



OPACIDADE DOS ALGORITMOS E A NECESSIDADE DE TRANSPARÊNCIA: GARANTINDO EXPLICABILIDADE¹

OPACITY OF ALGORITHMS AND THE NEED FOR TRANSPARENCY: ENSURING EXPLAINABILITY

Lucas Moreschi Paulo²

Resumo: A partir do método de pesquisa lógico-dedutivo, de caráter bibliográfico e multidisciplinar, o tema abordado é a opacidade dos algoritmos e a necessidade de transparência para garantir a explicabilidade para os cidadãos. O objetivo é fornecer uma reflexão crítica sobre a opacidade dos algoritmos jurídicos e a necessidade de transparência para garantir a justiça e a equidade na tomada de decisão automatizada. A tomada de decisão automatizada já se tornou uma realidade em diversas áreas, incluindo o acesso à justiça. No entanto, a falta de transparência e explicabilidade na forma como essas decisões são tomadas levanta questões sobre a justiça e a equidade. O trabalho será dividido em três partes. Na primeira parte, será abordada a autodeterminação informativa e as dimensões humanas em que a privacidade e os dados podem afetar a dignidade humana. A segunda parte explorará a dimensão metafísico-constitucional do direito à participação do ser humano no mundo, abordando o conceito de autonomia e sua relação com a dignidade humana. Será estabelecido um paralelo entre a obnubilação do ser em regimes autoritários e a nova realidade algorítmica, que guardam fina relação. Na terceira e última parte, serão discutidos os temas das decisões automatizadas e a necessidade de transparência e explicabilidade algorítmica. Será analisada a forma como esses dois antídotos podem levantar as cortinas da opacidade algorítmica e garantir a dignidade

¹ Este artigo é resultante das atividades do projeto de pesquisa “Teoria da essencialidade” (*Wesentlichkeitstheorie*) e discriminação algorítmica: *standards* protetivos em face do Supremo Tribunal Federal e da Corte IDH – proposta de parâmetros de controle”, financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Bolsa de Produtividade em Pesquisa – Processo 309115/2021-3). A pesquisa é vinculada ao Grupo de Pesquisa “Jurisdição Constitucional aberta” (CNPq) e desenvolvida junto ao Centro Integrado de Estudos e Pesquisas em Políticas Públicas – CIEPPP e ao Observatório da Jurisdição Constitucional Latino-Americana (ambos financiados pelo FINEP e ligados ao Programa de Pós-Graduação em Direito – Mestrado e Doutorado da Universidade de Santa Cruz do Sul – UNISC). Também se insere no âmbito do projeto de cooperação internacional “Observatório da Jurisdição Constitucional Latino-Americana: recepção da jurisprudência da Corte Interamericana de Direitos Humanos e sua utilização como parâmetro para o controle jurisdicional de Políticas Públicas pelos Tribunais Constitucionais”, financiado pela Capes (Edital PGC1 02/2015 – Processo 88881.1375114/2017-1 e Processo 88887.137513/2017-00).

² Advogado. Doutorando em Direito no Programa de Pós-Graduação em Direito – Mestrado e Doutorado da Universidade de Santa Cruz do Sul (UNISC), bolsista do Programa de Suporte à Pós-Graduação de Instituições Comunitárias de Educação Superior (PROSUC) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Mestre e graduado em Direito pela Fundação Escola Superior do Ministério Público (FMP), foi bolsista institucional do PPGD da FMP. Pesquisador do Grupo de Pesquisa Colisão de Direitos Fundamentais e o Direito como Argumentação, coordenado pelo Prof. Dr. Anizio Pires Gavião Filho, e pesquisador do Grupo de Pesquisa Teoria do Direito: Academia à Prática, coordenado pelo Prof. Dr. Francisco José Borges Motta, ambos do PPGD – Mestrado da FMP e vinculados no CNPq ao Grupo de Estudos Tutelas à Efetivação dos Direitos Transindividuais. Integrante do Grupo de “Pesquisa Jurisdição Constitucional Aberta”, coordenado pela Prof.^a Dr.^a Mônia Clarissa Hennig Leal, vinculado ao PPGD – Mestrado e Doutorado da UNISC, financiado pelo CNPq. Membro da *Argumentation Network of the Americas* - ANA. Currículo Lattes: <http://lattes.cnpq.br/4330914363996350>. ORCID: <https://orcid.org/0000-0003-4583-4853>. E-mail: lucasmoreschipaulo@gmail.com.



humana em relação às decisões tomadas por máquinas. Para tanto, também será visualizado de que forma pode-se exigir explicabilidade a título de tutela jusfundamental, tanto como parâmetro quanto como objeto.

Palavras-chave: Algoritmos, Explicabilidade, Opacidade, Transparência, Tomada de decisão automatizada.

Abstract: Through a logical-deductive research method, with a bibliographic and multidisciplinary approach, the theme addressed is the opacity of algorithms and the need for transparency to ensure explainability for citizens. The aim is to provide a critical reflection on the opacity of legal algorithms and the need for transparency to ensure justice and equity in automated decision-making. Automated decision-making has become a reality in various fields, including access to justice. However, the lack of transparency and explainability in the way these decisions are made raises questions about justice and equity. The work will be divided into three parts. The first part will address informational self-determination and the human dimensions in which privacy and data can affect human dignity. The second part will explore the metaphysical-constitutional dimension of the right to human participation in the world, addressing the concept of autonomy and its relationship to human dignity. A parallel will be established between the obfuscation of being in authoritarian regimes and the new algorithmic reality, which are closely related. In the third and final part, the themes of automated decisions and the need for transparency and algorithmic explainability will be discussed. The way in which these two remedies can raise the curtains of algorithmic opacity and ensure human dignity in relation to decisions made by machines will be analysed. For this purpose, it will also be examined how Explainability can be demanded as a means of safeguarding fundamental rights, as parameter and object itself.

Keywords: Algorithms, Explainability, Opacity, Transparency, Automated decision-making.

1. Introdução

A opacidade algorítmica começa a se tornar crescente preocupação na sociedade contemporânea. Ferramentas algorítmicas e de Inteligência Artificial começam a fazer parte cada vez maior da vida hodierna. A tomada de decisão automatizada já se alavancou ao ponto de estruturar espaço próprio – de difícil remoção – de áreas como o mercado de emprego, planos de saúde, acesso ao crédito e mesmo acesso à justiça. Nesse sentido, a falta de transparência e explicabilidade na forma como se dão tais decisões se mostra desarrazoada, literalmente, visto que joga sombras sob as próprias decisões que, mesmo contra a vontade do usuário impactado, conta principalmente com a tomada de decisão pela máquina e, quando muito, com a confirmação humana, essa mesma já enviesada pelo viés de confirmação comum aos humanos atarefados que somos.

