-se em fase de dissertação. O escrito que ora se apresenta corresponde, *mutatis mutandis*, ao relatório de Mestrado apresentado no âmbito da disciplina de Direito do Trabalho, sob a regência da Senhora Professora Doutora Maria do Rosário Palma Ramalho.

de uma decisão, permitindo imputar responsabilidades, reagir fundamentadamente, desta forma, impedir esvaziamentos legislativos de conteúdo e normas simbólicas perante uma (nova) realidade que necessita de urgente adaptação uma vez que veio para ficar.

Palavras-chave: *decisões individuais automatizadas, algoritmos, discriminação e arbitrariedade, prova algorítmica, responsabilidade, RGPD, ser humano no comando, direito à explicação, direitos de PI e segredos comerciais*

Abstract: The proliferation of artificial intelligence in decision-making contexts is hailed as a silver bullet that promises to replace human subjectivity by objective and infallible decisions, covering up greater evils that reflect the permissibility of discriminatory and arbitrary algorithmic decisions, taking the proportion of real WMDs. Bearing in mind the exponential growth of the adoption of ML algorithms in labour relations, we intend to show a clearer image of the potential and effective harm together with questions that lack clear answers as a result of the advent and the gradual implementation of the fourth Industrial Revolution, trying to disarm something that, at first sight, seems to be a noxious weapon that acts silently, by seeking to restrict it and instruct it in order to (re)balance the scale and to provide an effective protection to job applicants and workers on the verge of another (big) challenge that appears to defy the conditions of their hiring, provision of their work and termination of their professional activity, either with regard to traditional work or the new ways of providing it. To this end, we advocate a reform of the use of ML algorithms based on ethical-legal regulation, a complete and structured auditing process, an effective right to human intervention and informed access to the criteria underlying the ratio of a decision, allowing responsibilities to be imputed, a well-founded and solid reaction and, in this way, preventing legislative emptiness of content and symbolic provisions in the face of a (new) reality that needs urgent adaptation once it's here to stay.

Keywords: *automated decision making; algorithms; discrimination and arbitrariness; algorithmic proof; accountability; GDPR; human in command; right to explanation; IP rights and trade secrets.*

1. Introdução

A importância da digitalização levou ao domínio da inteligência artificial (IA) em quase todas as esferas da nossa vida, onde se incluem as relações laborais, foco do nosso estudo. Mais concretamente, o algoritmo de *machine learning* (ML) apresenta traços de grande utilidade e genialidade no seio das empresas, porém, esse toque de magia vem rapidamente ofuscar os inúmeros problemas sensíveis que representa para o trabalhador ao trazer uma mão cheia de desafios que parecem não ter uma resposta clara à vista, pois a verdade é que os algoritmos parecem estar atualmente a substituir o empregador através da gestão digital do trabalho no que diz respeito a decisões chave da carreira do trabalhador que incluem a contratação, avaliação e despedimento. Algo que tem despertado a preocupação por parte de algumas entidades cujo juízo de prognose faz crer que tais práticas se tornarão ainda mais comuns, constituindo formas mais intensas na maioria das empresas espalhadas por todo o mundo. Tendo isto presente, propomo-nos trazer à discussão a exposição e tratamento de certas questões algorítmicas que nos parecem mais críticas no que concerne quer ao candidato a emprego, quer ao trabalhador, numa perspetiva internacional, europeia e comparatística, ao mesmo tempo que colocamos o ordenamento português a teste.

Num primeiro momento, fazemos um breve enquadramento acerca da relação do algoritmo de *ML* com a IA, no que consiste e quais as suas principais qualidades e defeitos, tratando de aferir, como ponto de partida, qual dos lados da balança tende a pesar e a ter mais impacto na relação laboral, i.e., se as vantagens que lhes estão associadas valem o sacrifício dos potenciais efeitos perniciosos e se, de facto, vale a pena auxiliarmo-nos da sua presença num contexto de emprego. Em seguida, fazemos uma análise teórico-prática na qual se dá ênfase e destaque a um punhado de casos reais da maior relevância que se inserem nos dois principais campos problematizantes do processo interno do

algoritmo e refletem questões ético-jurídicas: a discriminação e a parcialidade e arbitrariedade.

Ulteriormente, abordamos as decisões individuais automatizadas e o *profiling* no Regulamento Geral sobre a Proteção de Dados (RGPD) ao debruçarmo-nos sobre o seu art. 22º, norma inédita no que diz respeito a esta temática que a Lei n.º 58/2019, de 8 de agosto - Lei de Execução portuguesa - não foi capaz de dar seguimento, não deixando de tratar questões derivadas igualmente importantes e com soluções legislativas duvidosas que se colocam a montante e a jusante da relação laboral como: o direito à intervenção humana traduzido no facto da última palavra (dever) ser de um interveniente humano e não da máquina, bem como do direito do indivíduo a reagir e contestar a decisão tomada (pela máquina) perante um humano; o direito de acesso à informação e o nível de transparência possível de alcançar espelhado na difícil conciliação entre direitos de autor, *trade secrets* e *IP rights* que muitas vezes impedem que a *source code* algorítmica seja conhecida, pelo que, inserindo um algoritmo num contexto de *black box*, aferimos da possibilidade de se encontrar o melhor equilíbrio possível uma vez que, para uma pessoa alvo de *automated decision making* (ADM) poder reagir, fundamentando devidamente uma contestação informada, precisará de ter um mínimo de conhecimento acerca do critério subjacente a tal decisão que lhe foi desfavorável sob pena de não ter qualquer oportunidade de se defender.

Nesta esteira, caminhando para o final, o ordenamento português é também ele colocado a teste no que diz respeito às duas principais questões basilares do nosso estudo. Assim, após uma breve contextualização da tutela empregue no combate à discriminação, procura-se saber, por um lado, quem deverá ser responsável por uma decisão discriminatória tomada por um algoritmo de modo a evitar que a imputabilidade se dilua pelo facto da decisão ter sido tomada por uma máquina, e que tipo de responsabilidade deverá ser acionada perante tal situação caso não sejam criadas normas e institutos específicos para tratar destas novas questões que têm a IA como atriz principal; por outro lado, procura-se ainda saber como poderá a pessoa visada por tal decisão discriminatória reagir tendo em conta a nossa legislação laboral, destacando os casos de prestação de trabalho nas plataformas digitais onde o

problema parece colocar-se de forma ainda mais intensa.

Em jeito de conclusão, face a esta nova realidade algorítmica que tem ocupado cada vez mais e intensamente o seu espaço nas relações laborais, tentamos desarmar os algoritmos enquanto *weapons of “math” destruction (WMDs)* e contribuir para uma discussão tão atual e premente dos nossos dias ao propor um conjunto de *guidelines* que se focam nos quatro pontos críticos que, no nosso entender, exigem um olhar atento e uma abordagem alternativa célere.

2. O Algoritmo e a sua ligação com a Inteligência Artificial.

2.1. *Conceptualização e importância*

Em termos generalistas, a Comissão Europeia entende por IA “os sistemas que dispõem de comportamento inteligente ao analisar o ambiente que os rodeia e, por conseguinte, tomar ações de modo a atingir propósitos específicos, sendo dotados de algum grau de autonomia”¹. De uma forma mais restrita, poderíamos qualificá-la como a ciência e a engenharia de produzir máquinas inteligentes capazes de resolver tarefas que tipicamente requerem inteligência humana reproduzindo as suas idiossincrasias, daí que o termo “algoritmo” se identifique com algo que tem a capacidade de aprender a resolver determinado problema para o qual foi programado, apresentando respostas através de um conjunto de instruções e comandos que lhe são fornecidos², sintetizando e automatizando tarefas, sendo precisamente esta a ideia chave. A partir daí, podemos centrar-nos numa das suas modalidades, o *Machine Learning* enquanto terminologia utilizada para descrever o processo de fornecimento aos sistemas da capacidade de aprender e melhorar automaticamente a partir da

1 CE, Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the regions, “Artificial Intelligence for Europe”, Brussels, 25.4.2018, COM (2018) 237 final, 2018. De igual modo, vide FIDALGO, Vítor Palmela, “Inteligência artificial e direitos de imagem”, in: ROA, Ano 78, n.º 3-4, 2018, Lisboa, pp. 879-900, ao distinguir os conceitos de IA e tecnologia robótica usualmente confundidos, de onde poderemos concluir metaforicamente que a IA é a mente e a robótica o corpo.

2 Neste sentido, EUAFR, “Big data, algorithms and discrimination”, 2018, disponível em: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-in-brief-big-data-algorithms-discrimination_en.pdf

experiência, sem ser explicitamente programado, onde se encontram os nossos algoritmos. JAMES ZOU utiliza uma frase muito simples, mas precisa, para os definir, referindo que “um algoritmo de *ML* é como um bebé recém-nascido a quem foi dado milhões de livros para ler sem lhe ensinarem o alfabeto”³, o que certamente antecipa alguns problemas de relevo, levando CATHY O’NEIL a qualificá-los como autênticas Armas de Destruição Matemática⁴ (*WMD - Weapons of Math Destruction*) cujos três elementos característicos assentam na opacidade, no efeito de escala e nos danos provocados, que se traduzem **(a)** no respetivo alcance (*Widespread*), ao atingir um número considerável de pessoas e estando no centro das grandes e mais importantes decisões da vida de cada uma delas, como são a contratação e o despedimento; **(b)** no secretismo (*Mysterious*), pelo facto das pessoas não saberem o que são algoritmos, qual o seu papel na empresa ou como estão a ser avaliadas; **(c)** no efeito destrutivo (*Destructive*), pelo impacto arrasador que causam e por fomentarem a injustiça, dotados de critérios inquinados e poluídos.

De um modo simplificado, a conceptualização de um algoritmo assenta em duas características cruciais de modo a compreendermos o seu propósito: **(i)** A *data*⁵, i.e., a reunião de informação e grandes quantidades de dados e

3 ZOU, James, “Removing gender bias from algorithms”, *The Conversation Journal*, 2016, disponível em: ZOU, James, “Removing gender bias from algorithms”, *The Conversation Journal*, 2016, disponível em: <<https://theconversation.com/removing-gender-bias-from-algorithms-64721>>

4 Acompanhamos de perto o valioso contributo da matemática O’NEIL, Cathy, *Weapons of Math Destruction - How Big Data Increases Inequality and Threatens Democracy*, Crown Publishers, New York, 2016, pp. 15-32, que, enquanto uma das grandes pioneiras do estudo dos efeitos perniciosos do algoritmo, podemos defini-la como o nosso Neo ao nos lembrar de observar os “senhores” digitais invisíveis que nos rodeiam uma vez que vivemos cada vez mais numa *Matrix* de que a maioria de nós não (se a) percebe. A sigla utilizada pela A. deriva de um trocadilho na medida em que, originalmente, *WMD* é utilizada como forma de qualificar as *Weapons of Mass Destruction*, i.e., Armas de Destruição Maciça, e a A. reinventa-a, com o intuito de tomar uma posição.

5 Cfr. GOLDENFEIN, Jake, “Algorithmic Transparency and Decision-Making Accountability: Thoughts for Buying Machine Learning Algorithms”, in: *Closer to the Machine: Technical, Social, and Legal aspects of AI*, Office of the Victorian Information Commissioner (ed.), 2019, pp. 41-60, (p. 42). Como *e.g.*, vide o Report: *Use of Workforce Analytics for Competitive Advantage*, Society for Human Resource Management Foundation, “What’s next: future global trends affecting your organization”, SHRM Foundation, 2016, p. 24 ao referenciar que a empresa *Nielson* usa a retenção do primeiro ano como métrica chave para avaliar se uma contratação foi bem-sucedida. Contudo é importante lembrar, tendo em conta a realidade do RGPD, o princípio da minimização dos dados (cfr. alínea c) do n.º 1 do

correspondência estatística padronizada, o *input* enquanto dados de entrada que vai espelhar o que aconteceu no passado, sendo por isso mesmo a característica *ex libris* da sua constituição e (ii) uma definição para o sucesso, i.e., qual o objetivo a atingir, o *output* enquanto gerador de um determinado resultado e decisão sem interferência humana, o que estamos à procura e o que queremos alcançar. Pois bem: na relação jurídico-laboral, a parte mais vulnerável (*in casu*, o candidato a emprego ou o trabalhador) não escolhe a dita definição do sucesso na empresa, sendo o seu contributo irrelevante para o conteúdo de um algoritmo que consiste sim em opiniões (*in casu*, do empregador) embutidas num código, que refletem a ideologia do seu criador dotado dos seus respetivos valores e desejos de modo a alcançar determinado resultado, ocorrendo, por isso, uma grande assimetria informativa (*e.g.*, em relação ao critério de contratação fornecido).

2.2. Da ilusão vantajosa aos reais efeitos perniciosos do seu uso

Os seus defensores pregoam alguns benefícios da sua presença cada vez mais acentuada nas relações laborais, pelo que destacamos os essenciais. *Primo*, integrando as novas tecnologias digitais, revelam-se como um recurso precioso na fase de recrutamento e seleção de trabalhadores, quer quando é feita diretamente pelo contratante, quer quando este recorre a prestadores de serviços externos, especialistas em recrutamento, auxiliando e permitindo que os recursos humanos (RH) tenham uma função cada vez mais estratégica e decisiva dentro das empresas⁶, assistindo-se, por isso, a uma verdadeira gestão

art. 5º RGPD) devendo ser adequados, relevantes e limitados ao necessário em relação aos fins para os quais são processados sob pena do tratamento ser ilícito e constituir contraordenação muito grave prevista e sancionada nos termos da alínea a) do n.º 5 do art. 83º RGPD e da alínea a) do n.º 1, e do n.º 2 do art. 37º da Lei de Execução. Discutivelmente, isto não significa que os responsáveis pelo tratamento de dados devem recolher sempre o mínimo de dados possível, significa antes que a quantidade deve estar relacionada com a finalidade, desde que os dados sejam adequados. Não obstante, o efeito parece ser paradoxal: *Primo*, quanto mais dados forem utilizados para treinar o algoritmo, mais preciso poderá ser o output; *Secondo*, o processamento de uma baixa quantidade de dados, levando a um output inexato ou deficitário na qualidade seria algo inadequado se se tivesse de tomar uma decisão com consequências legais ou que afetasse significativamente determinado indivíduo.

6 Neste sentido, CLARK, Melissa A., MAKOWSKY, Libby & ZUKIEWICZ, Marykate, “Implementation of the Teach For America Investing in Innovation Scale-Up”, in: Mathematica Policy Research, Princeton, 2015, pp. 1-56, (pp. 11-16) e TODOLÍ-SIGNES,

automatizada do trabalho com cada vez maior expressão, de que é exemplo a contratação, a aferição dos tempos de produção, a designação e distribuição de tarefas, a avaliação do desempenho e até o despedimento dos trabalhadores. *Secondo*, é vista enquanto garante de ganhos de produtividade ao replicar operações naturais, o que permite desenvolver, produzir e acumular raciocínios aprendidos ao longo do tempo pelo ser humano com alto grau de velocidade e eficiência, otimizando o tempo dos recrutadores, tornando o processo de recrutamento menos custoso e mais eficiente, e reduzindo o tempo⁷ e o risco de más contratações como demonstram os casos bem-sucedidos da *IBM*⁸, da *Pymetrics*⁹ e da *HireVue*¹⁰. *Terzo*, no que diz respeito ao *profiling* ou

Adrian, “Algorithms, Artificial Intelligence and Automated Decisions Concerning Workers and the Risks of Discrimination: The Necessary Collective Governance of Data Protection”, in: *European Review of Labour and Research*, Vol. 25, n.º 4, 2019, pp. 1-17, (pp. 3-5).

7 Como indicam BOGEN, Miranda & RIEKE, Aaron, “Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias”, in: *Upturn*, 2018, pp. 1-47, (pp. 6-7), “a maioria dos empregadores quer reduzir o tempo de contratação dado que quanto mais tempo for necessário para encontrar um candidato adequado, mais tempo e recursos são desviados de outras prioridades”, receando perder candidatos para os seus concorrentes.

8 Segundo dados estatísticos do THE ECONOMIST, “Big data and hiring - Robot recruiters: How software helps firms hire workers more efficiently”, 2013, disponível em: <<https://www.economist.com/business/2013/04/06/robot-recruiters>>, a IA permitiu à empresa poupar cerca de 266 milhões de euros na retenção de trabalhadores através de fidelização e retenção dos profissionais à empresa ao desenvolver um algoritmo capaz de prever, com um grau de exatidão de cerca de 95%, quais os trabalhadores que ponderam (ou poderão vir a ponderar num futuro próximo) abandonar a empresa, orientando os gestores na melhor estratégia para os reter.

9 Enquanto plataforma de matching de aquisição de talentos, o seu sistema apresentou um tempo de redução do processo de recrutamento das empresas de 75%, e o dobro do rendimento a nível de contratação. Acerca do funcionamento do seu algoritmo patenteado, *vide* POLLI, Frida & YOO, Julie, “Systems and methods for data-driven identification of talent”, 2019 e DENCİK, Lina, EDWARDS, Lilian & SÁNCHEZ-MONEDERO, Javier, “What does it mean to ‘solve’ the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems”, in: *Conference on Fairness, Accountability, and Transparency*, 2020, pp. 458-468, (p. 462).

10 Cfr. HIREVUE, “Bias, AI Ethics, and the HireVue Approach”, 2019, disponível em: <<https://www.hirevue.com/why-hirevue/ethical-ai/>>, onde na mesma linha da *Pymetrics*, possibilitou que uma cadeia de hotéis de grande dimensão reduzisse o tempo de contratação de mês e meio para uma semana. Desenvolvendo exemplos comprovados de plataformas neste contexto, *vide* a preciosa análise de DENCİK, Lina, et al., op. cit., pp. 458-468. Para uma perspetiva comparada, *vide* BESSE, Philippe, “Détecter, évaluer les risques des impacts discriminatoires des algorithmes d’IA”, 2020, disponível em: <<https://hal.archives-ouvertes.fr/hal-02616963/document>>, pp. 14-15 acerca das principais plataformas de contratação enquanto ferramentas que afastam a existência de discriminação, como a *JobiJoba*, a *Assessfirst*, a *Clémentine*, a *365talents* e a *easyrecrue*.

definição de perfis¹¹, favorece a recolha e tratamento dos dados pessoais dos trabalhadores por parte das empresas e instituições permitindo-lhes otimizar as opções e estratégias comerciais e de intervenção, podendo inclusive conferir benesses para as próprias pessoas como é disso exemplo os dados de saúde, mantidos nos hospitais ou nos centros de saúde, cujo acesso rápido facilita o tratamento do doente, quando necessário¹². *Quarto*, afasta o preconceito e a subjetividade humana substituindo-os por um processo técnico neutro¹³ por ser mais objetivo e assertivo, o que possibilita recrutar pessoas da mais variada índole apelando à diversidade do local de trabalho¹⁴, impedindo que certas características humanas tendenciosas, preconceituosas e injustas, como contratar alguém que partilhe dos mesmos gostos do recrutador ou alguém pertencente à mesma *Alma Mater* ou à mesma família que o empregador sejam fatores determinantes¹⁵, daí que a contratação possa servir como apoio a uma maior igualdade no local de trabalho em comparação com os métodos de contratação mais tradicionais. *Quinto*, poderá ser mais minucioso nas análises que faz, na atenção ao pormenor que muitas vezes escapa ao falível olho humano.

Fruto deste cenário idealista e perfeito existe um número razoável de empresas¹⁶ que ao dia de hoje recorre a algoritmos, porém, a verdade é que não

11 Que nos termos do número 4 do art. 4º do RGPD é entendido como “qualquer forma de tratamento automatizado de dados pessoais que consista em [utilizá-los] para avaliar certos aspetos pessoais de uma pessoa singular, nomeadamente para analisar ou prever aspetos relacionados com o seu desempenho profissional, a sua situação económica, saúde, preferências pessoais, interesses, fiabilidade, comportamento, localização ou deslocações”.

12 Na esteira de PALMA RAMALHO, Maria do Rosário, “A Economia Digital e a Negociação Coletiva”, in: CRL, Ministério do Trabalho, Solidariedade e Segurança Social, 2018, pp. 1-165, (p. 65).

13 Cfr. AJUNWA, Ifeoma, “Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work”, in: Work and Labor in the Digital Age, Vol. 33, Emerald Publishing Limited, 2019, pp. 61-91.

14 Cfr. BORNSTEIN, Stephanie, “Antidiscriminatory Algorithms”, in: Alabama Law Review, Vol. 70, n.º 2, 2019, pp. 518-572, (pp. 532-533).

15 Desenvolvendo este ponto, vide ADELMANN, Lena, & WIEDMER, Jennifer, “Der Einsatz von Künstlicher Intelligenz in der Rekrutierung”, in: Universität Basel Wirtschaftswissenschaftliche Fakultät, 2020, pp. 1-7, (p. 5) e TEETZ, I., “Künstliche Intelligenz im Recruiting”, in: Digital HR: Smarte und agile Systeme, Prozesse und Strukturen im Personalmanagement, Petry, T.; Jäger, W. (Hrsg.), Haufe Lexware GmbH, 2018, pp. 225-240.

16 Como os reconhecidos exemplos da *Microsoft*, *J.P. Morgan*, *Goldman Sachs*, *Credit Suisse*, *Netflix*, entre outros.

estamos perante algo que apresenta soluções mágicas ou infalíveis, até pelo contrário. Deste modo, repudiamos este fenómeno como sendo uma “aura de objetividade e infalibilidade”¹⁷ ou até de *mathwashing*¹⁸, pelo que se afasta a ideia que nos encontramos ante uma espécie de oráculo¹⁹ profético que contém nada mais que a verdade. Inversamente, é necessário não nos deixarmos iludir com a ideia de que os números, *rectius*, os algoritmos não mentem, antes atendendo efetivamente ao conteúdo, à substância que transparece uma realidade bastante diferente e, até certo ponto, assustadora. Aliás, esta ideia de confiança implícita que a maior parte das pessoas deposita no resultado dos algoritmos denomina-se por <<*algorithmic authority*>> na medida em que, a nível social, é conferido um valor superior às decisões que toma, o que muitas vezes dá azo a um certo tipo de culto, como se tratassem de “deuses” cujos comandos devem ser seguidos sem serem questionados²⁰ e este parece ser precisamente o problema: os algoritmos que estão na base do modelo de negócio de muitas empresas geram injustiças porque se baseiam em modelos matemáticos que reproduzem preconceitos e vieses humanos, o que se repercute na tomada de decisões e na automatização dos processos, i.e., a jusante. Repare-se que não surgem do nada, são criados pelos seres humanos, são ferramentas projetadas por pessoas que têm controlo sobre o seu funcionamento, daí que possam servir, ou para quebrar injustiças e promover a igualdade, ou para perpetuar a exclusão e discriminação ao espelhar o *status quo* do ser humano, o que mais se observa infelizmente na maioria das vezes ainda que provocado de forma não consciente e accidental²¹, pois para mal dos nossos pecados a verdade é que não somos agentes económicos racionais, somos tendenciosos por natureza

17 Cfr. OSOBA, Osonde A., & WELSER IV, William, An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence, RAND Corporation, 2017, pp. 1-29.

