

Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal

Preliminary considerations on artificial intelligence, ethics and personal autonomy

Danilo Cesar Maganhoto Doneda*
Laura Schertel Mendes**
Carlos Affonso Pereira de Souza***
Norberto Nuno Gomes de Andrade****

Resumo

Aplicações concretas de inteligência artificial tornam-se cada vez mais frequentes e, com elas, a discussão sobre seu impacto para a personalidade e autonomia pessoal. A utilização intensa de dados pessoais em processos e algoritmos capazes de tomar decisões proporcionam avanços, ao mesmo tempo em que podem discriminar e causar danos em situações concretas que demandam a atenção do Direito e da comunidade científica. A sofisticação de tais modelos decisórios chega ao ponto de que se vislumbrem até alterações em características subjetivas das relações jurídicas, como demonstra o debate sobre a personalidade jurídica de robôs. O recurso a elementos de ética de dados (*data ethics*), ao cabo, vislumbra-se como uma possibilidade concreta, tanto de amortizar alguns dos riscos concretos na implementação de sistemas de inteligência artificial, preservando os benefícios destes sistemas e, ao mesmo tempo, resguardando direitos e garantias, como de servir como parâmetro regulatório.

Palavras-chave: Inteligência artificial. Autonomia pessoal. Personalidade. Ética de dados.

Abstract

Concrete applications of Artificial Intelligence become more frequent and, with them, the discussion about its impact on personality and personal autonomy. The intense use of personal data in decision-making processes and algorithms enables progress as well as conditions to discriminate and cause damages, creating situations that arise awareness to the law and the scientific community. The decision-making models' sophistication may even allow alterations in subjective characteristics of legal relationships, as demonstrated by the issue of the robot's legal personality. The deploying of data ethics elements is a possibility to amortize risks of the implementation of Artificial Intelligence systems, to preserve their benefits and to safeguard rights, guarantees and compliance to regulation.

Keywords: Artificial intelligence. Autonomy. Discrimination. Ethics. Robots. Legal personality.

* Mestre e Doutor em Direito pela Universidade do Estado do Rio de Janeiro. Professor visitante na Faculdade de Direito da Universidade do Estado do Rio de Janeiro. Foi Coordenador-Geral de Estudos e Monitoramento de Mercado na Secretaria Nacional do Consumidor do Ministério da Justiça. Pesquisador visitante na Università degli Studi di Camerino e na Autorità Garante per la Protezione dei Dati Personali, ambas na Itália. Rio de Janeiro – RJ- Brasil. E-mail: danilo@doneda.net.

** Doutora em Direito Privado pela Universidade Humboldt de Berlim e Mestre em Direito, Estado e Constituição pela UnB. Professora adjunta de Direito Civil na Universidade de Brasília (UnB) e do Mestrado em direito constitucional do Instituto Brasileiro de Direito Público (IDP). Diretora do Centro de Direito, Internet e Sociedade do IDP e Diretora da Associação Luso-Alemã de Juristas (DLJV). Brasília –DF- Brasil. E-mail: lauraschertel@hotmail.com.

*** Doutor e Mestre em Direito Civil na Universidade Estadual do Rio de Janeiro. Professor da Faculdade de Direito da Universidade do Estado do Rio de Janeiro (UERJ). Diretor do Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS). Pesquisador Visitante do Information Society Project, da Faculdade de Direito da Universidade de Yale. Professor dos cursos de graduação e pós-graduação da UERJ e da PUC-Rio. Membro da Comissão de Direito Autoral da OAB/RJ. Policy Fellow da ONG Access. Conselheiro do Instituto NUPEF. Rio de Janeiro –RJ- Brasil. E-mail: caffsouza@gmail.com.

**** Mestre em Relações Internacionais e Estudos Europeus pela Central European University (CEU, Budapeste – Hungria). Doutor em Direito pelo Instituto Universitário Europeu (Florença – Itália). Pós-Doutoramento na Faculdade de Direito da Universidade de Berkeley (EUA). Especialista em Privacidade e Políticas Públicas no Facebook. Professor Assistente em Ética de Dados, Privacidade e Novas Tecnologias na IE Law School. Bacharel em Direito pela Faculdade de Direito da Universidade de Lisboa. Foi Pesquisador Científico no Centro Comum de Investigação da Comissão Europeia, e no Hague Institute for the Internationalization of Law (HILL). Membro acadêmico do Center for Internet and Society da Universidade de Stanford (EUA). Berkeley - CA – Estados Unidos. E-mail: norbertonuno@gmail.com.

1 Introdução

Diversas garantias fundamentais e direitos da personalidade vêm tendo seu perfil modificado, há um bom tempo, por efeito ou em diálogo direto com o desenvolvimento tecnológico. Desde as tecnologias que passaram a permitir um maior fluxo de informação e a ampliação das possibilidades de liberdade de expressão até o fortalecimento e objetivação de garantias referentes à privacidade, devido ao tratamento de dados pessoais, entre outras tantas situações, o diálogo constante e cada vez mais intenso entre tecnologia e proteção da personalidade é um dos temas fundamentais de que se ocupa, hoje, o jurista.

Até um determinado momento, foi possível observar esses efeitos do desenvolvimento tecnológico a partir de um vetor quantitativo. A tecnologia costumava possibilitar “mais”, ou seja, estender vetores da atuação de uma determinada atividade para (muito) além das possibilidades humanas. Assim, mais informações poderiam ser enviadas a mais pessoas, mais cálculos poderiam ser executados por uma máquina do que legiões de matemáticos jamais poderiam manualmente, mais imagens poderiam ser capturadas, e daí em diante. Esses efeitos, por maior impacto que possam ter, ao serem decompostos revelam, basicamente, um componente objetivo, qual seja, maior força ou eficiência ao executar determinadas funções. Geralmente, algo capaz de ser mensurado.

Mais recentemente, o desenvolvimento e a implementação de tecnologias de inteligência artificial (IA) proporcionou efeitos que, muitas vezes, não podem mais ser compreendidos em termos meramente quantitativos, e que implicam uma mudança na subjetividade das relações entre as pessoas e a tecnologia. Essas novas tecnologias possibilitam a automatização da tomada de decisão em diversas situações complexas, executando tarefas que estávamos habituados a considerar como prerrogativas humanas, derivadas da inteligência - a ponto de que diversas manifestações dessas tecnologias foram denominadas como realizações de uma “inteligência artificial”. Assim, os computadores passaram a não ser vistos somente como dispositivos destinados a fazer cálculos, sistematizações ou classificações, porém a deter, em algum grau, algo passível de ser comparado às ações humanas autônomas.

A bem da verdade, a percepção desse cenário foi antecedida em muito pela literatura científica, o imaginário popular e a cultura. Mesmo a gigantesca capacidade de cálculo, que hoje vemos como intrínseca aos computadores ao se demonstrar uma realidade concreta, foi de início percebida em contraposição direta às capacidades humanas (que eram, afinal, o padrão de comparação por excelência), abrindo a possibilidade de novas comparações possíveis entre o humano e a máquina, entre criação e criatura. Em um ensaio clássico de 1945, que antecipa cenários que se tornariam cotidianos nas décadas seguintes, Vannevar Bush (1945) conjecturou que o tratamento automatizado da informação viria a se tornar um adendo ou apêndice da memória humana. A intuição de Bush veio em época na qual a tecnologia para o processamento automatizado de informações começava a se tornar realidade, após ter seus fundamentos estabelecidos em modelos que, até hoje, são referência para a computação, como a Máquina de Turing, concebida, em meados da década de 1930, por Alan Turing (1948).

A iminência de que a tecnologia permitiria, efetivamente, às máquinas realizar ações antes privativas de humanos fomentava também uma retomada definitiva de uma tradição, até então puramente especulativa, a respeito de “autômatos”. Estes seriam engenhos concebidos pelo homem, em geral com características antropomorfizadas e que poderiam apresentar algo parecido (ou equivalente) ao livre-arbítrio, sendo capazes de tomar decisões. A presença de elementos que, de alguma forma, poderiam mimetizar a tomada de decisões a partir de um raciocínio lógico-dedutivo, abriu espaço para as primeiras indagações de caráter ético acerca da atuação desses entes, indagações estas que são antecedentes diretas do debate contemporâneo, mas continuam se projetando diretamente sobre ele, como o demonstra a persistência do debate em torno das três leis da robótica formuladas por Isaac Asimov, no conto “Runaround”, no algo longínquo ano de 1942.¹

¹ As três leis da robótica de Asimov são: “1. um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal; 2. um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que entrem em conflito com a Primeira Lei; 3. um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira ou Segunda Leis”. O conto Runaround foi publicado pela primeira vez em 1942, na revista *Astounding Science Fiction*, nos Estados Unidos. Posteriormente, veio a integrar o volume *I, robot*, de 1950, editado no Brasil em ASIMOV, Isaac. *Eu, robô*. Aleph: São Paulo, 2014.