Além disso, sistemas que decidem de maneira autônoma e sigilosa, obnubilando a participação do ser do jogo decisório, de seu potencial contraditório, acaba incrementando a



discriminação e a desigualdade que, se já não bastassem ocorrer de modo estrutural e adicional na atual conjuntura, ainda são potencializados quando vem a ocorrer em meio digital ou a partir da interação com *softwares* que são abastecidos com dados pessoais que, anonimizados ou não, conseguem triangularizar perfis, tendências e riscos, incrementando ainda mais a prejudicialidade de minorias e grupos historicamente marginalizados. Não só a esses, mas a esses especialmente, é necessária, portanto, reestruturar a leitura a partir do convite democrático que nos faz a Constituição da República Federativa do Brasil quando, em seu texto constitucional, emancipa o ser-cidadão como um ser participante da coisa pública, participante da tomada de decisão da sua vida, que, ainda que ocorram decisões contrárias e alheias às vontades individuais, ainda assim guardam uma explicabilidade, seja por empresas seja pelo poder público.

Nesse sentido, a partir de uma metodologia lógico-dedutiva de caráter bibliográfico e multidisciplinar, em um primeiro momento se enfrentará o tema da autodeterminação informativa para compreender em quais dimensões humanas a privacidade e os dados podem ir de encontro à dignidade humana ao se imiscuir de demonstrar seu adequado uso em prejuízo, ou mesmo benefício, do cidadão. Ato seguinte será explorada a dimensão metafísico-constitucional do direito à participação do ser-humano no mundo, o autonomizando enquanto um ser que guarda dignidade em si e que, só por isso já é merecedor de uma tutela que lhe garanta a visualização dos processos que lhes dizem respeito. Nesse sentido, um paralelo entre a obnubilação do ser em regimes autoritários servirá de um ótimo paralelo para se analisar o que é a nova realidade algorítmica.

Por fim, se adentrará nos temas candentes das decisões automatizadas e a sua necessária fenomenologia conjunta com a transparência e o direito à explicabilidade algorítmica. Em conjunto, será advogada a forma como esses dois antídotos devem agir para levantar as cortinas do embrutecimento algorítmico, aquela que desconsidera o ser em sua dignidade e o subjugam a uma passividade animalésca.

2. AUTODETERMINAÇÃO INFORMATIVA, PRIVACIDADE E DADOS: Entre o opaco e o transparente

Se há algo que nos distingue das máquinas, dos robôs e das artificialidades esse algo é a consciência. A consciência humana é o fator determinante de categorização entre uma inteligência do tipo humana e uma inteligência do tipo máquina. Essa consciência é marca



determinante das espécies racionais, de alocar-se em percepção de perspectiva sobre panoramas gerais (RUSSEL, 2021, p. 25). Em termos de capacidade de processamento de dados, estamos para trás: em 2016 a Tianhe-2 processava 55 quatrilhões de operações (conexões entre células de processamento) por segundo enquanto a raça humana consegue cerca de 10 quatrilhões de operações (conexões entre neurônios) por segundo. A única vantagem humana, por enquanto, é a energia gasta para tanto. O Tianhe-2 utiliza 1 milhão de vezes mais energia que o cérebro humano (NICOLELIS, 2020, p. 137-138).

Outra característica distingue homem de máquina: a capacidade de situação, isto é, a capacidade de se situar no tempo e de sub-alocar energia e dedicação a uma tarefa não principal. As máquinas, mesmo aquelas chamadas de Baby-AGI (recentes tentativas de Inteligência Artificial Geral³), não conseguem por si só determinar que devem ser criados novos *softwares*, algoritmos e *bots* para a solução de problemas que lhes são apresentados. Para tanto, é necessário que nas linhas de código esteja presente a possibilidade dessa sub-criação. O cérebro humano não é assim, o cérebro humano enquanto *locus* de processamento informativo e determinante do agir humano, um ser situacional, é distinguido (NICOLELIS, 2020, p. 138-139).

A autonomia privada é o poder que os particulares têm de regular, pelo exercício de sua própria vontade, as relações de que participam, estabelecendo-lhe o conteúdo e a respectiva disciplina jurídica interrelacional. O ser humano, sozinho ou em comunhão (pacífica ou litigiosa) tem total condição de situar-se no mundo e exigir comportamentos, ou mesmo identificar quando não se age pela regra geral. Se alguém trapacear de uma máquina de *narrow AI* no Xadrez, a primeira reação da máquina é buscar a lógica por detrás daquele movimento estranho; e é capaz de encontrar e alterar dentro de sua “consciência” para uma regra de um novo jogo, um Xadrez’, enviesado. Ao contrário, o impulso do ser humano será de gritar: HEY, ESTÁS ME ROUBANDO!

Essa é uma diferença estrutural do funcionamento dos cérebros, de um lado o orgânico e de outro o artificial. Seres humanos não são feitos de *bits* (NICOLELIS, 2020, p. 143). E, assim como uma máquina nunca conseguirá replicar o processo de fotossíntese das plantas, ainda que se siga a risca todo o processo bioquímico-físico de quebra do dióxido de carbono em açúcar, e consiga efetivamente açúcar “para as plantas”, não se conseguirá produzir oxigênio, o que falta é o fluxo orgânico e natural que, ao longo de centenas de milhões de anos,

³ Não se recomenda a instalação, apenas indica-se a fonte <https://github.com/yoheinakajima/babyagi>



moldou a evolução para que de maneira integrada – e até mesmo mágica – o mesmo processo que alimenta e gera energia às plantas, gera o elemento tão essencial à sobrevivência humana.

A consciência humana é uma característica distintiva que separa a humanidade das máquinas e robôs, permitindo-lhes compreender e analisar perspectivas e visões gerais. Embora as máquinas possam processar mais informações do que o cérebro humano, elas carecem da capacidade de contextualização e de uma autêntica autonomia, que se funda na dignidade humana, que é para os seres humanos, e que poderia ser estendida para seres animais, mas nunca para um carro, para uma geladeira, ainda que *smart* e, portanto, nunca a poderia ser reconhecida às máquinas. A autonomia é uma característica distintiva dos seres humanos, permitindo-lhes regular as relações nas quais participam.

A liberdade individual, a proteção dos dados pessoais e a privacidade são garantias constitucionais que, em uma via de mão dupla, fundamentam e são fundamentadas pela autodeterminação individual. Isso implica, na modernidade, o reconhecimento do direito à explicabilidade das decisões automatizadas, como forma de garantir a proteção dos dados e da privacidade dos indivíduos, não apenas para garantir-lhes o direito, mas para os tornarem conscientes sobre os motivos determinantes de decisões tomadas a seu benefício ou prejuízo.

