

PRINCÍPIOS DE INTELIGÊNCIA ARTIFICIAL
*PRINCIPLES OF ARTIFICIAL INTELLIGENCE***Zulmar Fachin¹****Jéssica Fachin²****Deise Marcelino da Silva³**

RESUMO: O texto é uma reflexão sobre os princípios inspiradores do uso da Inteligência Artificial (IA). Considera que esse novel campo do conhecimento deve ser desenvolvido com a observância de matrizes axiológicas, diante da racionalidade imposta pelas tecnologias que não param de avançar. Tem por objetivo abordar princípios que, embora desenvolvidos no campo da bioética, também devem ser aplicados à IA. Reconhece que a IA impacta em uma pluralidade de direitos fundamentais das pessoas, especialmente, os relativos à sua personalidade. Adotou-se o método hipotético-dedutivo, utilizando-se de textos doutrinários nacionais e estrangeiros. O artigo aponta para a necessidade de se usar a IA de acordo com os princípios que a regem, tendo em vista a simbiose homem-máquina.

Palavras-Chave: Princípios; Inteligência artificial; Direito.

ABSTRACT: The paper is a discussion about the inspiring principles of Artificial Intelligence (AI) usage. It considers that this new field in knowledge should be developed based on the observation of axiological sources facing the rationality imposed by technologies that will not cease advancing. This research addresses principles that, although developed in the bioethical field, also must be applied in AI. This paper affirms that AI impacts a plurality of people's fundamental rights, especially, the ones related to their

¹ Doutor em Direito Constitucional (UFPR). Mestre em Direito (UEL). Mestre em Ciência Política (UEL). Bacharel em Direito (UEM). Licenciado em Letras (Unicesumar). Professor na UEL e no Programa de Doutorado e Mestrado em Ciência Jurídica na Universidade Cesumar - Unicesumar. Coordenador do Mestrado Profissional em "Direito, Sociedade e Tecnologias" da Escola de Direito das Faculdades Londrina. Membro eleito da Academia Paranaense de Letras Jurídicas. Presidente do IDCC - Instituto de Direito Constitucional e Cidadania. Pesquisador do Instituto Cesumar de Ciência, Tecnologia e Inovação. Bolsista Produtividade em Pesquisa do ICETI. E-mail: zulmarfachin@uol.com.br. Lattes: <http://lattes.cnpq.br/8640721822545057>. ORCID - 0000.0001.5514.5547.

² Doutora em Direito Constitucional (PUCSP). Mestre em Ciência Jurídica (UENP). Especialista em Direito Constitucional Contemporâneo e Direito Processual Civil (IDCC). Graduação em Direito e Letras. Professora no Programa de Mestrado Profissional em "Direito, Sociedade e Tecnologia" da Escola de Direito das Faculdades Londrina. Coordenadora de Pós-Graduação *Lato Sensu* do IDCC – Instituto de Direito Constitucional e Cidadania. Advogada. E-mail: jessicaafachin@gmail.com. Lattes: <http://lattes.cnpq.br/5305097400126062>. ORCID: 0000-0003-0486-7309.

³ Doutora em Direito Ambiental Internacional pela UNISANTOS (2017). Professora do Programa de Mestrado Profissional em Direito, Sociedade e Tecnologias da Escola de Direito das Faculdades Londrina/PR. E-mail: deise.marcelino@hotmail.com. Lattes ID: <http://lattes.cnpq.br/2466797631971309>. ORCID ID: <http://orcid.org/0000-0002-7938-0813>.

personality. It is used the hypothetical-deductive method based on foreign and national doctrinaire texts. This article points to the necessity to use AI according to the principles with lead it considering the symbiosis humankind-machine.

Key-words: Principles; Artificial intelligence, Law.

Sumário: 1. Introdução; 2. Princípio da beneficência; 3. Princípio da não maleficência; 4. Princípio da autonomia; 5. Princípio da justiça; 6. Princípio da explicabilidade; 7. Princípio da privacidade; 8. Princípio da responsabilidade; 9. Considerações finais; 10. Referências.

1. INTRODUÇÃO

A IA está em nossa vida cotidiana e pode ser vista ou sentida em cada ato, a cada instante. Ela está em todo o lugar, em nossos telefones e carros, nas experiências de compras, encontros, hospitais, bancos e em todos os tipos de mídias (AGRAWAL; GANS; GOLDFARB, 2018). Sua presença proporciona bem-estar, mas também gera riscos. Posto isto, o Direito busca conhecer a IA para estabelecer a melhor forma de regulação, visto que ela deve estar a serviço da vida, cumprindo o papel de garantir, promover e efetivar direitos e liberdades fundamentais da pessoa humana. No entanto, os desafios da regulação são enormes.

O desenvolvimento da IA remonta à década de 1950, época em que surgiu o computador. John MacCarthy, Marvin Minsky, Alan Newell e Herbert Somin são considerados os pais da IA. O vocábulo foi criado em 1956, por John MacCarthy.

O objetivo desta pesquisa é refletir sobre os princípios que inspiram o uso da IA, para melhor compreender esse novo campo do conhecimento humano. Nessa caminhada, é como se estivéssemos sob as luzes do alvorecer matinal, descortinando um longo alvorecer. Temos a necessidade de acelerar no caminho, mas não podemos ignorar as advertências de Lee (2019) “quando se trata de entender nosso futuro com a IA, somos todos crianças no jardim de infância”. Esse ambiente pode ser também um tempo de oportunidades, se houver a consciência de que o “nosso futuro com a IA será criado por nós e refletirá as escolhas que fizermos e as ações que tomarmos” (LEE, 2019, p. 10-11). Não se está diante de algo decorrente da natureza, mas produzido pelo intelecto humano.

O estudo justifica-se em razão do caráter desafiador que permeia esse vasto campo de utilização das mais avançadas tecnologias. A IA artificial desempenha múltiplas tarefas,

entre as quais podem ser mencionadas as de armazenar e processar dados de forma rápida e barata; interpretar exames médicos; solucionar questões jurídicas; oferecer (sugerir) produtos; deslocar veículos por ruas e estradas; realizar o *machine learning*, tarefas que podem ser realizadas sem a interferência específica da ação humana.