18 Como esclarece JOH, Elizabeth E. “Feeding the Machine: Policing, Crime Data, & Algorithms”, in: William & Mary Bill of Rights Journal, Vol. 26, Issue 2, 2017, pp. 287-302, (p. 292), pode ser definido como a (errónea) perceção de que modelos algorítmicos não possuem a subjetividade inata ao ser humano, suprimindo-a por se tratar de um sistema matemático.

19 Cfr. AJUNWA, Ifeoma, “The Paradox of automation as anti-bias intervention”, in: Cardozo Law Review, Vol. 41, Issue 5, 2020, pp. 1671-1742, (pp. 1688-1690).

20 Cfr. HARARI, Yuval Noah, Homo Deus: A Brief History of Tomorrow, Harper, 2017.

21 Não temos dúvidas que a regra geral é depararmo-nos com formas de discriminação não intencionais, em que os empregadores têm o sincero e legítimo interesse em selecionar os melhores candidatos para a vaga disponível, contudo o algoritmo não tem consciência para distinguir o certo do errado.

e intolerantes, de maneiras que não queríamos ser, somos preconceituosos e estamos a injetar esse preconceito no algoritmo.

Um dos principais problemas é, por isso, a tomada de decisões discriminatórias pelo algoritmo que se poderá dever: **(i)** à possibilidade dos algoritmos reproduzirem, de facto, o preconceito humano onde releva a possibilidade de surgimento de novas fontes de discriminação no acesso ao emprego. A este respeito, veja-se as decisões de recrutamento que são por vezes tomadas com base numa variedade de atributos não intuitivos como é exemplo a variável que os cientistas de *data* descobriram ter valor preditivo: a distância do domicílio até ao trabalho²². De acordo com esta variável, pessoas com viagens mais longas sofrem maiores taxas de desgaste, porém, a distância pendular está intimamente ligada ao local onde uma pessoa mora que, por sua vez, é conexas com preços da habitação que estão relacionados com a renda e, conseqüentemente, com a etnia. Deste modo, possibilitar que o algoritmo contrate com base no local onde moram as pessoas – ainda que de forma indireta não intencional - tem provavelmente um impacto adverso em grupos protegidos como as minorias raciais, sendo este um exemplo de um raciocínio algorítmico²³; **(ii)** à própria construção do algoritmo que requer dados que contêm fatores de discriminação. Com efeito, o algoritmo toma a realidade como fator de aprendizagem na hora de processar dados, o que implica que os resultados extraídos venham a perpetuar discriminação na nossa sociedade, daí que, enquanto não se proceder a uma reforma a nível de coleta do histórico e da definição do sucesso de cada empresa, o argumento de que a máquina é imparcial não passará de uma utopia, pois a verdade é que, o facto dos dados processados serem opiniões do contratante (*in casu*, o empregador) apresentam uma (intensa e clara) componente subjetiva²⁴, precisamente o inverso da

22 Cfr. CAPPELLI, Peter, “Your Approach to Hiring Is All Wrong”, in: Harvard Business Review, 2019, pp. 48-58.

23 Poderíamos mencionar outros exemplos, como a religião que pode estar estatisticamente relacionada com o código postal da cidade onde a pessoa vive, ou mesmo o tempo dedicado a redes sociais e profissionais - como o *FB* ou o *LinkedIn* – que podem indicar a correspondente afiliação política ou sindical e com isso prejudicar alguém numa determinada decisão. Nesta linha, com outros exemplos semelhantes, vide KULLMANN, Miriam, “Discriminating Job Applicants Through Algorithmic Decision-Making”, in: Ars Aequi, Vol. 68, n.º 1, 2019, pp. 45-53.

24 MILLER, Alex P., “Want Less-Biased Decisions? Use Algorithms”, in: Harvard,

ideia que querem transmitir i.e., de que são objetivos e justos atuando como profetas da imparcialidade; **(iii)** ao facto de que, quando um algoritmo está no comando, as minorias ficam geralmente em desvantagem devido à falta de informação disponível acerca das mesmas, ou seja, a ideia é de que nas minorias – pertencentes a um qualquer fator discriminatório - haverá menos dados disponíveis, o que implica que o algoritmo entenda que tomar uma decisão favorável a um grupo minoritário seja mais arriscado do que fazê-lo a favor de um grupo maioritário, ou seja, para selecionar um candidato de um grupo minoritário o algoritmo exigirá mais qualidades, habilitações e conhecimento comparativamente a um grupo maioritário simplesmente por ser mais fácil de prever estatisticamente o comportamento de alguém pertencente a este do que àquele grupo²⁵; **(iv)** à adoção de regras que, embora aparentem neutralidade, indiretamente revelam discriminação tendo um impacto desproporcional em determinado grupo, *e.g.*, eliminar candidatos que tiveram interrupções superiores a um ano na sua carreira, o que vai prejudicar de forma desproporcional as mulheres que muitas vezes se afastam temporariamente da vida profissional em razão da maternidade.

Assim, os algoritmos condenam-nos a repetir o passado do qual queríamos fugir ao replicarem preconceitos mas com uma capacidade muito superior de discriminação²⁶, de uma maneira silenciosa e muito menos transparente uma vez que, fruto de questões relacionadas com a programação e design, funcionam

Business Review, Vol. 26, 2018, dá-nos uma perspetiva interessante ao alegar que os que tecem críticas à utilidade do algoritmo esquecem-se de responder à questão de saber se, mesmo com as suas debilidades, o desempenho algorítmico se compara ao status quo humano, enunciando alguns estudos suportados por casos reais que auxiliam a conclusão de que os algoritmos são menos tendenciosos e mais precisos em termos comparativos. Pessoalmente, não podemos concordar com esta linha argumentativa uma vez que, no final do dia, o que se revela mais determinante é aferir do balanço a nível da magnitude das consequências perante um cenário negativo, e nesse campo, as inúmeras incertezas aliadas à falta de respostas inequívocas acerca dos problemas relacionados com esta questão tão sensível, leva-nos a não nos comprometer com a adoção pura de um sistema algorítmico dotado de tanto poder, não valendo a pena esse risco.

25 Cfr. TODOLÍ-SIGNES, Adrian, *op. cit.*, p. 6.

26 Cfr. LA DIEGA, Guido Noto, “Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information”, in: JIPITEC, Vol. 9, n.º 1, 2018, pp. 1-33 e COELHO MOREIRA, Teresa, “Igualdade de género no Trabalho 4.0”, in: APODIT 5: A igualdade nas relações de trabalho (coord. Maria do Rosário Palma Ramalho e Teresa Coelho Moreira), AAFDL, Lisboa, 2019, pp. 45-68, (pp. 63-67).

como uma *black box*²⁷, em que por vezes nem o próprio programador conseguirá aferir do porquê do algoritmo ter chegado a um determinado resultado que se revela controverso, enquanto reflexo da infra informação tecnológica inerente ao homem médio, pois a maioria dos programadores possui um conhecimento muito insuficiente e diminuto acerca das vicissitudes do contexto de emprego o que faz com que criem os seus modelos com parca ou pouca informação útil²⁸. Todas estas *red flags*, enquanto fatores de risco, são ainda mais sérios quando tais práticas se baseiam na IA de autoaprendizagem, com o software a ser capaz de reprogramar os seus próprios critérios e métricas para alcançar um resultado predefinido²⁹ tornando os efeitos nefastos em larga escala.

3. Indícios práticos de vicissitudes no contexto atual: os problemas ético-

27 Como apontam ANALIDE, Cesar e REBELO, Diogo Morgado, “Inteligência Artificial na era data-driven: a lógica fuzzy das aproximações soft computing e a proibição de sujeição a decisões tomadas exclusivamente com base na exploração e prospeção de dados pessoais”, in: Fórum de Proteção de Dados, n.º 6, 2019, pp. 60-91, (p. 87), “trata-se de um conceito decorrente da ciência da computação utilizado para referir sistemas de que se conhece somente os dados de entrada e de saída, sem possibilidade de acesso ao seu funcionamento interno”, existindo duas configurações comuns de testagem do software: *white box* e *black box*. Nas primeiras, o analista tem acesso ao *source code*, nas segundas, o analista está limitado a observar apenas as entradas (*inputs*) e saídas (*outputs*) do sistema, mas não o seu funcionamento interno (cfr. DESAI, Deven R., & Kroll, Joshua A., “Trust but verify: A guide to algorithms and the law”, in: Harvard Journal of Law & Technology, Vol. 31, n.º 1, 2017, pp. 1-64, (p. 36). Estamos perante a polémica discussão sobre saber se se deve abrir ou não a Caixa de Pandora que, in casu, significa aceder ou não ao código algorítmico, entrando conseqüentemente no problema de confronto com outros interesses constitucionalmente protegidos, podendo colocar em causa a segurança, o interesse económico e comercial das empresas e os *IP rights* detidos pelas entidades públicas e privadas, temática que iremos abordar no Cap. III, ponto §2.

28 Quando têm (apenas) em conta, *e.g.*, os melhores desempenhos de trabalhadores no passado de determinada empresa identificando, a partir daí, quais os candidatos a emprego que têm os mesmos atributos, o que provoca discriminação *ab initio* a montante, inquinando todo o processo decisório. Neste sentido, vide GOLDENFEIN, Jake, *op.cit.*, p. 49.

29 Acerca da (não) programação de sistemas inteligentes e as questões laterais que fazem surgir, *vide* FESTAS, David. “A contratação eletrónica automatizada”, in: Direito da Sociedade da Informação – Vol. VI, Coimbra Editora, Coimbra, 2006, pp. 411-461. Nesta linha, KULLMANN, Miriam, *op. cit.*, pp. 47-48, distingue entre modelos determinísticos ou modelos de autoaprendizagem “supervisionados” nos quais existe sempre uma decisão consciente (humana) que afeta o algoritmo e um modelo de dados que este utiliza para resolver o problema específico que podem ser ajustados e, por outro lado, os modelos “não supervisionados” que são capazes de se adaptar e determinar por si próprios quais os dados que irão utilizar para o efeito.

-jurídicos das decisões algorítmicas

A questão passa então por saber se, na prática, com os algoritmos, somos capazes de eliminar o preconceito humano ou se simplesmente o camuflamos com a tecnologia a partir do momento em que estamos perante uma intensificação do controlo eletrónico praticamente ilimitado e ainda mais intrusivo para a privacidade, chegando a falar-se em *algocracy*³⁰ ou <<trabalhador algoritmo>> podendo correr-se o risco de uma coisificação da pessoa humana. Isto leva alguma doutrina³¹ a recear pela ausência de liberdade e democracia num futuro próximo sendo este o (infeliz) Estado de Arte dos dias correntes, em que os preocupantes efeitos perniciosos do algoritmo ofuscam as vantagens da implementação de uma sociedade pós-industrial informatizada³².

3.1. A discriminação

Primo, o primeiro caso real relatado de um algoritmo discriminatório que remonta à década de 1970, em que a Faculdade de Medicina do Hospital de *St. George*, no sul de Londres, recebia mais de doze candidaturas para cada uma das suas 150 vagas anuais. Ora, à data, grandes organizações como o Pentágono, já utilizavam sistemas inteligentes para tratar de candidaturas, mas para um Hospital criar o seu próprio programa de avaliação automatizada no final dos anos 70 representou uma experiência ousada, o que se revelou um fracasso total. *St. George* foi o pioneiro involuntário e inconsciente destas *WMDs* em

30 Ao fim ao cabo, pegando nas palavras de ROTA, Anna, “Rapporto di lavoro e big data analytics: profili critici e risposte possibili”, in: *Labour & Law issues*, Vol. 3, n.º 1, 2017, pp. 32-52, (p. 35), assiste-se a um conjunto de dinâmicas cuja força disruptiva alimentada por sistemas computacionais e algorítmicos fomenta a difusão crescente de um novo modelo de governação.

31 Cfr. TEJERO, Emilio L., “Algoritmos. El totalitarismo determinista que se avecina. ¿La pérdida final de libertad?”, in: *Revista de Pensamiento Estratégico y Seguridad CISDE*, Vol. 5, n.º 1, 2020, pp. 85-101.

32 Termo utilizado por COELHO MOREIRA, Teresa, “Algumas notas sobre as novas tecnologias de informação e comunicação e o contrato de teletrabalho subordinado”, in: *Scientia Iuridica*, Vol. 64, n.º 335, 2014, pp. 323-343, (pp. 323-324) para caracterizar esta nova Era de mutação económica radical fruto da inserção das NTIC ao provocar um gradual desaparecimento das fronteiras técnicas que anteriormente permitiam distinguir os diferentes sectores da economia, uma vez que se assiste a uma perda de protagonismo dos sectores primário, secundário e terciário para o setor quaternário da informação.

que os objetivos de eficiência e justiça não foram cumpridos. Tendo o Hospital registos volumosos de candidaturas rejeitadas dos anos anteriores, o trabalho era ensinar ao sistema algorítmico como replicar os mesmos procedimentos que os seres humanos estavam a seguir, sendo que os respetivos *inputs* foram precisamente o problema pois foi através deles que o algoritmo aprendeu como discriminar e realizou esse trabalho de maneira extremamente eficiente na medida em que, para além da discriminação contida no *input* ser racista e xenófoba, um número considerável de candidaturas com nomes ou endereços estrangeiros veio de pessoas que claramente não dominavam o idioma inglês, e em vez de considerar a possibilidade de bons médicos aprenderem a língua, a tendência era simplesmente de os rejeitar automaticamente. De igual modo, interligaram as candidaturas rejeitadas do passado com locais de nascimento e apelidos fazendo com que pessoas oriundas de certos locais, como África, Paquistão ou bairros de imigrantes, recebessem uma pontuação mais baixa sendo, conseqüentemente, excluídas do processo³³. Na mesma linha, também se rejeitavam as candidatas com a justificação muito comum à época de que as suas carreiras provavelmente seriam interrompidas pelos deveres da maternidade, e o algoritmo, conseqüentemente, fez o mesmo juízo. Por todo este conjunto de razões, a Comissão de Igualdade Racial do governo britânico considerou, em 1988, o Hospital culpado de discriminação racial e de género na sua política de admissões³⁴, estimando a Comissão que, no mínimo, cerca de sessenta dos dois mil candidatos viram, anualmente, uma entrevista ser-lhes rejeitada derivado de um fator de discriminação.

33 A este respeito, *vide* o estudo de BANERJEE, Rupa, OREOPOULOS, Phil, & REITZ, Jeffrey G., “Do Large Employers Treat Racial Minorities More Fairly? A New Analysis of Canadian Field Experiment Data”, in: Canadian Public Policy, University of Toronto Press, Vol. 44, n.º 1, 2018, pp. 1-12, cuja investigação demonstrou que, candidatos com nomes chineses, indianos ou paquistaneses tinham cerca de 28% menos probabilidades de obter um convite para uma entrevista de emprego do que os candidatos que tinham nomes com sonoridade inglesa, independentemente de terem as mesmas qualificações. Na mesma linha, *vide* o interessante estudo de BERTRAND, Marianne & MULLAINATHAN, Sendhil, “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination”, in: American Economic Review Vol. 94, n.º 4, 2004, pp. 991-1013.

34 A nível de conseqüências do caso, com maior desenvolvimento, *vide* BAROCAS, Solon & SELBST, Andrew D., “Big Data’s Disparate Impact”, in: California Law Review, Vol. 104, 2016, pp. 671-732, (p. 682); GARCIA, Megan, “Racist in the machine: The disturbing implications of algorithmic bias”, in: World Policy Journal, Vol. 33, n.º 4, 2016, pp. 111-117 e O’NEIL, Cathy, *op. cit.*, pp. 115-122.

Secondo, o enigmático caso recente da *Amazon* e do seu algoritmo sexista³⁵ que, ao replicar as tendências passadas de contratação na empresa excluía as mulheres. Isto sucedia porque os modelos algorítmicos foram treinados para vetar candidatos através da observação de padrões nos *CVs* submetidos à empresa num período temporal de cerca de 10 anos que apontava para um universo de trabalhadores predominantemente do sexo masculino enquanto reflexo do mercado quase monopolizado por este género na indústria tecnológico-digital³⁶. Desta forma, o modelo afeto ao recrutamento aprendeu a penalizar currículos que, *e.g.*, continham palavras ou expressões associadas ao sexo feminino, como “capitã de voleibol”, e concedeu classificações mais baixas a licenciados de universidades com nomenclaturas femininas como *Maryland University* ou *University of North Carolina*³⁷, fazendo notar que a ferramenta que pode fazer milagres pelas empresas no momento de identificar talento, pode também ser uma fonte de discriminação para a qual a maioria dos sistemas jurídicos ainda não está preparado para responder. A falta de imparcialidade na abordagem de contratação serviu assim como um *wake up call* ao demonstrar o quão arriscado é para as empresas deixar as decisões de contratação exclusivamente nas mãos da tecnologia.

35 A respeito do fator género, é de mencionar o Relatório OIT, “Trabalhar para um futuro melhor”, Genebra, 2019, disponível em: <https://www.ilo.org/wcmsp5/groups/public/-/europe/---ro-geneva/---ilo-lisbon/documents/publication/wcms_677383.pdf>, pp. 35-36, ao apelar a “uma agenda transformadora e mensurável para a igualdade de género no contexto do futuro do trabalho, [na medida em que] os algoritmos utilizados no preenchimento das vagas de emprego [se têm] revelado perpetuadores de assimetrias de género [podendo reproduzir tendências e preconceitos históricos, pelo que] recomendamos a adoção de medidas específicas para garantir a igualdade de oportunidades (...) nos empregos que serão criados com as novas tecnologias”.

36 Aliás, em termos globais, tal disparidade é bem notória na contratação de pessoal para cargos técnicos, tal como apresentado nos dados estatísticos por DASTIN, Jeffrey, “Amazon scraps secret AI recruiting tool that showed bias against women”, Reuters, 2018, disponível em: <<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>>, relativamente a cargos exercidos em grandes empresas como a *Microsoft*, a *Google* e a *Apple*.

37 Para maiores desenvolvimentos acerca desta forma de discriminação com origem pura e simplesmente nas expressões recolhidas e interpretadas pela *ML*, *vide* BOLUKBASI, Tolga, CHANG, Kai-Wei, KALAI, Adam, SALIGRAMA, Venkatesh & ZOU, James, “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings”, in: NIPS’16: Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, 2016, pp. 4356-4364.

Terzo, a discriminação de *target ads*³⁸, i.e., quando a publicidade é direcionada especificamente para determinado público-alvo, deixando outros de fora. Cabe formular uma questão de fundo: quando procuramos emprego online podemos ter a certeza de que estamos a visualizar todas as ofertas disponíveis para as nossas qualificações? Podemos confiar na tecnologia e ter a garantia de que não seremos excluídos de um processo de seleção por algum tipo de discriminação tecnológica invisível a olho nu? A resposta é negativa, todo o ser humano tem um determinado comportamento que leva o algoritmo a mostrar-lhe o que poderá ser relevante subsequentemente a ter sido criado um perfil sobre si, tendo sido uma situação semelhante que levou um grupo de mulheres a interpor, no final de 2019, um processo contra o *FB* no Tribunal Federal de São Francisco nos EUA³⁹, no qual, alegadamente, nove empresas contratantes discriminavam por género, idade e etnia nas ofertas de emprego divulgadas através da rede social. Essencialmente, discutiu-se a possibilidade que o *FB* conferiu às empresas de publicar estes *targeted ads*, que mesmo não contendo na descrição da oferta qualquer fator discriminatório eram exibidos, por decisão dolosa ou negligente do próprio *FB*, apenas a um determinado perfil de utilizadores, consoante a respetiva etnia, idade ou género, o que se revelou preocupante inclusivamente pela dificuldade probatória, tendo o *FB* eventualmente alterado os termos de serviço e erro na programação dos *target ads*, restringindo a sua utilização e tornando mais transparente o modelo de funcionamento na sequência de um acordo extrajudicial alcançado.

38 Para maior desenvoltura neste campo, *vide* os estudos de BLASS, Joseph, “Algorithmic Advertising Discrimination”, in: *Northwestern University Law Review*, Vol. 114, Issue 2, 2019, pp. 416-468 e de LAMBRECHT, Anja & TUCKER, Catherine, “Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads”, in: *Management Science*, Vol. 65, Issue 7, 2019, pp. 2966-2981, (pp. 2974-2975).

39 U.S. District Court for the Northern District of California, Facebook, Inc. Consumer Privacy User Profile Litigation, n.º 18-md-02843-VC, 13.12.2019. Mas existem outros, *e.g.*, o caso *Bradley v. T-Mobile US, Inc.*, n.º 5:17-cv-07232 (N.D. California. Aug. 20, 2018), onde os queixosos alegaram que empresas como a *T-Mobile* e a *Amazon* agiram de modo semelhante ao direcionarem os seus anúncios de emprego para o *FB* de uma forma que exclui os trabalhadores mais velhos.

3.2 A parcialidade e arbitrariedade

Veja-se o curioso caso do algoritmo de avaliação de professores⁴⁰. No Estado de NY (e em grande parte dos Estados norte-americanos⁴¹), os professores são avaliados pelo denominado *Value-Added Model* (MVA), sendo que a cada professor é dada uma pontuação no final do ano letivo de 0 a 20 pontos consoante – e esta é a pedra de toque – o crescimento previsível da prestação dos alunos. Contudo, este modelo algorítmico revelou-se arbitrário, injusto e cego no que diz respeito a recursos e apelações, tanto que foi intentada uma ação por uma professora da quarta classe no Supremo Tribunal de NY em 2016⁴² uma vez que, no ano letivo 2012-2013 tinha recebido uma pontuação de crescimento previsível de 14 em 20 pontos, ou seja, uma classificação designada como “Eficaz”, significando que tinha atingido a média do Estado para estudantes semelhantes. Mas no ano letivo seguinte, de 2013-2014, tinha recebido uma pontuação de 1 em 20 pontos, ou seja, uma classificação designada como “Ineficaz”, significando que estava muito abaixo da média do Estado para estudantes semelhantes. A autora confirmou com o Departamento de Educação do Estado de NY que nem um recurso administrativo nem um recurso ao Comissário da Educação estavam disponíveis para contestar a pontuação que lhe foi atribuída pelo algoritmo e a essência do seu argumento

40 Para melhor contextualizarmos este importante caso real por se tratar de um dos pioneiros em matéria de caso julgado, salienta-se que o MVA teve a sua origem em 1983 numa altura em que a administração Reagan emitiu um alarme sobre o estado das escolas americanas em que num relatório da National Commission on Excellence in Education, “A Nation at Risk: The Imperative for Educational Reform”, 1983, disponível em: <https://edreform.com/wp-content/uploads/2013/02/A_Nation_At_Risk_1983.pdf>, o painel presidencial alertou que existia uma maré crescente de mediocridade nas escolas que ameaçava o próprio futuro como nação e povo, sendo que um dos sinais mais notáveis de fracasso foi o número de classificações em queda nos *SAT's* entre 1963 e 1980, cujo relatório não deixava dúvidas de que os culpados seriam os professores. Nesta linha, *vide* o igualmente curioso caso *Houston Federation of Teachers Local 2415, et al. v. Houston Independent School District, Civil Action H-14-1189* (S.D. Tex. 2017). Acerca de outros casos de utilização deste modelo, *vide* O'NEIL, Cathy, op. cit., pp. 135-140 e TOYAMA MIYAGUSUKU, Jorge e RODRÍGUEZ LEÓN, Ariana, “Algoritmos laborales: big data e inteligencia artificial”, in: *Revista Themis*, Vol. 75, 2019, pp. 255-266, (p. 262).