É possível compreender que, desde logo, do art. 5º, *caput*, da Constituição, decorre a liberdade individual, do inciso X decorre a inviolabilidade da intimidade e da vida privada e do inciso XII decorre o livre desenvolvimento, a proteção dos dados pessoais (recentemente com a aprovação da EC 115/2022) e, portanto, a autodeterminação individual. Pode-se afirmar que a soma interpretativo-constitutivista resulta no reconhecimento do direito implícito à explicabilidade das decisões automatizadas fundada na autodeterminação individual, na proteção dos dados e na privacidade como uma garantia que decorre do texto constitucional, como também o era antes também com os dados antes de 2022 (MENDES, 2011).

Há, portanto, se não de modo formal, um direito fundamental material e implícito à explicabilidade e à autodeterminação na CF/88, isto é, um direito fundamental associado. Segundo Gavião Filho e Freitas (2020, p. 240-241), os direitos fundamentais implícitos são aqueles que não estão expressamente dispostos no constitucional, mas dependem de uma formulação interpretativo-argumentativa para existir. Como se propõe. Veja-se em dimensões a autonomia privada e a autodeterminação existem e se complementam.

Na Alemanha há muito já se foi decidido acerca da autodeterminação informativa. No clássico julgamento do Tribunal Constitucional Federal Alemão (*Bundesverfassungsgericht*) *BVerfG* (conhecido como «*Census Judgment*»), por meio do direito fundamental de



personalidade, deduz-se a existência de uma proteção à autodeterminação informativa, proporcionando aos indivíduos um campo de salvaguarda contra a ocorrência de discriminações por sistemas (*machine learning*, *big data*, inteligência artificial, v.g.) geridos por algoritmos.

A falta de controle dos indivíduos acerca de seus dados, gerada pela sua comercialização desenfreada e não consentida, despertou nos legisladores de diversos países o desejo de criar regras sobre o tratamento de dados. Mesmo antes da propagação de leis de proteção de dados, o Tribunal Constitucional Federal alemão (*BVerfG*), em 1983, já havia fixado histórico precedente em favor da «autodeterminação informativa» (*Informationelle Selbstbestimmung*), entendendo ser um direito derivado da personalidade e da dignidade humana (SCHWABE, 2005, p. 233-234).

Portanto, desde a aludida data, passou a vigorar na Alemanha veredito capaz de permitir a judicialização de discriminações geradas pelo tratamento de dados – certamente que, naquela época, em meio físico. Em que pese o Brasil ter sistematizado e incorporado uma proteção robusta à privacidade e aos dados pessoais e sensíveis, por meio da Lei nº. 13.709/18 (Lei Geral de Proteção de Dados, 2018), apenas permitindo que seu tratamento ocorra caso autorizado por alguma das suas dez bases legais (artigo 7º), ainda assim existe uma insuficiência, na legislação regente, para a efetiva tutela judicial de discriminações geradas por algoritmos no âmbito digital.

A decisão do *BVerfG* ficou conhecida como «*census judgment*» (julgamento do censo), visto que se referia a queixas da população alemã contra a recém aprovada Lei do Censo («*1983 Federal Census Act*»), que previa que, para a realização da contagem populacional, seriam coletadas diversas informações pessoais dos cidadãos, pelas quais eram indesejadas de serem fornecidas por alguns deles. A despeito do veredito ter sido, parcialmente, desfavorável aos requerentes, o Tribunal aproveitou o contexto para determinar medidas acerca do tratamento de dados, inclusive anulando partes da lei. Entendeu a Corte que o que prevê o direito ao livre desenvolvimento da personalidade, estava diretamente correlacionado ao artigo que dispõe acerca da cláusula geral da proteção da dignidade da pessoa humana, e, a partir de sua leitura conjunta, haveria uma noção associada de garantia à autodeterminação do indivíduo acerca das informações compositivas de sua personalidade.

Isso autoriza ao entendimento de que autodeterminação informativa, além de existir enquanto um direito fundamental, ainda não se lastreia como um direito de propriedade do indivíduo, podendo ser disposto em qualquer circunstância, mas como um direito fundamental



instrumental, cujo papel na formatação constitucional é de parametrizar a interpretação e aplicação como escopo protetivo autônomo e associável.

A autodeterminação informativa engloba duas posições que o indivíduo pode ter em relação às informações processadas sobre si: o direito de saber (*right to know*) e o direito de não saber (*right to not know*). O primeiro consiste no direito do sujeito afetado em conhecer os resultados do tratamento de seus dados, ou seja, na necessidade de transparência. Já o segundo indica que, caso não queira conhecer os resultados, a pessoa não deve ser informada sobre eles ou, se já os conhece, deve ter o poder de excluí-los. Sem a liberdade de exercer esses direitos de saber e de não saber, não é possível garantir a verdadeira autodeterminação informativa.

A autodeterminação informativa é um direito fundamental que abrange tanto o direito de saber quanto o direito de não saber sobre as informações processadas a respeito de uma pessoa. Esses direitos são fundamentais para garantir transparência e privacidade, e a falta de liberdade para exercê-los compromete a autodeterminação informativa. No entanto, a aplicação desses direitos pode gerar conflitos com outros direitos fundamentais, como a liberdade religiosa e de expressão, como se verá no caso do padeiro Jack Phillips. Esse conflito de direitos evidencia a complexidade na aplicação dos direitos fundamentais, incluindo a autodeterminação informativa, e a necessidade de encontrar um equilíbrio que respeite todos os direitos envolvidos. Veja-se o caso.

Em 2012, o casal Charlie Craig e David Mullins foram a uma loja de bolos em Colorado, nos Estados Unidos, para encomendar um bolo de casamento personalizado. No entanto, o proprietário da loja, Jack Phillips, se recusou a criar o bolo, alegando que isso violaria suas crenças religiosas. O casal processou Phillips por discriminação, e o caso foi parar na Suprema Corte dos Estados Unidos. Em 2018, a Corte decidiu que Phillips não violou as leis antidiscriminatórias do Colorado, alegando que a Comissão de Direitos Cíveis do estado demonstrou hostilidade em relação às crenças religiosas de Phillips ao lidar com o caso. A decisão da Suprema Corte gerou um intenso debate sobre o equilíbrio entre o direito à igualdade e à não discriminação e a liberdade religiosa e de expressão.

O caso mencionado refere-se à recusa do confeitiro Jack Phillips em criar um bolo personalizado para um casamento entre pessoas do mesmo sexo, o que gerou uma acusação de discriminação. Em 2018, a Suprema Corte dos Estados Unidos absolveu Phillips da acusação, alegando que a decisão do confeitiro se baseava em suas crenças religiosas. Veja-se, mesmo um ato discriminatório precisa ser explicado.