Com o propósito de estabelecer uma base ética a ser observada por profissionais e empresas que trabalham com IA, foram publicados diversos documentos jurídico-normativos. Podem ser mencionados os “Princípios de Asilomar” (2017), a Declaração de Montreal para o Desenvolvimento Responsável da Inteligência Artificial (2017), o Design Éticamente Alinhado (2017), a Declaração sobre Inteligência Artificial (2018), os Princípios da Parceria em Inteligência Artificial (2018) e o Guia Ético (2019). A quantidade e a denominação dos princípios da IA, tratados nesses documentos, oscilam.

Nesse campo em construção, todos os envolvidos devem estar submetidos a diversos princípios. Esse texto é uma reflexão sobre os princípios da beneficência, da não maleficência, da autonomia, da justiça, da explicabilidade, da privacidade e da responsabilidade. Portanto, o trabalho foi estruturado em sete itens, sendo um para cada princípio.

2. PRINCÍPIO DA BENEFICÊNCIA

Benevolência, do latim *benevolentia*, é um substantivo feminino com múltiplos significados. Ele designa bondade de ânimo em face de algo ou alguém; complacência, transigência e magnanimidade em relação às pessoas que se encontram sob sua orientação ou comando; disposição favorável; boa vontade; manifestação de afeto; benquerença; afabilidade (DICIONÁRIO ON LINE). O vocábulo pode, ainda, traduzir amizade, fraternidade, altruísmo, tolerância, estima e bondade. Trata-se de uma atitude ou comportamento direcionado ao bem próprio e alheio.

Uma pessoa, empresa, governo ou sistema podem ter atitudes benevolentes ou não, podem incorporar em suas atuações o objetivo de realizar benevolência, mas podem se negar a fazê-lo. É certo, porém, que o agir benevolente engrandece o benévolo, visto que o ato praticado, em regra, revela o que o seu autor é.

Na concepção aristotélica do que é o justo, os homens devem praticar o bem, visto

que, construindo, tornam-se construtores; tocando a cítara tornam-se citaristas; realizando ações justas, tornam-se justos; praticando ações moderadas, tornam-se moderados; e realizando ações corajosas, tornam-se corajosos (ARISTÓTELES, 2016).

Praticar bons atos deve ser elevado a ideário de vida. Ronald Dworkin (2014) afirma que as pessoas devem conduzir suas vidas da forma mais valiosa possível, o que implica em ter atitudes benevolentes.

[...] defendo a ideia de que cada um de nós tem a responsabilidade ética e soberana de transformar sua vida em algo de valor [...] nossas diversas responsabilidades e obrigações para com os outros defluem dessa responsabilidade pessoal que temos pela nossa própria vida [...]. Temos de tratar a construção da nossa vida como um desafio que podemos enfrentar bem ou mal. Temos de reconhecer a ambição de tornar boa a nossa vida – torná-la autêntica e valiosa, não mesquinha ou degradante – como o primeiro dos nossos interesses particulares. Em específico, temos de preservar nossa dignidade (DWORKIN, 2014, p. 22).

Na perspectiva da IA, não se deve deixar de reconhecer que qualquer pessoa pode ter conduta maléfica, seja essa pessoa má ou boa. Amy Webb (2020, p. 8) observa que, “mesmo as pessoas com boas intenções podem, sem querer, causar grande mal. Dentro da tecnologia, e sobretudo quando se trata de IA, devemos sempre nos lembrar de planejar tanto a sua aplicação pretendida quanto sua aplicação abusiva”, ainda que a aplicação abusiva não seja desejada pelo autor do ato.

O princípio da benevolência pode ser compreendido como a obrigação ética de buscar o máximo de benefício com a redução dos riscos ao mínimo possível. A IA deve ser utilizada para realizar o bem. O profissional ou a empresa que a utiliza deve impor a si mesmo o dever de proporcionar máximo benefício possível e reduzir os riscos ou os danos ao mínimo possível ou mesmo não acarretar nenhum risco ou dano a outros indivíduos ou à comunidade.

Ressaltamos que a benevolência é um princípio que busca valorizar a pessoa humana. De forte conteúdo ético, volta-se para a realização de benefícios à humanidade.

Significa dizer que a IA deve necessariamente ser centrada na pessoa humana, a ela direcionada e por ela supervisionada, engajando os sistemas para servir à coletividade, no sentido de ampliar não só a segurança, mas também garantir a autonomia e a capacidade de decidir das pessoas. Considerando essas premissas, os sistemas de IA devem, a um só tempo, potencializar os benefícios que geram para a humanidade, e prevenir os riscos que venham a causar à coletividade

(MULHOLLAND; FRAJHOF, 2021, p. 74).

Assinalamos que o princípio da benevolência pode ser denominado princípio do bem-estar. A IA tem que estar vinculada a compromissos capazes de gerar o bem-estar às pessoas e à própria humanidade.

Nesse sentido, a IA deve ajudar os indivíduos a melhorarem suas condições de vida, saúde e trabalho; deve permitir que os indivíduos atinjam seus objetivos, sem causar danos a outros seres conscientes; deve permitir que as pessoas exerçam suas capacidades mentais e físicas; não deve ser uma fonte de mal-estar, a menos que seja uma forma de se criar um bem-estar superior e que seja absolutamente necessário (não atingível de outra maneira); e, por fim, não deve contribuir para aumentar o estresse, a ansiedade e um sentimento de perseguição ou preocupação relacionado ao ambiente digital de alguém (PEIXOTO, 2020, p. 120).

A IA deve estar a serviço das pessoas e seu uso, comprometido com a realização do bem.

3. PRINCÍPIO DA NÃO MALEFICÊNCIA

O princípio da não maleficência guarda semelhanças com o da benevolência, porém, é distinto deste. Segundo o princípio aqui analisado, o uso da IA deve causar o menor prejuízo. Desse modo, quem utiliza da IA deve prever possíveis riscos a outras pessoas ou à sociedade, devendo tomar medidas eficazes para que isso não aconteça.