O desenvolvimento dos computadores, a partir da década de 1950, teve como um de seus panos de fundo a eventualidade de que eles proporcionassem, efetivamente, o surgimento de alguma espécie de “inteligência”. A viabilidade do desenvolvimento efetivo de uma inteligência artificial recebeu uma espécie de “certidão de nascimento” em 1956, quando da realização de uma (longa) conferência em Dartmouth sobre o tema (McCARTHY et al., 1956). Na década de 1960, o tema começou a ser objeto de mais ampla discussão a partir de diversos marcos, como a utilização de linguagem “natural” para a comunicação com computadores, implementada e teorizada por Joseph Wiezenbaum (1976) com o programa ELIZA, e também, a partir da década de 1970, com a introdução dos chamados “sistemas especialistas”.

Uma boa parcela da problemática que hoje circunda a utilização de sistemas automatizados capazes de tomar decisões teve sua gramática formulada nessas primeiras discussões sobre o impacto da automação e da inteligência artificial. Sua implementação, no entanto, não foi de forma alguma imediata: apesar das bases teóricas dos seus elementos computacionais poderem ser, ao menos, antevistos nesse primeiro momento, sua utilidade e aplicabilidade se revelou bastante limitada por motivos que podemos, com o risco da simplificação, resumir a dois: o primeiro, limitações na capacidade de processamento dos computadores; o segundo, os limites das primeiras abordagens da implementação de sistemas de inteligência artificial. Considerando que a capacidade computacional vem sendo progressivamente incrementada, vale destacar que as novas abordagens e técnicas utilizadas nos mais recentes sistemas de inteligência artificial revelam-se fator determinante, principalmente ao abrir mão de utilizar um conjunto de regras previamente instruídas em prol da utilização de algoritmos que “aprendem” a partir da observação e análise de bases de dados.

Em síntese, a disponibilidade de mais recursos computacionais e também de mais informação, resultantes tanto do desenvolvimento tecnológico como do paradigma do *Big data*, proporcionaram o surgimento de sistemas de inteligência artificial que não fossem total ou prioritariamente baseados em um modelo de regras pré-fixadas, mas que fossem alimentados com grandes quantidades de dados, para que pudessem neles fundamentar suas decisões e formar seus padrões decisoriais.

A estreita relação entre o desenvolvimento mais recente dos mecanismos de inteligência artificial com a maior disponibilidade de informação deixou seus reflexos na regulação que começou a ser concebida em relação à proteção de dados pessoais. Assim, consolidando uma tradição presente na legislação de diversos países, a Diretiva 95/46/CE da União Europeia, sobre proteção de dados, estabeleceu como a regra que decisões tomadas por meios automatizados não podem ser a base exclusiva para a tomada de decisão que afete significativamente um cidadão sem que sejam estabelecidas garantias.

A utilização de dados pessoais para alimentar os novos sistemas de inteligência artificial e a sua utilização para tomar decisões proporcionam uma acurácia bastante significativa para um número crescentes de aplicações. Isto abre espaço para, ao menos, dois temas centrais para os debates sobre autonomia e direitos fundamentais nos próximos anos: os efeitos que a utilização desses sistemas causarão para a pessoa e sua autonomia pessoal, bem como a necessidade de qualificar a natureza desses instrumentos e sistemas de inteligência artificial. Nesse debate, a necessidade de que sejam proporcionadas soluções que preservem os direitos fundamentais, dentro de um quadro de intenso desenvolvimento tecnológico e mesmo de questionamento de alguns institutos centrais do ordenamento jurídico, sugere a necessidade de recorrer à ética como instrumento capaz de encaminhar soluções que, eventualmente, e se for o caso, possam consolidar-se em alternativas legislativas posteriormente.

2 Impactos da introdução de IA em mecanismos decisoriais: personalidade, autonomia e riscos de discriminação

A inteligência artificial, operada por meio da utilização de algoritmos, tem como importante função produzir previsões (ZARSKY, 2013, p. 1505). Ela pode auxiliar a prever quem será o próximo presidente do país, quais regiões estão mais propensas à propagação do vírus da zika, quando será o próximo terremoto no México, qual a probabilidade de termos uma crise econômica na próxima década, quem é o favorito para a Copa do Mundo, qual a probabilidade de determinado prisioneiro reincidir em um crime, a

probabilidade de uma pessoa gostar de determinado filme e qual a capacidade de pagamento dos clientes de uma empresa, por exemplo.

Como se vê, a aplicação cotidiana da inteligência artificial pode ser usada tanto para previsões de fatos gerais sobre a economia, natureza ou política, como também para prever o comportamento individual. Para o tema ora abordado, interessa-nos essa última função: como as previsões sobre o comportamento de indivíduos determinados pode fundamentar decisões sobre as suas vidas, seja no mercado, seja nas suas relações com o Estado, influenciando diretamente o seu acesso a bens, serviços e mesmo ao mercado de trabalho.

As decisões automatizadas, referentes a um indivíduo determinado, que se baseiam em um método estatístico para análise de grande volume de dados e informações, podem ter grande impacto sobre os direitos individuais, especialmente no que se refere à autonomia, igualdade e personalidade. Afinal, na sociedade atual, caracterizada pelas relações remotas, os dados pessoais acabam por se constituir na única forma de representação das pessoas perante as mais diversas organizações estatais e privadas, sendo determinantes para “abrir ou fechar as portas de oportunidades e acessos” (LYON, 2003, p. 27). Dessa forma, uma eventual representação equivocada em determinados contextos sociais – por meio de um equívoco do algoritmo ou dos dados em que o algoritmo se baseou – afetaria tanto a forma como o indivíduo se percebe como também o modo como a sociedade o enxerga e o avalia, afetando a sua integridade moral e a sua personalidade (BRITZ, 2008, p. 179). Ademais, se essa representação, conforme alertado por Lyon (2003), acarretar a perda de chances e oportunidades do indivíduo na sociedade, dar-se-á uma restrição indevida à sua autonomia, limitando a sua liberdade de ação, suas escolhas econômicas e até mesmo existenciais. Por fim, destaca-se também a possibilidade de violação do princípio da igualdade, na hipótese de que a classificação e seleção operada por algoritmos produza resultados desiguais para pessoas em situações semelhantes, afetando negativamente as suas oportunidades de vida na sociedade (LYON, 2003, p. 27).

Não há dúvidas quanto à relevância econômica e social das decisões tomadas com base em algoritmos, e seria mesmo difícil se pensar o funcionamento da sociedade contemporânea sem a utilização de algoritmos. Uma das funções mais importantes do processamento de dados por meio de algoritmos é fornecer a base para decisões econômicas, de modo a contribuir para a mitigação de riscos nos negócios (BUCHNER, 2006, p. 138; MENDES, 2014, p. 94). Isto se torna ainda mais importante quando um setor econômico é caracterizado por um alto grau de assimetria de informação, como os setores de crédito e seguros privados, que há muito já trabalham com a análise da dados por meio de algoritmos, expresso na figura do *credit score*. Mas sua relevância também está aumentando no setor privado como um todo, tendo em vista que o processamento de informações e dados por meio de algoritmos pode ser um meio de simplificar decisões econômicas, aumentar a eficiência e personalizar a produção, bem como ampliar as possibilidades de monetização de serviços digitais a partir da personalização do conteúdo e do marketing digital.

Também o Estado há muito se utiliza de decisões automatizadas para o cumprimento das funções dos diversos órgãos públicos. Como bem identificado por Citron (2008, p.1259) em sua descrição do “Governo 2.0”, o aumento da utilização de decisões automatizadas por agências de governo nos últimos anos tem diversos fundamentos, tais como a diminuição os custos e melhoria da performance de sistemas automatizados, o surgimento da Internet, bem como a redução de custos e de recursos humanos. Nesse contexto, atualmente, inúmeras decisões tomadas pelo Estado são efetuadas por meios automatizados, seja no auxílio ao combate à evasão fiscal (ZARSKY, 2013, p. 1511) e no combate ao terrorismo (PASQUALE, 2011, p. 236), seja na determinação das empresas que receberão licenças para funcionamento, das crianças que receberão assistência médica e dos passageiros que poderão embarcar em determinados vôos (CITRON, 2008, p.1259).