Agora, necessário se faz relacionar o caso *Masterpiece Cakeshop v. Colorado Civil Rights Commission*, 584 U.S. ____ (2018) ao direito à privacidade, mais especificamente ao direito de ser deixado em paz (*right to be left alone*). O confeitiro alegou que criar um bolo para um casamento entre pessoas do mesmo sexo violaria suas crenças religiosas, o que o levou a se recusar a fornecer o serviço. Embora essa decisão possa ter sido motivada por suas crenças pessoais, ela também pode ser vista como uma afirmação de seu direito à privacidade, ou seja, o direito de se afastar de situações que violem suas convicções pessoais.

Esse caso pode ser visto como um exemplo de como as decisões automatizadas podem afetar o direito à privacidade e ao direito de saber o motivo da recusa de certo acesso a bens ou serviços. Os algoritmos que usam dados pessoais para tomar decisões automatizadas podem afetar o direito à privacidade, já que os dados pessoais dos indivíduos são usados para tomar decisões que afetam suas vidas. Além disso, a falta de transparência sobre como esses algoritmos tomam decisões pode afetar o direito de saber o motivo da recusa de certo acesso a bens ou serviços.

Portanto, é importante que haja transparência e explicabilidade nas decisões automatizadas, de modo que os indivíduos afetados possam entender como essas decisões são tomadas e contestar decisões que possam ser discriminatórias ou violar seus direitos. Além disso, é importante que sejam estabelecidos critérios claros para o uso de algoritmos em decisões importantes, para garantir que esses sistemas sejam desenvolvidos com transparência e equidade.

Assim, percebe-se com facilidade que a falta de transparência pode conduzir a situações de violação da privacidade e discriminação. Quando os dados são tratados de maneira opaca, os indivíduos não têm controle sobre o que está sendo coletado, como é utilizado e quem tem acesso a esses dados. De igual forma, quando esses dados são utilizados para a tomada de decisão para a formação de *scoring* de crédito, *profiling* para *microtargeting* e *micromarketing*, ou ainda a recusa de um serviço ou emprego, ficando apenas nesse campo em que os dados pessoais são centralmente utilizados, se percebe uma visível necessidade de se poder sindicar a funcionalidade destes dados, isto é, se jogar luzes e vencer a opacidade da cortina de ferro, ou da *blackwall* da *blackbox*.

Em rápidas linhas, é interessante explicar que a perfilização (*profiling*) consiste na elaboração de perfis de comportamento de uma pessoa (ou de um grupo de pessoas) a partir de dados pessoais, disponibilizados por ela ou colhidos. Para Clarke (1993, p. 403), o “*profiling is a technique whereby a set of characteristics of a particular class of person is inferred from past*



experience". Assim, através de algoritmos aliados à ciência da economia comportamental, é possível prever condições, decisões ou comportamentos futuros de uma pessoa. Isso certamente beneficia as empresas que trabalham com risco, ou que querem evitar o risco demais, com a técnica do *Credit Scoring*, conhecida dos brasileiros, e julgada lícita pelo STJ nos autos do RESP nº 1.419.697/RS, afetado para julgamento pelo rito dos recursos repetitivos. Os usos dos dados pessoais, além disso, tem sido utilizados para o direcionamento do marketing ostensivo e personalizado. Todas essas questões são tomadas a partir de uma automatização da decisão.

Quando os dados são tratados de maneira opaca, os indivíduos não têm controle sobre o que está sendo coletado, como é utilizado e quem tem acesso a esses dados. Além disso, a falta de transparência no tratamento de dados pode levar à violação da privacidade dos indivíduos, o que vai contra a nossa atual ordem constitucional vigente. Por isso, é importante que as empresas e governos sejam transparentes em relação ao tratamento de dados, garantindo que os indivíduos tenham controle sobre suas informações pessoais e que sejam tomadas medidas para proteger a privacidade e prevenir a discriminação.

Para que se possa entender os verdadeiros riscos da nocividade da opacidade no tratamento de dados, é importante entender como e de quais formas a opacidade pode levar à discriminação e violação da privacidade, não apenas de modo direto e imediato, mas a partir de uma matiz filosófica considerada.

3. A RETOMADA DO PARADIGMA DO SER NÃO-PARTÍCIPE NA REALIDADE ALGORÍTMICA

A Magna Carta foi o primeiro documento formal e juridicamente relevante que submeteu, não só os súditos a um texto legal vinculante, mas também os barões da nobreza e, em especial, o soberano maior do Estado, o Rei, que até então tinha poder absoluto sobre todas as esferas da vida civil, militar, religiosa e política. Este momento histórico é dado como o marco principal, ainda que em fase embrionária, do movimento chamado de constitucionalismo, que lutou, e ainda luta, pela concretização de direitos e garantias inalienáveis, certas e universais.

Infelizmente não é estranho começar o presente capítulo pela Magna Carta. De modo totalmente infeliz, os debates que a instruíram, no longínquo 1215, quando Robin Hood era, talvez, ainda um soldado e não uma lenda, foram estabelecidas certas bases do direito de defesa do homem em relação ao Estado, o que vem a ser mais bem definido a partir do



constitucionalismo de 1ª geração, dos chamados direitos de defesa, em que o Estado tinha limites à imposição de sua vontade ao cidadão, pois este agora tinha uma poderosa arma: direitos de proteção, de liberdade. Uma liberdade galgada na possibilidade de se autodeterminar enquanto indivíduo, povo ou nação.

O tom aqui é, como se vê, de retomada de Hobbes (2002, p. 25-28), para quem a concepção do estado absolutista era a de que a natureza humana era a da desordem, para quanto “o homem é o lobo do homem”, de modo que “a origem de todas as grandes e duradoras sociedades não provém da boa vontade recíproca que os homens tiveram uns para com os outros, mas do medo recíproco que uns tinham dos outros”. Assim, no célebre *O Leviatã*, Thomas Hobbes (2008, p. 147) argumenta que:

a única maneira de instituir um tal poder comum, capaz de os defender das invasões dos estrangeiros e dos danos uns dos outros, garantindo-lhes assim uma segurança suficiente para que, mediante o seu próprio labor e graças aos frutos da terra, possam alimentar-se e viver satisfeitos, é conferir toda a sua força e poder a um homem, ou a uma assembleia de homens, que possa reduzir todas as vontades, por pluralidade de votos, a uma só vontade. Isso equivale a dizer: designar um homem ou uma assembleia de homens como portador de suas pessoas, admitindo-se e reconhecendo-se cada um como autor de todos os atos que aquele que assim é portador de sua pessoa possa praticar ou levar a praticar, em tudo o que disser respeito à paz e à segurança comuns; todos submetendo desse modo as suas vontades à vontade dele, e as suas decisões à sua decisão. [...]. Feito isso, à multidão assim unida numa só pessoa se chama *República*, em latim *Civitas*, é esta a geração daquele grande Leviatã, ou antes (para falar em termos mais reverentes), daquele *deus mortal*, ao qual devemos, abaixo do *Deus imortal*, a nossa paz e defesa.