O princípio da não maleficência deve guiar a tomada de decisão para se evitar que se cause danos voluntários ou involuntários aos demais seres humanos; devendo-se, portanto, avaliar os riscos das condutas para se escolher a melhor alternativa para a sociedade. Desse modo, embora conectado com o princípio da beneficência, o princípio da não maleficência é diverso, pois exige que se adotem medidas eficazes para a prevenção dos riscos advindos com o desenvolvimento da IA (LIMA, OLIVEIRA, RUIZ, 2021, p. 122).

Assim, podemos identificar semelhança entre o princípio da não maleficência com o da prevenção, que inspira o Direito Ambiental, visto que ambos impõem que se adote prevenção em face do risco. Em matéria ambiental, por força do princípio da prevenção, é preciso antever o risco de causar prejuízo a outrem, visto que ele traduz a ideia de prevenir

contra possível fato danoso ao meio ambiente (MILARÉ, 2020), ao passo que, no campo da IA, o princípio da não beneficência exige a escolha da melhor alternativa, ou seja, aquela que não cause danos às pessoas ou à sociedade.

Os riscos gerados pelas máquinas inteligentes são potencialmente causadores de danos às pessoas, pois “O futuro da Inteligência Artificial traz consigo grandes promessas e ameaças: o que definirá sua realização somos nós, no tempo presente” (STEIBEL; VICENTE; JESUS, 2020, p. 60). Embora os danos sejam causados pela máquina, não se pode ignorar que ela é acionada e gerenciada por pessoas, as quais devem zelar para que o risco do dano seja eliminado ou minorado.

A IA proporciona benefícios e riscos de malefícios. Essas duas realidades, que podem se apresentar como verso e reverso de uma mesma medalha, precisam ser levadas em consideração (variantes) pelos que manuseiam o sistema. Neste sentido, a lição de Kai-Fu Lee (2019, p. 10-11) explica que:

Grandes avanços teóricos em IA finalmente têm produzido aplicações públicas que estão prestes a mudar nossas vidas. A IA já alimenta muitos de nossos aplicativos e sites favoritos, e nos próximos anos dirigirá nossos carros, gerenciará nossos portfólios, fabricará muito do que compramos e potencialmente tirará nossos empregos. Esses usos estão repletos de riscos promissores e perigos potenciais, e devemos nos preparar para as duas coisas.

O princípio da não maleficência também se aplica a questões bélicas, ou seja, no caso de uso da IA para produzir armas e artefatos utilizados para destruir vidas, como ocorre nos casos de guerras entre nações. O mesmo se pode dizer em relação aos genocídios e à eliminação de pessoas (minorias) desprovidas de condições de defesa ou de contra-ataque.

Destacamos que o princípio da não maleficência pode ser traduzido pela primeira das Três Regras da Robótica criadas por Issac Asimov: “um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano venha a ser ferido” (ASIMOV, 2014, p. 65). Assim, o risco de ferir ou prejudicar outrem precisa ser identificado e evitado.

4. PRINCÍPIO DA AUTONOMIA

A autonomia designa a capacidade que uma pessoa tem de se autogovernar, tomando as decisões que pretende, de acordo com sua livre escolha e com os fins almejados.

Autônomo é o sujeito que tem a faculdade para determinar seu próprio comportamento, desvinculado de qualquer condicionamento advindo de outro núcleo de poder. No espaço virtual, especialmente em face da IA, devemos respeitar a autonomia em favor de cada pessoa.

A máquina pode desempenhar tarefas laborais, inclusive, substituindo as atividades humanas. Contudo, o homem deve permanecer no controle da máquina. Ainda que toda a tarefa seja desempenhada pela máquina, o homem detém a autonomia para decidir. Tem se afirmado que, em face da IA, “não se deve reduzir a liberdade humana em benefício da rapidez das máquinas em realizar tarefas específicas”, visto que “é necessário garantir que os indivíduos possam recobrar o poder de decisão ao utilizarem aplicações de IA para determinadas tarefas que tenham escolhido” (LIMA; OLIVEIRA; RUIZ; 2021, p. 122).

Nessa perspectiva, para Fabiano Hartmann Peixoto, “Um sistema de IA deve ser desenvolvido e usado respeitando a autonomia das pessoas e com o objetivo de aumentar o controle das pessoas sobre suas vidas e seu ambiente” (PEIXOTO, 2020, p. 120).

Klaus Schwab (2019) adverte que a tecnologia é uma força da natureza externa, sobre a qual não temos nenhum controle. Diante dos seus avanços, nós não estamos limitados por uma escolha binária entre 'aceitar e viver com ela' ou, de outro lado, 'rejeitar e viver sem ela'. As mudanças tecnológicas nos conduzem a refletir sobre quem somos e como vemos o mundo que se transforma constantemente (SCHWAB, 2019). Nesse contexto, a autonomia de escolher e tomar decisões sofre restrições.

Todavia, assinalamos, diante das novas tecnologias, que a pessoa humana é o *sujeito* e, como tal, deve manter o protagonismo. Em nenhuma hipótese, a pessoa pode ser tratada como objeto em relação às tecnologias, especialmente, em face da IA.