41 Atualmente, estima-se que são cerca de 40 Estados segundo O'NEIL, Cathy, op. cit., pp. 135-140.

42 *State of NY Supreme Court, Sheri G. Lederman v. John B. King, Jr., & Candace H. Shyer*, RJI n.º 01-14-ST6183, 2016.

assentava no facto do sistema “castigar” injustamente os professores cujos alunos obtêm de forma consistente classificações substancialmente acima dos padrões do Estado, salientando que, para o ano letivo onde obteve 1 em 20 pontos (de 2013-2014), cerca de 67% dos seus alunos cumpriram ou excederam os padrões da disciplina de Matemática e Inglês que lecionava, sendo que a média do Estado de NY foi de 31%, pelo que ficou estupefacta como seria possível ter recebido esta pontuação. Mais incrédula ficou quando observou que, para o ano letivo anterior os seus alunos pontuaram cerca de 69% para ambas as disciplinas, o que não justificava a descida pontual escandalosa que obteve, sendo aliás elogiada no seio académico pelos seus pares bem como pelos seus discentes, pela qualidade do seu ensino.

Na sua decisão, o Supremo Tribunal de NY decretou que, primariamente, o ónus de estabelecer a existência de um padrão arbitrário da parte do algoritmo recaía sobre a autora em razão de constituir precedente norte-americano⁴³, acabando por concordar com a mesma de que, de facto, a pontuação que lhe foi atribuída de 1 em 20 pontos foi arbitrária por uma pluralidade de razões⁴⁴ sendo de extrema importância para o nosso estudo salientar apenas uma que apresenta, conseqüentemente, várias ramificações: a existência de uma prova convincente e detalhada da existência de preconceito do modelo algorítmico contra alguns professores em ambos os polos do espectro (com alunos de alto desempenho e com alunos de baixo desempenho), que se traduziu: **(i)** na incapacidade dos estudantes de alto desempenho demonstrarem crescimento semelhante a estudantes de baixo desempenho (i.e., se já estavam no topo das suas habilitações, não conseguiram melhorar, sendo exatamente isto que é exigido por este modelo algorítmico); **(ii)** em toda a oscilação não explicada na pontuação da autora, de um ano letivo para outro, de 14 para 1 em 20 pontos possíveis apesar da presença de alunos com pontuação estatisticamente semelhante nas suas respetivas aulas; **(iii)** principalmente, no facto de ter ficado provado que, segundo o sistema algorítmico de NY, as percentagens de professores que se enquadram em quatro categorias diferentes de pontuação se encontrarem predeterminadas: 7% estarão na categoria “Altamente Eficaz”,

43 Cfr. Johnson Elec. Const. Corp. v. NY State Dept. of Transp, 124^a.D.3d1199.1200 (3rd Dept. 2015).

44 Id. 43, pp. 11-12.

77% na categoria “Efetiva”, 9% na categoria “Desenvolvimento” e 7% na categoria “Ineficaz”, sendo que estas quatro categorias permanecerão intactas de ano para ano, independentemente do desempenho dos alunos ter aumentado ou diminuído drasticamente em relação ao ano precedente, concluindo que o referido modelo algorítmico seria, portanto, irracional.

Porém, existiu um *twist* que nos parece um problema grave que se deveu ao facto de, quando o tribunal teve a oportunidade de colocar um entrave, balizar e estabelecer regras relativamente à discricionariedade e arbitrariedade de um algoritmo (especialmente nos EUA fruto da extrema importância dos precedentes), não o fez. É que apesar de decidir a favor da autora, o tribunal não deixou de realçar o facto de que não possui formação educacional, recursos ou tempo para propor uma alternativa plausível ao modelo algorítmico ou correções sólidas para quaisquer falhas do referido sistema e, como tal, não fez qualquer análise crítica sobre o assunto⁴⁵, acabando, aliás, por não apresentar quaisquer outras consequências fruto da decisão que tomou, continuando o algoritmo a vigorar normalmente.

4. As decisões individuais automatizadas no RGPD

O art. 22º RGPD, complementado pelo Considerando (71), consagra o direito à não sujeição a *ADMs* baseadas nos dados pessoais do titular⁴⁶ procurando combater a discriminação algorítmica, algo reconhecido como direito fundamental da UE nos termos do n.º 1 do art. 21º CDFUE. Salientamos uma primeira nota: a Lei de Execução manteve-se silente quanto ao assunto não se pronunciando, e uma vez que os regulamentos europeus são atos legislativos dotados de obrigatoriedade em todos os seus elementos, revelam-se como diretamente aplicáveis na ordem interna e, por isso, vinculativos, de acordo com o art. 288º TFUE, pelo que o RGPD desempenhará um papel importante na forte harmonização dos países⁴⁷ enquanto único regime *ad*

45 Id. 43, p. 11.

46 Como resultado da ratificação do RGPD, a noção de dados pessoais em vigor no ordenamento jurídico português consiste na “informação relativa a uma pessoa singular identificada ou identificável”, nos termos do n.º 1 do art. 4º RGPD.

47 Podemos identificar diferentes abordagens legislativas no seio interno dos países acerca das *ADMs* na sequência da entrada em vigor do RGPD nos termos da alínea b) do n.º 2 do

hoc contra decisões algorítmicas. Tendo isto presente, é de realçar que o termo “direito”, empregue nesta disposição, não significa que o preceito seja aplicável somente quando ativamente invocado pelo titular dos dados, pelo que entendemos que o preceito estabelece uma verdadeira proibição geral⁴⁸ da

art. 22º, que refletem a cultura histórica da temática da proteção de dados de cada ordenamento, cuja intensidade varia de acordo com o uso, maior ou menor, das tecnologias. *Primo*, verifica-se uma abordagem negativa em que o E-M não prevê nenhum caso específico de tomada de decisão automatizada permitida, sendo o caso mais usual e exemplo disso Portugal, Itália, Roménia, Suécia, Dinamarca, Polónia, Finlândia, Chipre, Grécia, República Checa, Estónia, Lituânia, Bulgária, Letónia, Croácia, Luxemburgo, Malta, Lituânia, Eslováquia ou Espanha. *Secundo*, uma abordagem neutra, na qual o E-M implementou algo, mas não propõe nenhuma medida específica adequada para salvaguardar os direitos, liberdades e interesses legítimos da pessoa em causa, como a Alemanha e parcialmente, a Áustria e a Bélgica. *Terzo*, uma abordagem mais procedimental, em que alguns E-M proporcionam salvaguardas específicas que se baseiam principalmente numa descrição dos procedimentos que os responsáveis pelo tratamento de dados devem adotar quando executam *ADMs* sobre indivíduos (e.g., notificação, revisão, etc.) ou algumas formas de avaliação do impacto do algoritmo, sendo disso exemplo o UK, a Irlanda e parcialmente, a Eslovénia. Por fim, temos uma abordagem pró-ativa, na qual alguns E-M propõem novas salvaguardas e mais específicas, como o direito de conhecer os parâmetros de ponderação dos algoritmos, sendo disso exemplo o caso Francês e Húngaro. Reunindo a mais recente das legislações apontadas que comprovam a nossa afirmação, *vide*, por todos, MALGIERI, Gianclaudio, “Automated Decision-Making in the EU Member States: The Right to Explanation and Other ‘Suitable Safeguards’ for Algorithmic Decisions in the EU National Legislations”, in: *Computer Law & Security Review*, Vol. 35, Issue 5, 2018, pp. 1-40, (pp. 8-10; pp. 28-29 e notas rodapé 42-60).

48 Acolhemos de perto a interpretação preconizada pelo GT29º, “Guidelines on Automated Individual Decisionmaking and Profiling for the purposes of Regulation 2016/679”, Adopted on 3 October 2017, last Revised and Adopted on 6 February 2018, disponível em: <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053>, pp. 19-20. Aliás, a sua redação parece deixar entender, de forma implícita, que é esse o objetivo sendo complementada pelo Considerando (71) ao mencionar que, “no entanto, a tomada de decisões com base nesse tratamento, incluindo a definição de perfis, deverá ser permitida se expressamente autorizada pelo direito da União ou [E-M] (...), ou se for necessária para a celebração ou execução de um contrato (...), ou mediante o consentimento explícito do titular”. Contudo, não é unânime na doutrina, HORDERN, Victoria & USTARAN, Eduardo, “Automated Decision-Making Under the GDPR - A Right for Individuals or A Prohibition for Controllers?”, in: *Chronicle of Data Protection*, 2017, rejeitam que esta secção deva ser interpretada como uma proibição com o argumento de que se os legisladores da UE pretendessem que esta disposição se tratasse de uma proibição geral, poderiam (e deveriam) ter incluído uma linguagem mais clara neste preceito tendo inversamente utilizado uma linguagem mais ambígua de que “o titular dos dados tem o direito de não ser sujeito a uma decisão”, o que não indica se o direito a estar livre deste tipo de decisão se aplica *ex ante* ou *ex post*. De igual modo, é invocado um argumento histórico dado que, em 2016, o texto final do RGPD seguiu a abordagem do Conselho qualificando-a como uma exceção, o que faz desta alteração uma rejeição deliberada da proibição. Nesta linha, MEYER, David, “Did the WP29 Misinterpret the GDPR on Automated Decision-Making?”, in: *IAPP*, 2017, adiciona um argumento sistemático baseado no facto do art. 22º ser colocado no Cap. III centrado nos

tomada de decisões com base exclusivamente no tratamento automatizado⁴⁹ “que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar” (cfr. art. 22º/1) e sem prejuízo das exceções contempladas no RGPD (cfr. art. 22º/2), ou seja, caso não haja nenhuma intervenção humana no processo decisório e, por exemplo, a intervenção humana se limite à aplicação da decisão tomada pela máquina sem qualquer influência sobre o resultado⁵⁰, daí que os trabalhadores se encontrem automaticamente protegidos dos

direitos, o que sugere que os legisladores pretenderam que fosse interpretado como um direito que pode ser reivindicado pelas pessoas ao invés de proibir determinadas decisões. Porém, os críticos deste último ponto de vista podem argumentar que existem várias disposições no Cap. III que simultaneamente impõem obrigações aos responsáveis pelo tratamento e concedem direitos às pessoas em causa, como é exemplo disso o art. 12º segundo o qual, “o responsável pelo tratamento tomará as medidas adequadas para fornecer qualquer informação”, sendo que os arts. 13º e 14º explicam o que o responsável pelo tratamento “deve” fazer. Para mais, é de relembrar, a par de PEHRSSON, Emily, “The Meaning of the GDPR Article 22”, *in: Stanford-Vienna European Union Law, Working Paper*, n.º 31, 2018, pp. 1-32, (p. 21), que o RGPD se baseou no art. 15º da DPD de 1995 cuja maior diferença assenta nas possibilidades de interrogações ao direito, e embora existissem pequenas variações entre os E-M, foram vários os que interpretaram este art. como uma proibição geral, entre os quais Portugal, apoiando a nossa interpretação.

49 O responsável pelo tratamento não pode eximir-se do disposto no art. 22º fabricando uma intervenção humana, pelo que se uma pessoa acabar por examinar e ponderar outros fatores ao tomar a decisão final, já não se preencherá o preceito. Assim, para que se considere haver uma intervenção humana, o responsável pelo tratamento tem de garantir que qualquer supervisão da decisão seja relevante e não um mero gesto simbólico, daí a restrição do escopo do preceito. Essa supervisão deve ser levada a cabo por alguém com autoridade e competência para alterar a decisão e que, no âmbito da análise, deverá tomar em consideração todos os dados pertinentes. Neste sentido, FLORIDI, Luciano, MITTELSTADT, Brent, & WACHTER, Sandra, “Why a Right to Explanation of Automated Decision Making Does Not Exist in the General Data Protection Regulation”, *in: International Data Privacy Law*, Vol. 7, Issue 2, 2017, pp. 1-47, (p. 34), invocam um argumento histórico de que existiu uma clara intenção dos legisladores de excluir a palavra “predominantemente” e adotar uma interpretação restrita da palavra “exclusivamente” (*solely*). Partilhando desta opinião, *vide* GT29º, *op. cit.*, p. 23, a fim de saber quando é que o nível de intervenção humana é considerado significativo, e também os pareceres do ICO, “Feedback request - profiling and automated decision-making”, 2017, disponível em: <<https://ico.org.uk/media/2013894/ico-feedback-request-profiling-and-automated-decision-making.pdf>> e “Guide to the General Data Protection Regulation”, 2018, disponível em: <<https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf>>. Contra: BAYAMLIOGLU, Emre, “Transparency of Automated Decisions in the GDPR: An Attempt for systemisation”, *in: SSRN*, 2018, pp. 1-50, (pp. 41-48), defendendo que aferir acerca do grau de intervenção humana que pode ser permitido não se encontra bem clarificado o que impossibilita de dar uma resposta única.

50 Na esteira de TODOLÍ-SIGNES, Adrian, *op. cit.*, pp. 7-8, pode adotar-se a abordagem filosófica de que um algoritmo não tem qualquer vontade real, logo, não toma decisões, mas produz resultados, pelo que será sempre um ser humano a tomar (no sentido de aplicar) as decisões.

possíveis efeitos deste tipo de tratamento. Resumidamente, o art. 22.º dispõe o seguinte: (i) em regra, existe uma proibição geral das decisões individuais totalmente automatizadas, incluindo a definição de perfis com efeitos jurídicos ou similarmente significativos⁵¹ (n.º 1); (ii) há exceções a essa regra (n.º 2); (iii) sempre que se aplique uma dessas exceções, devem existir medidas para salvaguardar os direitos e liberdades, assim como os legítimos interesses do titular dos dados (n.º 3); (iv) existe uma espécie de exceção à exceção⁵², i.e., o titular pode estar sujeito se tiver dado o seu consentimento explícito⁵³ ou se

51 Embora o RGPD não defina estes dois conceitos, o GT29º, *op. cit.*, pp. 21-23 entende que o art. 22.º abrange apenas os efeitos com impactos graves, ou seja, os seus efeitos devem ser suficientemente grandes ou relevantes para merecerem atenção. Quer isto dizer que a decisão deve ser suscetível de (i) afetar significativamente as circunstâncias, o comportamento ou as escolhas das pessoas em causa; (ii) ou ter um impacto prolongado ou permanente no titular dos dados; (iii) ou nos casos mais extremos, dar origem a uma exclusão ou discriminação das pessoas, sendo que um dos exemplos dados pelo GT29º é precisamente o de decisões que impeçam o acesso de uma pessoa a uma oportunidade de emprego ou a coloquem em séria desvantagem. Relativamente à situação “que o afete significativamente de forma similar”, o Considerando (71) apresenta alguns exemplos típicos, sendo de realçar a expressão, para o que nos interessa, “práticas de recrutamento eletrónico sem qualquer intervenção humana”.

52 Neste sentido, GONÇALVES, Carlos Jorge, e PINHEIRO, Alexandre Sousa, em anotação ao art. 22º RGPD, *in: Comentário ao Regulamento Geral de Proteção de Dados* (coord. Alexandre Sousa Pinheiro), Almedina, Coimbra, 2018, pp. 389-390. De igual modo, *vide* a interpretação restrita do GT29º, *op. cit.*, p. 24 quanto às *ADMs* que digam respeito a categorias especiais de dados pessoais invocando o preenchimento cumulativo de determinados pressupostos. Contra esta redação, *vide* BALDINI, Davide, “Article 22 GDPR and prohibition of discrimination: An outdated provision?”, *in: Privacy & GDPR*, 2019, ao defender que o seu elemento literal fica aquém de proteger de forma eficaz o direito à não discriminação uma vez que o simples facto de remover categorias especiais de dados não impediria a ocorrência de discriminação, propondo uma solução que acompanhamos que assenta numa interpretação baseada nos direitos na medida em que esta disposição deve ser interpretada como abrangendo não só os dados que imediatamente revelam pertencer a uma categoria especial (*e.g.*, o país de nascimento, que pode revelar diretamente a origem étnica), mas também os dados que indiretamente revelam essa pertença no contexto do processo de decisão algorítmico. No seguimento desta interpretação mais ampla, este último tipo de dados deve ser excluído ou tratado de forma a diminuir a sua correlação com categorias especiais a fim de evitar qualquer resultado discriminatório, o que parece estar de acordo com a jurisprudência do TJUE que tem sustentado, de forma reiterada, que as disposições de proteção de dados devem ser interpretadas de modo a alcançar a sua (maior) eficácia na defesa dos direitos fundamentais (cfr. TJUE, *Google Spain SL, Google Inc. v. AEPD, Mario Costeja González*, de 13.05.2014, Case C-131/12).

53 Parece suficiente a definição de consentimento prevista no ponto 11) do art. 4º e no Considerando (32), não esquecendo que o silêncio não vale como declaração de vontade segundo o RGPD. Criticando a declaração de consentimento do candidato a emprego enquanto fundamento adequado garante da licitude do tratamento ao abrigo do art. 6º RGPD por parte do responsável pelo recrutamento, *vide* DUARTE, Tatiana, em anotação ao art. 88º RGPD, *in:*

o tratamento for necessário por motivos de interesse público de relevo e sejam aplicadas as medidas adequadas, baseadas em tratamentos automatizados (n.º 4). Contudo, o que é certo é que alguma doutrina ainda se encontra desconfiada relativamente à clareza da interpretação do preceito, existindo quem teça severas críticas ao próprio articulado⁵⁴.

De destacar ainda o *profiling*⁵⁵ que consiste na avaliação dos aspetos pessoais de uma pessoa singular envolvendo algum tipo de apreciação ou juízo

Comentário ao Regulamento Geral de Proteção de Dados (coord. Alexandre Sousa Pinheiro), Almedina, Coimbra, 2018, p. 671, chegando a referir em PINHEIRO, Alexandre Sousa, *Privacy e proteção de dados: a construção dogmática do direito à identidade informacional*, AAFDL, Lisboa, 2015, p. 812, que “a sacralização do consentimento constitui uma das ilusões mais correntes na história da proteção de dados [adquirindo] características de puro logro” quando aplicado às NTIC. Nesta linha, acerca das especificidades relativas ao consentimento e à sua relação com as ADM e com o *profiling* a partir de uma análise minuciosa interpretativa do RGPD, relacionadas com as orientações do GT29º, úteis relativamente ao candidato a emprego acerca de saber quando e como deve ser usado, *vide* por todos, BAYAMLIOGLU, Emre, *op. cit.*, pp. 33-35 e DENCİK, Lina, *et al.*, *op. cit.*, p. 466. Contudo, é de realçar que o Considerando (43) afirma não ser possível aceitar a legalidade do processamento de dados com base no consentimento numa relação em que existe um forte desequilíbrio de poder entre as partes e embora RE se refira especificamente à relação entre as Administrações Públicas e os cidadãos, tal consideração parece ser perfeitamente aplicável à relação de emprego. De facto, nas suas orientações sobre o consentimento, o GT29º, “Guidelines on Consent under Regulation 2016/679”, Adopted on 28 November 2017, last Revised and Adopted on 10 April 2018, disponível em: <<https://ec.europa.eu/newsroom/article29/items/623051>>, p. 8 realça que é difícil que ele seja dado em conformidade com os requisitos, principalmente o que exige a forma livre, pelo que dada a natureza da relação entre empregador e trabalhador, o consentimento dado por este não deve ser entendido como válido enquanto regra geral, devendo apenas ser aceite como tal em circunstâncias excecionais. Sobre esta problemática, de forma bastante esclarecedora, *vide* GT29º, *op. cit.*, e a Deliberação interpretativa CNPD 494/2019, de 3 setembro.

54 Por todos, *vide* a análise de EDWARDS, Lilian & VEALE, Michael, “Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling”, *in: Computer Law & Security Review*, Vol. 34, Issue 2, 2018, pp. 398-404 ao contraporem os principais argumentos invocados pela doutrina europeia e, da mesma maneira a análise de HINTZE, Mike, “Automated Individual Decisions to Disclose Personal Data: Why GDPR Article 22 Should Not Apply”, *in: SSRN*, 2020, pp. 1-15, (p. 3), para quem esta redação levanta mais questões do que aquelas a que responde, e também, FERREIRA, Afonso José, “Profiling e algoritmos autónomos: um verdadeiro direito de não sujeição?”, *in: Anuário da Proteção de Dados*, (coord. Francisco Pereira Coutinho e Graça Canto Moniz), CEDIS, 2018, pp. 35-43 ao densificar o porquê deste “direito” não ter qualquer utilidade prática na realidade.

55 São três as formas possíveis de o utilizar: (i) definição geral de perfis, (ii) tomada de decisões com base na definição de perfis, e (iii) decisões exclusivamente automatizadas, onde se inclui a definição de perfis como disposto no n.º 1 do art. 22º, sendo que a diferença residirá no nível de intervenção humana consoante determinado caso (cfr. Considerando (30)).

sobre a mesma, cuja aceção inscrita no ponto 4) do art. 4º RGPD utiliza a expressão “qualquer forma de tratamento automatizado” e não um tratamento “exclusivamente” automatizado como no art. 22º podendo implicar algum tipo de intervenção humana, pelo que se revela importante lembrar que existem diferenças entre ambos os regimes⁵⁶.