Deste modo, *autorictas, non veritas facit legem* (HOBBS, 2008, p. 230). Estar-se-ia vivendo, agora, em um *potestas economicus datum, non veritas facit legem?*, isto é, o poder econômico do dado, e não a verdade, é que faz a lei/o direito? A comparação Estado absolutista com o poder exercido pelas *Big Techs* e demais companhias de tecnologia pode até parecer, para alguns leitores, algo totalmente ilógico e despreocupante. Para estes, a questão é como está acessando esse texto?

Os *cookies*, *caches* e metadados de navegação estão todos sendo nesse exato momento sendo coletados, armazenados e utilizados para gerar valor econômico para empresas que vendem tais dados, ou meramente os utilizam, e para auxiliar outras empresas que, ao adquirir esses dados, poderão ter um real controle de influência sobre o comportamento daquele ser-alvo. Assim, uma notícia impactante: o e-mail da promoção do livro, da geladeira, da camiseta ou da bicicleta que talvez nem se tenha chegado a buscar em provedores de pesquisa como o



da Google, não chegou por acaso; e não chegou por acaso naquele exato minuto em que se sabe que a caixa de e-mail individual do caro leito é acessado com maior frequência.

Dados são poderosos. E são mais poderosos que o petróleo. O petróleo tem um valor inexoravelmente ligado à sua disponibilidade, visto que é finito e de certa forma difícil de ser extraído. Os dados, por sua vez, estão à disposição, são infinitos e mais: enquanto houver homem, haverá economia, haverá consumo, haverá dados e haverá economia de dados. E, por trás dela: decisões automatizadas, cujo combustível são os dados.

Assim, ao desestruturar o cidadão do meio de sua própria determinação, hoje, na realidade da internet das coisas, do Big Data e da Economia Comportamental e dos Dados, o homem passa a ser objeto de sua própria autodeterminação. Paradigmaticamente, passa a sofrer influxos personalizados pelas redes e começa a inferir suas predileções, vontades e atitudes em um mito moderno, que é o da consciência.

Isso não lembra muito a Rousseau (2017, p. 23-25), segundo o qual todo homem nasce livre no estado natural, sendo a natureza humana boa, pacífica e próspera. A questão está na falta de autoridade central que concentre o poder da vontade e das possibilidades. Eis a fundamentação da celebração do contrato social, para que os homens possam ter segurança contra os inconvenientes da natureza e contra a arbitrariedade de homens mal-intencionados. Hoje é necessário um (re)pacto 2.0, ou melhor, 4.0. Hoje o contrato social não servirá para a proteção da natureza, que era a fonte emissora do risco, mas deverá ser das *Big Techs* e demais companhias, que passam a ter a centralidade sob o comando dos rumos de vidas e nações⁴.

Veja-se, na sequência dessa espécie de reconstrução, tem-se Kant (2017, p. 73-93), que afirmava serem os homens livres no domínio da razão prática, isto é, os seres humanos são dotados de liberdade e consciência moral sendo ele, o ser-humano, o “ente” habilitado a realizar distinções entre as muitas possibilidades de seu desenvolvimento, isto é, da promoção de sua vida guiada pelo exercício de uma racionalidade livre, de um livre-arbítrio, livre de arbítrios que advinham de fora do ser, impostos de modo escrachado ou sutilmente. A liberdade, para Kant, é ser livre para optar mesmo em contradição a seus próprios impulsos. Fazer aquilo que não se quer, ou que não se tem o impulso a realizar é a liberdade em si e de si que Kant percebe.

⁴ Como aconteceu com o caso da Cambridge Analytica que, ao que se sabe, e o que uma parte ficou demonstrado no documentário *The Great Hack*, do Netflix, a empresa britânica de análise de dados, contratada, dentre outros, pelo então candidato à presidência norte-americana Donald Trump, influenciou eleitores indecisos em 2016. A estratégia era a de avaliar os dados existentes em redes sociais, como o Facebook, montando perfis de eleitores, os influenciando através do marketing dessas plataformas através de estratégias de postagens que cada perfil receberia, conquistando eleitores pelo perfil. A mesma empresa teria atuado no *British Exit from EU* (BREXIT). Tudo isso possibilitado e potencializado pelo *Big Data* (NICOLELIS, 2020, p. 153).



O constitucionalismo amadurecido pela Guerra de Independência Norte Americana, de 1776, e pela Revolução Francesa, de 1789, se deu a partir de uma noção de liberdade enquanto uma nova visão de vida e relação jurídico-econômica do ser-humano para com o Estado (GRIMM, 2016, p. 41-42). Hoje se necessita de uma nova relação liberal para que se estabeleça o entendimento normativo de defesa para com a tecnologia? Assim como o Estado pode ser um grande aliado da nação e do povo, também o pode a tecnologia, mas não desprovida de obrigações, limites e deveres instrumentais. A questão, portanto, é da junção entre Estado, Mercado (tecnológico) e Sociedade Civil, mas de um modo não baseado no medo. Veja-se, o medo é avesso à condição conceitual de tratarmos de liberdade:

Apenas aqueles que conhecem a liberdade em relação à necessidade podem apreciar por completo o significado da liberdade em relação ao medo, e só aqueles que estão livres de ambos – necessidade e medo – têm condições de conceber uma paixão pela liberdade pública (ARENDDT, 2019).

Para Arendt, o medo ou a ignorância fundamentam a existência de uma massa que, ao lado da propaganda e do uso do terror, formam a receita completa para o surgimento e o funcionamento de regimes autocráticos totalitários. Não se está advogando a existência de um governo central da internet, longe disso. A analogia passa por compreender como que, do mesmo modo que as massas funcionaram para legitimar regimes nas décadas de 30 e 40 do século passado, também a ignorância e o comodismo tem o condão de fazer algo muito semelhante no atual tempo: legitimação de um descontrole e uma desregulamentação da aplicação diária e cotidiana das novas tecnologias. O ponto de Arendt (2012, p. 439), como se percebe, é, também, da indiferença, para além da ignorância e do medo:

[...] lidamos com pessoas que, simplesmente devido ao seu número, ou à sua indiferença, ou a uma mistura de ambos, não se podem integrar numa organização baseada no interesse comum, seja partido político, organização profissional ou sindicato de trabalhadores. Potencialmente, as massas existem em qualquer país e constituem a maioria das pessoas neutras e politicamente indiferentes, que nunca se filiam a um partido e raramente exercem o poder de voto (ARENDDT, 2012, p. 439).