Sublinhamos que a Declaração de Montreal pelo Desenvolvimento da Inteligência Artificial, inserida no Relatório da Declaração de Montreal pelo Desenvolvimento Responsável da Inteligência Artificial, aprovado em 2018, prevê que os sistemas de IA devem ser desenvolvidos e usados com respeito à autonomia de cada pessoa e com o objetivo de aumentar o controle, por parte dos indivíduos, de sua vida e também de seu meio ambiente. Neste sentido, de acordo com a Declaração, os sistemas de IA devem proporcionar algumas vantagens: a) capacitar os indivíduos a realizarem seus próprios objetivos morais e

sua concepção de uma vida digna de ser vivida; b) desenvolver sistemas para prescreverem aos indivíduos um modo particular de vida, direta ou indiretamente, implementando mecanismos restritivos de monitoramento, avaliação ou incitação; evitar a propagação de informações não confiáveis, mentiras e propaganda; serem projetados com o objetivo de reduzir a propagação informações não confiáveis, mentiras e propaganda; evitar criar dependências por meio de técnicas de captação da atenção e de imitação da aparência humana, que possam induzir a uma confusão entre SIAs e seres humanos. Acrescentamos, ainda, que as instituições públicas não devem usar os SIAs para promover ou para desvalorizar uma concepção do que seja uma boa vida. Neste sentido, é essencial capacitar os cidadãos em relação às tecnologias digitais, garantindo-se o acesso aos diferentes tipos de conhecimento relevantes, o desenvolvimento de competências estruturantes (alfabetização digital e midiática) e a formação do pensamento crítico (art. 2º).

5. PRINCÍPIO DA JUSTIÇA

A justiça é um valor universal elevado a princípio. Podemos afirmar que ela sempre foi revestida de certa mística. Tem sido representada por símbolos: uma deusa (Deusa Themis), mas também pela espada e a balança. Assystematicamente, utilizam-se as seguintes expressões para designar alguma espécie de justiça: legal, social, ambiental, universal, particular, corretiva, relativa, comutativa, distributiva, restaurativa, preditiva e dedutiva.

A justiça é um daqueles valores que atravessam os tempos. Na tradição grega, Platão e Aristóteles a desenvolveram. Platão (2016) tratou de diversas concepções de justiça, por meio de Sócrates, o personagem principal. Os demais personagens que participaram do diálogo sobre a justiça foram Adimanto, Céfalo, Clitofon, Gláucon, Polemarco e Trasímaco. Aristóteles, por sua vez, concebeu diversas espécies de justiça: universal, particular (distributiva ou corretiva), econômica, política, natural e legal. Para ele, a justiça é “uma disposição de caráter pela qual os homens praticam coisas que são justas, e pela qual agem de maneira justa e desejam coisas justas”. Virtude revestida de grandeza ímpar, “a justiça parece ser às vezes a mais forte dentre as virtudes, e nem ‘a estrela da noite’, nem a ‘estrela da manhã’ são assim admiráveis” (ARISTÓTELES, 2015, p. 123 e 125).

Na Era Moderna, a justiça foi objeto de muitos estudos. Blaise Pascal ensinou que a

justiça, para imperar, deve estar acompanhada da força, advertindo também que a força sem justiça conduz à tirania. “A justiça sem a força é impotente; a força sem a justiça é tirânica [...]. É preciso, pois, colocar juntas a justiça e a força e, para isso, fazer com que aquilo que é justo seja forte ou que o que é forte seja justo” (PASCAL, 2000, p. 36).

Na sociedade contemporânea, a justiça tem sido concebida a partir de diversos significados. Trata-se de vocábulo plurissignificativo e dotado de grande riqueza semântica, visto que ele pode traduzir variadas concepções desse ideal tão almejado pelas pessoas e pela própria humanidade. São incontáveis os autores – localizados em diversos campos do conhecimento humano – que se dedicaram à compreensão desse tema tão complexo.

A partir de três abordagens distintas, Michael J. Sandel (2011) atribui à justiça diversos significados: a) maximizar a utilidade ou o bem-estar, isto é, buscar a máxima felicidade para o maior número possível de pessoas; b) respeitar a liberdade de escolha, o que abrange as *escolhas reais* que as pessoas fazem e as *escolhas hipotéticas* que as pessoas deveriam fazer; c) envolve o cultivo e a preocupação com o bem comum. O autor acrescenta que, para alcançar uma sociedade justa, é preciso que as pessoas raciocinem juntas não apenas sobre o significado da vida boa, mas também para criar uma cultura pública capaz de conviver com as divergências que inevitavelmente ocorrerão (SANDEL, 2011).

Nessa perspectiva, podemos fazer referência a uma definição bastante comum, segundo a qual a justiça consiste em *dar a cada um o aquilo que é seu*. Parece ser nessa perspectiva que Chaïm Perelman, advertindo ser ilusório enumerar todos os sentidos da noção de justiça, destacou as concepções mais estudadas, reconhecendo que tais concepções tem o caráter inconciliável entre si. Para ele, a justiça pode ser concebida em qualquer das seguintes definições: dar a cada qual a mesma coisa; dar a cada qual segundo seus méritos; dar a cada qual segundo suas obras; dar a cada qual segundo suas necessidades; dar a cada qual segundo sua posição; dar a cada qual segundo o que a lei lhe atribui (PERELMAN, 1996).

Os tempos mudam, as tecnologias revolucionam, as pessoas evoluem – mas o ideário de justiça continua a desafiar os espíritos. No campo da IA, especificamente, ela mantém-se como princípio inspirador e condicionante.

6. PRINCÍPIO DA EXPLICABILIDADE

Trata-se de princípio desenvolvido especificamente na IA. Leva ao reconhecimento de que os domínios da IA não podem estar em segredo. As tecnologias que a conduzem devem estar ao alcance do conhecimento das pessoas.

O princípio da *explicabilidade* [...] consiste na compreensão pelos seres humanos do que está sendo oferecido a eles no que se relaciona à aplicação da IA. Desse modo, ao se adquirir determinados produtos que são desenhados para automatizar as tarefas por meio do emprego da IA, deve-se compreender como as decisões são formadas, como se desenvolve uma *machine learning*, quem é responsável por eventuais danos etc. (LIMA; OLIVEIRA; RUIZ, 2021, p. 122).