Percebe-se, a partir de tais exemplos, o potencial de violação aos direitos fundamentais das decisões automatizadas quando tomadas sem o cumprimento de determinados parâmetros éticos e legais que assegurem a sua transparência e controle individual, a participação do indivíduo no âmbito do processo decisório, bem como a correção e atualização das informações que servem como *input* do algoritmo.

Em especial, é preciso ficar atento à principal promessa que as decisões baseadas em algoritmos costumam suscitar: a obtenção de um maior grau de objetividade das decisões e a possibilidade de fundamentá-las em critérios estatísticos, eliminando o risco dos vieses (“*bias*”) típicos das decisões tomadas pelos seres humanos. Ao contrário, grande parte da literatura tem demonstrado o risco de discriminação inerente aos processos de tomada de decisão baseados em algoritmos, a qual ocorre, muitas vezes, de forma involuntária e sem o conhecimento dos tomadores de decisão (BAROCAS, 2015, p. 3). As razões pelas quais os algoritmos podem produzir resultados discriminatórios e prejudicar determinadas pessoas ou grupos sociais dividem-se em dois grandes grupos.

Primeiramente, qualquer algoritmo só é tão bom quanto os dados que lhe servem como base (BAROCAS, 2015, p. 1). Isto é: a qualidade da decisão automatizada (“*output*”), baseada em um algoritmo, tem uma correlação direta com a qualidade dos dados que ele processa (“*input*”). Por isso, se o algoritmo se baseia em dados históricos repletos de preconceitos, ele reproduzirá, de forma automatizada, os mesmos padrões preconceituosos utilizados como base de seu processamento.

Assim, é fácil perceber que, se forem utilizados no modelo estatístico dados com alto potencial discriminatório, tais como dados raciais, étnicos ou de orientação sexual, haverá um grande risco de que a decisão que resultará do processo automatizado (*output*) também seja discriminatória. Esses dados são os chamados dados sensíveis, cujo processamento é limitado pelas legislações de proteção de dados de vários países, assim como pelo Regulamento Europeu de Dados Pessoais.

Em segundo lugar, é preciso observar que o próprio método utilizado nas decisões automatizadas – por meio da classificação e seleção dos indivíduos – gera um risco de se produzirem resultados discriminatórios, ainda que de forma não intencional. Isto pode ocorrer porque, na discriminação estatística, teoria econômica que se tornou conhecida a partir dos textos de Edmund Phelps (1972) e Kenneth Arrow (1973), os indivíduos são diferenciados com base em características prováveis de um grupo, no qual esse indivíduo é classificado. Essa prática se baseia em métodos estatísticos, que associam esses atributos a outras características, cuja identificação pelo tomador de decisão é mais difícil, como nível de renda, risco de inadimplência, produtividade no trabalho, etc. (BRITZ, 2008, p. 15). Nesse contexto, é possível a ocorrência da discriminação por erro estatístico, o que decorreria tanto de dados incorretamente capturados como também de modelo estatístico de bases científicas frágeis (BRITZ, 2008).

Resultados discriminatórios também são possíveis por meio da generalização, prática muito utilizada nas decisões automatizadas, o que levou Gabriele Britz (2008, p. 134) a cunhar a expressão “injustiça pela generalização”. A discriminação estatística se dá por meio da classificação de pessoas com determinadas características em certos grupos – isto é, por meio da generalização de que pessoas com tais características têm maior probabilidade de agir de certa maneira ou de apresentar determinadas qualidades. A generalização, nesse caso, embora o modelo possa funcionar bem e seja estatisticamente correto, pode levar à discriminação das pessoas que configuram os casos atípicos, não se enquadrando nas características do grupo geral. É o caso, por exemplo, da pessoa que, apesar de morar em determinada região, considerada de baixa renda e, portanto, classificada como de maior risco de inadimplência em modelos de risco de crédito, aufere na realidade renda superior à de seus vizinhos. A discriminação, nesse caso, dar-se-ia, porque, em um modelo em que a informação sobre endereço tem peso fundamental, o caso atípico seria tratado conforme o grupo em que está inserido, e não conforme as outras pessoas de sua faixa de renda.

Nesse contexto, é fácil perceber que o pressuposto para esse processo automatizado de tomada de decisão é o acesso a uma enorme quantidade de dados e, por isso, o debate sobre a utilização de algoritmos está quase sempre associada à discussão sobre *big data*. O termo *big data* refere-se às possibilidades de acesso a grandes quantidades de dados de diferentes tipos, qualidade e formas de coleta (“*volume*”), bem como alta velocidade de processamento (“*velocity*”). Além disso, o *big data* é a base de novos modelos de negócios e possibilidades de várias criações de valor (“*value*”), na medida em que pode ser usado em conjunto com outras tecnologias, como a Internet das coisas ou o *Cloud Computing* (MAYER-SCHÖNBERGER, 2001).

Ocorre que o *big data* acaba por apresentar ainda mais desafios no que diz respeito aos riscos de discriminação algorítmica por uma série de motivos. Primeiramente, como afirma Mayer-Schönberger,

a imensa quantidade de dados utilizada nas análises de *big data*, assim como a diversidade de fontes, torna difícil assegurar a correção de todos os dados utilizados na análise. Nas palavras do autor, em muitas situações, as enormes bases de dados na análise de *big data* são caracterizados pela confusão (“*messy*”) e imprecisão. Assim, como bem analisa Buchner, diante da presença de dados desatualizados ou equivocados, um sistema aparentemente objetivo, incorruptível, acaba se transformando em um sistema rígido e imutável, deixando desorientada e desamparada a pessoa afetada pelo processamento equivocado de dados pessoais (BUCHNER, 2006, p. 125).

Outro grave problema diz respeito à falta de transparência, sendo comum que os algoritmos baseados no processamento de *big data* sejam chamados de *blackbox*, o que denota a obscuridade de seu processo decisório, não só para o indivíduo, como também para autoridades reguladoras e supervisoras. O problema da falta de transparência é extremamente relevante para a presente discussão, visto que a verificação da ocorrência de eventual discriminação depende de se saber qual é o *input* do algoritmo ou qual é método estatístico utilizado, o que, muitas vezes, pode estar protegido pelo sigilo empresarial.

Por fim, também se caracteriza como uma fonte de preocupação, em relação às análises baseadas em *Big data*, a busca incessante por correlações e não por causalidades. Em outras palavras, em face da imensa quantidade de dados hoje disponível, é possível que se encontrem infinitas correlações estatísticas sobre os fatos de interesse dos tomadores de decisão, sem que tais informações tenham necessariamente qualquer relação de causa e efeito com a área em que a decisão será tomada. Dessa forma, uma decisão poderia ser tomada em relação a um determinado indivíduo com base em dados aleatórios, sem qualquer relação com o assunto objeto da decisão, com a simples justificativa de guardar uma correlação estatística – e não causal – com a informação buscada.

Nesse contexto, é interessante mencionar a polêmica na Alemanha envolvendo a SCHUFA (empresa alemã que presta serviços de proteção ao crédito), a qual, no âmbito da avaliação de risco do consumidor, classificava como critério negativo o seu pedido de acesso a seus próprios dados. Isto se deve a uma correlação estatística que foi estabelecida no sentido de que consumidores que acessavam mais o seu score tinham maior chance de serem inadimplentes. A empresa sofreu inúmeras críticas em razão dessa conduta, que penalizava aquele que queria contratar um crédito com um *scoring* mais baixo, exclusivamente, em razão do exercício de um direito. Diante dessa prática, o legislador alemão passou a vedar especificamente essa prática na reforma da lei federal de proteção de dados de 2009 (BUCHNER, 2006, p. 123).

Assim é que dados aparentemente inofensivos também podem ser usados como base de condutas discriminatórias, em razão da grande capacidade atual de processamento e cruzamento de informações. Um interessante exemplo desse fato é reportado pelo Financial Times, segundo o qual um bem sucedido empresário negro americano recebeu uma carta de sua operadora de cartão de crédito, informando-lhe que o seu limite de crédito tinha sido reduzido de \$10.800,00 para \$3.800,00. A razão para isso, segundo a carta, era o fato de que o Kevin tinha comprado em período recente em muitas lojas frequentadas por pessoas com um histórico de crédito ruim (BIG..., 2015). Resta claro, portanto, no caso analisado, que ocorreu a chamada discriminação por associação ou por *proxy*, visto que dados raciais não foram utilizados diretamente como *input* do algoritmo, mas certamente o foram de modo indireto, isto é, por meio de um dado aparentemente inofensivo (compra em determinada loja).