4.1. O direito a obter uma intervenção humana

Recuperando alguns exemplos reais, o que é facto é que empresas como a *Amazon* ou a *Uber* não têm apenas o problema da discriminação na fase de entrada, *ab initio*, sendo de igual modo notória a carência de uma *human in command approach*, i.e., da necessidade da última palavra ser da pessoa e não da máquina na fase final da relação jurídico-laboral. Quanto ao primeiro⁵⁷, constatou-se que o algoritmo distribuía tarefas aos trabalhadores, media a sua velocidade e diligência por via de um rastreio das taxas de produtividade⁵⁸ e, muito importante, gerava automaticamente avisos ou mesmo despedimentos consoante a qualidade ou produtividade do trabalhador, sem qualquer intervenção dos supervisores. Alguns trabalhadores evitavam deslocar-se à casa de banho - fazendo as necessidades em garrafas de plástico - de modo a manter o seu tempo em paridade com as expectativas e cerca de 55% reportaram sofrer a determinado ponto de depressão. Quanto ao segundo⁵⁹, verificou-se que

⁵⁶ Embora tenham vários pontos de contacto entre si, para um desenvolvimento das disparidades das duas figuras, *vide* os pontos enunciados pelo GT29º, *op. cit.*, *maxime* o âmbito de aplicação de cada um, dado que o regime do art. 22º pode acabar por se sobrepor parcialmente à definição de perfis ou resultar da mesma.

⁵⁷ De realçar a este respeito um relato inédito do jornalista britânico BLOODWORTH, James, *Hired: Six Months Undercover in Low-Wage Britain*, Atlantic Books, 2018, que atuou disfarçadamente como um trabalhador no armazém da *Amazon* e como condutor da *Uber*, “desmascarando” e trazendo à discussão as condições em que são prestados os respetivos trabalhos associados ao poderio tecnológico digital.

⁵⁸ Aliás, o sistema foi mais longe ao rastrear o *time off task* na medida em que, se os trabalhadores fizessem descansos e intervalos prolongados, o sistema gerava avisos automaticamente e, eventualmente, o trabalhador poderia ser despedido (cfr. *id.* 57).

⁵⁹ *Id.* 58. A este respeito, *vide* a importante investigação levada a cabo por ROSENBLAT, Alex, *Uberland: How Algorithms Are Rewriting the Rules of Work*, University of California Press, 2018, no qual documentou inúmeras queixas dos motoristas da *Uber* por mau funcionamento da plataforma. Especificamente, pela empresa utilizar um sistema de resposta automática contra incidentes e reclamações de trabalhadores, chamando a atenção para o fraco desempenho algorítmico em responder às questões dos trabalhadores - quando o faz - de forma

uma avaliação mais baixa realizada nestas novas formas de prestar labor tem consequências diretas em termos de manutenção do posto de trabalho⁶⁰, em que o algoritmo permite que o condutor aceite a viagem (num curto espaço de 15 segundos) existindo no final uma avaliação através da plataforma cabendo ao sistema automático aferir quais os condutores que poderão ser suspensos ou mesmo convidados a deixar a empresa por não terem aceite um número suficiente de viagens, ou por terem uma pontuação baixa dada pelos utentes.

Face a esta (infeliz) realidade, encontramos algo de essencial no n.º 3 do art. 22º RGPD⁶¹ e complementarmente no Considerando (71), nos termos do qual as “garantias deverão incluir a informação específica ao titular dos dados e o direito de obter a intervenção humana⁶², de manifestar o seu ponto de vista, de obter uma explicação sobre a decisão tomada na sequência dessa avaliação e de contestar a decisão”. Deste modo, compete ao responsável pelo tratamento (*in casu*, empregador) colocar à disposição do titular dos dados uma forma simples de exercer esses direitos o que realça a necessidade de transparência quanto ao tratamento na medida em que apenas poderá contestar uma decisão ou manifestar o seu ponto de vista se compreender plenamente como foi tomada e com que fundamento⁶³, e com isto transmitir-se a ideia

deficiente, desprovidas de lógica, semnexo, ou não darem efetivamente uma resposta, o que demonstra uma total falta de sensibilidade (e de condições) para com os trabalhadores.

60 Pense-se, *e.g.*, no despedimento de um funcionário porque a sua reputação digital foi inferior a 4.6 em 5 “estrelas”, como exige a *Uber* (cfr. BAROCAS, Solon, HWANG, Tim, LEVY, Karen, & ROSENBLAT, Alex, “Discriminating Tastes: Uber’s Customer Ratings as Vehicles for Workplace Discrimination”, *in: Policy and Internet*, Vol. 9, n.º 3, 2017, pp. 256-279, (p. 259)). Trata-se de uma situação abrangida pelo âmbito de aplicação do art. 22º RGPD por consistir numa intervenção não humana.

61 Como observável pelas expressões “designadamente” e “pelo menos” do preceito, as garantias estão mencionadas a título meramente exemplificativo, o que significa que poderão ser adotadas outras.

62 Existem três tipos diferentes de envolvimento humano com IA: *human-in-the-loop*, onde um sistema de IA fornece informação a um humano para que este tome uma decisão (IA → Humano → decisão); *human on-the-loop*, onde um humano supervisiona um sistema de IA que toma uma decisão (IA → Decisão → Humano); *human out-of-the-loop*, onde um sistema de IA toma uma decisão sem qualquer envolvimento humano (IA → Decisão). Pregoamos pela primeira opção. Sobre o assunto, *vide* GOLDENFEIN, Jake, *op. cit.*, p. 48.

63 Neste sentido, *vide* o GT29º, *op. cit.*, p. 16 que vai mais longe ao recomendar que, a fim de tornar essas informações úteis e compreensíveis, devem ser dados exemplos reais dos tipos de repercussões possíveis. Contra: MITTELSTADT, Brent, RUSSEL, Chris & WACHTER, Sandra, “Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR”, *in: Harvard Journal of Law & Technology*, Vol. 31, n.º 2, 2018, pp. 842-887,

de que se trata de uma das mais importantes salvaguardas explicitamente mencionadas no seio de alguns Estados-Membros (E-M)⁶⁴ atuando, em teoria, como uma verdadeira válvula de segurança de forma a impedir decisões arbitrárias, injustas e discriminatórias.

4.2. O direito (de acesso) à informação

Como entende HILDEBRANDT⁶⁵ e, entre nós, GRAÇA MONIZ⁶⁶, o RGPD “expande as categorias de informação a prestar com o fim de garantir um tratamento equitativo e transparente”, assumindo particular importância as informações úteis relativas à lógica subjacente prestadas sobre a existência de ADMs⁶⁷, conferindo um direito *ex post* ao titular de dados através da interpretação conjunta do disposto nos arts. 13º/2, f), 14º/2, g) e 15º/1, h) auxiliados pelos Considerandos (60) e (63) do RGPD⁶⁸, pelo que se exige que os responsáveis pelo tratamento prestem informações específicas e de fácil

(p. 874).

64 Curiosamente, os E-M utilizam redações diferentes quando versam sobre esta salvaguarda. Como aponta MALGIERI, Gianclaudio, *op. cit.*, pp. 34-35, a lei alemã utiliza “contestação” (*contest*), enquanto a lei holandesa utiliza “*challenge*”, a lei irlandesa utiliza “recurso” (*to appeal*) e a lei britânica utiliza “o direito de solicitar ao controlador que reconsidere a decisão”, sendo que todos eles acabam por ser sinónimos com nuances ligeiramente diferentes que provavelmente revelam abordagens distintas adotadas. Na Hungria, o responsável pelo tratamento deve informar o sujeito sobre “os métodos e critérios utilizados no mecanismo de tomada de decisão”, contudo, não é claro se a explicação deve ser *ex ante* (baseada na funcionalidade geral do algoritmo) ou *ex post* (baseada na fundamentação da decisão). Paralelamente, no caso francês, como denota CASTETS-RENARD, Céline, “Comment construire une intelligence artificielle responsable et inclusive?”, *in: Recueil Dalloz*, n.º 4, 2020, pp. 225-230, a lei menciona estes dois elementos: informação *ex ante* sobre as regras gerais e específicas que definem o tratamento de dados e as principais características da implementação *ex post*.

65 HILDEBRANDT, Mireille, “The new imbroglio - living with machine algorithms”, *in: The art of ethics in the information society*, Amsterdam University Press, 2016, pp. 55-60.

66 MONIZ, Graça Canto, “Direitos do titular dos dados pessoais: o direito à portabilidade”, *in: Anuário da Proteção de Dados*, (coord. Francisco Pereira Coutinho e Graça Canto Moniz), CEDIS, 2018, pp. 11-34.

67 Em termos da duração da prestação das mesmas, *vide* os n.º 1 e 3 do art. 12º e alíneas a), b) e c) do n.º 3 do art. 14º, a par do Considerando (61).

68 Contra: PEHRSSON, Emily, *op. cit.*, p. 27, defendendo que os arts. 13º, 14º, e 15º criam direitos à informação, mas não conferem um direito *ex post* para o titular de dados obter informações sobre uma determinada decisão após o facto, o que parece ser reforçado pelos Considerandos (60) e (63) ao confirmarem (apenas) um direito *ex ante* à lógica e às consequências do tratamento.

acesso sobre tais *ADMs*, nomeadamente os fatores tidos em conta no processo decisório, os critérios e a relevância dos mesmos em termos de consequências previstas para o titular dos dados que não seja uma explicação complexa⁶⁹, de modo a cumprir as garantias a que se refere o já mencionado no n.º 3 do art. 22º.

Os preceitos mencionam a expressão “informações úteis relativas à lógica subjacente”, o que se revela um ponto fulcral uma vez que a complexidade implícita ao processamento operativo do algoritmo torna difícil de compreender o funcionamento do processo interno da decisão automatizada, atuando como uma *black box*⁷⁰ pois à medida que os algoritmos se tornam mais precisos e eficientes, tornam-se também mais complexos e, portanto, mais opacos e resistentes ao escrutínio, sendo apenas possível conhecer os *inputs* e *outputs* do processamento. Todo este secretismo e sofisticação que impede o acesso por parte do homem médio deve-se a motivos de concorrência e inovação, encontrando-se o código algorítmico geralmente protegido por *trade secrets* e *IP rights*⁷¹, pelo que se deve procurar equilibrar e conciliar os interesses entre o atual binómio⁷² transparência/*IP rights* aliás, enquanto o Considerando (63) RGPD afirma que os direitos de proteção de dados “não devem afetar (...) o segredo comercial ou a [IP] e, particularmente, o direito de autor que protege o software”, o Considerando (35) da Diretiva relativa aos *trade secrets* sublinha, de uma forma radical, que os direitos nela consagrados “não devem afetar os direitos e obrigações estabelecidos na Diretiva 95/46/CE [que

69 Neste sentido, GT29º, *op. cit.*, p. 25.

70 Como refere BAROCAS, Solon & SELBST, Andrew D., “The Intuitive Appeal of Explainable Machines”, *in: Fordham Law Review*, Vol. 87, Issue 3, 2018, pp. 1085-1139, (p. 1085), mesmo com conhecimento especializado, a base que está no fundo de uma decisão ainda é muitas vezes inescrutável.

71 É possível qualificar quase todas as informações como *trade secret* e argumentar que a sua divulgação é prejudicial para a vantagem competitiva do seu detentor legal, mas para isso, o algoritmo tem de satisfazer os requisitos cumulativos do art. 2º da *Trade Secret Directive* (EU) 2016/943, de 8 junho 2016, que define amplamente *trade secret* como qualquer informação que não é geralmente conhecida, tem valor comercial derivado deste segredo, e tem sido sujeita a medidas razoáveis para garantir que permaneça em segredo. O objetivo desta Diretiva assenta na proteção contra a aquisição, utilização e divulgação ilegal de segredos comerciais, contudo, como indica KULLMANN, Miriam, *op. cit.*, pp. 50-51, nada é regulamentado sobre a situação em que o acesso a um segredo comercial, como um algoritmo, é exigido devido a uma decisão (alegadamente) discriminatória.

72 Ambos são garantidos pela CDFUE, respetivamente, nos art. 8º e no n.º 2 do art. 17º.

precede o RGPD], em particular os direitos dos titulares de dados aos seus dados pessoais”. Para mais, o art. 5(d) estipula que é possível uma derrogação à proteção do *trade secret* como forma de “proteger um interesse legítimo reconhecido pela legislação da União ou do E-M”, como pode ser o direito de obter uma explicação racional, lógica e clara acerca de determinada decisão⁷³, tendo sido perante este contexto que o GT29^o declarou que a proteção dos *trade secrets* não pode ser uma desculpa para recusar a partilha de informações com os titulares dos dados⁷⁴ e, na mesma linha, o EDPS declarou que a proteção dos *trade secrets* não prevalece sobre a proteção dos dados pessoais, devendo haver antes um equilíbrio delicado entre os dois direitos⁷⁵, equilíbrio este discutido pelo BGH na decisão SCHUFA, de 28.01.2014⁷⁶, tendo-se o tribunal pronunciado a favor de que não havia a necessidade de partilhar o algoritmo, o *source code* do software de cálculo ou outros valores estatísticos relacionados com a fórmula de pontuação, baseando-se na proteção do *trade secret*. Na mesma linha argumentativa encontra-se a solução francesa, tendo o *Conseil Constitutionnel*⁷⁷ decidido que, quando os princípios do funcionamento interno de um algoritmo não possam ser comunicados sem infringir um segredo ou um interesse de *IP*⁷⁸, nenhuma decisão individual pode ser tomada com base exclusiva nesse algoritmo, procurando desta forma conciliar por um lado, o

73 Neste sentido, BRKAN, Maja, “Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond”, in: *International Journal of Law and Information Technology*, Vol. 27, Issue 2, 2019, pp. 91-121, (p. 98).

74 Neste sentido, GT29^o, *op. cit.*, p. 17 e, na mesma linha, COMANDÉ, Giovanni, & MALGIERI, Gianclaudio, “Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation”, in: *International Data Privacy Law*, Vol. 7, n.º 4, 2017, pp. 243-265, (p. 264).

75 EDPS, “Opinion 7/2015’ - Meeting the Challenges of Big Data: A call for transparency, user control, data protection by design and accountability”, 2015, disponível em: <https://edps.europa.eu/data-protection/our-work/publications/opinions/meeting-challenges-big-data_en>, pp. 7-10.

76 BGH, Urteil vom 28.01.2014 - VI ZR 156/13. Embora neste caso o poder de decisão residisse num humano, o que difere do caso das *ADMs*, não deixa de servir como paralelo.

77 *Conseil Constitutionnel*, du 12.06.2018, n.º 2018-765DC.

78 Sendo que a bitola deve assentar na condição de que, se a descoberta de informação exigir um esforço considerável, é provável que seja considerada como não sendo facilmente acessível, premissa esta explorada por SURBLYTE, Gintare, “Enhancing TRIPS: Trade Secrets and Reverse Engineering”, in: *TRIPS plus 20 - From Trade Rules to Market Principles (MPI Studies on Intellectual Property and Competition Law, 25)*, Hanns Ullrich, et al. (ed.), Springer, Berlin, 2016, pp. 725-760, (pp. 725-738).

sigilo e a propriedade intelectual e, por outro, a transparência e a atribuição de responsabilidade.

Porém, a verdade é que a opacidade do algoritmo⁷⁹ pode facilmente ser uma cobertura para uma nova forma de dissimulação de padrões de discriminação. Por esta razão, doutrinários como PASQUALE, embora reconheçam que tais medidas poderiam tornar os algoritmos ineficazes, defendem que o código deve ficar disponível e acessível para escrutínio através de meios regulamentares se necessário, sugerindo ainda a utilização de um auditor independente que possa manter o segredo ao mesmo tempo que serve o interesse público⁸⁰. Tomando a mesma posição relativamente a esta questão da legitimidade de acesso ao código algorítmico, encontramos a jurisprudência inédita do TAR da Lazio, de 22.03.2017⁸¹ que decidiu que um algoritmo é um ato administrativo digital e, por isso, sob o regime do direito fundamental (da liberdade) de informação, os cidadãos têm o direito de aceder ao mesmo⁸² esclarecendo que, em princípio, se o direito de acesso (liberdade de informação) e o direito à privacidade (entendida aqui como os *trade secrets*, a *IP* e os direitos de autor) entrarem em conflito, o primeiro prevalecerá. Realçando a argumentação do tribunal, este acabou por concluir que o facto do código estar protegido não

79 Acompanhamos BURRELL, Jenna, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”, *in: Big Data & Society*, Vol. 3, n.º 1, 2016, pp. 1-12, que distingue três tipos de opacidade algorítmica: segredo empresarial ou estatal, iliteracia técnica, e a resultante de características de *ML*.

80 A favor: PASQUALE, Frank, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, Londres, 2015, p. 141; BURRELL, Jenna, *op. cit.*, p. 4. Contra: PARECER CESE, “Inteligência artificial: antecipar o seu impacto no trabalho para assegurar uma transição justa”, de 19.09.2018, disponível em: <<https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52018IE1473>> e o GT29º, *op. cit.*, p. 25 onde apesar de entender que as informações prestadas devem ser suficientemente completas para permitir ao titular dos dados compreender os motivos da decisão, não considera a divulgação do algoritmo necessária no que respeita às obrigações de transparência mencionadas.

81 TAR Lazio, Chamber III bis, *Gilda vs. MUIR*, de 22.03.2017, n.º 3769/2017. O caso tratava da distribuição de um corpo docente pelas províncias italianas que foi determinada exclusivamente por um algoritmo encomendado a uma empresa privada, onde o requerente procurou exercer o direito de acesso mediante o regime de liberdade de informação.

82 Este entendimento pode hoje ser encontrado na Lei de Procedimentos Administrativos Italiana (arts. 22º a 28º), segundo a qual os cidadãos têm o direito de aceder a documentos administrativos se possuírem um “interesse direto, específico e real, correspondendo a uma situação legalmente protegida e vinculada ao documento que se deseja aceder”, citando a alínea a) do n.º 1 do art. 22º desta Lei.

preclui a possibilidade de se abrir esta Caixa de Pandora⁸³, na medida em que, sem o próprio algoritmo, torna-se extremamente difícil compreender a lógica subjacente à decisão final, servindo este acesso para que o trabalhador seja capaz de fundamentar devidamente a sua reclamação caso a decisão afete os seus direitos ou interesses legítimos. Uma interpretação mais restrita que obstaculize a visualização (*disclosure*) do próprio algoritmo não estaria em conformidade com o direito a um recurso efetivo e a um julgamento justo consagrado na CDFUE e na CEDH. Todavia, o tribunal não deixa de referir algo muito importante: é que uma solicitação de FOIA⁸⁴ traz uma grande falha ao regime italiano que é a sua limitação ao setor público, contudo não deixou de mencionar que um dos critérios de maior relevância que permite o acesso será o interesse público da questão apesar da relação poder ser entre privados, tal qual o caso em discussão⁸⁵. Embora seja uma abertura muito significativa,

83 Referiu o tribunal: “Imaginemos o que aconteceria se todos os procedimentos fossem manipulados por algoritmos e as solicitações de liberdade de informação não fossem aplicáveis a documentos algorítmicos [por não terem cariz administrativo], o referido regime ainda existiria nos livros, mas não mais na prática”. Tecendo severas críticas acerca desta posição do tribunal, *vide* TABARRINI, Camilla, “Understanding the big mind: Does the GDPR Bridge the Human-Machine Intelligibility Gap?”, *in: Journal of European Consumer and Market Law*, Vol. 9, Issue 4, 2020, pp. 135-143 ao defender a irrelevância da evidente violação dos direitos de autor uma vez que ter acesso à *source code* para chegar a uma determinada decisão, dificilmente ajudaria os titulares dos dados a exercer o seu direito de contestar pela sua complexidade inerente que se revela como incompreensível para a pessoa comum. Na mesma linha, encontra-se os organismos europeus na sequência dos estudos do MSINET (Committee of experts on internet intermediaries) do CE, “Algorithms and human rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications”, DGI (2017)12, 2018, disponível em: <<https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html>>, VILLANI REPORT, “Donner un sens à L’intelligence Artificielle: Pour une stratégie nationale et européenne”, 2018, disponível em: <<https://www.aiforhumanity.fr>> e HOUSE OF LORDS REPORT, “AI in the UK: ready, willing and able?”, 2018, disponível em: <<https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>>, (especialmente) par. 92 (p. 36) e par. 96-99 (p. 38), ao inviabilizarem uma abordagem de “transparência técnica completa” de todo o algoritmo ou *source code* por enfrentarem inevitavelmente a oposição dos detentores de *IP rights*, comprometendo as escassas hipóteses de uma transparência eficaz. Para uma perspetiva comparada, *vide* CORREIA, Catarina Camacho, “Inteligência Artificial e Propriedade Intelectual”, *in: CEDIS Working Papers*, n.º 2, NOVA School of Law, Lisboa, 2021, pp. 1-23, (pp. 8-19).

84 Introduzida em Itália pelo Decreto Legislativo n.º 97 de 2016, é parte integrante do processo de reforma da administração pública definido pela Lei n.º 124, de 7 de agosto de 2015, cujos remédios operam *ex post*, posteriormente à tomada de decisão.

85 O que concordamos, aliás, nem faria sentido restringir a sua divulgação à função

muito dependerá de como a jurisprudência europeia se vier a pronunciar num futuro próximo, na procura de um equilíbrio entre informação e privacidade⁸⁶ que para já se revela (quase que) inexistente, não tendo ainda tecido qualquer decisão vinculativa.

Nesta esteira, o RGPD desencadeou um vivo debate em torno da questão de saber se existe um efetivo direito de explicação das *ADMs*⁸⁷. A nosso ver, e reforçando a ideia, revela-se premente que o responsável pelo tratamento encontre formas simples de comunicar ao titular dos dados algum tipo de justificação (*e.g.*, o raciocínio subjacente, ou os critérios aplicados para tomar a decisão) de modo a que seja cognoscível, minimamente clara e compreensível pelo titular dos dados e lhe permita contestar a *ADM* de que foi alvo. Daí que defendamos que a informação significativa sobre a lógica envolvida seria idealmente composta pela informação sobre os dados que serviram de base (*input*) à decisão automatizada, a informação sobre os fatores que influenciaram a decisão, a sua importância e peso, e uma explicação razoável sobre a razão pela qual uma determinada decisão foi tomada, utilizando uma forma compreensível, nunca matemática e complexa⁸⁸. De igual modo, a proteção dos

pública: quereria então dizer que a nível jurídico-laboral, os profissionais (*e.g.*, professores) que ocupam a função pública estariam abrangidos, mas os privados não? Não cremos.

86 Na Itália, o primeiro prevalece, mas *e.g.*, o UK tende a favorecer o segundo, e entre nós, nada é dito a respeito de um cenário de intervenção algorítmica. Para maiores desenvolvimentos acerca desta questão a nível comparado, *vide* LA DIEGA, Guido Noto, *op. cit.*, pp. 1-33.

87 Para uma visão exaustiva das posições doutrinárias existentes sobre esta temática, *vide* CALDAS, Gabriela, “O direito à explicação no Regulamento Geral sobre a Proteção de Dados”, *in: Anuário de Proteção de Dados*, (coord. Francisco Pereira Coutinho, Graça Canto Moniz), CEDIS, 2019, pp. 37-53, (pp. 40-45).