Os conteúdos democrático e tecnológico são, com certa frequência, incompatíveis. A norma técnica tem força coercitiva, impositiva, arbitrária, mas garante um grau variável de cientificidade. O conteúdo democrático, muitas vezes, exige o afastamento ou o abrandamento da técnica. Isto ocorre em regimes ditatoriais, mas também se verifica, ainda que em menor escala, nos regimes democráticos. Segundo Pierre Lévy (2010, p. 8), “apesar de vivermos em um regime democrático, os processos sociotécnicos raramente são objeto de deliberações coletivas explícitas, e menos ainda de decisões tomadas pelo conjunto dos cidadãos”. Diante de um quadro assim específico, o detentor da tecnologia precisa explicar a tomada de decisão que impacta na esfera dos indivíduos. Em outras palavras, quem faz uso da IA não pode reivindicar para si o segredo dos seus atos.

É importante notar que o princípio da explicabilidade comporta os da inteligibilidade e o da *accountability*, embora alguns doutrinadores tratam de forma autônoma esses três princípios. Por outro lado, ele também pode ser compreendido como um princípio de participação democrática, na medida em que os cidadãos devem ser consultados e informados das decisões tomadas pelos operadores do sistema de IA, ou como princípio da transparência, visto que o sistema de IA não pode estar coberto pelo segredo.

Nesse contexto, a atuação dos operadores do sistema da IA deve ser transparente, pois “A questão da transparência é uma condição necessária para [que] tais agentes inteligentes possam argumentar e explicar as decisões por eles tomadas”. Como resultado dessa atuação, “Os chamados sistemas de IA explicáveis incorporam processos de explicação que permitem aos usuários obter informações sobre os modelos de decisões do

sistema” (SICHMAN, 2021, p. 43).

Não se pode ignorar que um sistema de IA concentra poder e, em razão da natureza deste, precisa estar visível e ser compreendido pelas pessoas. Já o poder invisível é potencialmente mau pela sua própria essência. Lembra Norberto Bobbio que “o poder invisível se forma e se organiza não somente para combater o poder público, mas também para tirar benefícios ilícitos e extrair dele vantagens que não seriam permitidas por uma ação à luz do dia” (BOBBIO, 2015, p. 33).

A LGPD assegura, além de outros princípios, o da transparência, consistente na garantia, em favor dos titulares de dados pessoais, de informações claras, precisas e facilmente acessíveis no que tange à realização do tratamento e aos respectivos agentes de tratamento, observados os segredos comercial e industrial (art. 6º, inciso VI).

7. PRINCÍPIO DA PRIVACIDADE

O princípio da privacidade assume elevada importância em face do deslumbramento do ciberespaço. A IA, proporcionadora de inúmeros benefícios sociais, acarreta riscos à privacidade das pessoas. Basta imaginar sua utilização para o reconhecimento facial, o que, aliás, deixou de ser novidade. Há bastante tempo tem sido observado que “a tecnologia e o tratamento automatizado de informações pessoais mudaria o balanço de poder entre cidadãos, corporações e Estado em relação à privacidade” (DONEDA, 2015, p. 369).

O espaço da vida privada tem sido protegido em face dos atos praticados na esfera pública. Essa preocupação não deixou de existir. Contudo, em razão das facilidades tecnológicas, “a esfera *pública* é que se encontra hoje inundada e sobrecarregada, invadida pelos exércitos da *privacidade*” (BAUMAN, 2011, p. 41). Com frequência inédita e assustadora, as pessoas deixam seu recôndito para se expor em espaços públicos, sobretudo, os que são oferecidos no mundo virtual.

No espaço virtual, especificamente em relação à IA, a proteção jurídica da privacidade ocupa um ponto central. As novas tecnologias oferecem instrumentos para a classificação, seleção, triagem e controle de indivíduos em qualquer parte do mundo, dificultando a realização de controle adequado por parte das autoridades públicas, quer nacionais, quer internacionais. Não se pode mais achar que os riscos à privacidade advêm

apenas do poder público, já que o poder privado, principalmente por meio das big tech ou de grandes organizações, representa ameaças permanentes à privacidade. Nesse contexto, a pessoa humana se encontra em desvantagens em relação a esses núcleos de poder, visto que é evidente a enorme defasagem de poder que existe entre ele e o poder público ou as grandes organizações coletoras, administradoras e manipuladoras das mais sofisticadas tecnologias (RODOTÀ, 2008)

Os sistemas de IA têm significativo poder de analisar e utilizar os dados pessoais que estão à sua disposição. Em face desse poder cada vez mais agigantado, as pessoas que titularizam dados pessoais têm o direito de acessar, gerenciar e controlar o uso desses dados que, na maioria das vezes, elas mesmas fornecem. Por outro lado, a IA deve resguardar o direito à privacidade frente aos sofisticados mecanismos de vigilância, que são potencialmente violadores desse bem jurídico tão essencial às pessoas. Isto se faz necessário, inclusive, porque “A vigilância extrema, pública ou privada, e o uso de dados para fins secundários, sem consentimento, consubstanciam práticas, moral e juridicamente, reprováveis” (FREITAS; FREITAS, 2020, p.76).

A IA trabalha com dados pessoais e eleva os resultados do trabalho. Quanto mais dados pessoais, melhor será o desempenho e maior a produtividade. Isto gera um dilema entre desempenho versus privacidade. Observam os economistas que a IA funciona melhor com mais dados e a utilização de dados pessoais acarreta a redução do espaço da privacidade (AGRAWAL, 2018). Cria-se um mercado em que a privacidade passa a ter valor econômico. Logo, proteger a privacidade significa estabelecer obstáculos ao lucro.

No âmbito da sociedade informacional, com predomínio do digital e a expansão das comunicações, a privacidade não é apenas um bem individual inerente à personalidade humana, mas um patrimônio da humanidade. Para Ulrich Beck (2018), neste início de século, “O direito de proteger a privacidade combinado com o dever de proteção de dados é o supremo direito humano fundamental” (BECK, 2018, p. 187). Esse elevado grau de importância assumido pela privacidade fez com que ela passasse a receber proteção jurídica nas mais variadas escalas.

A privacidade tem proteção no Direito Internacional, no Direito Comunitário e no Direito interno de cada País. No Brasil, a privacidade tem proteção na Constituição Federal

e no direito infraconstitucional.

