



**Rúbia Luana Carvalho Viegas Schmall**

**A supervisão humana das decisões  
autônomas de IA como instrumento de  
tutela da autonomia existencial.**

**Dissertação de Mestrado**

Dissertação apresentada à banca examinadora  
como exigência parcial para obtenção do título  
de Mestre em Direito Civil Contemporâneo e  
Prática Jurídica da Pontifícia Universidade Católica  
do Rio de Janeiro (PUC-Rio).

Orientadora: Prof.<sup>a</sup> Adriana Vidal de Oliveira.

Rio de Janeiro,  
maio de 2023.



**Rúbia Luana Carvalho Viegas Schmall**

**A supervisão humana das decisões  
autônomas de IA como instrumento de  
tutela da autonomia existencial.**

**Dissertação de Mestrado**

Dissertação apresentada à banca examinadora  
como exigência parcial para obtenção do título  
de Mestre em Direito Civil Contemporâneo e  
Prática Jurídica da Pontifícia Universidade Católica  
do Rio de Janeiro (PUC-Rio).

**Prof<sup>a</sup>. Adriana Vidal de Oliveira**

Orientadora

Departamento de Direito – PUC-Rio

**Prof<sup>a</sup>. Caitlin Sampaio Mulholland**

Departamento de Direito – PUC-Rio

**Prof. Carlos Affonso Pereira de Souza**

Departamento de Direito – PUC-Rio

Rio de Janeiro, 05 de maio de 2023.

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, da autora e da orientadora.

### **Rúbia Luana Carvalho Viegas Schmall**

Graduou-se como Bacharel em Direito na Faculdade de Direito da Universidade Federal da Bahia (UFBA) em 2001. Especializou-se em Direito Processual Civil e do Trabalho na Fundação Faculdade de Direito da Bahia em 2007. Graduou-se como Mestre em Direito na Universidade do Texas em Austin em 2011. Advogada inscrita na OAB desde 2001. Atua como advogada de uma empresa pública federal desde 2012.

Ficha Catalográfica.

Schmall, Rúbia Luana Carvalho Viegas

A supervisão humana das decisões autônomas de IA como instrumento de tutela da autonomia existencial / Rúbia Luana Carvalho Viegas Schmall ; orientadora: Adriana Vidal de Oliveira. – 2023.

125 f. : il. ; 30 cm

Dissertação (mestrado)—Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Direito, 2023.

Inclui bibliografia

1. Direito – Teses. 2. Inteligência artificial. 3. Decisões autônomas. 4. Autonomia existencial. 5. Supervisão humana. I. Oliveira, Adriana Vidal de. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Direito. III. Título.

CDD: 340

Para os meus pais, Wilma e Sady;  
para o amor da minha vida, Luiz Otávio;  
para o meu anjo de luz, Cloe.

## **Agradecimentos**

Sonhos não são realizados na solidão. Eles requerem parceria, incentivo, inspiração e apoio. Assim, agradeço aos meus pais, Wilma e Sady, por terem me ensinado a importância do conhecimento e o valor do trabalho; aos meus irmãos, Neca e Ronaldo, pela parceria e afeto; ao meu marido, Luiz Otávio, pelo incentivo, companheirismo e apoio incondicionais; à minha filha, Cloe, pela doçura e inspiração; à querida Conceição, pela ajuda imensurável; aos professores do mestrado, pelo valioso aprendizado; à Profa. Caitlin Sampaio Mulholland e ao Prof. Carlos Affonso Pereira de Souza, pela iluminação do caminho; à minha orientadora, Profa. Adriana Vidal de Oliveira, pela disponibilidade, auxílio, confiança e autonomia; aos meus colegas, pela ajuda e carinho; à minha psicanalista, Alessandra Tavares Silva, pela companhia nessa jornada. Obrigada a todos com amor e deferência.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## Resumo

SCHMALL, Rúbia Luana Carvalho Viegas. **A supervisão humana das decisões autônomas de IA como instrumento de tutela da autonomia existencial**. Orientadora: DE OLIVEIRA, Adriana Vidal. 2023. 125 p. Dissertação de Mestrado – Departamento de Direito, Pontifícia Universidade Católica do Rio de Janeiro.

A sociedade passa por um momento de intensas transformações pautadas no uso de novas tecnologias. Nesse contexto, a inteligência artificial (IA) assume o papel de protagonista gerando debates acerca dos limites da sua aplicação. Cada vez mais presente na vida dos indivíduos e instituições, a IA participa de processos de tomada de decisão que, não raro, têm reflexos em direitos fundamentais e questões existenciais dos indivíduos. Tendo em vista a busca de soluções pautadas na centralidade humana no uso da tecnologia, o presente estudo se propõe a investigar a supervisão humana como instrumento de tutela da autonomia existencial no contexto de tomada de decisões automatizadas por sistemas de IA. Por meio da análise da relação entre a autonomia da máquina e a autonomia existencial, o texto traça o caminho para a investigação dos princípios e dispositivos legais vigentes que legitimam a necessidade da supervisão humana sobre os resultados decisórios gerados por sistemas de IA de alto risco, bem como o tratamento do tema nas propostas legislativas em trâmite no Brasil e União Europeia que visam regular os usos e aplicações da IA.