88 Cfr. BRKAN, Maja, *op. cit.*, pp. 104-111. Neste sentido, chamamos a atenção para o *draft* de relatório explicativo da Convenção n.º 108 do CE, “For the protection of individuals with regard to automatic processing of personal data”, *draft explanatory report* (da denominada “versão modernizada”), 2016, disponível em: <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806b6ec2>, p. 13 (ponto 75.), segundo o qual “os titulares de dados devem ter o direito de conhecer a fundamentação subjacente ao tratamento dos seus dados, incluindo as consequências dessa fundamentação, que tenha resultado em quaisquer conclusões, em especial nos casos que impliquem o recurso a algoritmos para as decisões automatizadas, incluindo a definição de perfis. (...) [D]evem poder conhecer a lógica subjacente ao tratamento dos seus dados resultantes de uma decisão de «sim» ou «não», e não somente as informações relativas à própria decisão, [pois] sem a compreensão destes elementos, não é possível exercer efetivamente outras garantias essenciais, entre as quais o direito de oposição e o direito de reclamação junto de uma autoridade competente”. Todavia,

IP rights e dos *trade secrets* não pode privar o titular dos dados do seu direito a ser informado sobre a lógica de uma *ADM*, mas pode justificar a limitação do âmbito da informação, pelo que um nível alto, não técnico, de uma descrição da decisão do algoritmo tem maior probabilidade de ser significativa⁸⁹. No fundo, a prestação deste tipo de informação, a par do acesso aos dados pessoais, são condições indispensáveis para que o seu titular possa exercer os demais direitos apreciando a licitude do tratamento sujeita aos critérios do art. 6º RGPD, sendo por esta mesma razão que estes direitos foram descritos, por parte da doutrina inglesa⁹⁰, como o núcleo da proteção de dados pessoais. Um contra-argumento poderia ser o de que a expressão “informação útil relativa à lógica subjacente” pode ser encontrada no Considerando (71), mas não no art. 22º e, não tendo o preâmbulo força jurídica vinculativa, a inserção de um direito de explicação na sua redação poderá ter sido algo propositado e intencional da parte do legislador face à discórdia que já se verificava quanto a este assunto na altura da realização dos trabalhos preparatórios, invocando-se assim um argumento histórico⁹¹. Porém, em primeiro lugar entendemos que está disposto nos já referidos arts. 13.º ao 15.º RGPD capazes de extrair a *ratio* enunciada⁹² e, em segundo lugar, esta aposição numa parte não vinculativa do regulamento, como é o Considerando, tem vindo a ser interpretada pela Comissão, esta sim, como

o que é certo é que as opiniões se dividem entre a doutrina. A favor: COMANDÉ, Giovanni, & MALGIERI, Gianclaudio, *op. cit.*, pp. 243-265; HODSON, Hal & POWLES, Julia, “Google DeepMind and healthcare in an age of algorithms”, *in: Health and Technology*, Vol. 7, Issue 4, 2017, pp. 351-367. Contra: EDWARDS, Lilian & VEALE, Michael, “Slave to the algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for”, *in: Duke Law & Technology Review*, Vol. 16, n.º 1, 2017, pp. 18-64.

89 Acompanhamos de perto CATE, Fred H., KUNER, Christopher, LYNSKEY, Orla, MILLARD, Christopher, & SVANTESSON, Dan Jerker B., “Machine learning with personal data: is data protection law smart enough to meet the challenge?”, *in: International Data Privacy Law*, Vol. 7, Issue 1, 2017, pp. 1-3 e ARACI, Gözde, “A Quest for Fair Balance: Testing the Right of Access against IP Rights and Trade Secrets”, *in: Munich Intellectual Property Law Center (MIPLC)*, Master’s Thesis, 2019, pp. 1-70, (pp. 13-16).

90 Cfr. L’HOIRY, Xavier & NORRIS, Clive, “The honest data protection officer’s guide to enable citizens to exercise their subject access rights: lessons from a ten-country European Study”, *in: International Data Privacy Law*, Vol. 5, Issue 3, 2015, pp. 190-204.

91 Cfr. FLORIDI, Luciano, *et al.*, *op. cit.*, pp. 41-43 e MITTELSTADT, Brent, *et al.*, *op. cit.*, pp. 842-887. Acompanhando este raciocínio, PEHRSSON, Emily, *op. cit.*, pp. 25-26 indica que será muito provável que um tribunal determine que o direito à explicação é aqui fortemente encorajado, mas não exigido.

92 Cfr. POWLES, Julia & SELBST, Andrew D., “Meaningful information and the right to explanation”, *in: International Data Privacy Law*, Vol. 7, Issue 4, 2017, pp. 233-242, (p. 233).

uma “aposição propositadamente deliberada” e, por sua vez, reconhecido o seu papel essencial na interpretação das disposições de um ato da UE⁹³, pelo que a sua referência ao direito de explicação deve ser usada para interpretar adequadamente o art. 22.º ao refletir o contexto da disposição e a principal finalidade do RGPD, que consiste em aumentar a proteção do titular dos dados pessoais⁹⁴.

5. O confronto com a discriminação de decisões individuais automatizadas: em especial, no panorama português.

5.1. O quadro legal no fomento da igualdade e no combate à discriminação

Em Portugal, o uso de algoritmos em contexto laboral ainda não é uma realidade ou, pelo menos, uma realidade transversal a um conjunto alargado de entidades empregadoras. Porém, num contexto internacional e realizando um exercício de prognose futura a curto-médio prazo, importa antecipar os desafios e questões da mais difícil resposta que a utilização de algoritmos poderá colocar, de que são exemplo as práticas discriminatórias que implicam uma série de problemas éticos e colocam em causa as liberdades fundamentais, justificando o enquadramento legal exigente e restrito desta matéria.

No que se refere à recolha e tratamento dos dados pessoais e, por sua

93 É verdade que os Considerandos não podem ser utilizados para uma interpretação *contra legem* das disposições do direito da UE, porém, contar com uma abordagem de interpretação conjunta de várias disposições do RGPD, tal como sugerido acima, e a utilização do Considerando (71) para reforçar a interpretação de apoio à existência do direito à explicação, não conduziria a uma interpretação *contra legem*, pelo contrário, serviria como um meio para resolver ambiguidades resultantes da leitura de disposições relevantes do RGPD, como é reforçado por BRKAN, Maja, *op. cit.*, pp. 104-111. Na mesma linha, *vide* KLIMAS, Tadas & VAICIUKAITE, Jurate, “The Law of Recitals in European Community Legislation”, *in: ILSA Journal of International & Comparative Law*, Vol. 15, 2008, pp. 86-88 ao indicarem que, quando o Considerando for claro, controlará uma disposição operacional ambígua (como poderá ser o exemplo do art. 22º), o que significa que a disposição operacional, no que toca à prestação de informações, será interpretada à luz do Considerando.

94 Nesta linha, o Considerando (58) é claro, ao indicar que o princípio da transparência é “especialmente relevante em situações em que a proliferação de operadores e a complexidade tecnológica das práticas tornam difícil que o titular dos dados saiba e compreenda se, por quem e para que fins os seus dados pessoais estão a ser recolhidos”, daí que a complexidade do algoritmo de *ML* não sirva de justificação para que não sejam fornecidas as informações ao titular dos dados.

vez, à igualdade e não discriminação, o regime legal ordinário do Código do Trabalho (CT) deve ser articulado com o RGPD, cujas normas deverão ser interpretadas em consonância com as diretivas e convenções europeias nesta matéria⁹⁵, não esquecendo os reconhecidos instrumentos europeus⁹⁶ ao constituir direito universal⁹⁷. De forma mais concreta, a nível interno, tendo sempre subjacente uma ideia de dignidade da pessoa humana a Constituição (CRP) consagra uma robusta tutela de direitos de personalidade, fomentando a igualdade de oportunidades no acesso ao emprego e proibindo privilégios e discriminações negativas, ou seja, situações de vantagem infundadas e situações de desvantagem, impondo regras de neutralidade no contexto dos processos de candidatura de acesso ao emprego, nomeadamente quanto aos critérios de seleção dos trabalhadores⁹⁸ elencando um conjunto de fatores discriminatórios (cfr. art. 13º CRP), que se espelham no campo específico do Direito Constitucional do Trabalho, daí que a Subsecção III do Título II do nosso CT, i.e., do art. 23º ao art. 32º, se revele bastante completa neste campo, nunca esquecendo que o art. 10º CT estende a aplicação das normas respeitantes aos direitos personalidade, igualdade e não discriminação e segurança e saúde no trabalho à prestação de serviços⁹⁹. Desta forma, para efeitos do nosso estudo, apresentam maior relevo os arts. 23º, 24º/1 e 2, a) e c) *in fine*, 25º, 28º e 30º/1 e

95 De salientar a Diretiva 2000/43/CE de 29 de junho, a Diretiva 2000/78/CE de 27 de novembro, a Diretiva 2005/54/CE de 25 de janeiro e a Diretiva 2006/54/CE de 5 de julho, a Convenção das NU sobre a eliminação de todas as formas de discriminação contra as mulheres, a Convenção para a proteção dos direitos do homem e das liberdades fundamentais, de que Portugal é signatário, e de igual modo, de assinalar a Convenção n.º 111 da OIT que proíbe a discriminação em matéria de emprego e atividade profissional.

96 Designadamente a DUDH, a CDFUE, o PIDCP e PIDESC, a CEDH e o protocolo 12 em anexo à mesma.

97 Porque o problema se coloca em grande escala nos EUA dado ser o país que mais utiliza estes sistemas algorítmicos, o ordenamento apresenta como instrumentos principais de combate à discriminação algorítmica o Título VII da Lei dos Direitos Cíveis de 1964, que proíbe a discriminação no emprego com base numa característica proibida, e o *Age Discrimination in Employment Act of 1967* que proíbe a discriminação no emprego com base na idade de pessoas com mais de quarenta anos.

98 Cfr. PALMA RAMALHO, Maria do Rosário, *op. cit.*, p. 50.

99 Quanto ao facto dos trabalhadores por conta própria parecerem excluídos da lei de igualdade de tratamento da UE, *vide* FREDMAN, Sandra, “Equality Law: Labour Law or an Autonomous Field?”, *in: The Autonomy of Labour Law*, (Alan Bogg, Cathryn Costello, Acl Davies and Jeremias Prassl (eds.)), Hart Publishing, Oxford, 2015, pp. 257-274, (p. 260), que aborda as consequências destes trabalhadores face a decisões algorítmicas.

2 todos do CT, atribuindo-se ao candidato a emprego e ao trabalhador o direito de controlo dessas informações, interessando aqui o disposto nos arts. 26º/1 e 35º/1 da CRP conexos com os arts. 17º/3 e 4 CT, podendo este último preceito revelar-se importante ao permitir que se extraia dele a obrigatoriedade do titular de dados vir a tomar conhecimento no que respeita às decisões algorítmicas, pelo que pensamos que poderá servir como norma remissiva para essa mesma questão, mesmo sendo a letra da lei limite de interpretação faltando, no entanto, balizá-la devidamente, por isso mesmo entendemos que tal corpo normativo poderá e deverá ser adaptado em face da nova realidade da IA¹⁰⁰ hoje com grande influência na relação laboral, embora naturalmente, ao dia de hoje, este regime legal se imponha às novas tecnologias digitais quando estas sejam utilizadas no contexto de processos de recrutamento e mesmo durante a prestação do labor. Contudo, como refere PALMA RAMALHO¹⁰¹, “face ao manancial de informação sobre as pessoas, que é hoje facilmente acessível nas plataformas digitais e que consta de armazenamentos de *Big Data*, na prática poderá ser difícil de garantir que, quando usadas no contexto dos processos de recrutamento, as tecnologias digitais se mantenham dentro dos limites impostos pelo sistema legal”, aliás, se um sistema algorítmico se diferencia com base em classes recém-inventadas, poderia permanecer fora do âmbito da lei da não discriminação, dado que tem pouco a dizer sobre as previsões algorítmicas¹⁰². Por último, fazer menção ao facto dos E-M poderem, por lei ou por convenção coletiva de trabalho (CCT),

100 Em Portugal, o n.º 1 do art. 25º do DL 260/2009, de 25 setembro que regula o regime jurídico do exercício e licenciamento das agências privadas de colocação e das empresas de trabalho temporário, estabelece que “o candidato a emprego tem o direito de ser informado, por escrito, sobre: os métodos e técnicas de recrutamento aos quais se deve submeter e as regras relativas à confidencialidade dos resultados obtidos”. Parece-nos que esta espécie de lugar paralelo, pode ser um ponto de partida face ao caminho a seguir face ao atual silêncio da Lei de Execução e, desta forma, colmatava-se ou mitigava-se a assimetria informativa e a ausência de transparência à falta de melhor solução sobre esta nova realidade algorítmica que veio para ficar, cujo CT deverá acompanhar nos próximos tempos de maneira urgente ao suscitarem-se questões sensíveis que carecem de resposta e orientação.

101 PALMA RAMALHO, Maria do Rosário, *op. cit.*, p. 51.

102 Neste sentido, BORGESIU, Frederik J. Zuiderveen, “Strengthening legal protection against discrimination by algorithms and artificial intelligence”, *in: The International Journal of Human Rights*, Vol. 24, Issue 10, 2020, pp. 1-23, (p. 13). Como refere KULLMANN, Miriam, *op. cit.*, p. 53, para além da dificuldade de obter prova, as leis de não discriminação da UE, em combinação com o RGPD, proporcionam ao candidato a emprego poucos mecanismos legais que lhe permita reagir em face de decisões discriminatórias tomadas pelo empregador.

adotar regras mais específicas em relação ao tratamento dos dados pessoais dos trabalhadores no contexto do emprego, e em particular para efeitos do processo de recrutamento e da igualdade e diversidade no local de trabalho, nos termos do art. 88º RGPD.

5.2. Da aferição de accountability

Quem deve ser responsável se um algoritmo tomar uma decisão discriminatória? São muitas as indagações e, ao dia de hoje, poucas respostas e certezas. Este grande desafio de apuramento de responsabilidade surge associado aos elevados níveis de heterogeneidade existente no processo de decisão algorítmica, dificultando o estabelecimento donexo causal ao respetivo dano de modo a apurar e imputar responsabilidades, daí que este assunto seja referido como “o problema de muitas mãos”¹⁰³ por existirem várias camadas na construção de um algoritmo e precisamente porque se baseiam em práticas constantes de inovação e atualização¹⁰⁴, denotando-se, por isso, diferentes entendimentos sobre a matéria.

Num primeiro momento, a responsabilidade terá de ser necessariamente da pessoa, pois só perante ele se poderá reagir devidamente e imputar civilmente responsabilidades pelos atos ou omissões que a máquina causou a terceiros¹⁰⁵,

103 Cfr. DOORN, Neelke, FAHLQUIST, Jessica Nihlén, ROYAKKERS, Lambèr, VAN DE POEL, Ibo, & ZWART, Sjoerd, “The problem of many hands: Climate change as an exemple”, in: *Science and Engineering Ethics*, Vol. 18, n.º 1, 2012, pp. 49-67.

104 Neste sentido, MEDON, Filipe, & TEFFÉ, Chiara Spadaccini de, “Civil Liability and Regulation of new technologies: issues about the usage of AI in Business decision-making”, in: *Journal of Institutional Studies*, Vol. 6, n.º 1, 2020, pp. 301-333, (p. 305). Como refere SANTOS GONZÁLEZ, María José, “Regulación legal de la robótica y la inteligencia artificial: Retos de futuro”, in: *Revista Jurídica de la Universidad de León*, n.º 4, 2017, pp. 25-50, (pp. 37-38), quanto mais aumenta a autonomia da IA, mais a responsabilidade é diluída entre os múltiplos atores envolvidos no processo.

105 Cfr. PIRES, Thatiane Cristina Fontão, e SILVA, Rafael Peteffi da, “A responsabilidade civil pelos atos autônomos da inteligência artificial: notas iniciais sobre a resolução do Parlamento Europeu”, in: *Revista Brasileira de Políticas Públicas*, Vol. 7, n.º 3, 2017, pp. 239-255, (pp. 247-248); TOBOSO SÁNCHEZ, Belén, *El impacto disruptivo de la Inteligencia Artificial: Estudio prospectivo sobre su configuración en la sociedad y la categorización de su responsabilidad jurídica*, Trabajo de Fin de Grado, Universidad de Girona 2019, pp. 48-50 e GONZÁLEZ MARTÍNEZ, Andrea de los Ángeles, *El Derecho de los robots con inteligencia artificial, una nueva disciplina jurídica?*, Trabajo de Fin de Grado, Universidad de La Laguna, 2019, p. 23, embora este último A. não descure a imputação de responsabilidade à máquina num futuro próximo à medida que aumenta a sua autonomia e capacidade de decisão independente

único através do qual é exigível o reparo pelo dano causado, conferindo pretensões indemnizatórias que nunca deverão ficar limitadas no seu tipo ou na extensão dos danos a indemnizar, nem nas formas de compensação à parte lesada¹⁰⁶, assente na pobre justificação de que os danos foram causados por um agente não humano porque em momento algum pode ser atribuída à *ML* a noção de culpa ou de dolo na prática dos seus atos lesivos uma vez que partimos de uma vontade que não foi criada de forma totalmente livre, mas sempre sujeita à sua condição de dependência e vontade de um humano¹⁰⁷. Porém, a verdade é que recentemente, o Parlamento Europeu propôs a criação de uma nova figura jurídica¹⁰⁸ que assenta num *tertium genus* entre pessoas físicas e legais sugerindo a designação de pessoa eletrónica¹⁰⁹ configurada como um sujeito moral dotado de direitos e obrigações por manter capacidades similares (mas nunca iguais) às das pessoas¹¹⁰, o que se veio a repercutir no Parecer dirigido à Comissão dos Assuntos Jurídicos da União Europeia que contém recomendações à Comissão

da programação humana. Neste mesmo sentido, *vide* a Resolução 2015/2103(INL) do PE, de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre robótica, disponível em: <https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html#title1>, sendo de salientar o ponto AD. e o ponto AF.

106 Neste sentido, *vide* os pontos AF. e 52. da Resolução 2015/2103(INL), *id.* 106.

107 Cfr. NÚÑEZ ZORRILLA, María del Carmen, “Los nuevos retos de la Unión Europea en la regulación de la responsabilidad civil por los daños causados por la inteligencia artificial”, *in: Revista Española de Derecho Europeo*, n.º 66, 2018, pp. 9-53, (pp. 9-10).

108 Ponto 59(f) da Resolução 2015/2103(INL), *id.* 106. Apologista desta ideia, *vide* SANTOS GONZÁLEZ, María José, *op. cit.*, pp. 37-38 e pp. 40-42, que trata de enumerar as semelhanças entre a IA e uma pessoa jurídica (singular e coletiva), bem como as diferenças que podem contribuir para a formação de uma nova categoria jurídica.

109 Cfr. BONALDO, Arianna, & CUGINI, Gianvirgilio, “Intelligenza artificiale: responsabilità nella progettazione e utilizzo di sistemi: Analisi della tematica e riflessi legali, fiscali ed etici”, *in: Diritto tributario internazionale e dell’UE*, 2020, pp. 39-44, (pp. 40-41). Todavia, como alerta GARCÍA-MICÓ, Tomás Gabriel, “Electronic personhood: a *tertium genus* for smart autonomous surgical robots?”, *in: Data Science, Machine Intelligence, and Law*, Vol. 1, Algorithmic Governance and Governance of Algorithms, (Ebers M., Cantero Gamito M. (eds.)), Springer, 2021, pp. 87-108, (pp. 87-108), a personalidade não é apenas um pré-requisito para a responsabilidade, mas uma construção jurídica complexa que requer um maior desenvolvimento na medida em que se abre uma janela que será difícil de controlar.

110 Ao que cerca de duas centenas de peritos em IA vieram contestá-la, tendo sido submetida uma carta aberta ao PE (cfr. OPEN LETTER: To the European Commission Artificial Intelligence and Robotics, de 05.04.2018, disponível em: <<https://g8fip1kplyr33r3krz5b97d1-wpengine.netdna-ssl.com/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf>>), referindo-se que deverá ser a defesa do ser humano a estar no centro do debate e não a atribuição de direitos à máquina.

sobre disposições de Direito Civil sobre robótica, de 09.11.2016¹¹¹, onde se reconheceu que se deve utilizar processos que assegurem o controlo humano e a reversibilidade das operações pelos sistemas inteligentes e que, atendendo ao seu crescente nível de autonomia, deverá ser acompanhada pela adaptação das regras de responsabilidade no que respeita às consequências associadas às suas ações ou inações. De igual modo, doutrina italiana¹¹² e francesa¹¹³, fala-nos acerca da necessária complementaridade (no futuro) entre o ser humano e a máquina, propondo que a responsabilidade não seja exclusiva do ser humano, devendo acompanhar-se o progresso científico e tecnológico decisivo para um futuro de novas categorias jurídicas, sugerindo uma responsabilidade mútua quer do humano, quer da *AI*. Tendo esta realidade presente, a par de VÍTOR FIDALGO¹¹⁴, entendemos que esta rejeição deverá ser categórica, pois ao contrário do que sucede com o ser humano, qualquer personalidade jurídica fonte de direitos e deveres atribuída à máquina será sempre legal e não inerente a ela, pelo que se as pessoas são o princípio e o fim do Direito, o Direito limita-se a constatar a existência de personalidade jurídica no ser humano, o que já não acontece com as pessoas coletivas ou qualquer outro tipo de entidades em que o Direito imputa subjetivamente situações jurídicas à semelhança das pessoas humanas¹¹⁵, daí que, mesmo que se denote um nível elevado de autonomia na

111 CEAS, Parecer dirigido à CAJUE que contém recomendações à Comissão sobre disposições de direito civil sobre robótica, de 09.11.2016, disponível em: <https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_PT.html>, ponto 4.

112 Cfr. CASONATO, Carlo, “Potenzialità e sfide dell’intelligenza artificiale”, *in: BioLaw Journal*, n.º 1, 2019, pp. 177-182, (pp. 181-182).

113 Cfr. SOULEZ, Marie, “Questions juridiques au sujet de l’intelligence artificielle”, *in: Enjeux numériques - L’intelligence artificielle: un enjeu d’économie et de civilisation?*, n.º 1, 2018, pp. 81-85, (pp. 84-85).