No Direito Internacional, a privacidade encontra proteção jurídica na Declaração Universal de Direitos Humanos (art. 12) e no Pacto Internacional de Direito Cívico e Político (art. 17). No Direito Comunitário, a privacidade é protegida em diversos documentos jurídicos, tais como a Convenção Americana de Direitos Humanos (art. 5º), a Carta dos Direitos Fundamentais da União Europeia (art. 7º) e a Declaração Africana de Direitos Humanos e dos Povos (art. 4º).

No Direito brasileiro, a privacidade recebe forte proteção na Constituição e no direito infraconstitucional. Na seara constitucional, ela tem status formal e material de direito fundamental (art. 5º, inciso X), reforçada pelo qualificativo de cláusula pétrea (art. 60, § 4º, inciso IV). No direito infraconstitucional, a privacidade aparece como um dos fundamentos da Lei Geral de Proteção de Dados Pessoais (art. 2º, inciso I) e como princípio na Lei do Marco Civil da Internet (art. 3º, inciso II).

Trata-se, portanto, de um princípio que tem expressa proteção em diversos planos normativos.

8. PRINCÍPIO DA RESPONSABILIDADE

Este princípio impõe responsabilidade a quem manuseia a IA. Os *designers* e os construtores de sistemas de IA devem ser responsabilizados por suas condutas lesivas. Embora esse sistema possa violar direitos de pessoas, “somente seres humanos podem ser responsabilizados por decisões decorrentes de recomendações feitas pelo sistema e pelas ações que daí decorrem” (PEIXOTO, 2020, p. 126). Nesse campo, podem ser identificadas violações éticas, mas também jurídicas, acarretando responsabilidades administrativa, civil e criminal. Podemos falar, então, em múltiplas responsabilidades.

Em uma perspectiva ética, é preciso reconhecer que a tecnologia não é neutra, pois o desenvolvimento e a implementação da IA (robôs) têm implicações de diversas espécies, inclusive éticas. Gianmarco Veruggio observa que qualquer pessoa, mesmo não dotada de conhecimentos técnicos, “pode vislumbrar seu potencial e imaginar os mil caminhos que essa disciplina e suas aplicações podem tomar, invadindo campos tradicionais de conhecimento e abrindo novos e complexos problemas de natureza ética, filosófica, social,

jurídica” (VERUGGIO).

Em face da ausência de regulação jurídica específica e eficaz da IA, as normas éticas, ainda que não sejam dotadas de força coercitiva que caracterizam as normas jurídicas, passam a desempenhar papel relevante na proteção das pessoas. Em outras palavras, as normas éticas impõem condutas, as quais, se observadas, cumprem papel similar ao desempenhado pelas normas jurídicas. Contudo, não se pode deixar de reconhecer que as normas jurídicas são bem mais eficazes para a regulação das condutas humanas.

O uso da IA deve se dar em consonância com normas éticas. Isso implica em reconhecer que não é possível tratar de IA sem compreender a dimensão ética desse espaço da virtualidade. A exigência de conteúdo ética, contudo, deve ocorrer em relação às pessoas que usam a IA e não em relação às máquinas. Segundo a doutrina atualizada, “O uso da inteligência artificial perpassa aspectos éticos desde o plano da sua configuração, dos sistemas de recomendação automáticos, da mineração de dados, de clusterização ou classificação” (MELLO; DRESCH, 2021, p. 148).

A responsabilidade administrativa pode ocorrer na administração pública e no âmbito do poder privado. As empresas, utilizando-se de IA no âmbito de sua gestão, assumem, com seus atos gestão, a responsabilidade de não causarem danos a terceiros. No mesmo sentido, as tecnologias têm sido cada vez mais utilizadas pelos poderes públicos. A IA, especificamente, vem desempenhando, sempre em expansão, papel importante no planejamento e na execução de serviços públicos, como é o caso destacado do Poder Judiciário. O dano causado pelo poder público faz nascer a imposição da responsabilidade administrativa. Então, podemos afirmar que a IA contribui para a elevação da qualidade dos serviços públicos e privados.

O uso da IA pode causar às pessoas dano material, moral ou à imagem, os quais podem atingir diversos bens jurídicos, tais como a integridade física, a honra, a imagem, a propriedade, a privacidade, a igualdade, a liberdade e, até mesmo, a vida.

Homem e máquina não se confundem. Máquina – mesmo diante das mais avançadas tecnologias – não causam danos às pessoas físicas. Mas quem manuseia a máquina pode causar danos, o que enseja o dever de reparação. Na perspectiva extracontratual, o *robot* não pode ser o garante do ser humano que atua por detrás de si, visto que não é responsável pela

pessoa, mas, ao contrário, o homem é responsável pelo uso do ente dotado de inteligência artificial. Ademais, por não deter patrimônio para solver crédito de terceiros, o *robot* não tem como assumir o ônus da indenização, o que deverá ser suportado pela pessoa física que está por detrás do ente dotado de inteligência artificial ou por um fundo especial constituído para essa finalidade (BARBOSA, 2021)

É importante observar que os modelos tradicionais de responsabilidade civil são insuficientes para a reparação dos danos causados pela IA. Atenta a essa deficiência, a doutrina tem buscado construir novas soluções para a responsabilidade dos entes dotados de inteligência artificial. Neste sentido, Mafalda Miranda Barbosa aponta duas hipóteses. A primeira seria a criação de fundos de compensação, os quais poderiam tomar duas formas: a) um fundo comum, subsidiado por todos os cidadãos; b) um fundo mantido pelos utilizadores da IA ou por um grupo de pessoas que, direta ou indiretamente, obtêm benefícios dessa tecnologia. A segunda hipótese seria tratar os entes dotados de inteligência artificial como *agentes morais*, pois, com isso, eles teriam personalidade jurídica. Em ambas as hipóteses, a responsabilização civil tornar-se-ia possível e as vítimas não correriam o risco de quedarem-se irressarcidas (BARBOSA, 2021)

É imprescindível que os danos sejam indenizados. Mas, quem irá indenizá-los, já que os robôs de IA não tem personalidade jurídica própria? Em outras palavras, “como não se reconhece hoje a Inteligência Artificial como um sujeito de direito, ela não pode ser considerada individualmente responsável pelos potenciais danos que pode causar” (MAGRANI; SILVA; VIOLA, 2020, p. 113).