Como visto acima, apesar dos benefícios das decisões automatizadas, como a maior agilidade no processamento de informações e a redução dos custos associados à gestão e à contratação de funcionários, tais decisões podem afetar significativamente os direitos fundamentais do indivíduo, em especial a sua autonomia, personalidade e igualdade. Embora a disciplina da proteção de dados pessoais por meio da regulação do fluxo de informação na sociedade possa auxiliar a endereçar os desafios debatidos acima, é preciso ir além das regras de privacidade e se construir uma verdadeira agenda relativa aos princípios éticos da inteligência artificial e das decisões automatizadas amparadas no uso de algoritmos, como será abordado ao final deste artigo.

3 Autonomia, personalidade jurídica e responsabilidade de robôs

Conforme visto no item anterior, as aplicações de inteligência artificial, gradativamente, fazem parte da rotina de cada pessoa. Se a primeira imagem com a qual se relaciona o conceito de inteligência artificial é a de um robô humanoide, a realidade apresenta versões menos ancoradas em visões de ficção científica, mas nem por isso menos surpreendentes.

Grande parte da personalização proporcionada por algoritmos nas aplicações de Internet mais populares adota uma forma de inteligência artificial. A recomendação de novas músicas (de acordo com o histórico de audições), a seleção de quais postagens irão aparecer primeiro no *feed* de notícias de uma rede social (com base nas interações prévias) e mesmo o melhor caminho a seguir em um aplicativo de trânsito (com base nas informações sobre engarrafamentos na cidade) são aplicações de inteligência artificial.

Essas aplicações se tornam ainda mais interessantes para o debate sobre ética e implicações legais quando incorporadas em um robô, servindo como um verdadeiro *hardware* que interage com o ambiente externo. Os questionamentos trazidos pela popularização dos robôs inteligentes desafia (e desafiará) o Direito de forma diferente daquela experimentada com a expansão da Internet nas últimas duas décadas.

Ryan Calo (2015, p. 513-515) aponta três características peculiares da robótica que se divorciam do debate havido nos últimos vinte anos sobre a interface entre Direito e Internet. Segundo o autor, robôs são marcados por: (i) sua materialidade, (ii) um comportamento emergente (geralmente denominado de “autonomia”) e (iii) o seu valor social.

A *materialidade* representa uma característica importante da robótica, já que aplicações de inteligência artificial passam a interagir com o ambiente de modo mais evidente. Um robô que possui corpo – independentemente do seu formato, se humanoide ou não – é mais facilmente percebido. Ao dispor de um corpo, o robô acrescenta uma nova camada de questionamentos na interação com o ambiente e com humanos, já que ele pode vir a causar danos até então desconhecidos no ambiente simplesmente digital. Os danos causados por um robô não são nada virtuais.

Além dos questionamentos sobre os danos físicos causados por robôs, a materialidade contém uma segunda particularidade que pode dar ensejo a um novo tratamento jurídico de robôs inteligentes: a sua aproximação com o trato e com a fisionomia humana. Robôs humanoides tendem a mais facilmente gerar empatia, e esse sentimento poderia levar ao reconhecimento de que uma aplicação de inteligência artificial é algo mais do que simplesmente uma programação embarcada em um componente físico. Ela mereceria alguma forma de tutela especial. Para essa conclusão, as outras duas características da robótica são ainda mais relevantes.

Por *comportamento emergente* se entende a capacidade do robô de aprender e de se adaptar às circunstâncias. Na medida em que aplicações de inteligência artificial são treinadas para reagir aos estímulos que recebem, é cada vez mais importante saber o grau de previsibilidade do resultado de sua resposta.

Cada vez mais se noticia sobre o desenvolvimento de aplicações de inteligência artificial cujo comportamento não havia sido previamente imaginado pelos programadores responsáveis. No momento em que o resultado não foi expressamente antevisto, será correto dizer que a máquina tomou uma decisão autônoma? Aqui reside o debate sobre a chamada autonomia da inteligência artificial. Como ainda não se está em um momento em que se possa falar em autonomia total de comportamento por parte das máquinas, Ryan Calo (2015) prefere chamar essa característica de “comportamento emergente”, de modo a prestigiar o fato de que o elemento principal no debate não é “autonomia”, mas o fato de que o robô passa a se comportar de modo distinto a partir do *input* que recebe do ambiente.

A terceira característica da robótica que vale a pena mencionar é o *valor social*. Como nenhuma outra tecnologia, os robôs despertam - muitas vezes por possuírem características antropomórficas – uma resposta e um envolvimento social distinto daquele dedicado a outros bens. A construção de que a inteligência artificial, especialmente quando corporificada em um robô, parece merecer uma tutela distinta de uma simples coisa é maior quanto mais parecido com um ser humano for a máquina. Mas se o robô tem

um *corpo*, possui um *comportamento emergente* e gera *envolvimento social*, seria correto então conceder ao mesmo alguma forma de personalidade jurídica?

Ao responder à tese apresentada por Ryan Calo e Jack Balkin (2015, p. 56) argumenta que o conceito de “valor social” precisa passar por uma reflexão mais aprofundada. Não se trata de imaginar que as pessoas vão simplesmente confundir robôs com seres humanos (e nessa direção acabar dando aos mesmos uma forma de personalidade). No entender de Balkin, o que existe é um “efeito de substituição”, através do qual as pessoas trocam um agir humano por uma atuação de uma máquina, mas sempre para funções bem específicas.

Dito de outra forma, o robô jamais substituiria o humano integralmente, mas apenas faria em seu lugar uma dada função: abriria a porta, dirigiria o carro ou marcaria hora no cabeleireiro. Estaríamos ainda muito longe do momento em que um robô pudesse realizar todas essas funções.

Robôs foram criados por pessoas. Sendo assim, tudo o que a máquina faz seria uma mera execução de comandos e de programação prévia. Todavia, quanto mais complexas são as soluções apresentadas pelas máquinas para os dilemas com as quais são confrontadas, é se de se esperar que o Direito avance também para buscar compreender o que são robôs inteligentes e como o ordenamento jurídico deve reagir à sua progressiva inserção na sociedade.

Ao estabelecer as suas estratégias nacionais sobre inteligência artificial, diversos países comentam sobre a necessidade de a população se acostumar a se relacionar com (e através de) máquinas cada vez mais inteligentes. No caso do Japão, por exemplo, o plano nacional de robótica menciona, explicitamente, como objetivo a criação de uma “sociedade livre de barreiras para robôs”.²

Essa rápida inserção de robôs nos relacionamentos sociais, naturalmente, aciona uma série de questionamentos jurídicos. Qual é a natureza jurídica dos robôs inteligentes? Teriam eles direitos próprios? Seriam dotados de personalidade jurídica? No início de 2017, o Parlamento Europeu adotou uma resolução com recomendações sobre regras de Direito Civil e Robótica. O texto aponta para a necessidade de regular o desenvolvimento de robôs autônomos e inteligentes, com a recomendação de que se considere a criação de uma espécie de personalidade jurídica para robôs (PARLAMENTO EUROPEU, 2017).

A leitura do estudo (NEVEJANS, 2016) que serviu de base para a Resolução do Parlamento revela como aspectos ligados à responsabilidade civil foram relevantes para a adoção da medida proposta. Os danos decorrentes do desenvolvimento de carros autônomos e outros robôs inteligentes serviram de ponto de partida para o questionamento: quem responde quando um robô causa um dano? Responderia a empresa que fabricou o robô? Ou aquela responsável por sua programação? E se o robô for, na verdade, utilizado por uma terceira empresa, que contrata diretamente com o consumidor, não seria melhor que essa fosse responsável?

A solução aventada pelo Parlamento Europeu seria criar uma espécie de personalidade jurídica para o robô em si, chamada, por vezes, de “*e-personality*” ou “personalidade eletrônica”. O nome aproxima o debate sobre a personalidade jurídica dos robôs inteligentes daquele já enfrentado em outras situações pelos mais diversos ordenamentos jurídicos.

Por que o ordenamento jurídico confere personalidade? Ao lado das pessoas físicas, que naturalmente a possuem, é comum encontrar situações em que o ordenamento confere a diversos entes uma personalidade jurídica autônoma. Pessoas jurídicas das mais distintas (como sociedades e associações) e fundações são os melhores exemplos. Se o ordenamento jurídico confere personalidade jurídica à fundação, resultante do destacamento de um patrimônio, por que não conferiria a um robô inteligente?³

² De acordo com os termos da *New Robot Strategy*, elaborada pelo Ministério da Economia, Comércio e Indústria japonês: “*a society and structure must undergo a transformation to take full advantage of robots. [...] Having said that, it is crucial to meet the requirements for a society where human and robots can coexist and cooperate on a daily basis for the maximization of robot capacities. Such kind of society can be called ‘robot barrier-free society’ which should realized.*” Disponível em: <http://www.meti.go.jp/english/press/2015/pdf/0123_01b.pdf>. Acesso em: 18 jun. 2018.