114 FIDALGO, Vítor Palmela, *op. cit.*, pp. 879-900.

115 Esta ideia está presente em NAUCIUS, Mindaugas, “Should fully autonomous artificial intelligence systems be granted legal capacity?”, *in: Teisés Apžvalga Law Review*, Vol. 17, n.º 1, 2018, pp. 113-132, (pp. 116-122), ao mencionar, e bem, que em contraste com a “pessoa natural”, a designação “pessoa jurídica” é utilizada para se referir a uma entidade que não é um ser humano, mas para a qual a sociedade opta por conceder algumas das mesmas proteções e direitos legais que às pessoas naturais, tratando-se, por isso, de uma entidade fictícia criada para realizar determinados atos. Nesta linha, a par de DYSCHKANT, Alexis, “Legal personhood: how we are getting it wrong”, *in: University of Illinois Law Review*, Vol. 2015, n.º 5, 2015, pp. 2075-2109, (pp. 2084-2085), poder-se-ia argumentar que associando os sistemas de IA à categorização de “pessoas jurídicas”, dependendo do seu desenvolvimento, gozariam de direitos e responsabilidades que poderiam ser atribuídos a outras pessoas jurídicas, mas não teriam direitos e responsabilidades inerentes aos seres humanos.

ML, tal não será sinónimo de liberdade¹¹⁶, e não sendo livre, nunca poderá ser titular de responsabilidade pelos atos que pratique.

O próximo passo será aferir que pessoa (i.e., quem) deverá responder pelo sucedido. Para um sector doutrinário, a responsabilidade será do empregador¹¹⁷ sendo que é ele quem determina o *how, when and what* do tratamento de dados mesmo que utilize terceiros para realizar tais tarefas, tendo uma responsabilidade acrescida, designada por *information fiduciary*¹¹⁸, de verificar e assegurar que as práticas laborais não prejudicam terceiros, a par de ações ou omissões referentes à implementação de medidas que evitem alcançar resultados tendenciosos, arbitrariedades e discriminações, pois são eles quem têm a última palavra decisiva¹¹⁹. Para outro sector, a responsabilidade deverá ser atribuída ao programador ou mesmo à entidade que cria e comercializa o algoritmo, uma vez que são os mais capazes (ou mesmo os únicos) de decretar a mudança interna no algoritmo, ou seja, por outras palavras, a ideia é que ao criarem um algoritmo que funciona de uma forma particular, tornam-se

116 Neste sentido, BARBOSA, Mafalda Miranda, “Inteligência Artificial, E-PERSONS e Direito: Desafios e perspetivas”, in: *Revista Jurídica Luso Brasileira*, Vol. 3, n.º 6, 2017, pp. 1475-1503, (p. 1482) e SOULEZ, Marie, *op. cit.*, pp. 82-83, ao salientarem que a autonomia tecnológica da IA está longe do agir ético dos humanos em que radica o ser pessoa, falhando na relação de cuidado com o outro.

117 Cfr. CALO, Ryan, “Open Robotics”, in: *Maryland Law Review*, Vol. 70, n.º 3, 2011, pp. 571-613, (pp. 604-605).

118 Conceito originalmente proposto por BALKIN, Jack, “Information fiduciaries and the first amendment”, in: *UC Davis Law Review*, Vol. 49, n.º 4, 2016, pp. 1183-1234, (pp. 1183-1234). Nesta linha, MEDON, Filipe, & TEFÉ, Chiara Spadaccini de, *op. cit.*, pp. 317-325 estabelecem o paralelo com os deveres de cuidado, zelo e diligência que adstringe o administrador em prol de uma empresa, não se podendo afastar a responsabilidade pessoal dos administradores por danos decorrentes de sistemas de IA em razão da *culpa in eligendo* pela escolha da tecnologia a partir do momento em que o administrador delega parte dos processos decisórios da empresa num sistema de *AI*.

119 Cfr. TRINDADE, Beatriz Santiago, “Two years in: does the GDPR already need updates?: a question brought by algorithmic decision-making”, in: *Anuário de Proteção de Dados* (coord. Francisco Pereira Coutinho, Graça Canto Moniz), CEDIS, 2020, pp. 79-103, (p. 86, nota rodapé (26)). Nesta linha, pode recorrer-se a teorias da negligência como já entendia OPPENHEIMER, David Benjamin, “Negligent Discrimination”, in: *University of Pennsylvania Law Review*, Vol. 141, Issue 3, 1993, pp. 899-972, ao defender que num sistema imprevisível como a proliferação de dados, a maneira correta de pensar sobre a solução poderia passar por um dever de cuidado traduzindo na ideia de responsabilidade sem culpa. Contra este critério, vide SANTOS GONZÁLEZ, María José, *op. cit.*, p. 39 ao seguir uma orientação assente na responsabilidade de gestão de riscos que não se centre na pessoa que agiu negligentemente, mas na pessoa capaz, em determinadas circunstâncias, de minimizar os riscos e de gerir o impacto negativo.

voluntariamente parte do sistema de decisão e assumem a responsabilidade pelos danos criados, princípios violados e direitos diminuídos do seu sistema de decisão¹²⁰. Para mais, mesmo que o algoritmo aprenda a modificar o próprio código, o facto da intervenção humana ser prévia e a vontade negocial abstrata determinar um afastamento aos princípios clássicos nesta matéria significa que a máquina, por mais sofisticada e complexa que seja, sempre fará aquilo para que foi programada, tendo como fonte a vontade humana, sendo o programador que estabelece os critérios de determinação logo, deve ser-lhe imputada a responsabilidade¹²¹. Por último, existe um sector que parte da teorização aristotélica¹²² de que existem pelo menos duas condições tradicionais para a atribuição de responsabilidade por uma ação, a chamada condição “de controlo” e a condição “epistémica”, sendo que uma pessoa é responsável se (i) o fizer (ou o tiver feito) enquanto agente da ação, se tiver causado a ação ou se tiver um grau de controlo suficiente sobre a ação e (ii) se souber e tiver consciência do que está a fazer (ou do que fez). Tendo presente esta orientação filosófica, há quem defenda a repartição de responsabilidades, *maxime* entre o criador, programador do algoritmo e empregador¹²³ numa espécie de responsabilidade

120 Cfr. MARTIN, Kirsten, “Ethical Implications and Accountability of Algorithms”, *in: Journal of Business Ethics*, Vol. 160, n.º 4, 2019, pp. 835-850 e COECKELBERGH, Mark, “Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability”, *in: Science and Engineering Ethics*, Vol. 26, n.º 4, 2020, pp. 2051-2068, (p. 2059).

121 Neste sentido, de forma mais exaustiva, *vide* FESTAS, David, *A declaração negocial automatizada*, Relatório de Internet, Direito de Autor e Comércio Eletrónico, FDUL, Lisboa, 2003, pp. 26-41 e pp. 48-72 e, de forma mais sintética, *vide* FESTAS, David, *op. cit.*, pp. 417-425 e pp. 433-460 ao mencionar que podemos chegar a uma situação de nos encontrarmos fora dos quadros negociais, por impossibilidade de recondução da declaração automatizada à vontade humana (sendo que sem vontade, pode haver imputação e vinculação com aplicação direta ou analógica das disposições do NJ, mas não há NJ), indo mais longe, ao desenvolver a questão acerca da diferença entre os agentes eletrónicos autónomos projetados e os agentes eletrónicos convencionais, na medida em que, o facto dos agentes eletrónicos autónomos projetados poderem aprender com a sua experiência, modificar instruções que lhe tenham sido introduzidas e, no limite, criar instruções próprias, significa que, no extremo, os agentes eletrónicos podem ser programados para se auto-programarem, contudo, a possibilidade de criarem instruções próprias não deixa de resultar da programação originária implementada pelos seus criadores. De igual modo, debruçando-se sobre esta questão, *vide* MEDON, Filipe, & TEFFÉ, Chiara Spadaccini de, *op. cit.*, pp. 317-325.

122 Cfr. ARISTOTE, *Éthique de Nicomaque* (trad. Jean Voilquin), Garnier-Flammarion, Paris, 1965.

123 Cfr. TOBOSO SÁNCHEZ, Belén, *El impacto disruptivo de la Inteligencia Artificial: Estudio prospectivo sobre su configuración en la sociedad y la categorización de su responsabilidad jurídica*, Trabajo de Fin de Grado, Universidad de Girona 2019, pp. 48-50

solidária, sendo possível exercer ações de regresso entre as partes responsáveis, posição esta que acompanhamos no seguimento da entrada em vigor do RGPD por via de um critério de finalidade material da recolha e tratamento dos dados pessoais, na medida em que, *e.g.*, numa prestação de serviços de recrutamento e seleção, caso o prestador (exterior) aplique os seus métodos e estratégias de recrutamento com autonomia técnica em relação ao cliente (empregador), *i.e.*, se não receber quaisquer instruções quanto aos métodos e técnicas de recrutamento, exceto quanto ao perfil de candidato a selecionar, não poderá ser entendido como subcontratante no âmbito do tratamento de dados pessoais, mas sim como responsável pelo tratamento (*cfr.* arts. 28º/3, a) e 29º ambos do RGPD). Para mais, caso o cliente (empregador) determine um perfil de candidato e posteriormente venha a escolher os candidatos a contratar, entende-se que este define as finalidades do tratamento, sendo, por isso, também responsável pelo tratamento. Neste exemplo, o enquadramento da responsabilidade da empresa de recrutamento no âmbito do art. 28º RGPD como subcontratante não se afigura adequado, pois nenhuma das entidades se limita a tratar dados pessoais por conta da contraparte¹²⁴, ora se as partes tratam dados em momentos distintos do processo de recrutamento, faz sentido que as

e GORDILLO RODRÍGUEZ, Cristhian Fabián, *La responsabilidad civil en el derecho privado colombiano de la robótica con inteligencia artificial*, Trabajo de Fin de Grado, Fundación Universitaria Los Libertadores, 2020, p. 81. Como apontam KING, Allan G., & MRKONICH, Marko, “Big Data and the Risk of Employment Discrimination”, *in: Oklahoma Law Review*, Vol. 68, n.º 3, 2016, pp. 554-584, (p. 583), apesar do empregador que subscreve a *Big Data* ser responsável pelas consequências mesmo que da sua parte exista falta de intenção ou conhecimento discriminatório do algoritmo, é provável que o seu criador tenha uma perceção mais precisa do risco, uma vez que desenvolveu o produto. Chegando à mesma conclusão pelo caminho da responsabilidade por produtos defeituosos através do regime nacional (*in casu*, Espanhol) de adaptação da Diretiva 85/374, de 25 julho, *vide* SANTOS GONZÁLEZ, María José, *op. cit.*, pp. 43-46; RAMÓN FERNÁNDEZ, Francisca, “Robótica, inteligencia artificial y seguridad: ¿Cómo encajar la responsabilidad civil?”, *in: La Ley: Revista jurídica española de doctrina, jurisprudencia y bibliografía*, n.º 9365, Wolters Kluwer, 2019, pp. 1-13, (pp. 7-8) e GARCÍA-MICÓ, Tomás Gabriel e GÓMEZ LIGUERRE, Carlos, “Responsabilidad por daños causados por la Inteligencia Artificial y otras tecnologías emergentes”, *in: InDret*, Vol. 1, 2020, pp. 501-511, propugnando que a origem do defeito pode estar no fabrico, conceção ou informação sobre o consumo, utilização ou manuseamento do produto fazendo o paralelo para o algoritmo.

124 Atente-se o n.º 10 deste preceito, segundo o qual, “(...) o subcontratante que, em violação do presente regulamento, determinar as finalidades e os meios de tratamento, é considerado responsável pelo tratamento (...) em questão”, em conjugação com a definição de responsável pelo tratamento prevista no n.º 7 do art. 4º RGPD.

suas obrigações e responsabilidades sejam repartidas em função da influência no tratamento em determinada ocasião, de maneira proporcional, sendo que o próprio art. 26º RGPD pretende evitar que a diluição da *accountability* pelo tratamento impeça a sua imputação pelo incumprimento dos seus princípios e normas, entendimento que encontra apoio na jurisprudência europeia¹²⁵.

Por fim, questiona-se com frequência se seria necessário criar normas e institutos específicos para tratar de matérias atuais relativas à IA ou se deveriam ser aplicadas as normas já existentes, i.e., saber se os regimes de responsabilidade civil atuais seriam suficientemente maleáveis para lidar com os novos conflitos resultantes da relação entre o ser humano e a IA. Seguindo de perto a Resolução 2015/2103(INL), entendemos que o (futuro) instrumento legislativo deverá basear-se numa avaliação aprofundada da Comissão que determine se a abordagem a aplicar deve ser a da responsabilidade civil aquiliana por factos ilícitos ou a da gestão de riscos, consoante exista ou não culpa, dado que esta última modalidade prescinde do elemento subjetivo e tendo em mente que aquele instrumento incentiva – e bem - a adoção de uma responsabilidade objetiva ao ter a benesse de se exigir “apenas” a prova de que o dano ocorreu e o estabelecimento de um nexo de causalidade entre o funcionamento prejudicial e os danos sofridos pela parte lesada¹²⁶ sendo que o regime de gestão do risco é dirigido à pessoa capaz de minimizar o respetivo risco e de gerir o seu impacto. Nesta esteira, poderá fazer sentido propor-se o paralelo com a responsabilidade de danos causados por animais (cfr. art. 502º do Código Civil - CC) na linha da teoria do risco, no sentido de que, quem for responsável pelo algoritmo,

125 Veja-se o Ac. TJUE, *Jehovan todistajat - uskonnollinen yhdyiskunta*, de 10.07.2018, Case C-25/17, ponto 65. e 66.: “Os referidos intervenientes podem estar envolvidos em diferentes fases [do] tratamento e em diferentes graus, pelo que, a responsabilidade das partes deverá ser proporcional ao nível de intervenção no tratamento de dados (...). Uma pessoa singular ou coletiva que, para fins que lhe são próprios, influencia o tratamento de dados pessoais e contribui assim para a determinação da finalidade e dos meios do tratamento pode ser considerada responsável pelo tratamento”. Nesta linha, chama-se a atenção para o ponto 56. da Resolução 2015/2103(INL), *id.* 106.

126 Ponto 53. e 54, *id.* 106. Fazendo um apanhado das restantes pronúncias europeias sem, contudo, resultar das mesmas alguma proposta ou clarificação acerca do tipo de responsabilidade que pode estar em causa, *vide* JÚNIOR, José Luiz de Moura Faleiros, “Discriminação por algoritmos de Inteligência Artificial: A responsabilidade civil, os vieses e o exemplo das tecnologias baseadas em luminância”, *in: Revista de Direito da Responsabilidade*, Ano 2, 2020, pp. 1007-1043, (pp. 1034-1039).

suporta o risco pelos danos que eventualmente venha a provocar tendo que lidar com o perigo decorrente da sua imprevisibilidade¹²⁷, isto porque, como vimos, muitas das vezes o fundamento da responsabilidade não assenta num ato culposo, devendo, por isso, ter o seu alicerce no controle de um risco, ou talvez, com maior rigor, no controle de potenciais danos aliado ao princípio da justiça distributiva, segundo a qual quem tiver o lucro ou, em todo o caso, o benefício de uma certa coisa (*in casu*, de um algoritmo decisório), deve suportar os correspondentes encargos (*ubi commodum ibi incommodum*) apesar da nossa restrita solução legislativa que acolhe, como regra, a responsabilidade por factos ilícitos com base na culpa, uma vez que só existe a obrigação de indemnizar independentemente da mesma nos casos especificados na lei (cfr. n.º 2 do art. 483º CC). Não obstante, é ainda merecida uma menção honrosa à ideia defendida por alguns de imputar responsabilidades de acordo com a decorrente de produtos defeituosos em paralelo com o que é aplicado aos veículos autónomos¹²⁸, também ela se mantendo na alçada da responsabilidade objetiva e beneficiando das características presentes no velhinho e ainda em vigor DL n.º 131/2001, de 24 abril (na sua versão mais recente e atualizada) que transpõe a Diretiva n.º 1999/34/CE, de 10 maio, sendo de realçar o facto de ter um conceito amplo de produtor e de produto, a par de uma eventual responsabilidade solidária, ainda que a aplicação deste regime contenha algumas lacunas que nos retraem e impede que esta seja a melhor opção face às idiossincrasias do nosso caso concreto, desde logo, a dificuldade de ligar o empregador a este tipo de responsabilidade.

5.3. Da reação perante a discriminação

Premente na discussão sobre as questões aqui em análise consiste

127 Cfr. MEDON, Filipe, & TEFFÉ, Chiara Spadaccini de, *op. cit.*, pp. 317-325.

128 *Id.* 124 *in fine*. Para uma explicação do regime invocado e de um possível paralelo, *vide*, entre nós, COELHO, Vera Lúcia Paiva, “Responsabilidade do produtor por produtos defeituosos. Teste de resistência ao DL n.º 383/89, de 6 de novembro, à luz da jurisprudência recente, 25 anos volvidos sobre a sua entrada em vigor”, *in: Revista Eletrónica de Direito*, n.º 2, CIJE, Porto, 2017, pp. 1-53 e SILVA, Gonçalo Viana da, “Veículos autónomos: um novo desafio para o direito português”, *in: Data Venia – Revista Jurídica Digital*, Ano 8, n.º 11, s/1, 2020, pp. 5-98.

em saber como proceder¹²⁹ diante de uma efetiva discriminação algorítmica direta¹³⁰ ou indireta¹³¹, o que contestar e contra quem, perante a existência de uma aparente prova diabólica que deriva do facto de nem o RGPD, nem a própria lei de execução conterem qualquer orientação relativamente ao direito de contestar *ADMs* e ao tipo e grau de envolvimento humano necessário para tal, dificuldade probatória esta que já se revela um problema quanto mais neste contexto algorítmico¹³².

Estas vicissitudes são transpostas para o trabalho tradicional, mas também para as novas formas de prestar labor que decorrem das novas tecnologias de informação e comunicação (NTIC), como sucede no interessante caso das plataformas digitais. Tomando como exemplo a *Uber* na qual o algoritmo possui o poder imenso (de entre vários) de despedir trabalhadores consoante o *rating* dado pelos utilizadores, o problema assenta na eventual *score* que não é objetiva, i.e., podemos estar a falar do melhor condutor, amável, preocupado com o bem-estar dos seus clientes e dotado de habilidades ímpares de condução, mas que por ser de determinada nacionalidade, etnia ou género recebe baixas pontuações reiteradas, fazendo com que seja automaticamente

129 Recorde-se que o direito de ação também é conferido à associação sindical que represente o trabalhador (cfr. n.º 1 do art. 5º CPT). Para mais, o art. 80º RGPD concede ao titular o direito de mandar uma entidade não lucrativa, organização ou associação para apresentar uma reclamação ou exercer um direito de proteção de dados em seu nome, o que significa que qualquer sindicato pode, em nome dos seus membros, exercer um direito de acesso, retificação, eliminação ou oposição (entre outros) não só em tribunal, mas também perante o empregador ou uma agência nacional de proteção de dados.

130 Seguimos a definição da alínea a) do n.º 1 do art. 23º CT, semelhante à anglo-saxónica *disparate treatment*.

131 Seguimos a definição da alínea b) do n.º 1 do art. 23º CT, semelhante à anglo-saxónica *disparate impact*. Por levantar mais problemas que a sua congénere, é relevante deixar a nota de que, aqui não é relevante o *animus* discriminatório, sendo antes o efeito de determinada conduta que deve relevar (cfr. TEDH, *Biao v. Denmark*, de 24.05.2016, application n.º 38590/10, par. 91, 92 e 103) ao visar disposições, critérios ou práticas que parecem ser neutras, mas que, de facto, discriminam, podendo ser o exemplo das situações já mencionadas em que os processos de *ML* utilizam dados que incorporam preconceitos do passado, levando assim a resultados imprecisos e não fiáveis. Sobre o assunto, vide KAMARINOU, Dimitra, MILLARD, Christopher & SINGH, Jatinder, “Machine Learning with Personal Data”, in: *Queen Mary School of Law Legal Studies*, Research Paper n.º 247, 2016, pp. 1-23, (p. 16), acerca da implementação desta lógica face a indivíduos pertencentes a grupos minoritários.

132 Cfr. KAJTÁR, Edit e MESTRE, Bruno, “Redes sociais e o direito à privacidade dos trabalhadores na fase pré-contratual: algumas questões e considerações comparativas”, in: *Prontuário de Direito do Trabalho*, CEJ, n.º 2, 2016, pp. 219-243, (pp. 220-221).

desativado da plataforma. Dito isto, como é que se poderá provar o impacto discriminatório do sistema de classificação atribuído? É que as classificações são intencionalmente subjetivas e visam captar o nível geral de satisfação de um cliente com o serviço prestado, mas devido ao seu cariz arbitrário e pessoal, torna-se complicado aferir das efetivas razões que levaram os utentes a atribuir uma classificação considerada fraca (i.e., abaixo de 4.6 “estrelas”¹³³), crucial em termos de manutenção do posto de trabalho. Nos EUA, o Supremo Tribunal mantém um precedente importante em *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971) observando que “[t]he touchstone is business necessity. If an employer practice (...) cannot be shown to be related to job performance, the practice is prohibited”, cujo *standart* é de tal forma amplo que concede ao empregador uma fácil escapatória e uma justificação suficiente e simples, possibilitando que a *Uber* recuse reter condutores sujeitos a avaliação discriminatória por integrar o sistema classificatório que lhe permite gerir uma grande população geograficamente distribuída e transitória de milhares de trabalhadores em todo o mundo, e muitas alternativas menos discriminatórias levariam à imposição de custos significativos suportados pela *Uber* fruto de uma eventual alteração do sistema que prevê o critério de decisão algorítmico¹³⁴, o que por sua vez impediria a empresa de fornecer o seu serviço principal, sendo incapaz de realizar o seu objetivo comercial ao utilizar um sistema de avaliação alternativo.

Face a esta realidade, defendemos a abolição de um critério assente em necessidades empresariais que se revela demasiado amplo, discricionário e facilmente atingível, desequilibrando uma balança que por defeito se revela deficitária e, *ipso facto*, pouco eficiente para o problema sensível que representa, sendo ao invés necessária uma maior objetividade materialmente mais acessível à parte precária, munindo-a dos meios próprios de modo a que consiga atingir o seu propósito ou no limite, que lhe dê essa oportunidade. Para tal, plataformas como a *Uber* poderiam recolher dados sobre as características demográficas dos condutores e associá-las às classificações que são atribuídas, incluindo informação acerca do nível de probabilidade que determinados condutores têm de receber classificações baixas comparativamente a outros e,

133 Cfr. BAROCAS, Solon, *et al.*, *op. cit.*, p. 259.

134 *Id.* 134, pp. 268-269.

consequentemente de receberem avisos de potencial desativação da plataforma. Mais: deve exigir-se uma mínima justificação dos clientes acompanhada de determinada classificação, principalmente para as mais baixas, *e.g.*, se se deveu a questões relacionadas com a condução, personalidade ou higiene, permitindo fornecer orientações específicas ao condutor e mesmo à plataforma de modo a aferir da veracidade destes dados mais objetivos, o que poderia levar a uma mudança concreta no comportamento do condutor e/ou a demonstrar ao seu superior que a respetiva fundamentação se mostra infundada, não justificando o seu despedimento, sendo indicadores como estes que podem servir melhor os interesses dos trabalhadores de plataforma ao facilitar a identificação de padrões problemáticos no decurso da atividade e permitindo uma efetiva defesa da sua parte. Por fim, deverá repensar-se no critério de tomada de decisão acerca de algo tão sensível como é o despedimento, podendo as classificações atribuídas ser um fator a considerar para esse apuramento juntamente com outras técnicas de avaliação não devendo recair puramente sobre aquelas, toda a decisão¹³⁵.