Na Alemanha do início da década de 1930 pode-se perceber isso. Nutridos pelo revanchismo, não se opuseram a tomada de poder pelo Partido Nacional-Socialista que, quando não foi eleito, em abril de 1932, obtendo apenas 36,8% dos votos, realizaram golpes institucionais e crimes políticos para que chegassem ao poder em 1933, sendo referendado como chanceler com 90% dos votos no plebiscito realizado em agosto de 1933, com o Auto de Autorização (*Gesetz zur Behebung der Not von Volk und Reich*). Tal Auto permitia, por uma



brecha da constituição de Weimar, que o chanceler pudesse elaborar leis à margem da constituição. Assim, com base nessa legislação, o Reich com Hitler no comando iniciou uma expansão aos territórios vizinhos, uma forte e monstruosa política de segregação, e posteriormente eliminação, com os campos de concentração, a Regulamentação 11 da Lei da Cidadania do Reich – que retirou a cidadania dos judeus – e com a criação da Gestapo, tudo pelo *Gemeinnutz von Eigennutz* (o bem comum acima de todos!) (SHIRER, 2017, p. 168-315).

A chave para o sucesso alemão, se não uma, passa pelas 3 comentadas acima com aporte no pensamento de Arendt. Assim, parece evidente que em nada se pode deixar cair em obnubilação e em tudo deve-se buscar a luz, a verdade e a transparência. É o ato de lançar luzes sobre a decisão automatizada, exigindo-lhe explicabilidade que se poderá conformar a realidade estruturada a partir dos avanços tecnológicos – retoricamente sempre ótimos para o desenvolvimento humano – com um tom mais crítico e exigente, em um timbre que eleve os direitos fundamentais para o seu necessário papel de destaque frente aos desatinos da atual realidade algorítmica.

4. DECISÕES AUTOMATIZADAS E A NECESSÁRIA TRANSPARÊNCIA E EXPLICABILIDADE COMO ANTÍDOTOS À OPACIDADE

A utilização das Inteligências Artificiais para a tomada de decisões é um feito realizado desde os anos 80, quando empresas do mercado de ações passaram a contar com as ferramentas para conseguir resultados muito acima da concorrência. A automatização da tomada de decisão para comprar, vender, alocar, realocar ou *opt out* demonstrava uma certeza e uma aptidão quase absoluta para os melhores resultados acionários. De lá para cá, as IAs se expandiram, sendo aplicadas em muitos campos da vida cotidiana, como na agricultura, na gestão de consumo de energia, no controle de vazão de água, ou mesmo para solucionar problemas antes impensáveis. Eis um aspecto interessante dos algoritmos: eles são uma cláusula aberta, e sua capacidade dependerá da habilidade que seu programador der à programação inaugural (STEINER, 2012).

Estamos prestes a passar de um ponto substancial. Nunca antes na história da humanidade foi permitido que uma máquina decidisse de forma autônoma e baseada meramente em dados e em sua base de treino e de fluxo neural quem deveria viver e quem deveria morrer, algo que a qualquer momento irá acontecer em múltiplas situações, como a da tomada de decisão de carros autônomos em situação de escolha, entre a colisão e a morte de seu proprietário ou a colisão e morte de pedestre, e a conservação de seu usuário; em uma



impossibilidade de outras respostas igualmente seguras sob ponto de vista epistêmico, nesse revivido e atualizado *Trolley problem* (EYSENCK; EYSENCK, 2023, p. 184-185 e 193-195).

Talvez seja preferível que carros sejam todos autônomos, visto que 94% dos acidentes são ocasionados por erros humanos (EYSENCK; EYSENCK, 2023, p. 184). Mas, talvez a responsabilização ficasse mais complicada. Afinal, um erro humano é facilmente sindicável e identificável. Um erro não humano pode ser fruto de múltiplos fatores, sobretudo quando em uma colisão entre dois veículos, apenas um deles for gerenciado por IA. A resposta, portanto, não parece ser meramente baseada na efetividade da máquina ou na falibilidade humana. Mas então, o que de fato é importante?

Eysenck e Eysenck (2023, p. 187) asseveram que para que os carros autônomos sejam considerados seguros, precisariam percorrer aproximadamente 17,7 bilhões de quilômetros e atingir números infinitamente menores do que em comparação ao motorista humano. Quantos quilômetros foram percorridos por seres humanos antes que aceitássemos a condução de humanos nos carros? A pergunta sequer guarda lógica com a forma de raciocínio humana, correto?

Isso se dá pelo simples fato de que o ser-humano confia mais no ser-humano do que viria a confiar numa máquina. Tal paradigma talvez esteja a ponto de ser substituído, com uma maior confiança depositada nas máquinas. Esse, contudo, não é o caso até a atual quadra histórica.

Portanto, se é o caso de exigirmos explicações, fundamentações e motivações para que sejam aceitas as decisões, sejam administrativas, sejam judiciais, é de ainda maior e vital importância que a existência destas autonomizações esteja necessariamente vinculada com o direito à explicações, isto é, que os programadores e desenvolvedores ponham ao público aplicações aptas à justificar de modo claro, transparente e escalonado todos os passos prejudiciais para que a decisão fosse XYZ e não ABC. Para dar um exemplo extremo: caso um carro vá colidir, algo que tenha 100% de certeza epistêmica, e opte por sofrer a pancada do lado em que está um passageiro humano feminino, ao invés de optar pelo passageiro humano masculino, poderia se avaliar os motivos que o levaram a determinar que tal escolha trágica fosse tomada. As respostas poderiam variar, entre questões de gênero, estatística ou as mais esdrúxulas tentativas de racionalização e de justificação do injustificável.

Agora, é a partir da resposta, dessa tentativa de justificação – e que para isso esse direito à explicabilidade deve existir – que se poderá identificar os erros de avaliação e, ao longo do tempo, corrigir via programação tais defeitos, que vão além de defeitos técnicos e, no



caso do exemplo dado, poderiam ser de origem ética, moral ou jurídica. Assim, talvez para os humanos do exemplo dado, a explicação não seja muito útil, até pela marcha lenta e unidirecional do tempo, mas esse direito é essencial e central para o desenvolvimento ético, técnico e jurídico das aplicações de Inteligência Artificial.

Assim, exigir explicabilidade, ao lado de transparência, é lidar com a caixa preta dos algoritmos, isto é, com a opacidade com a qual os algoritmos são hermeticamente fechados e considerados em sua programação. A *black box* de Pasquale (2015), na verdade é uma situação criada pelo segredo comercial e industrial dos algoritmos, visto que, querendo ou não, hoje o ser-humano está sendo objeto de desejo de *Big Techs* que promovem uma verdadeira guerra fria e desenvolvimentista das IAs. Quem colocará primeiro o homem no espaço? E o que é colocar o homem no espaço no atual cenário? Questões certamente relevantes, que não necessariamente precisam ser respondidas, visto que não influem na centralidade da questão objeto: a dignidade humana na dimensão de sua percepção acerca das decisões automatizadas tomadas contra si. Se grandes empresas se digladiam na arena econômica e tecnológica, o ser-humano-cidadão-usuário não pode ser uma minúscula formiga à mercê de ser pisado; ao contrário, merece lugar de destaque na tribuna da contenda, sendo o juiz da disputa que, em verdade, é uma disputa por seu uso das aplicações.