A impossibilidade de atribuir responsabilidade civil à IA (sistemas, robôs) tem sido sustentada pela doutrina.

Afirma-se que, para que seja conferida a responsabilidade – sob qualquer fundamento teórico –, é necessário identificar uma pessoa a quem se pode imputar a obrigação de reparar o dano. A atribuição de personalidade jurídica é um pressuposto, portanto, para a responsabilização. Ainda quando o dano for ocasionado por uma coisa – isto é, um bem, desprovido, portanto, de personalidade jurídica –, a obrigação de indenizar deve ser deferida a uma pessoa física ou jurídica, únicas figuras do Direito Civil a quem se concede a capacidade de atuação jurídica (MULHOLLAND, 2020, p. 334).

É oportuno registrar que a Europa pode estar construindo um caminho diferente. Em

16 de fevereiro de 2017, o Parlamento Europeu editou uma resolução que contém recomendações à Comissão de Direito Civil sobre Robótica (2015/2013 (INL)) no sentido de criar um estatuto jurídico específico para que os robôs autônomos mais sofisticados possam vir a ter o status de pessoas eletrônicas responsáveis por qualquer dano que causarem. O reconhecimento dessa personalidade *sui generis* poderia ser adotado nos casos em que o robô toma decisão autônoma ou que venha a interagir com pessoas físicas de forma independente (art. 59) (2015/2013 (INL))

O uso da IA, por fim, pode resultar na prática de crime. Alguns bens jurídicos mais suscetíveis a esse tipo de crime – porém não os únicos – são a vida, a propriedade e a honra. A prática delitiva pode se dar por dois caminhos distintos: a IA artificial é concebida como instrumento, por meio do qual o delito é concretizado; a IA, por si só, desenvolvendo atividade autônoma, realiza a conduta delitiva, quando, por exemplo, a sequência algorítmica se desloca da sua trilha programada na origem (FREITAS, FREITAS, 2020).

Ressaltamos, ainda, que, no campo das tecnologias em geral e da IA em particular, utiliza-se da *soft law*, ou seja, dispositivos legais com baixa densidade normativa, sinalizando caráter pedagógico e não propriamente sancionatório.

9. CONSIDERAÇÕES FINAIS

A benevolência tem muitos significados, tais como magnanimidade, complacência, benquerença, amizade, fraternidade, tolerância e bondade. Refere-se ao bem próprio ou alheio. Cada pessoa tem o compromisso ético de buscar o máximo de benefício com a redução dos riscos ao mínimo possível. Nesta linha, a IA dever ser utilizada para o bem das pessoas.

A não maleficência exige que o uso da IA cause o menor prejuízo possível. Quem se utiliza da IA deve prever possíveis riscos a outras pessoas e à sociedade. Diante do risco identificado, a pessoa deve tomar as medidas eficazes para que os danos não se concretizem.

A autonomia designa a capacidade que uma pessoa tem de se autogovernar, tomando as decisões que pretende, de acordo com sua livre escolha e seus fins. No espaço virtual, especialmente em face da IA, deve-se respeitar a autonomia em favor de cada pessoa, de modo que ela não seja dominada pela máquina.

A justiça é um valor universal elevado a princípio. Revestida de uma certa mística, ela é representada por símbolos, tais como a Deusa Themis, a espada e a balança. O vocábulo justiça é plurissignificativo e dotado de grande riqueza semântica, visto que ele pode traduzir variadas concepções desse ideal tão almejado pelas pessoas e pela própria humanidade. São incontáveis os autores – localizados em diversos campos do conhecimento humano – que se dedicaram à compreensão desse tema tão complexo. Fim último a ser alcançado pelo Direito, ela pertence também a outros campos do conhecimento humanos.

O princípio da explicabilidade foi desenvolvido especificamente no campo da IA. Os domínios da IA não podem estar em segredo. As tecnologias que a conduzem devem estar ao alcance do conhecimento das pessoas. Relacionado com inteligibilidade, *accountability*, transparência e democracia, o princípio da explicabilidade busca garantir que as pessoas possam compreender que o que está sendo oferecido diga respeito à IA. Diante de um quadro assim específico, o detentor da tecnologia precisa explicar a tomada de decisão que impacta na esfera dos indivíduos, ou seja, quem faz uso da IA não pode reivindicar para si o segredo dos seus atos.

O princípio da privacidade assume elevada importância em face do ciberespaço, visto que a IA acarreta riscos à privacidade das pessoas. O espaço da privacidade tem sido protegido diante dos atos praticados na esfera pública. Contudo, em uma pluralidade de situações, as pessoas deixam seu recôndito para se exporem nos espaços públicos, sobretudo, os que são oferecidos pelo mundo virtual. Nesse campo, a proteção jurídica da privacidade ocupa um ponto central. Os riscos à privacidade advêm não apenas do poder público, mas também do poder privado, principalmente, por meio das *big tech* ou de grandes organizações.

O uso da IA, portanto, pode gerar não apenas responsabilidades éticas, ensejando a aplicação de sanções sociais, mas também responsabilidades jurídicas, estas de natureza administrativa, civil e/ou penal, o que pode acarretar a vedação à prática de atos, o dever de reparar o dano ou o cumprimento de sanção penal.

10. REFERÊNCIAS

AGRAWAL, Ajay; GANS, Joshua; GOLDFARB, Avi. **Prediction Machines**: the simple economics of artificial intelligence. Boston, Massachusetts: Harvard Business Review

Press, 2018.

ARISTÓTELES. **Ética a Nicômaco**. Tradução de Luciano Ferreira de Souza. São Paulo: Martin Claret, 2015.

ASIMOV, Isaac. **Eu, Robô**. Tradução de Aline Storto Pereira. São Paulo: Aleph, 2014.