³ Sobre o debate envolvendo a personalidade jurídica de robôs em uma perspectiva nacional, vale destacar o trabalho de Marco Aurélio de Castro Júnior (2009).

Vale questionar, de início, se a solução de se conceder uma personalidade jurídica seria mesmo a resposta adequada. No cenário europeu, impulsionado por indagações sobre responsabilidade, a questão da personalidade aparece muito mais ligada à construção de um mecanismo de reparação à vítima de danos do que como resultado de uma discussão sobre o que é um robô inteligente e seu estatuto jurídico de forma mais abrangente.

Não parece que criar uma personalidade jurídica autônoma seja a única (e quiçá a melhor) forma de direcionar a questão dos danos causados por robôs inteligentes. Quem vai gerir o patrimônio do robô? Não seria uma forma ampla de seguro mais eficiente para amparar a vítima do que criar uma nova categoria de pessoas jurídicas?

A resolução do Parlamento Europeu chega então a mencionar duas iniciativas relacionadas ao desenvolvimento de robôs inteligentes: (i) a adoção de um registro obrigatório desses robôs e (ii) a criação de um seguro que possa então fazer frente às hipóteses de danos causados pelos mesmos.

São, assim, questões estritamente patrimoniais que levam à criação da figura da personalidade jurídica dos robôs. Em grande medida, também foram questões patrimoniais que levaram à criação de pessoas jurídicas como sociedades e fundações. Todavia, diferente do que ocorre com essas entidades, robôs inteligentes são capazes de desenvolver uma especial interação com humanos. Essa é a característica do “valor social” mencionada acima.

Quais são as consequências dessa interação entre robôs e humanos? Diferente das fundações, que não possuem materialidade, os robôs inteligentes, na definição do Parlamento Europeu, precisam ter uma presença física. Percebe-se, assim, como também a característica da materialidade é levada em consideração pelo Parlamento. Não necessariamente a representação física reproduzirá traços humanos, mas a interação com um robô material (e não simplesmente com uma voz programada), cria uma série de novas situações.

Uma delas, por exemplo, são os robôs cuidadores, especializados no acompanhamento de pessoas idosas ou enfermas. Existiria um direito a não ser tratado ou cuidado por um robô? Esse questionamento começa no eventual desconforto com a companhia não humana, mas atinge, certamente, inquietações urgentes quanto mais avança a formação de diagnósticos realizados por inteligência artificial.

A doutrina de Direito Civil está acostumada a tratar os animais como coisas. Na classificação de bens semoventes, geralmente são inseridos todos os animais. Mas o que dizer da recente evolução no tratamento do tema na França (SALICETI, 2015) e em Portugal (ALEXANDRE, 2016), que transforma os animais em “seres sensíveis” à luz de suas respectivas legislações? Se animais foram destacados da categoria de bens para ocupar uma nova posição, o mesmo raciocínio poderia ser aplicado aos robôs inteligentes?

Outra particularidade parece tornar ainda mais complexo o debate sobre a personalidade jurídica dos robôs: o sexo. Ao criar um *hardware* dotado de inteligência que pode interagir com um parceiro humano para a satisfação de necessidades sexuais, o desenvolvimento de robôs inteligentes enfrenta outras discussões. Por um lado, se argumenta como a adoção de robôs para finalidades sexuais pode ser útil para satisfazer pessoas com sérios problemas de locomoção (NUCCI, 2016). Por outro, existe a preocupação de que, quanto mais parecidos com um ser humano forem, mais a produção e a utilização de robôs sexuais pode reforçar conceitos sobre o papel da mulher na sociedade e ideias de subordinação.⁴

O debate sobre robôs inteligentes, cada vez mais, importa para o Direito. Conceder à máquina uma personalidade jurídica autônoma, nem que seja para dotá-la de patrimônio para compensar eventuais danos, é uma solução que desponta seriamente no horizonte. Todavia é importante ir além da dinâmica da responsabilidade civil e investigar o que significa dotar robôs inteligentes de personalidade à luz do ordenamento jurídico.

⁴ Para diferentes visões sobre o debate envolvendo a criação e o uso de robôs para finalidades sexuais, vide Danaher e McArthur (2017). Como curiosidade, vale também acompanhar a discussão sobre a noção de robôs sexuais na perspectiva do Direito Islâmico. O tema é abordado por Amuda e Tijani (2012).

Questionar a natureza dos relacionamentos humanos, curiosamente, é um caminho que leva a compreender o futuro dos robôs. Esse debate, como visto, vai além da simples dinâmica da responsabilidade civil.

4 A ética de dados como estrutura analítica e operacional para a compreensão e a aplicação de inteligência artificial e algoritmos

Tecnologias, como a inteligência artificial, a robótica e a biotecnologia, não só afetam - inibindo ou induzindo - comportamentos individuais e sociais, como também têm o potencial de alterar intrinsecamente os próprios indivíduos e a sociedade. Na fase de criação, essas tecnologias são desenvolvidas com valores específicos incorporados e, quando implementadas e adotadas, elas carregam esses valores, moldando e mudando indivíduos, comunidades e sociedades em conformidade. Como qualquer outra tecnologia, mas em maior escala, com maior velocidade e âmbito mais amplo, essas tecnologias apresentam vários desafios e acionam várias preocupações legítimas; e também apresentam oportunidades para trazer benefícios sem precedentes aos indivíduos e à sociedade em geral.

Com esse “pano de fundo tecnológico”, esta seção irá, em primeiro lugar, analisar o campo emergente da ética de dados (*data ethics*), definindo e explicando seu âmbito. A ética de dados, nesse contexto, surge como um instrumento analítico através do qual podemos avaliar e entender melhor os desafios apresentados por essas tecnologias, e também como uma estrutura operacional que nos permite enfrentar esses desafios e chegar a decisões moralmente boas e justificáveis. Em segundo lugar, e restringindo a nossa análise ao campo da inteligência artificial e aprendizagem de máquina (ou *machine learning* - ML), identificaremos e avaliaremos os principais riscos e preocupações apresentados por essas tecnologias, juntamente com os seus benefícios e oportunidades. Em terceiro lugar, como uma maneira viável e promissora de abordar as preocupações levantadas, assim como de ampliar os seus benefícios, explicaremos a crescente relevância e importância dos quadros éticos corporativos na governança das tecnologias digitais.

Os dados estão transformando a sociedade, sublinhando o que muitos chamam Quarta Revolução Industrial (SCHAWAB, 2016). Enquanto a Primeira Revolução Industrial foi baseada em água e vapor para mecanizar a produção, e a Segunda usou energia elétrica para criar produção em massa; a Terceira Revolução Industrial alavancou a eletrônica e a tecnologia da informação para automatizar a produção. Agora, devido ao número de tecnologias que, progressivamente, estão se integrando no nosso dia a dia (variando de robótica e inteligência artificial à biotecnologia e à Internet das coisas), estamos na Quarta Revolução Industrial que, alimentada por dados, está eliminando as fronteiras entre a Física e a Biologia. A Ética, como ramo do conhecimento que lida com a moralidade e, no sentido mais amplo, com a preocupação que os humanos sempre tiveram em descobrir a melhor maneira de viver, e de distinguir entre o bem e o mal, o certo e o errado, não ficou para trás, mas também está evoluindo.

É aqui que encontramos o campo da ética de dados que, com base na fundamentação fornecida pela *computer and information ethics*, sinaliza uma mudança das consultas éticas centradas na informação para as centradas nos dados. Esse ramo específico da ética marca uma mudança de foco sobre o conteúdo (informação) que pode ser criado, registrado, processado e compartilhado através de meios tecnológicos para um foco nos dados (FLORIDI, 2016).

Desta forma, a *data ethics* permite uma melhor análise de temas éticos, como privacidade, anonimato, transparência, confiança e responsabilidade, concentrando-se nos seus aspectos procedimentais (como coleta, curadoria, filtragem, criação de dados) e algorítmicos (aqueles que alimentam IA e ML), que não se traduzem, direta ou necessariamente, em informação, mas ainda assim exercem um impacto significativo na ação e comportamento humanos (FLORIDI, 2016).