Assim, é importante ter mapeada características que devem ser reforçadas junto às IAs que tomam decisão (FRAZÃO, 2020, p. 122-123). A primeira delas é a Rastreabilidade, que diz respeito ao dever de documentar na maior extensão possível os dados e os processos que produzem a decisão do sistema de IAs, incluindo os processos de recolha e etiquetagem dos dados, bem como os algoritmos utilizados, de modo a possibilitar a identificação dos motivos da decisão de IA que se quer analisar. A rastreabilidade é essencial para a posterior auditabilidade e a explicabilidade.

A questão central está na explicabilidade que se refere à capacidade de explicar tanto os processos técnicos de um sistema de IA como as decisões humanas relacionadas a ele, incluindo os domínios de aplicação de um sistema de IA. A explicabilidade técnica exige que as decisões tomadas por um sistema de IA possam ser compreendidas e rastreadas por seres humanos. Em algumas situações, pode ser necessário buscar soluções de compromisso entre o aumento da explicabilidade de um sistema (que pode reduzir sua precisão) e o aumento da sua precisão (à custa da explicabilidade). Sempre que um sistema de IA tiver um impacto significativo na vida das pessoas, deverá ser possível solicitar uma explicação adequada do respectivo processo de tomada de decisões, que deve ser oportuna e adaptada ao nível de



especialização da parte interessada em questão (por exemplo, leigo, regulador ou investigador). Além disso, devem ser disponibilizadas explicações sobre o grau de influência e de intervenção de um sistema de IA no processo decisório da organização, as opções de concepção do sistema e os fundamentos da sua implantação, assegurando assim a transparência do modelo de negócio. Além disso, quanto mais intensa a intervenção, mais transparente deverá se fazer a explicação. O grau de necessidade da explicabilidade depende em grande medida do contexto e da gravidade das consequências de um resultado potencialmente errado, inexato ou nocivo à direitos.

Por fim, a comunicação é o que estabelece que os sistemas de IA não devem se apresentar como seres humanos aos utilizadores e que estes têm o direito de serem informados de que estão interagindo com um sistema de IA. Isso implica que os sistemas de IA devem ser identificáveis como tal. Além disso, deve ser fornecida a opção de escolher entre essa interação e a interação humana, sempre que necessário, a fim de garantir que os direitos fundamentais sejam respeitados. Além disso, as capacidades e limitações do sistema de IA devem ser comunicadas aos profissionais no domínio da IA ou aos utilizadores finais de forma adequada ao caso de utilização em questão. Essa comunicação pode incluir o nível de precisão do sistema de IA, bem como suas limitações.

Assim, a XAI/xAI (*Explainable Artificial Intelligence*) se constitui do campo destinado a fazer inteligível os resultados dos sistemas de IA, dizendo respeito à compreensibilidade dos dados, dos processos e dos comportamentos dos algoritmos que atuam em um dado sistema decisório autônomo. A XAI é um marco a ser perseguido para uma IA ética, que possa ser, inclusive, alvo de determinações judiciais no futuro, na tutela de direitos atingidos em virtude de decisões algorítmicas automatizadas. É a busca por uma XAI que promove um equilíbrio entre a total desaparecimento de interesses privados no desenvolvimento de IAs, por conta do segredo comercial, e de uma total clausura das caixas-pretas (BONAT; PEIXOTO, 2020, p. 54-56).

Algo é explicável quando consegue demonstrar todo o trajeto da tomada de decisão, possibilitando sua própria sindicabilidade. O que se precisa entender, no entanto, é que uma qualquer explicação nem sempre bastará. O grau de explicabilidade deve estar relacionado com o grau de transparência necessária para que se possa promover ajustes ou reparos diante de situações concretas em que os resultados, sobretudo quando estes forem decisões automatizadas, firam direitos fundamentais. Assim, pode-se defender uma XAI em grau, em



que quanto maior for o grau de afetação de direito por conta do algoritmo, maior deverá ser o grau de explicabilidade da decisão ou do resultado dado ou entregue pela IA.

A explicabilidade das IAs já é lei. Na França, o Decreto 2016-1321, de 7 de outubro, sobre a República Digital, estabelece que os cidadãos devem ser informados quando uma decisão individual é adotada por meio de um algoritmo. O Decreto 2017-330, de 14 de março, sobre direitos das pessoas sujeitas a decisões individuais adotadas com base em tratamentos algorítmicos, obriga a administração a informar de forma inteligível sobre o grau e o modo em que o algoritmo contribuiu para o tratamento automatizado da tomada de decisões, incluindo os dados tratados, a sua fonte, os parâmetros de tratamento e a sua ponderação, bem como as operações realizadas. Já na Itália, o Conselho de Estado exige um prévio conhecimento pleno do módulo utilizado e dos critérios empregados para adotar a decisão. Esse conhecimento do algoritmo deve ser garantido em todos os aspectos, desde seus autores até o procedimento empregado para sua elaboração, o mecanismo de decisão, as prioridades atribuídas no procedimento de avaliação e os dados selecionados como relevantes. Isso permite verificar se os critérios, condições e resultados do procedimento automatizado cumprem as prescrições e finalidades estabelecidas pela lei ou pela própria administração, além de esclarecer as regras e as modalidades de programação. No Brasil, a previsão está no art. 20 da LGPD.

Há, no entanto, maneiras distintas de se exigir tais explicações, visto que muitas são as formas de XAI. Se pode ter capacidade de **i)** explicação para apenas justificar a decisão – a partir de resultados específicos ou da lógica do raciocínio interno do funcionamento, ou **ii)** para conhecer os caminhos tomados pela aplicação e, com isso, promover aprendizado de máquina para outras aplicações ou para que seja de conhecimento geral, ou **iii)** ainda para incrementar a lógica da aplicação ou deixá-la mais segura, ou ainda, e por fim, para **iv)** controlar de maneira mais eficientes os resultados, tornando mais aprimorado o controle das vulnerabilidades (ADADI; BERRADA, 2018).

Assim, como visto, a transparência e da explicabilidade deve ser levada em consideração não apenas na área dos carros autônomos, mas em todas as aplicações de Inteligência Artificial que envolvam tomada de decisões automatizadas. É preciso que os desenvolvedores e programadores sejam responsáveis e éticos em relação aos algoritmos que criam, garantindo que eles sejam justos, confiáveis e seguros para todos os usuários. Nesse sentido, é importante que as empresas invistam em pesquisas e estudos que busquem compreender e mitigar os possíveis impactos negativos das IAs nas diversas áreas em que elas



são aplicadas e que mantenham controles e deixem o funcionamento da máquina rastreável. Eis o caminho para uma IA sustentável e ética, que prima pelo respeito aos direitos.