## Palavras-chave

Inteligência artificial; decisões autônomas; autonomia existencial; supervisão humana.

## **Abstract**

SCHMALL, Rubia Luana Carvalho Viegas. **Human supervision of automated AI decisions as a legal tool for the protection of human autonomy**. Orientadora: DE OLIVEIRA, Adriana Vidal. 2023. 125 p. Dissertação de Mestrado – Departamento de Direito, Pontifícia Universidade Católica do Rio de Janeiro.

The society is going through a moment of intense transformations based on the use of new technologies. In this context, artificial intelligence (AI) takes on the role of protagonist, generating debates about the limits of its application. Increasingly present in the lives of individuals and institutions, AI participates in decision-making processes that, not infrequently, have an impact on fundamental rights and existential issues of individuals. In view of the search for solutions based on the human centrality in the use of technology, the present study proposes to investigate human supervision as an instrument to protect human autonomy in the context of automated decision-making by AI systems. Through the analysis of the relation between machine autonomy and existential autonomy, the text outlines the path for the investigation of the current legal principles and laws that legitimize the need for human supervision over the decision-making results generated by high-risk AI systems, as well as the treatment of the subject in the legislative proposals in progress in Brazil and the European Union that aim to regulate the uses and applications of AI.

## **Keywords**

Artificial intelligence; automated decisions; human autonomy; human supervision.

## Sumário

1 Introdução .....	1
2 Decisões autônomas e a autonomia da IA .....	6
2.1 Aspectos preliminares relevantes na conceituação, caracterização e regulação da IA. ....	6
2.1.1 Conceituando a “inteligência” da IA .....	6
2.1.2 Os insumos da IA e a sua regulação .....	9
2.1.3 Uso de probabilidade – inexistência de neutralidade – potencial para a ocorrência de discriminação .....	12
2.2 Processo autônomo de tomada de decisões por mecanismos de IA: funcionamento, finalidade e riscos .....	18
2.2.1 Ciclo de vida da IA – fases de utilização – controle <i>ex post</i> .....	19
2.2.2 O funcionamento da IA por meio de regras e padrões .....	24
2.2.3 O uso da tecnologia com centralidade no humano .....	26
2.2.4 Usurpação de papéis humanos – valor social – efeito substituição .....	30
2.3 A capacidade de tomada de decisões automatizadas conduz à autonomia da IA? .....	35
2.3.1 Conceito de singularidade tecnológica .....	36
2.3.2 Níveis de automação da IA: exemplo dos carros autônomos .....	39
2.3.3 Autonomia ou comportamento emergente? .....	41
3 “Existirmos: a que será que se destina?” .....	44
3.1 A existência e a autonomia humana .....	44
3.1.1 Existência digna, solidária e igual .....	46
3.2 A relevância da tutela da autonomia existencial no contexto de tomada de decisões autônomas .....	48
3.2.1 Projeto de vida e pluralidade .....	48
3.2.2 Direitos fundamentais e questões existenciais .....	51
3.2.3 A delegação da autonomia no uso da IA .....	54
3.3 A questão do controle .....	56
3.3.1 Vigilância, classificação e punição .....	56
3.3.2 Mediação e persuasão .....	63
3.4 Reflexos das decisões autônomas de IA na autonomia existencial – análise casuística. ....	66
3.4.1 Exemplo de discriminação de gênero por biometria facial – contexto fático .....	67



3.4.2 Funcionamento de sistemas de detecção e reconhecimento facial e reconhecimento automatizado de gênero (AGR) .....	69
3.4.3 Análise de dois estudos que tratam dos impactos dos sistemas de AGR nos indivíduos transgênero e não-binários .....	71
3.4.4 Abordagem das propostas legislativas sobre a identificação biométrica e categorização de pessoas singulares .....	74
 4 A supervisão humana como instrumento de tutela da autonomia existencial no contexto de utilização da IA .....	 76
4.1 Finalidade e forma de implementação da supervisão humana .....	76
4.2 A relevância do controle humano no uso da tecnologia.....	78
4.3 A importância da promoção de valores humanos no uso da IA .....	83
4.4 A supervisão humana nos textos das propostas legislativas em andamento.....	89
4.4.1 Abordagem .....	89
4.4.2 Formas de implementação e operação .....	92
4.5 Dispositivos legais vigentes que justificam a necessidade de implementação da supervisão humana nas decisões autônomas de IA ...	96
 5 Conclusão .....	 102
 6 Referências bibliográficas .....	 106

## Tabela de figuras

Figura 1 - Representação esquemática de um sistema de IA .....	22
Figura 2- Representação esquemática da centralidade humana na IA de confiança .....	28
Figura 3- Planta do panóptico de Jeremy Bentham .....	60
Figura 4- Interior da penitenciária de Stateville, EUA, século XX.....	60

*“Quase sem tocar no corpo, a guilhotina  
suprime a vida, tal como a prisão su-  
prime a liberdade ou uma multa retira  
os bens. Aplica a lei menos a um corpo  
real susceptível de dor do que a um su-  
jeito jurídico, detentor, entre outros, do  
direito de existir. Esta máquina devia ter  
a abstração da própria lei.”*

Michel Foucault.