Nessa linha, a *data ethics* foi definida como um novo ramo da ética, que estuda e avalia problemas morais relacionados aos dados (incluindo geração, registro, curadoria, processamento, disseminação, partilha e uso), aos algoritmos (incluindo IA, agentes artificiais, *machine learning* e robôs) e à práticas correspondentes (incluindo inovação responsável, programação, *hacking* e códigos profissionais), a fim

de formular e apoiar soluções moralmente boas (por exemplo, condutas corretas ou valores adequados) (FLORIDI, 2016). Analisando cada uma das suas vertentes, a ética dos dados se concentra em problemas éticos colocados pela recolha e análise de grandes conjuntos de dados, abordando questões relacionadas à criação de perfis, publicidade, re-identificação de indivíduos, privacidade em grupo, discriminação e transparência, entre outras; já a ética dos algoritmos enfoca questões levantadas pela crescente complexidade e autonomia de algoritmos (incluindo inteligência artificial e agentes artificiais, por exemplo, como “bots” de internet e outras aplicações de *machine learning*), ou seja, temas relativos à responsabilidade moral e à responsabilidade de projetistas e cientistas de dados; à análise ética e à auditoria de algoritmos; à avaliação de possíveis resultados indesejáveis resultantes do desenvolvimento de produtos e pesquisas conduzidas por essas tecnologias. Por sua vez, a “ética das práticas” (incluindo ética profissional e deontologia) aborda as questões urgentes relativas às responsabilidades das pessoas e das organizações encarregadas de processos, estratégias e políticas de dados, nomeadamente em relação a questões de consentimento, privacidade dos utilizadores e uso secundário (FLORIDI, 2016).

Tendo definido e esclarecido o ângulo ético através do qual avaliar, revisar e governar essas tecnologias, a próxima seção aborda a IA em particular, olhando para os seus desafios, suas oportunidades, seus riscos e benefícios. Posteriormente, e para concluir, a última seção dedicar-se-á à importância dos quadros éticos corporativos como instrumento para maximizar as oportunidades e os benefícios dessa tecnologia, assim como para abordar e minimizar os seus riscos e preocupações.

5 Riscos e desafios éticos da IA⁵

Com foco na IA, que tem recebido recentemente atenção privilegiada por parte de meios de comunicação, acadêmicos e círculos políticos (principalmente devido a conquistas significativas em aplicações de *machine learning*), há uma série de desafios éticos que precisamos de enfrentar.

5.1 Redução do controle humano

A IA facilita a delegação de tarefas importantes em sistemas autônomos que - em circunstâncias específicas - devem permanecer, pelo menos em parte, sujeitos à supervisão humana, seja no “*loop*”, para fins de monitorização, seja no “*pós-loop*”, para corrigir erros ou danos que eventualmente surjam.

5.2 Remoção da responsabilidade humana

Os desenvolvimentos no campo da IA também podem aumentar a tendência para os desresponsabilizar sempre que um sistema de inteligência artificial puder ser responsabilizado por uma falha ou má conduta. Dada a complexa “cadeia” de *designers*, fornecedores e usos automatizados que essas tecnologias envolvem sem intervenção humana, essa tendência poderá tornar mais difícil a responsabilização das pessoas por falhas específicas da IA.

5.3 Desvalorização de competências humanas

Os desenvolvimentos tecnológicos da IA também podem desvalorizar competências humanas, nomeadamente em domínios sensíveis e intensivos, como o diagnóstico médico e a aviação. Se, por exemplo, daqui a algumas décadas, houver uma escassez de especialistas humanos capazes de diagnosticar o câncer, a sociedade estaria mal preparada e equipada para o mau funcionamento da IA e para lidar com eventuais ataques malévolos.

⁵ As seções 5.1, 5.2, 5.3, 5.4, 5.5 e 5.6 replicam, à exceção do risco relacionado com preconceito e (in)justiça, a identificação dos riscos da IA expostos em Yang, G.-Z., J. Bellingham, P. E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, R. Merrifield, B. J. Nelson, B. Scassellati, M. Taddeo, R. Taylor, M. Veloso, Z. L. Wang and R. Wood (2018). “The grand challenges of Science Robotics”. *Science Robotics* 3.

5.4 Erosão da autodeterminação humana

A IA pode corroer a autodeterminação pessoal, pois pode induzir a mudanças não planejadas e indesejadas nos comportamentos humanos para acomodar rotinas que facilitam a automação e a vida das pessoas. O poder de previsão da IA e o denominado “nudging”,⁶ mesmo que não intencional, devem fomentar, e nunca minar a dignidade humana e a autodeterminação (YANG, 2018).

5.4.1 Facilitação de condutas humanas controversas ou mesmo malévolas

Há também o potencial risco para que essa tecnologia caia nas mãos erradas e para usos mal-intencionados da IA, o que representaria grave ameaça à segurança e prosperidade de todos nós. Uma forma de impedir o uso malévolo da IA é adotar a consideração e o entendimento de que devemos tratar as pessoas como fins em si mesmas, e nunca apenas como um simples meio.

5.4.2 Preconceito e (in)justiça

Como referido no início, as tecnologias podem ser desenvolvidas com valores específicos embutidos nelas. Isto é particularmente verdadeiro no caso da IA, na qual os algoritmos são treinados de maneira a informar ou apoiar decisões que refletem (ou deixam de refletir) tantos os valores como os seus preconceitos humanos. Os sistemas de IA - sem salvaguardas adequadas - podem herdar e propagar preconceitos existentes nos dados em que foram treinados. É o caso de modelos treinados em dados em que as minorias são sub-representadas e, portanto, não aparecem bem refletidas, o que levanta questões de potencial preconceito e discriminação, nomeadamente em decisões automatizadas em matéria de emprego, habitação, crédito e justiça criminal.

5.4.3 Benefícios e oportunidades da IA

A IA está refazendo indústrias em todo o mundo. Mais do que uma ciência, os economistas consideram a IA uma tecnologia de propósito geral: ela transforma todos os campos em que entra e opera, da saúde ao transporte, da agricultura à educação. Como parte integrante e cada vez mais determinante da nossa vida quotidiana, é necessário assegurar que a IA nos protege e trabalha a nosso serviço. Apesar dos desafios listados acima, os aplicativos de IA têm enorme potencial para beneficiar, tanto os indivíduos como a sociedade, numa multiplicidade de sectores.

5.4.4 Cuidados de saúde

A aplicação de novos métodos de IA à assistência médica, por meio da análise de imagens médicas e da previsão de resultados de registros, é promissora e, se bem-sucedida, poderá salvar muitas vidas nos próximos 10 a 20 anos. Exemplos são praticamente infinitos, desde o “mobile health startup” que usa IA para sinalizar doenças cardíacas (HELFT, 2017) e IA usada para a reabilitação de movimento e linguagem,⁷ até ao uso de IA para detectar câncer de pulmão (METZ, 2017) e de pele (JACKSON, 2014).

5.4.5 Acessibilidade

A IA tem potencial, já demonstrado aliás, para promover novos avanços em acessibilidade. É o caso de aplicativos que ajudam pessoas com deficiência visual a melhorar a maneira como interagem com o mundo ao seu redor (ACCENTURE..., 2017) através da identificação e leitura de palavras em voz alta. Outro

⁶ “O fenômeno do ‘nudging’, que encontra as suas raízes na disciplina de economia comportamental, pode ser definido como qualquer aspecto da arquitetura de escolha que altere o comportamento das pessoas de uma maneira previsível sem proibir quaisquer opções ou alterar significativamente seus incentivos econômicos.” (THALER; SUNSTEIN, 2008).

⁷ Disponível em: <<http://www.reabilityonline.com/tele-language/>>. Acesso em: 18 jun. 2018.

tipo de aplicativos, também baseados em IA, ajudam crianças autistas a decifram expressões faciais ou identificam sinais de rua num determinado idioma, traduzindo-os para outro.

5.4.6 Agricultura e meio ambiente

A IA pode ajudar a sustentabilidade tanto da agricultura⁸ como da pesca oceânica. Tem também o potencial de otimizar a produção de alimentos em todo o mundo, analisando as regiões agrícolas e identificando o que é necessário para melhorar o rendimento das culturas. No total, quanto mais amplas e profundas forem as aplicações da IA em uma determinada região ou setor econômico, maior será o impacto econômico esperável (CHEN et al., 2016).

5.4.7 Transporte

Impulsionados pela IA, carros autônomos poderiam salvar 300.000 vidas por década nos EUA (BUSH, 2015). Os métodos de IA também estão sendo usados para o planejamento de transportes, ou seja, otimizar serviços como horários de ônibus e metrô, condições de tráfego (MCGOOCAN, 2017), entre muitos outros. A IA é igualmente responsável pelo aumento dos serviços de transporte sob demanda.⁹

A lista continua, abrangendo setores tão diversos como a educação, a segurança pública e a pesquisa, entre outros.