A transparência e a explicabilidade têm sido cada vez mais consideradas como medidas cruciais para lidar com a opacidade algorítmica. A crescente utilização de algoritmos em diversas áreas da sociedade tem gerado preocupações em relação às possíveis consequências negativas dessas ferramentas automatizadas, tais como a discriminação e a injustiça. Nesse contexto, a transparência e a explicabilidade tem o condão de destronar a opacidade enquanto a principal característica quando se pensa em sindicabilidade dos algoritmos. Será a partir do antídoto da transparência que a própria defesa do cidadão, a partir do Poder Judiciário ou mesmo de outras instâncias fiscalizatórias ou reguladoras poderão controlar e melhor readequar os mecanismos e as aplicações que tomam decisões de forma autônoma. Por meio da transparência e da explicabilidade, é possível deslacrar as *black boxes* e olhar para o verdadeiro componente de cada algoritmo, partindo dali ajustes ou postulações a reparações.

5. CONCLUSÃO

A automatização de processos de tomada de decisão tem sido amplamente utilizada em diversas áreas, trazendo benefícios, mas também apresentando desafios. A opacidade dos algoritmos pode afetar a justiça e a equidade nas decisões automatizadas. A transparência e a explicabilidade desses sistemas são fundamentais para garantir a compreensão e contestação das decisões tomadas, especialmente em casos em que as decisões podem ser discriminatórias ou violar os direitos dos indivíduos. Nesse sentido, a jurisdição constitucional, que tutelará tais direitos, deve ser apta tanto a se valer das explicações como a exigí-las, constituindo-se a si mesmo (o direito à explicabilidade) um direito fundamental associado.

Nesse contexto, a criação de regulamentações específicas para o uso de algoritmos em decisões importantes e o estabelecimento de critérios claros para o desenvolvimento desses sistemas são medidas essenciais para garantir a transparência e a equidade nas decisões automatizadas, enquanto isso não ocorre, tal abertura material para o reconhecimento do direito fundamental à explicabilidade será vital para a tutela constitucional dos direitos afetados. Afinal, é fundamental considerar as dimensões humanas em que a privacidade e o mal ou mau processamento de dados pode ocasionar, daí a importância da tutela à autodeterminação informativa, como escudo e controle pelos indivíduos perante as decisões automatizadas que se valem de seus dados.



A abordagem metafísico-constitucional do direito à participação do ser humano no mundo é relevante para a compreensão das questões éticas envolvidas na automatização de processos de tomada de decisão. A transparência e a explicabilidade são essenciais para garantir a justiça e a equidade na sociedade atual, protegendo-se a dignidade dos indivíduos afetados.

Aliás, o direito é também uma tecnologia. Uma técnica nada inovadora, mas bem inovativa e inventiva (*Ius est Ars inveniendi*) de solucionar conflitos humanos. E nem por isso deixa de ser explicável. Aliás, está aí a sua maior virtude: a capacidade de racionalmente demonstrar as premissas e suas conclusões com efeitos jurídicos e práticos, e os caminhos de sua fundamentação, seja em premissas fáticas ou normativas, sendo essas últimas mais ou menos concretas.

REFERÊNCIAS

ADADI, Amina; BERRADA, Mohammed. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). **IEEE Access**. v. 6, p. 52138-52160, 2018.

ARENDDT, Hannah. **As origens do totalitarismo**. Trad. Roberto Raposo. São Paulo: Companhia das Letras, 2012.

ARENDDT, Hannah. **Liberdade para ser livre**. Trad. Pedro Duarte. Rio de Janeiro: Bazar do Tempo, 2019.

BONAT, Débora; PEIXOTO, Fabiano Hartmann. **Racionalidade no direito: inteligência artificial e precedentes**. Curitiba: Alteridade, 2020.

CLARKE, Roger. Profiling: A hidden challenge to the regulation of data surveillance. **Journal of Law & Information Science**, v. 4, p. 403-419, 1993.

EYSENCK, Michael W.; EYSENCK, Christine. **Inteligência artificial x humanos: o que a ciência cognitiva nos ensina ao colocar frente a frente a mente humana e a IA**. Trad: Gisele Klein. Porto Alegre: Artmed, 2023.

FRAZÃO, Ana. Quais devem ser os parâmetros éticos e jurídicos para a utilização da inteligência artificial?. **Revista eletrônica [do] Tribunal Regional do Trabalho da 9ª Região**, Curitiba, v. 9, n. 86, p. 119-124, mar. 2020.

GAVIÃO FILHO, Anizio Pires; FREITAS, Luiz Fernando Calil de. Direitos fundamentais estatuidos não diretamente ou implícitos. **Revista Direitos Fundamentais & Democracia**, Curitiba, v. 25, n. 3, p. 232-257, set./out. 2020.

HOBBS, Thomas. **Do Cidadão**. São Paulo: Martins Fontes, 2002.

HOBBS, Thomas. **Leviatã**. São Paulo: Martins Fontes, 2008.



KANT, Immanuel. **Crítica da razão prática**. Rio de Janeiro: Nova Fronteira, 2017.

MENDES, Laura Schertel. O Direito Fundamental à proteção de dados pessoais. **Revista de Direito do Consumidor**. São Paulo, Revista dos Tribunais, v. 79, p. 45-81, jul./set. 2011.

NICOLELIS, Miguel. **O verdadeiro criador de tudo: como o cérebro humano esculpiu o universo como nós o conhecemos**. São Paulo: Planeta, 2020.

PASQUALE, Frank. **The black box society: the secret algorithms that control money and information**. Cambridge e Londres: Harvard University Press, 2015.

ROUSSEAU, Jean-Jacques. **O Contrato Social ou Princípios do Direito Político**. Trad. Edson Bini. São Paulo: Edipro, 2017.

RUSSEL, Stuart. **Inteligência Artificial a nosso favor: como manter o controle sobre a tecnologia**. Trad. Berilo Vargas. São Paulo: Companhia das Letras, 2021.

SCHWABE, Jürgen. **Cinquenta anos de jurisprudência do tribunal constitucional federal alemão**. Trad. B. Henning e L. Martins. Montevideo: Fundación Konrad-Adenauer, 2005.

SHIRER, William L. **Ascensão e queda do Terceiro Reich: Triunfo e consolidação (1933-1939)**. 2. ed. Rio de Janeiro: Nova Fronteira, 2017.

STEINER, Christopher. **Automate This: how algorithms came to rule our world**. New York: Penguin Group, 2012.