Perante o cenário português que contraria a regra geral no direito probatório de que quem alega um facto constitutivo de um direito que invoca deve prová-lo, em matéria de proteção e promoção da igualdade e não discriminação em sede laboral o legislador teve a noção da dificuldade existente neste campo, pelo que veio a consagrar mecanismos com vista à “inversão” do ónus da prova tendo por base o contributo europeu¹³⁶. Assim, nos termos do n.º 5 do art. 25º CT, o demandante tem o ónus de alegação e o empregador o ónus da prova, contudo, continua a não facilitar a tarefa da parte vulnerável uma vez que tem o ónus de demonstrar que é membro de uma classe protegida e que sofreu um tratamento prejudicial fruto da decisão algorítmica que teve um impacto díspar nas respetivas características protegidas, competindo-lhe

¹³⁵ *Id.* 134, pp. 271-273.

¹³⁶ Chamamos a atenção para as Diretivas Europeias (mais recentes) sobre a matéria do combate à discriminação, sendo de realçar o n.º 1 do art. 8º Diretiva 2000/43/CE, de 29 de junho, e o n.º 1 do art. 10º Diretiva 2000/78/CE, de 27 de Novembro, ao estabelecerem que “os E-M tomarão medidas necessárias para assegurar que incumba à parte demandada provar que não houve violação do princípio da igualdade de tratamento, cabendo ao demandante apresentar, perante o tribunal ou outra instância competente, elementos de facto constitutivos da presunção de discriminação direta ou indireta”. Destes enunciados há quem extraia a conclusão de que não se trata de uma inversão do ónus da prova, mas sim de aligeirar o ónus da prova a cargo do autor. Porém, nos respetivos preâmbulos, apesar do seu contributo tender

a prova do dano sofrido e pelo qual pretende ser indemnizado, situação em que opera a regra geral do n.º 1 do art. 342º CC porque constitutiva do direito, portanto, uma tarefa quase impossível sem acesso aos dados e ao *source code* do algoritmo. Assim, esta alegada “inversão”, ainda que louvável, não resolve as dificuldades do demandante, pelo que preferimos a qualificação de repartição do ónus da prova¹³⁷, pois apesar de não lhe competir provar que foi efetivamente discriminado, sendo antes da competência da parte passiva (empregador) que trará para a demanda um conjunto de factos dos quais se possa concluir que uma determinada decisão não é realizada com intuítos discriminatórios antes assentando em dados objetivamente fundados que terá de fazer prova de estarmos ante uma presunção de discriminação, a verdade é que consegue facilmente refutá-la ao demonstrar que a decisão está relacionada com a própria atividade profissional e se revela consistente com a necessidade empresarial (*business necessity*)¹³⁸, não obstante o queixoso ainda poder demonstrar que existe uma alternativa menos discriminatória que o empregador não adotou, portanto, queremos com isto evitar que o n.º 5 do art. 25º CT, quando confrontado com questões de prova referentes à nova realidade da IA nomeadamente de algoritmos discriminatórios, se transforme num preceito meramente simbólico, dotado de um vazio de conteúdo. Exemplificando, se um

para a conotação meramente interpretativa não vinculativa, partem do pressuposto de uma verdadeira presunção de discriminação, conforme se extrai dos respetivos Considerandos (21) e (31), nos quais se declara que se impõe a adaptação das regras do ónus da prova em caso de presumível discriminação com a correspondente exigência de que o mesmo incumba à parte demandada, o que leva FIALHO, Manuela Bento, “Igualdade no trabalho. Um caminho aberto, uma estrada por pavimentar...”, in: *Prontuário de Direito do Trabalho*, n.º 76-78 (número especial em homenagem à obra do Dr. Vítor Ribeiro), CEJ, 2007, pp. 91-103 a entender que toda esta formulação legal não deixa dúvidas de que se está perante uma verdadeira inversão do ónus da prova e não apenas de uma maior facilitação do mesmo. No direito comparado também se exige esta solução, vide KIM, Pauline T., “Big Data and Artificial Intelligence: New Challenges for Workplace Equality”, in: *University of Louisville Law Review*, Vol. 57, Issue 2, 2019, pp. 312-328, (p. 324), que pregoa pela existência de uma inversão do ónus da prova e, por sua vez, de uma presunção *iuris tantum* de culpa do empregador. Na mesma linha, vide a fundamentação de AJUNWA, Ifeoma, *op. cit.*, pp. 1726-1734.

137 Acompanhamos de perto CARVALHO, Paulo Morgado de, “Ónus da prova em caso de discriminação”, in: *Estudos dedicados ao Professor Doutor Bernardo da Gama Lobo Xavier*, Vol. 3, Lisboa, 2015, pp. 109-136, (pp. 119-122) e PALMA RAMALHO, Maria do Rosário, *Tratado de Direito do Trabalho: Parte II - Situações Laborais Individuais*, 7ª edição revista e atualizada, Almedina, Coimbra, 2019, pp. 180-181.

138 Porque a verdade é que temos de prevenir o que sucede, de forma muito usual nos EUA como relatado por BAROCAS, Solon, *et al.*, *op. cit.*, pp. 256-279 de que se uma empresa

candidato a emprego se considerar discriminado diretamente por determinado fator de discriminação terá que alegar qual a decisão que lhe foi imposta (*in casu*, de não o recrutar) e que o foi em função e devido a esse fator, indicando outro(s) candidato(s) a emprego relativamente aos quais tal medida não logrou. Já o empregador terá que alegar e provar que a decisão que impôs àquele candidato encontrava justificação em determinada circunstância¹³⁹, sendo neste pormenor que, mantendo-se silente a Lei de Execução, parece compactuar, *mutatis mutandis*, com a solução americana de que a sua justificação tende a prevalecer ao se invocarem *business necessities*, *e.g.*, alegando que o candidato não tem qualidades suficientes ou adequadas às necessidades da empresa após avaliação do *CV*. O mesmo se diga para as situações de discriminação indireta, cabendo, *e.g.*, ao candidato alegar a prática implementada e que essa prática coloca o grupo onde está inserido numa posição de desvantagem através de um exercício comparativo, do outro lado, resta ao demandado alegar e provar que a prática se alicerça em determinado fim legítimo e que os meios para alcançar tal fim são adequados e necessários. Ora, o empregador deve suportar o ónus de demonstrar que o modelo é estatisticamente válido e substantivamente significativo, em oposição a meramente relacionado com o trabalho *e.g.*, que foi construído utilizando dados que são precisos, imparciais e representativos permitindo um melhor contraditório, o que apenas será possível com uma certa abertura do *source code* algorítmico de modo a que o trabalhador seja capaz de rebater o que foi dito e apoiar-se em algo palpável ao fundamentar a sua contestação, pois se não se compreende os processos no *engine room* não será possível identificar violações da privacidade, as causalidades subjacentes e, *ipso facto*, a culpabilidade. Daí a necessidade de distribuir o ónus da prova de

decide quem contratar por via de um algoritmo que se revela claramente preconceituoso mas bem sucedido em prever (o tal “valor preditivo”) o desempenho do trabalhador, isso contaria como necessidade empresarial e iria enquadrar-se no argumento de resguardar os interesses da empresa e não como discriminação, sendo assim entendido que uma prática que tem efeito discriminatório é justificada por estar diretamente relacionada com o desempenho no trabalho. Noutras palavras, a lei existente em alguns países parece “apoiar” o uso de algoritmos preconceituosos. Utilizando outros exemplos recorrendo aos principais fatores de discriminação, *vide* KLEINBERG, Jon, LUDWIG, Jens, MULLAINATHAN, Sendhil & SUNSTEIN, Cass R., “Discrimination in the Age of Algorithms”, *in: Journal of Legal Analysis*, Vol. 10, 2018, pp. 113-174, (pp. 142-143).

139 Neste sentido, *vide* o Ac. STJ de 20.06.2018, processo n.º 31947/15.9T8LSB. L2.S1, (ANTÓNIO LEONES DANTAS).

uma forma que possibilite efetivamente a parte precária de se defender mais facilmente nos procedimentos legais oriundos desta nova realidade¹⁴⁰.

Por último, aplicando o regime atual a este contexto algorítmico, a prática de um qualquer ato discriminatório lesivo não deixa de conferir o direito a uma indemnização por danos patrimoniais e não patrimoniais nos termos do art. 28º CT, devendo ser eficaz, proporcional e dissuasiva nos termos do arts. 15º e 17º da Diretiva 2000/43/CE, de 29 de junho e da Diretiva 2000/78/CE, de 27 de novembro respetivamente. De igual modo, recorda-se que no caso de candidato ilicitamente excluído, a obrigação de o admitir será uma solução jurídica inviável e de difícil execução no nosso ordenamento ao contundir com a natureza jurídica da prestação laboral, nomeadamente o cariz *intuitu personae* e a liberdade de gestão empresarial, e ao limitar a autonomia privada e liberdade negocial do empregador, atentando contra a possibilidade de denúncia pelo mesmo durante o período experimental¹⁴¹. Diferentemente, no caso do trabalhador, o contrato já estaria em execução pelo que o empregador estaria obrigado à reconstituição natural da situação que existiria caso tal conduta não tivesse ocorrido.

140 Neste sentido, vide DATATILSYNET (Norwegian DPA), “Artificial intelligence and privacy”, 2018, disponível em: <<https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>>, p. 23 e MARTINI, Mario, *Fundamentals of a regulatory system for algorithm-based processes - Expert opinion prepared on behalf of the Federation of German Consumer Organisations*, Speyer, 2019, pp. 35-37, este último A. apologista de que os controladores de dados teriam então de refutar a presunção de prova apresentando registos das sequências de processamentos do algoritmo, fornecendo provas de uma supervisão adequada dos processos técnicos ou desafiando de outra forma a presunção de causalidade.

141 A nível comparado, o ordenamento espanhol apresentou uma solução distinta no caso STSJ Islas Canarias, Las Palmas, secção 1, de 22.12.2008, resolução n.º 1860/2008, (MARIA GARCIA HERNANDEZ) onde, para além da existência de indemnização durante um processo de seleção de pessoal, declarou a nulidade radical da conduta lesiva (*in casu*, a discriminação do candidato) e a reposição da situação no momento anterior a se produzir a violação do direito fundamental, impondo a obrigação de o contratar, ainda que a jurisprudência europeia nesta matéria (cfr. TJUE, *Von Colson e Kamann v. Land Nordrhein-Westfalen*, 10.04.1984, Case C-14/83) tenha vindo a defender que a contratação não deve ser exigida, sendo que o mais importante é que as indemnizações sejam eficazes, proporcionais e dissuasivas. Sobre este assunto, fazendo uma recensão de jurisprudência, vide MOLL NOGUERA, Rafael, “La discriminación en el acceso al empleo público por razón de edad y su relación con las medidas positivas: cuestiones procesales y sustantivas a raíz de la STSJ C. Valenciana de 20.12.2018 (rec. 3174/2018)”, in: *Labos*, Vol. 1, n.º 2, 2020, pp. 109-127.

6. Considerações finais

Tomando como emprestada a emblemática expressão de Ronald Reagan “*trust, but verify*”, que utilizou o provérbio do inimigo Russo num contexto de Guerra Fria para descrever a abordagem das relações entre os EUA e a URSS, a verdade é que é possível utilizá-la como paralelo para a atual revolução digital iminente e sem igual¹⁴² que traz questões sensíveis para as relações laborais, denotando-se uma verdadeira gestão digital do trabalho através do uso de algoritmos que acabam por estruturar silenciosamente a vida dos trabalhadores, pelo que apenas nos será possível confiar em sistemas automatizados decisores ao verificarmos e nos prevenirmos com o devido olhar atento que se exige. Tendo isto presente, de modo a eliminarmos as suas características de *WMDs* e fazer sobressair as suas qualidades, propõe-se atender ao a algumas ideias chave¹⁴³:

I. Criar regulamentação e legislação sobre o uso dos algoritmos - como a Proposta de Regulamento Inteligência Artificial¹⁴⁴ ou a própria Resolução 2020/2012(INL)¹⁴⁵ - é imperativo, dado o alcance, universalidade e sensibilidade dos problemas registados, uma vez que o principal argumento contra a IA é motivado, legitimamente, pelo medo do desconhecido em contraste com a falta de adaptação legislativa neste campo, daí que a nossa lei atual (tal como na grande maioria dos países), não está ainda preparada para abordar casos desta

142 Nas palavras de Coelho Moreira, Teresa, op. cit., pp. 245-264, “esta revolução [digital] é diferente já que se atravessa um período de uma evolução sistémica excecional e raramente comparada a qualquer outra prévia. Não se trata de crise, mas de uma verdadeira metamorfose, não de passagem entre dois estados mais sim de um salto para o desconhecido”.

143 Alguns dos pontos foram inclusive reforçados e enaltecidos nos arts. 9º e 14º da Carta Portuguesa de Direitos Humanos na Era Digital (Lei n.º 27/2021, de 17 maio), bem como pelo Livro Verde, “Sobre o futuro do Trabalho (trabalho, solidariedade e segurança social)”, versão preliminar, junho 2021, disponível em: <<https://www.portugal.gov.pt/pt/gc22/comunicacao/documento?i=livro-verde-sobre-o-futuro-do-trabalho>>, pp. 47-54 e pp. 59-69 nos dados estatísticos apresentados e linhas de reflexão sobre a matéria.

144 Proposta de Regulamento do Parlamento Europeu e do Conselho COM(2021) 206 final, que estabelece regras harmonizadas em matérias de Inteligência Artificial, de 21.04.2021, disponível em: <<https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52021PC0206>>.

145 Resolução (2020/2012(INL)) do Parlamento Europeu, que contém recomendações à Comissão sobre o regime relativo aos aspetos éticos da inteligência artificial, da robótica e das tecnologias conexas, de 20.10.2020, disponível em: <https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_PT.html>.

natureza, pelo que se defende a procura de uma garantia de que a busca por maior produtividade não resulte na violação de direitos fundamentais, em riscos ocupacionais e repercussões a nível de saúde para os trabalhadores envolvidos como vimos no caso da *Amazon* ou da *Uber*. Assim, parece-nos fundamental padronizar aquilo que é admissível que seja programado, criando uma espécie de ética uniforme para os algoritmos e uma lei-modelo que cada E-M implementaria na sua legislação nacional, seguida de um relatório revelador de soluções e justificações adotadas no seu território enquanto possível técnica legislativa¹⁴⁶. Nesta linha, as leis de combate à discriminação podem revelar-se nucleares ao fornecerem proteções importantes para os trabalhadores, sendo, por isso, necessária e urgente uma atualização e adaptação da legislação laboral¹⁴⁷ de modo a evitar que o lado mais frágil fique (ainda mais) vulnerável face à entrada rompanete da IA no campo das relações laborais. A nível europeu esteve bem o RGPD em trazer algo à discussão que a Lei de Execução não foi capaz de vingar, tratando-se quanto a nós, de uma clara falha legislativa incapaz de estar a par dos tempos modernos da economia digital que a cada dia que passa vai intensificando os seus (velhos) problemas e criando novos sem apresentar respostas. De igual modo, os avanços tecnológicos são mais eficazes quando acompanhados por uma maior clareza quanto ao apuramento da responsabilidade algorítmica, qual o tipo que estará em causa e como se poderá reagir na prática perante este novo contexto, daí a importância da negociação e acordos coletivos por parte das organizações de trabalhadores e empregadores enquanto recurso basilar para o efeito, podendo (e devendo) contribuir para a solução de assuntos relacionados com o uso de tecnologia digital, coleta de dados e algoritmos que direcionam e disciplinam a força de trabalho procurando assegurar transparência e sustentabilidade social, sendo que a conformidade de tais práticas com a regulamentação pode, *rectius*, deve tornar-se um objetivo crucial de diálogo entre empregadores e trabalhadores¹⁴⁸

146 Neste sentido, KAJTÁR, Edit e MESTRE, Bruno, *op. cit.*, pp. 241-243.

147 Neste sentido, BARZILAY, Arianne Renan, & BEN-DAVID, Anat, “Platform Inequality: Gender in the Gig-Economy”, *in: Seton Hall Law Review*, Vol. 47, n.º 2, 2017, pp. 393-431, (pp. 429-430).

148 Enaltecendo o importante papel da negociação coletiva na flexibilização da regulamentação algorítmica e da sua adaptação à inovação tecnológica, *vide* DE STEFANO, Valerio, “Negotiating the Algorithm: Automation, Artificial intelligence and labour

permitindo afastar a dimensão puramente unilateral da governação do trabalho perante esta nova realidade.

II. A necessidade de uma *human in command approach*. Embora seja possível ter a IA a decidir questões como contratar trabalhadores, ou despedilos, tais decisões devem ser sempre implementadas após uma abordagem de “ser humano no comando” defendida expressamente no seio comunitário¹⁴⁹ sendo, por isso, crucial que o ser humano permaneça legalmente responsável pela decisão e os consequentes efeitos nefastos¹⁵⁰, pelo que o fato das decisões

protection”, in: *Comparative Labor Law & Policy Journal*, Vol. 41, n.º 1, 2019, pp. 1-32 ao salientar que os acordos coletivos poderiam tratar do uso de tecnologia digital, coleta de dados e algoritmos que direcionam e disciplinam a força de trabalho, garantindo transparência, sustentabilidade social e a conformidade dessas práticas com a regulamentação. Na mesma linha, *vide* atentamente o próprio Relatório OIT, *op. cit.*, pp. 42-44 juntamente com TODOLÍ-SIGNES, Adrian, *op. cit.*, pp. 11-13 e MARTINI, Mario, *op. cit.*, pp. 39-41, sugerindo este último A., a criação de um Código de Responsabilidade Algorítmica que consistiria numa comissão governamental composta por representantes de trabalhadores e empregadores, empresas de software, administração pública, e cientistas com o intuito de formular recomendações sobre como os algoritmos devem ser utilizados em áreas sensíveis aos direitos fundamentais, como são exemplo as relações laborais.

149 Cfr. EESC, “Artificial Intelligence - The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society”, Adopted on 31.05.2017, disponível em: <<https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and>>, p. 43. Aliás, o Parecer CESE, *op. cit.*, indica expressamente que a “pré-condição de que o desenvolvimento da IA seja responsável, seguro e útil, onde as máquinas permanecem máquinas e as pessoas mantêm o controle sobre essas máquinas a todo o tempo devem ser rigorosamente seguidos também em relação ao [âmbito] laboral” e, na mesma linha se pronuncia o Relatório OIT, *op. cit.*, p. 13, de que “potenciar e gerir a tecnologia a favor do trabalho digno, significa também adotar uma abordagem da IA baseada no “ser humano no comando”, que garanta que decisões finais que afetem o trabalho sejam tomadas por seres humanos e não por algoritmos, [pois o] trabalho não é uma mercadoria”.

150 Segundo o GT29º, *op. cit.*, p. 27, “[deve primar-se pela existência de] meios [que permitam] ao titular dos dados manifestar o seu ponto de vista e contestar a decisão (...) e um mecanismo de intervenção humana em casos definidos, [como] a disponibilização de uma hiperligação para um procedimento de recurso no momento da entrega da decisão automatizada ao titular dos dados, com os prazos de revisão acordados e o nome de um ponto de contacto para quaisquer dúvidas”, e a verdade é que num estudo de investigação realizado por FISCHER, Sarah & PETERSEN, Thomas, *Was Deutschland über Algorithmen weiß und denkt: Ergebnisse einer repräsentativen Bevölkerungsumfrage*, Bertelsmann Stiftung, 2018, pp. 24-30, cerca de 79% da opinião pública revelou que se encontra desconfortável e reticente quanto ao facto dos algoritmos julgarem as pessoas e tomarem decisões de forma instantânea e automática, preferindo as decisões tomadas por pessoas por mais defeituosas e subjetivas que sejam. Nesta esteira, concedendo argumentos e razões do porquê de devermos confiar nas decisões humanas ao invés da AI, *vide* LA DIEGA, Guido Noto, *op. cit.*, pp. 1-33 e ODDENINO, Alberto, “Decisioni

serem tomadas após processos baseados em algoritmos nunca deve ser uma razão suficiente para excluir a responsabilidade humana nem que seja por uma questão de dignidade¹⁵¹, conferindo ao demandante o direito de entrar em contato com um responsável humano de forma a poder reagir perante uma decisão que lhe foi desfavorável, de justificar e fundamentar devidamente a sua contestação, podendo exigir um direito de explicação que muitas vezes se revela paradoxal¹⁵². No fundo, devemos evitar uma abordagem de desumanização do trabalho por uma máquina onnipotente e onnipresente que o instrumentaliza e desvaloriza os valores humanos¹⁵³, uma nova realidade que deixa cicatrizes em termos de saúde física e mental¹⁵⁴ levando ao aumento de pressão, stress, ansiedade, depressão e até *burnout*, refletindo-se na qualidade da própria prestação do labor, podendo as tecnologias digitais fazer surgir novos riscos ou intensificar riscos profissionais já existentes.

III. Assegurar transparência e o direito à explicação é um ponto que não deve ser subestimado. A infra informação tecnológica¹⁵⁵ faz-se notar nos

algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano”, in: *DPCE online*, Vol. 42, n.º 1, 2020, pp. 199-217.

151 Como refere ZARSKY, Tal Z., “Incompatible: The GDPR in the Age of Big Data”, in: *Seton Hall Law Review*, Vol. 47, Issue 4, 2017, pp. 995-1020, (pp. 1016-1018), quando confrontado com decisões cruciais, um ser humano deve ser tratado com a dignidade de ter um decisor humano a tratar do seu assunto pessoal, não uma máquina.

152 Assiste-se a este paradoxo no caso da *Uber*, na qual os termos e condições globais do condutor de *Uber* preveem que os despedimentos automatizados têm lugar, mas ao mesmo tempo no seio europeu pregoa-se por um direito de contestar perante um ser humano. Sobre o assunto, vide DENCİK, Lina, *et al.*, *op. cit.*, p. 467.