É importante notar que a IA poderá ser a solução para os desafios e preocupações que ela mesmo gera. Tomando o exemplo do preconceito e “*unfairness*”, os sistemas de inteligência artificial poderão ser usados para reduzir tanto os nossos preconceitos conscientes como os inconscientes. De fato, os algoritmos podem ajudar a eliminar o preconceito humano que afeta o modo como tomamos decisões. Estudos demonstram que os algoritmos podem melhorar a seleção de candidatos a emprego através de um melhor processamento de aspectos difíceis de medir, como adequação cultural, “*soft skills*”, etc. - em que o preconceito individual tende a prevalecer mais facilmente entre humanos (COWGILL, 2017). Outro estudo mostrou que preconceitos baseados em gênero podem aparecer nos algoritmos por causa de fatores atenuantes, como os custos mais altos de publicidade para mulheres jovens, uma vez que são um público-alvo mais procurado pelos anunciantes (LAMBRECHT, 2017). Estes tipos de preconceitos podem ser corrigidos e mitigados no processo de construção e avaliação de algoritmos.

6 O papel dos quadros éticos corporativos

Como explicado acima, a IA pode ser usada para o bem, e também para o mal. Isto levanta preocupações agudas, mas tem a virtualidade de trazer benefícios incriveis para os indivíduos e para a sociedade, assim como para resolver - ou pelo menos atenuar - essas mesmas preocupações. Como complemento à lei e à regulamentação, que já aborda e abrange muitas dessas preocupações (CNIL, 2017), as estruturas éticas corporativas podem ser ferramentas importantes para ajudar as empresas a maximizar os benefícios e oportunidades da IA, minimizando os seus riscos e preocupações. Esta última seção explora a relevância desses quadros éticos corporativos. A oportunidade de maximizar os benefícios que essas tecnologias podem trazer, garantindo que danos e resultados negativos sejam evitados, destaca a necessidade da ética como uma estrutura analítica e operacional destinada a orientar a estratégia dos atores corporativos e a moldar as suas práticas nesse domínio.

A ética e o reconhecimento de responsabilidades para com as comunidades servidas por empresas (tanto públicas como privadas) são importantes para definir o propósito organizacional, estabelecer um conjunto comum de valores, providenciar métodos de tomada de decisão consistentes e para aumentar a

⁸ Disponível em: <<http://gamaya.com/>>. Acesso em: 18 jun. 2018.

⁹ Os serviços de transporte sob demanda, como o Uber e o Lyft, surgiram como aplicações essenciais que utilizam sistemas de detecção, conectividade e inteligência artificial, com algoritmos que conectam os motoristas aos passageiros por localização e adequação (“*reputation modeling*”). Ver <<https://ai100.stanford.edu/2016-report/section-ii-ai-domain/transportation/demand-transportation>>.

confiança entre os usuários. A ética também é importante para a reputação, crescimento e *brand building* das empresas. O desenvolvimento de uma estrutura ética corporativa ajuda as empresas a demonstrar como os dados que lhes são confiados são usados de maneira responsável. A ética é também uma ferramenta flexível, permitindo que as empresas alinhem e atualizem, continuamente, as práticas no ritmo do desenvolvimento tecnológico, abordando questões que vão além da estrita conformidade legal.

O recurso à ética como quadro analítico e operacional reconhece que a legislação não cobre tudo (nem deveria), e que os quadros éticos corporativos podem desempenhar um papel importante na avaliação e decisão sobre o uso de dados e processos algorítmicos, nomeadamente em dois casos:

- Quando valores e interesses concorrentes precisam de ser equilibrados e a legislação não fornece uma resposta simples ou direta;

- Quando pode ser feito mais, para além do que as legislações linearmente exigem.

Neste último ponto, a ética justifica e facilita o desdobramento de um conjunto de práticas corporativas adicionais que não derivam de obrigações ou exigências legais, permitindo que as empresas sigam os valores que definiram e cumpram a sua missão integralmente. Tais práticas corporativas podem incluir, entre outras, investimentos em programas sociais, compromisso com iniciativas de transparência e apoio a campanhas educacionais.

A ética também pode ajudar as empresas a atender às expectativas e a seguir as recomendações dos órgãos reguladores, contribuindo para um relacionamento mais saudável e produtivo com essas entidades.¹⁰ Com efeito, pode fornecer a base para uma plataforma de diálogo sustentável e consequente com reguladores sobre regulamentação, especificamente sobre a necessidade e o valor de instrumentos jurídicos, códigos de conduta e outros mecanismos de co-regulação na governança do mundo digital. Em relação a esse ponto, a CNIL - Autoridade Francesa de Proteção de Dados - definiu a ética no seu relatório *How can humans gain the upper hand? Ethical issues of algorithms and artificial intelligence* como um processo de orientação em questões legais, e o padrão ético como uma prefiguração do padrão legal. As estruturas éticas corporativas podem ajudar a construir essa plataforma de diálogo e colaboração, promovendo uma cultura de engajamento construtivo entre órgãos reguladores e organizações responsáveis.

5 Conclusão

Neste artigo, procuramos traçar um panorama de alguns dos aspectos mais atuais da projeção de alguns sistemas de inteligência artificial que se vão inserindo cada vez mais no nosso cotidiano e que, necessariamente, suscitam a atenção pelo seus efeitos potenciais no ordenamento jurídico - como vimos, não somente pela necessidade de criação de novos institutos ou adaptação dos já existentes, mas também pela própria abertura de discussão sobre alguns de seus conceitos fundamentais que estão diretamente ligados às noções de subjetividade e autonomia.

O tema tem potencial de gerar efeitos quase horizontais para o ordenamento jurídico. Certamente, há outras particularidades que devem igualmente abordadas - por exemplo, o potencial das tecnologias de inteligência artificial para estender capacidades humanas ou mesmo para diminuir riscos de discriminação - ressoando Melvin Kranzberg, ao afirmar que “a tecnologia não é boa nem má, nem sequer é neutra” (CASTELLIS, 1996, p. 96).

¹⁰ O relatório da CNIL acima mencionado “*How can humans keep the upper hand? Report on the ethical matters raised by algorithms and artificial intelligence*” recomenda o fortalecimento das funções éticas nas empresas, enquanto que o relatório “*Big data, artificial intelligence, machine learning and data protection*”, publicado pelo Information Commissioner’s Office (IOC – Autoridade de Proteção de Dados do Reino Unido), recomenda o estabelecimento de comitês consultivos dedicados à análise de questões éticas: “*Larger organisations should create ethics boards to help scrutinise projects and assess complex issues arising from big data analytics*”. A criação e utilização de comitês éticos internos é igualmente incentivada pelo Consultative Committee do Conselho Europeu, estabelecido na Convenção para a Proteção das Pessoas relativamente ao Tratamento Automatizado de Dados de Carácter Pessoal (STE 108), nomeadamente no seu relatório “*Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data*”. Disponível em: <<https://rm.coe.int/16806ebe7a>>.

É claramente perceptível um vetor em várias abordagens da literatura e mesmo da regulação no sentido de preservar componentes “humanos” frente à introdução dessas novas tecnologias, por exemplo, como valor a ser salvaguardado como pela sua manutenção no elo da cadeia decisional. Não é possível, no entanto, determinar de forma concreta de que forma e em que situações isto deva ser feito, ainda mais em um cenário altamente dinâmico. Para essa tarefa, no entanto, afigura-se como fundamental a natureza maleável e modular do Direito Civil em possibilitar que a proteção da pessoa seja renovada e inserida em instrumentos novos e mesmo inovadores.

O recurso a elementos éticos pode ainda proporcionar a possibilidade de considerar com o devido cuidado situações que ainda não possam ser efetivamente objeto de regulação ou de atuação direta de institutos jurídicos, mas que, seja pela importância dos sujeitos e valores envolvidos, seja pelos seus potenciais efeitos, demandam uma resposta ágil e ponderada sobre opções a serem tomadas. Assim, verificamos a extrema pertinência da conjugação dos institutos fundamentais de proteção da personalidade com elementos de ética como componentes fundamentais para a devida recepção, que já estamos vivendo, de sistemas de inteligência artificial no nosso cotidiano.