BARBOSA, Mafalda Miranda. Responsabilidade Civil pelos Danos Causados por Entes Dotados de Inteligência Artificial. In: BARBOSA, Mafalda Miranda et al. **Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa**. Indaiatuba, SP: Foco, 2021, p. 157-158.

BECK, Ulrich. **A Metamorfose do Mundo: novos conceitos para uma nova realidade**. Tradução de Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar, 2018.

BAUMAN, Zygmunt. Estranhas Aventuras da Privacidade (2). In: BAUMAN, Zygmunt. **44 Cartas do Mundo Líquido Moderno**. Tradução Vera Pereira. Rio de Janeiro: Zahar, 2011.

BOBBIO, Norberto. **Democracia e Segredo**. Tradução Marco Aurélio Nogueira. São Paulo: UNESP, 2015.

BENEVOLÊNCIA. In: DICIO, Dicionário Online de Português. Porto: 7Graus, 2021. Disponível em: <https://www.dicio.com.br/benevolencia/>. Acesso em: 15 set. 2021.

DONEDA, Danilo. Princípios de Proteção de dados Pessoais. In: DE LUCCA, Newton; SIMÃO FILHO, Adalberto; LIMA, Cíntia Rosa Pereira. **Direito & Internet III – tomo I: Marco Civil da internet (Lei n. 12.965/2014)**. São Paulo: Quartier Latin, 2015, p. 369-384.

DWORKIN, Ronald. **A Raposa e o Porco-Espinho: justiça e valor**. Tradução de Marcelo Brandão Cipolla. São Paulo: Martins Fontes, 2014.

FREITAS, Juarez. FREITAS, Thomas Bellini. **Direito e Inteligência Artificial: em defesa do humano**. Belo Horizonte: Fórum, 2020.

LEE, Kai-Fu. **Inteligência Artificial: como os robôs estão mudando o mundo, a forma como amamos, nos comunicamos e vivemos**. Tradução de Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019.

LÉVY, Pierre. **As Tecnologias da Inteligência: o futuro do pensamento na era da informática**. Tradução de Carlos Irineu da Costa. 2. ed. São Paulo: Editora 34, 2010.

LIMA, Cíntia Rosa Pereira; OLIVEIRA, Cristina Godoy Bernardo de; RUIZ, Evandro Eduardo Seron. Inteligência Artificial e Personalidade Jurídica: aspectos controvertidos. In: BARBOSA, Mafalda Miranda et al. **Direito Digital e Inteligência Artificial: diálogos**

entre Brasil e Europa. Indaiatuba, SP: Foco, 2021.

MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael; **Novas Perspectivas sobre Ética e Responsabilidade de Inteligência Artificial**. 2. ed. São Paulo: Thomson Reuters Brasil, 2020, p. 111-143.

MAIA, Paulo. Inteligência Artificial: uma visão ética. IN: Prado, Gustavo Daniel et al (org). **Fonte: tecnologia da Informação na Gestão Pública**. ed. 17. Belo Horizonte: Prodemge, 2017, p. 66-67. Disponível em: https://www.prodemge.gov.br/images/com_arismartbook/download/19/revista_17.pdf. Acesso em: 15 set. 2021.

MELLO, Alexandre Schmitt da Silva; DRESCH, Rafael de Freitas Valle. Breves reflexões sobre Livre-arbítrio, Autonomia e Responsabilidade Humana de Inteligência Artificial. In: BARBOSA, Mafalda Miranda et al. **Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa**. Indaiatuba, SP: Foco, 2021, p. 148.

MILARÉ, Édís. **Direito do Ambiente**. ed. 12. São Paulo: Thomson Reuters Brasil, 2020.

MULHOLLAND, Caitlin; FRAJHOF, Isabella. Entre Leis da Robótica e Ética: regulação para o adequado desenvolvimento da inteligência artificial. In: BARBOSA, Mafalda Miranda et al. **Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa**. Indaiatuba, SP: Foco, 2021, p. 65-80.

PASCAL, Blaise. **Pensamentos**. Tradução de Mario Laranjeira. São Paulo: Martins Fortes, 2000.

PEIXOTO, Fabiano Hartmann. **Inteligência Artificial e Direito: convergência ética e estratégica**. Curitiba: Alteridade, 2020.

PLATÃO. **A Justiça**. Tradução de Edson Bini. São Paulo: EDIPRO, 2016.

RODOTÀ, Stefano. **A Vida na Sociedade da Vigilância: a privacidade hoje**. Tradução de Danilo Doneda e Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

SANDEL, Michael J. **Justiça: o que é fazer a coisa certa**. Tradução de Heloísa Matias e Maria Alice Máximo. Rio de Janeiro: Civilização Brasileira, 2011.

SCHWAB, Klaus. **A Quarta Revolução Industrial**. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2019.

SICHMAN, Jaime Simão. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**. São Paulo, v. 101, n. 35, p. 37-49, jan./abr. 2021.

STEIBEL, Fabro; VICENTE, Victor Freitas; JESUS, Diego Santos Vieira de.

Possibilidades e Potenciais da Utilização da Inteligência Artificial. In: FRANZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito: ética, regulação e responsabilidade**. 2. ed. São Paulo: Thomson Reuters Brasil, 2020, p. 51-62.

VERUGGIO, Gianmarco. **Roboetica**: la nuova etica applicata per affrontare la rivoluzione robótica. Itália: Società psicoanalitica Italiana – SPIWEB. Disponível em: <https://www.spiweb.it/dossier/umani-robot-una-relazione-pericolosa-ottobre-2017/gianmarco-veruggio-roboetica-la-nuova-etica-applicata-per-affrontare-la-rivoluzione-robotica/>. Acesso em 05 de jul. 2021.

WEBB, Amy. **Os Nove Titãs da IA**: como os gigantes da tecnologia e suas máquinas pensantes podem subverter a humanidade. Rio de Janeiro: Alta Books, 2020.

Data da submissão: 28/07/2022

Data da primeira avaliação: 17/08/2022

Data da segunda avaliação: 19/10/2022

Data da aprovação: 19/10/2022