153 O mesmo será dizer que a dignidade é violada quando a pessoa deixa de ser considerada como um sujeito individual e um fim em si mesmo, sendo, ao invés, tratada como um instrumento ou um meio de realização de fins alheios.

154 Quanto aos possíveis riscos para a saúde dos trabalhadores como consequência da algocracy, vide a nível europeu o Relatório EU-OSHA, “Foresight on new and emerging occupational safety and health risks associated with digitalisation by 2025”, 2018, disponível em:

<<https://osha.europa.eu/en/tools-andpublications/publications/foresight-new-and-emerging-occupational-safety-and-healthrisks/view>>, pp. 55-57 e o Relatório OIT, “Seguridad y salud en el centro del futuro del trabajo”, Ginebra, 2019, disponível em: <https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_686762.pdf>, p. 31. Na doutrina, vide Todolí-Signes, Adrian, “Complying with the first law of robotics: An analysis of the occupational risks associated with work directed by an algorithm/artificial intelligence”, in: SSRN, 2020, pp. 1-32, (pp. 14-19) e o interessante estudo de Popma, Jan, “The janus face of the “New ways of work”, Rise, risks and regulation of nomadic work”, in: SSRN, 2013, pp. 1-40, (p. 11), sobre o conceito de tecno-stress cindido em cinco elementos subjacentes: tecno-ansiedade, tecno-fadiga, tecno-adição, tecno-invasão e tecno-sobrecarga.

155 Aliás, num estudo levado a cabo por FISCHER, Sarah & PETERSEN, Thomas, *op. cit.*,

dias correntes, i.e., a privação, desconhecimento e compreensão dos critérios e técnicas que estão a ser aplicados em determinada decisão algorítmica, impedindo o demandante de avaliar e reclamar fundadamente, fomentando a sua inércia perante resultados arbitrários e/ou discriminatórios sendo, por isso, imperativo que possa obter as respostas que procura. Deste modo, é importante reforçar que “explicabilidade” não é sinónimo de “transparência”, na medida em que ser capaz de compreender o processo por meio do qual uma decisão foi tomada não é o mesmo que conhecer todos os passos tomados para atingir a mesma¹⁵⁶. Nesta linha, o dualismo entre o direito à explicação *ex post* e o direito à informação geral *ex ante* deve ser superado, pelo que defendemos que aqueles dois conceitos devem fundir-se num único denominado de “legibilidade” devendo os trabalhadores ser capazes de compreender a importância e implicações do processamento algorítmico de dados e por que razão um sistema automatizado chegou a uma determinada decisão, o que também se revela importante em termos de apuramento de responsabilidade e transparência, existindo quem proponha um modelo de explicações contra factuais, ou seja, o dever de esclarecer os indivíduos visados por *ADMs* acerca do que teria de mudar para receber um resultado desejado no futuro com base no atual modelo de tomada de decisão¹⁵⁷. Mais: a lei prevê um acesso a “informação significativa sobre a lógica envolvida”, quererá isto significar que a lei exige uma explicação da funcionalidade do sistema (i.e., da lógica, significado, consequências previstas e funcionalidade geral) ou uma explicação de decisões específicas (i.e., da fundamentação, razões e circunstâncias)? Entendemos que a última abordagem será preferível pois em contexto, tal acesso parece associado ao direito de contestar uma decisão do n.º 3 do art. 22º RGPD¹⁵⁸ e, para o efeito, poderia recorrer-se ao campo da AI explicável ou *XAI*

pp. 12-15, quase metade dos inquiridos (45%) não conseguiu dar uma resposta clara acerca do conceito de algoritmo ou da própria IA em geral. Revela-se, por isso, bastante importante combater esta ignorância do homem médio relativamente aos algoritmos enquanto arma cada vez mais utilizada contra si mesmo.

156 Neste sentido, IBNOUHSEIN, Mohamed Issam, & PÉGNY, Maël, “Quelle transparence pour les algorithmes d’apprentissage machine?”, 2018, pp. 1-34, disponível em: <<https://hal.inria.fr/hal-01791021>>, pp. 18-29.

157 A favor: MITTELSTADT, Brent, *et al.*, *op. cit.*, p. 4; Contra: TABARRINI, Camilla, *op. cit.*, pp. 135-143.

158 Neste sentido, POWLES, Julia & SELBST, Andrew D., *op. cit.*, p. 233.

cuja técnica apresenta um propósito de transparência, assumindo formas que possam ser compreendidas pelo ser humano¹⁵⁹.

Apesar dos algoritmos serem ferramentas úteis de auxílio ao empregador, não devem ser os fatores determinantes nas decisões que afetam os seus direitos vitais pelo que devemos ter em conta o explanado pela jurisprudência italiana ao elencar e munir o demandante de efetivas ferramentas de reação i.e., que um documento que inclui o algoritmo utilizado e a mera explicação da lógica em termos matemáticos e informações relativas à complexidade da construção e funcionamento do mesmo não atende, de *per si*, ao requisito legal, devendo antes ser interpretado como sendo necessária a divulgação do algoritmo com uma explicação em termos não técnicos da lógica da decisão e dos critérios em que se baseia, pois só desta forma o titular dos dados poderá contestar da melhor maneira a decisão e uma interpretação diferente não estaria em conformidade com o direito à ação e um processo equitativo, bem como o direito a um recurso efetivo nos termos dos par. 1º e 2º do art. 47º CDFUE e arts. 6º e 13º CEDH, que corresponde na nossa CRP, ao acesso ao direito e tutela jurisdicional efetiva do art. 20º CRP. Pregoamos assim por uma abertura segura da Caixa de Pandora, limitada, controlada, e acima de tudo, cognoscível na clareza do seu conteúdo pelo demandante. A não ser assim, sucede o que acontece em larga escala ao dia de hoje: o empregador, de forma consciente ou involuntária, subordina o seu controle e responsabilidade a um algoritmo com todos os riscos e perigos que isso representa, e os trabalhadores não podem exigir saber com que base, em que critério e com que peso e medida uma decisão algorítmica foi tomada contra si de modo a recorrer da mesma. No fundo, nunca devemos chegar à situação em que o empregador pode simplesmente encolher os ombros e dizer: “o algoritmo disse que o devia contratar ou despedir, mas não sei porquê”. Portanto, para que um direito à

159 Cfr. MITTELSTADT, Brent, RUSSEL, Chris & WACHTER, Sandra, “Explaining explanations in AI”, in: *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 279-288, (pp. 279-288). Para mais desenvolvimentos, vide VAQUERO DE MIGUEL, Guillermo, *Aproximaciones a la Explicación de Decisiones Algorítmicas: Inteligencia Artificial Explicable*, Trabajo Fin de Grado, Universidad Politécnica de Madrid, 2020, pp. 47-53, trazendo à colação as técnicas da metodologia *XAI*, que dentro da área da IA são capazes de “abrir” a *black box* de certos algoritmos, facilitando, desta forma, a transparência e a compreensão do seu funcionamento interno.

informação sobre a lógica envolvida seja significativo deve ser mais do que uma simples regurgitação do *source code*, rejeitando-se a divulgação formal do código algorítmico balizado por *IP rights* e *trade secrets*¹⁶⁰ pois só uma adequada prestação da informação descodificando a efetiva *ratio* subjacente à decisão se demonstra coerente com o RGPD¹⁶¹, conferindo materialmente informação útil que possibilite o demandante de reagir tornando o binómio transparência/privacidade numa dualidade.

IV. A premência da realização de auditorias ao algoritmo. A fiscalização algorítmica revela-se importantíssima para encontrar a justiça uma vez que permite “interrogar” os algoritmos de modo a alterá-los, aperfeiçoá-los e corrigi-los para serem melhores. Desta forma, propõe-se uma espécie de auditoria ao algoritmo¹⁶² cujo processo poderá consistir essencialmente em:

Numa fase (pré) inicial, (i) o algoritmo deverá ter em conta particularidades e características do ser humano, de modo a evitar a sua coisificação, ou seja, a subjetividade acaba por interessar e ser relevante mas na medida certa,

160 Neste sentido, POWLES, Julia & SELBST, Andrew D., *op. cit.*, pp. 233-242 e segundo a DATATILSYNET (Norwegian DPA), *op. cit.*, pp. 21-22, “independentemente do significado das diferenças linguísticas, o responsável pelo tratamento deve fornecer tantas informações quantas as necessárias para que o sujeito dos dados possa exercer os seus direitos. Isto significa que a decisão deve ser explicada de modo a que a pessoa em causa possa compreender o resultado. O direito a uma explicação não significa necessariamente que a *black box* deve ser aberta, mas tem de permitir à pessoa em causa compreender por que razão foi tomada uma determinada decisão, ou o que precisa de ser alterado para que uma decisão diferente possa ser tomada”. Entre nós, *vide* CABRAL, “AI and the Right to Explanation: Three Legal Bases under the GDPR”, in: *Data Protection and Privacy - Data Protection and Artificial Intelligence* (eds., Dara Hallinan, Ronald Leenes, Paul De Hert), Vol. 13, Hart Publishing, 2021, pp. 29-56.

161 Em França, a CNIL, “Comment Permettre à L’Homme de Garder La main? Les enjeux éthiques des algorithmes et de l’intelligence artificielle”, 2017, disponível em: <https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf>, defendeu este princípio da transparência. Em Itália, LA DIEGA, Guido Noto, *op. cit.*, pp. 27-32 propõe como alternativa que se a decisão algorítmica for baseada em dados pessoais por uma entidade privada, este caminho é preferível, já se o tomador de decisão for uma entidade pública, deve optar-se por uma solicitação de FOIA com a justificação de que o sector público tutela interesses superiores e de maior gravidade, pelo que se permite essa ponderação acrescida.

162 Apologista desta ideia, desenvolvendo-a com grande pormenor e detalhe, *vide* KIM, Pauline T., “Auditing Algorithms for Discrimination”, in: *University of Pennsylvania Law Review*, Vol. 166, Issue 1, 2017, pp. 189-203. De igual modo, trata-se de uma visão intensamente sugerida pela Administração Obama em 2016, quando se pronunciava sobre estes novos desafios relacionados com os sistemas algorítmicos (cfr. OBAMA ADMINISTRATION, “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights”, 2016, disponível em: <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf>, pp. 22-24).

devendo o algoritmo incorporar e ter em conta elementos que representem um risco para a segurança e a saúde dos trabalhadores e ser programado de forma a processar a informação que consiga reunir de modo a prevenir os possíveis riscos inerentes à respetiva prestação laboral da mesma maneira que um supervisor deve estar ciente dos perigos subjacentes ao posto de trabalho dos trabalhadores sob sua responsabilidade¹⁶³. Para tal, parece necessário que a avaliação de risco obrigatória seja inserida na programação do algoritmo devendo, se possível, consultar-se os representantes dos trabalhadores de forma a estabelecer os critérios que o algoritmo deve seguir para gerenciar esse trabalho; (ii) deve procurar-se testar várias versões do *software* com casos reais previamente à sua utilização com a específica finalidade de detetar preconceito e arbitrariedade¹⁶⁴ dado que nesse caso já haveria alternativas, devendo o responsável optar pela que fosse menos discriminatória e mais eficaz, permitindo clarificar a sua responsabilidade por falta desse cuidado pois a verdade é que o empregador ou mesmo a empresa que desenvolve o algoritmo muitas vezes não têm “incentivos” para proceder a estas avaliações minuciosas e morosas, ultrapassável por via da criação de um regime sancionatório por falta de cumprimento deste tipo de regras. Quem desenha o algoritmo precisa deste *feedback* de erro, de estudar os *outputs* maliciosos de modo a investigar e descobrir o que deu errado, o que foi mal lido, que dados foram ignorados, sendo desta maneira que os sistemas aprendem e se tornam mais inteligentes ao mesmo tempo que, paralelamente, precisam de *feedbacks* positivos de modo a terem conhecimento de que está efetivamente a resultar, de forma a melhorar

163 Neste sentido, TODOLÍ-SIGNES, Adrian, *op. cit.*, pp. 23-24 e como refere LA DIEGA, Guido Noto, *op. cit.*, p. 11, falta aos algoritmos a capacidade cognitiva e emocional de compreender uma situação da perspetiva do ser humano, uma vez que a empatia tempera os excessos legais.

164 Uma das estratégias da *HireVue* como forma de mitigar preconceitos consiste na eliminação de indicadores que têm um impacto adverso nos grupos protegidos, em que as características que provocam resultados tendenciosos são removidas e os modelos são re-treinados e reavaliados até que o viés não seja detetado (cfr. <<https://www.hirevue.com/why-hirevue/ethical-ai>>). Nesta linha encontra-se o GT29^o, *op. cit.*, p. 32 ao defender a “existência de controlos periódicos de garantia da qualidade dos respetivos sistemas a fim de assegurar um tratamento equitativo e não discriminatório das pessoas (...) e obter garantias contratuais (...) de que o controlo e os testes foram efetuados e que o algoritmo está conforme às normas aceites. Estas medidas deverão ser utilizadas de modo cíclico, ou seja, não apenas na fase de conceção, mas também permanentemente enquanto for aplicada uma definição de perfis às pessoas”.

e aprimorar o algoritmo¹⁶⁵; (iii) considerar o rigor e a *accuracy*, i.e., aferir com que frequência o algoritmo falha, com quem falha e qual é o custo dessa falha, devendo os empregadores de ser capazes de compreender e.g., a razão de um candidato ter sido rejeitado, e nesta linha, tomar em consideração os ciclos viciosos das decisões algorítmicas, os efeitos perniciosos a longo prazo, pará-los e modificá-los, aferindo se o algoritmo usado e desenvolvido pelo sistema de *ML* está a produzir resultados discriminatórios, arbitrários e injustificados.

Numa fase final, (iv) fiscalizar a integridade dos dados através de uma espécie de posição original de RAWLS¹⁶⁶ a partir do qual se levanta o “véu de ignorância” ao nos colocarmos numa “prova às cegas”¹⁶⁷ afastando qualquer preconceito *ab initio* na fase de recrutamento, daí a importância de reformular a coleta histórica de informação que serve de *input* uma vez que um algoritmo é tão bom quanto os dados com os quais trabalha; (v) garantir imparcialidade nessa auditoria - acolhendo a proposta da doutrina alemã e italiana¹⁶⁸ - por via da criação de um órgão independente que seja capaz de o fazer regularmente, se possível através da contratação de especialistas independentes, grupos de direitos humanos para o efeito ou atribuindo-se a empresas de fora o poder de recrutamento com a devida fiscalização por parte da Autoridade para as Condições do Trabalho (ACT). Por outro lado, poderá ainda desenvolver-se a criação de um sistema de deteção de discriminação incorporado na própria

165 Cfr. O’NEIL, Cathy, *op. cit.*, pp. 133 e 209; KLEINBERG, Jon, *et al.*, *op. cit.*, pp. 144-145 e EDWARDS, Lilian & VEALE, Michael, “Enslaving the algorithm: from a right to an explanation to a right to better decisions”?, in: *IEEE Security & Privacy*, Vol. 16, n.º 3, 2018, pp. 46-54.

166 RAWLS, John, *A Theory of Justice*, Oxford University Press, 1972, pp. 17-22.

167 Neste sentido, *vide* GOLDIN, Claudia & ROUSE, Cecilia, “Orchestrating Impartiality: The Impact of Blind Auditions on Female Musicians”, in: *American Economic Review*, Vol. 90, n.º 4, 2000, pp. 715-741, (pp. 737-738) e O’NEIL, Cathy, *op. cit.*, pp. 113-115, recorrendo ao exemplo das *blind orchestra auditions* que na década de 1970 estiveram em destaque ao iniciarem um movimento caracterizado pela realização de audições com o músico escondido atrás de uma partitura, levando a que reputações e interesses subjetivos que provocavam injustiças passassem a não ter qualquer valor na critério de recrutamento e avaliação, a par da etnia, do género ou mesmo da *Alma Mater* do músico, demonstrando que, desde então, a percentagem de mulheres que tocam em grandes orquestras deu um salto quantitativo aumentando em cerca de 5x recorrendo a este método, sendo deste modo que é invocada e utilizada como bom presságio para o paralelo *sub judice*.

168 Cfr. respetivamente, BAROCAS, Solon & SELBST, Andrew D., *op. cit.*, pp. 682-701 e FLORIDI, Luciano, *et al.*, *op. cit.*

ML de modo a preveni-la¹⁶⁹. Deste modo, relativamente à controversia questão que envolve os *trade secrets* e *IP rights*, reforçamos o entendimento de que, mesmo que uma empresa reclame que o algoritmo é propriedade sua e que, por isso, deve ser mantido em segredo, ainda assim deve-lhe ser exigido que ofereça testes robustos para que especialistas externos possam examinar o seu desempenho¹⁷⁰.

Tendo presente estas considerações, revela-se importante o contributo que o RGPD vem trazer com a criação da avaliação do impacto sobre a proteção de dados (AIPD), atuando como mecanismo *ex ante* ao permitir que os controladores avaliem e determinem se existe o risco de uma qualquer discriminação, parcialidade ou erro e, caso isso aconteça, desenvolver medidas para minimizar os danos potenciais, sendo também utilizada para informar a pessoa em causa sobre a lógica subjacente a um processo de *ADM* permitindo-lhe opor-se à decisão e/ou expressar o seu ponto de vista, enquanto processo contínuo e não isolado no tempo. De igual modo, a sua utilização obrigará as organizações a demonstrar a necessidade e proporcionalidade de um qualquer tratamento de dados pessoais relacionados com a IA, a prestar contas de algum prejuízo que possa resultar de uma certa parcialidade ou inexatidão de um sistema e a explicar a lógica subjacente a um necessário *trade-off* entre transparência/privacidade. As AIPDs podem também auxiliar organizações a ponderar sobre as melhores formas de prevenir os riscos mais amplos de danos a trabalhadores ou as implicações éticas para a sociedade em geral¹⁷¹, tentando

169 Neste sentido, propondo a criação de uma ferramenta algorítmica anti discriminatória incorporada no próprio sistema de *ML*, vide FELDMAN, Michael, FRIEDLER, Sorelle A., MOELLER, John, SCHEIDEGGER, Carlos, & VENKATASUBRAMANIAN, Suresh, “Certifying and Removing Disparate Impact”, in: *KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 259-268; BORNSTEIN, Stephanie, *op. cit.*, p. 1110 e MASSENO, Manuel David e SANTOS, Cristiana, “Assuring Compliance of European Smart Tourist destinations with the principles of the general data protection regulation: a roadmap”, in: *Anuário de Proteção de Dados*, (coord. Francisco Pereira Coutinho, Graça Canto Moniz), CEDIS, 2019, pp. 87-108, (pp. 95-107).

170 Veja-se o exemplo da *Pymetrics* que desenvolveu a ferramenta específica *audit-A* para realizar auditorias ao algoritmo e lançou-a como um *software* de código aberto (cfr. <<https://github.com/pymetrics/audit-ai>>).

171 Neste sentido encontra-se o GT29º, *op. cit.*, pp. 26-30, onde embora reforce que o “objetivo de tal processo seria ajudar a fomentar a confiança nas operações de processamento do controlador, e demonstrar responsabilidade e transparência”, entende que a publicação de AIPD’s não é um requisito legal do RGPD. Criticando esta posição, vide TUNCAY, Bilgesu,

corrigir as falhas do algoritmo, respeitando o disposto no art. 25º RGPD e promovendo a proteção de dados desde a conceção e por defeito. Nesta linha, a alínea a) do n.º 3 do art. 35º RGPD destaca a necessidade e obrigatoriedade de o responsável pelo tratamento proceder a uma AIPD (cfr. Considerando (90)) em caso de *ADMs* e definição de perfis, exigindo que o responsável afira dos riscos envolvidos neste campo, podendo igualmente constituir um meio útil de identificação das medidas a introduzir para combater os perigos em matéria de proteção de dados resultantes do tratamento¹⁷², entre as quais: **(a)** informar o titular dos dados da existência de um processo de *ADM* e da lógica subjacente; **(b)** explicar-lhe a importância e as consequências previstas do tratamento; **(c)** proporcionar-lhe meios de se opor à decisão; **(d)** e permitir que manifeste o seu ponto de vista. Porém, não obstante a introdução deste mecanismo, defendemos uma fiscalização a dois tempos de forma complementar¹⁷³, quer *ex ante* (informações genéricas, enquanto atuação preventiva em que se procura evitar ou minorar a ocorrência de danos), quer *ex post* (informações específicas, quando o dano já ocorreu), na medida em que uma intervenção prévia a ser tomada a decisão¹⁷⁴ se revela uma ferramenta importante a par da análise realizada *a posteriori*¹⁷⁵.

Scrutinizing Algorithms for Unlawful and Undesirable Discrimination in Automated Decision-Making: An Analysis of Algorithm Audits and the GDPR, Tilburg Institute for Law, Technology, and Society, 2019, pp. 37-38, afirmando que mesmo que as organizações ou empresas não revelem as suas AIPD's, o resultado das auditorias algorítmicas deve ser tornado público.

172 Nesta linha, em cumprimento do n.º 4 do art. 35º e ao abrigo da alínea k) do n.º 1 do art. 57º ambos do RGPD, a CNPD, por via do RE n.º 1/2018, DR 2ª Série, n.º 231, de 30.11.2018, veio estabelecer a lista de tratamentos de dados pessoais sujeitos a AIPD, devendo a avaliação incluir a descrição sistemática das operações de tratamento previstas e a sua finalidade, a avaliação da necessidade e proporcionalidade das operações de tratamento em relação aos objetivos, a avaliação dos riscos para os direitos e liberdades dos titulares dos direitos e as medidas previstas para fazer face aos riscos.

173 Neste sentido, BOURKE, Georgina, CARFANTAN, Paul-Marie, EICHHOLZER, Annalisa, HUTCHINSON, Ian, JOSHI, Arnav & POWELL, Alison, "Understanding and Explaining Automated Decisions", in: *SSRN*, 2019, pp. 1-28, (pp. 6-12 e 13-18); BESSE, Philippe, *op. cit.*, p. 6 e MEDON, Filipe, & TEFFÉ, Chiara Spadaccini de, *op. cit.*, p. 303.

174 Solução defendida e desenvolvida por BOGEN, Miranda & RIEKE, Aaron, *op. cit.*, p. 45 ao ser implementada no (seu) ordenamento esloveno.

175 Apologistas desta solução isolada, vide COMANDÉ, Giovanni, & MALGIERI, Gianclaudio, *op. cit.*, pp. 243-265; FLAXMAN, Seth & GOODMAN, Bryce, "European Union regulations on algorithmic decision-making and a right to explanation", in: *AI Magazine*, Vol. 38, n.º 3, 2017, pp. 50-57 e KIM, Tae Wan & ROUTLEDGE, Bryan, *Algorithmic Transparency, a Right to Explanation, and Trust*, Carnegie Mellon University, 2017, pp. 1-31.