Referências

- ACCENTURE develops AI-powered solution “Drishti” to help visually impaired. **BGR**, 2017. Disponível em: <<http://www.bgr.in/news/accenture-develops-ai-powered-solution-drishti-to-help-visually-impaired/>>. Acesso em: 15 jul. 2018.
- ALEXANDRE, João. Parlamento reconhece animais como seres sensíveis. **TSF**, 2016. Disponível em: <<https://www.tsf.pt/politica/interior/parlamento-reconhece-animais-como-seres-sensiveis-5567504.html>>. Acesso em: 18 jun. 2018.
- AMUDA, Yusuff Jelili; TIJANI, Ismaila B. Ethical and legal implications of sex robot: An Islamic Perspective. **OIDA International Journal of Sustainable Development**, v. 03, n. 6, p. 19-28, Feb. 2012. Disponível em: <<https://ssrn.com/abstract=2008011>>. Acesso em: 18 jun. 2018.
- ARROW, K. The theory of discrimination. In: ASHENFELTER, O.; REES, A. (Ed.). **Discrimination in Labor Markets**. Princeton: Princeton University Press, 1973. p. 3-32.
- ASIMOV, Isaac. **Eu, robô**. São Paulo: Aleph, 2014.
- BALKIN, Jack. The path of robotics law. **California Law Review Circuit**, Berkeley, v. 06, p. 45-60, jun. 2015.
- BIG data: Credit where credit's due. **Financial Times**, 2015. Disponível em: <<https://www.ft.com/content/7933792e-a2e6-11e4-9c06-00144feab7de>>. Acesso em: 18 jun. 2018.
- BRITZ, G. **Freie entaltung durch selbstdarstellung**. Tübingen: Mohr Siebeck, 2007.
- BUCHNER, Benedikt. **Informationelle selbstbestimmung im privatrecht**. Tübingen: Mohr Siebeck, 2006.
- BUSH, Vannevar. As we may think. **The Atlantic**, jul. 1945. Disponível em: <<https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>>. Acesso em: 18 jun. 2018.
- CALO, Ryan. Robotics and the lessons of cyberlaw. **California Law Review**, Berkeley, v. 103, n. 3, p. 513-563, jun. 2015.
- CASTELLIS, Manuel. **The rise of the network society**. Blackwell: Oxford, 1996.
- CASTRO JÚNIOR, Marco Aurélio. **A personalidade jurídica do robô e a sua efetividade no direito**. Salvador, Universidade Federal da Bahia, 2009. Disponível em: <<https://repositorio.ufba.br/ri/handle/ri/10719>>. Acesso em: 18 jun. 2018.
- CHEN, Nicholas et al. Global economic impacts associated with artificial intelligence. **Analysisgroup**, 2016. Disponível em: <<https://www.semanticscholar.org/paper/Global-Economic-Impacts-Associated->

with-Artificial-Chen-Christensen/ebc73c75f7aba486751a20257a3134c777afd255>. Acesso em: 18 jun. 2018.

COWGILL, Bo. Automating judgement and decision making: Theory and evidence from resume screening. Working Paper. **The society of labor economists**, 2017. Disponível em: <<http://www.sole-jole.org/17446.pdf>>. Acesso em: 18 jun. 2018.

DANAHER, John; McARTHUR (Ed.). **Robot Sex**. Cambridge: MIT Press, 2017.

DONEDA, D.; MENDES, L. Data protection in Brazil: New developments and current challenges. In: GURWIRTH, S.; LEENES, R.; HERT, P. de. (Ed.). **Reloading data protection: Multidisciplinary Insights and Contemporary Challenges**. Londres, Springer, 2014. p. 3-20.

FLORIDI, Luciano. Information ethics, its nature and scope. **SIGCAS Comput. Soc.**, New York, v. 36, n. 3, p. 21-36, 2016. Doi:10.1145/1195716.1195719. Acesso em: 18 jun. 2018.

FLORIDI, Luciano; TADDEO, Mariarosaria. Introduction: What is data ethics? **Phil. Trans. R. Soc.**, Londres, v. 374, n. 2083, 2016. DOI: 10.1098/rsta.2016.0360. Acesso em: 18 jun. 2018.

HELFT, Miguel. With \$30 million infusion, mobile health startup bets on ai to flag heart conditions. **Forbes**, 16 mar. 2017. Disponível em: <<https://www.forbes.com/sites/miguelhelft/2017/03/16/with-30-million-infusion-vic-gundotras-mobile-health-startup-unveils-ai-to-flag-heart-conditions/#3d847ca534f8>>. Acesso em: 18 jun. 2018.

HOW can humans keep the upper hand? Report on the ethical matters raised by algorithms and artificial intelligence. **CNIL**, 2017. Disponível em: <<https://www.cnil.fr/en/how-can-humans-keep-upper-hand-report-ethical-matters-raised-algorithms-and-artificial-intelligence>>. Acesso em: 18 jun. 2018.

JACKSON, Joab. IBM detects skin cancer more quickly with visual machine learning. **Computerworld**, 2014. Disponível em: <<https://www.computerworld.com/article/2860758/ibm-detects-skin-cancer-more-quickly-with-visual-machine-learning.html>>. Acesso em: 16 jun. 2018.

LAFRANCE, Adrienne. Self-driving cars could save 300,000 lives per decade in America. **The Atlantic**, 29 set. 2015. Disponível em: <<https://www.theatlantic.com/technology/archive/2015/09/self-driving-cars-could-save-300000-lives-per-decade-in-america/407956/>>. Acesso em: 18 jun. 2018.

LAMBRECHT, Anja; TUCKER, Catherine E. **Algorithmic Bias?** An empirical study into apparent gender-based discrimination in the display of STEM Career Ads, 2017. Disponível em: <<https://ssrn.com/abstract=2852260> or <http://dx.doi.org/10.2139/ssrn.2852260>>. Acesso em: 16 jun. 2018.

LYON, David. **Surveillance as social sorting: privacy, risk, and digital discrimination**. Routledge: New York, 2003.

MAYER-SCHONBERGER, V.; CUKIER, K. **Big Data: A revolution that will transform how we live, work, and think**. New York: First Mariner Books, 2014.

McCARTHY, J. et al. **A proposal for the Dartmouth summer research project on Artificial Intelligence**, 1956. Disponível em: <<http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>>. Acesso em: 16 jun. 2018.

MCGOOGAN, Cara. AI traffic lights to end rush hour jams in Milton Keynes. **Telegraph**, 2017. Disponível em: <<https://www.telegraph.co.uk/technology/2017/05/15/ai-traffic-lights-end-rush-hour-jams-milton-keynes/>>. Acesso em: 18 jun. 2018.

MENDES, Laura Schertel. **Privacidade, proteção de dados e defesa do consumidor: linhas gerais de um novo direito fundamental**. São Paulo: Saraiva, 2014.

METZ, Cade. Using AI to detect cancer, not just cats. **Wired**, 2017. Disponível em: <<https://www.wired.com/2017/05/using-ai-detect-cancer-not-just-cats/>>. Acesso em: 18 jun. 2018.

NEVEJANS, Nathalie. **European civil law rules in robotics**. 2016. Disponível em: <[http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU\(2016\)571379_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf)>. Acesso em: 18 jun. 2018.

NUCCI, Ezio Di. **Sexual rights, disability and sex robots**. 2016. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2814471>. Acesso em: 16 jun. 2018.

PARLAMENTO EUROPEU. **Resolution of 16 February 2017 with recommendations to the commission on civil law rules on robotics**. Disponível em: <<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//ep//text+ta+p8-ta-2017-0051+0+doc+xml+v0//en>>. Acesso em: 18 jun. 2018.

PHELPS, E. The statistical theory of racism and sexism. **American Economic Review**, Washington, v. 62, n. 4, p. 659-661, 1972.

SALICETI, Diana. Les animaux définitivement reconnus comme des «êtres sensibles». **Libération**, 2015. Disponível em: <http://www.liberation.fr/societe/2015/01/28/les-animaux-definitivement-reconnus-comme-des-etres-sensibles_1190862>. Acesso em: 18 jun. 2018.

SCHWAB, Klaus. **The fourth industrial revolution**. Geneva: World Economic Forum, 2016.

THALER, Richard H.; SUNSTEIN, Cass R. **Nudge: improving decisions about health, wealth, and happiness**. New York: Penguin Books, 2008.

TURING, Alan. **Intelligent machinery**, 1945. Disponível em: <http://www.alanturing.net/turing_archive/archive//I32/L32-001.html>. Acesso em: 18 jun. 2018.

WIEZENBAUM, Joseph. **Computer power and human reason**. San Francisco: Freeman, 1976.

YANHG, G.-Z. et al. The grand challenges of science robotics. **Science Robotics**, Washington, v. 3, n. 14, 2018. Disponível em: <<http://robotics.sciencemag.org/content/3/14/eaar7650>>. Acesso em: 13 jun. 2018.

ZARSKY, T. Transparent predictions. **University of Illinois Law Review**, Champaign, v. 2013, n. 4, p. 1503-1570, 2013.

Autores Convidados