## Introdução

Na letra da música “Cajuína”<sup>1</sup>, o cantor e compositor Caetano Veloso narra a história do encontro que teve com o pai de Torquato Neto, seu amigo, parceiro musical e letrista do “Tropicalismo”<sup>2</sup>, que havia cometido suicídio alguns anos antes, em 1972, aos 28 anos de idade<sup>3</sup>. Emocionado, em uma interação repleta de gestos e escassa em palavras, Caetano é consolado por aquele que havia perdido o único filho<sup>4</sup>.

O artista, que retrata em sucintos versos algumas das mazelas e belezas das relações humanas, inicia a composição com um questionamento - “Existirmos: a que será que se destina?”. Partindo desta premissa, o compositor aborda questões existenciais como a empatia, compaixão, solidariedade, igualdade, amor, laços familiares, saudade, amizade, afeto, imagem, depressão, vida, luto, experiências sensoriais e liberdade de expressão.

---

<sup>1</sup> “Existirmos: A que será que se destina?/ Pois quando tu me deste a rosa pequenina/ Vi que és um homem lindo e que se acaso a sina/ Do menino infeliz não se nos ilumina/ Tampouco turva-se a lágrima nordestina/ Apenas a matéria vida era tão fina/ E éramos olharmo-nos, intacta retina/A cajuína cristalina em Teresina” VELOSO, Caetano. **Cajuína**. LETRAS. Disponível em: <<https://www.lettras.mus.br/caetano-veloso/44704/>>. Acesso em: 22.11.2021.

<sup>2</sup> “O Tropicalismo foi um movimento de ruptura que sacudiu o ambiente da música popular e da cultura brasileira entre 1967 e 1968. Seus participantes formaram um grande coletivo, cujos destaques foram os cantores-compositores Caetano Veloso e Gilberto Gil, além das participações da cantora Gal Costa e do cantor-compositor Tom Zé, da banda Mutantes, e do maestro Rogério Duprat. A cantora Nara Leão e os letristas José Carlos Capinan e Torquato Neto completaram o grupo, que teve também o artista gráfico, compositor e poeta Rogério Duarte como um de seus principais mentores intelectuais.” DE OLIVEIRA, Ana. **Tropicalia**: Identifisignificados movimento. Disponível em: <<http://tropicalia.com.br/identifisignificados/movimento>>. Acesso em: 23.11.2021.

<sup>3</sup> CORREIO BRAZILIENSE. Publicado em: 08.11.2014. **Morto aos 28 anos, o piauiense Torquato Neto completaria 70 anos no domingo**. Disponível em: <[https://www.correiobraziliense.com.br/app/noticia/diversao-e-arte/2014/11/08/interna\\_diversao\\_arte,456582/morto-aos-28-anos-o-piauiense-torquato-neto-completaria-70-anos-no-domingo.shtml](https://www.correiobraziliense.com.br/app/noticia/diversao-e-arte/2014/11/08/interna_diversao_arte,456582/morto-aos-28-anos-o-piauiense-torquato-neto-completaria-70-anos-no-domingo.shtml)>. Acesso em: 23.11.2021.

<sup>4</sup> Em entrevista ao programa Altas Horas, Caetano Veloso conta que ao tomar conhecimento da morte de Torquato Neto não havia chorado. Cerca de dois anos depois, em visita a Teresina, foi levado pelo pai do amigo à sua casa. Vendo-se cercado por fotos, lembranças e pela presença de um homem sozinho, que havia perdido seu único filho de forma trágica, e cuja esposa estava hospitalizada, foi tomado de emoção. Debulhado em lágrimas, Caetano foi consolado por um homem calmo, doce e gentil, que lhe ofereceu cajuína, uma rosa menina, além do olhar silencioso em “intacta retina”. O episódio, que se tornou poesia, é um bom exemplo da dimensão e do alcance do afeto e da solidariedade, dos diferentes papéis que o ser humano desempenha no exercício da sua autonomia e nas relações interpessoais, de como a percepção e compreensão humana dos fatos e situações está muito além do que hoje as máquinas podem oferecer. VELOSO, Caetano. **Caetano Veloso conta a história da Música “Cajuína”**. Globoplay. Altas Horas. Serginho Groisman. Disponível em: <<https://globoplay.globo.com/v/3119899/>>. Publicado em: 01.12.2014. Acesso em: 22.11.2021.

Na toada da canção, o presente estudo utilizará a mesma indagação para abordar a necessidade de tutela do humano e da sua autonomia existencial em meio a uma intensa transformação tecnológica que vem sendo designada como a “quarta revolução industrial”<sup>1</sup> e na qual as máquinas assumem o papel de protagonismo.

Seja na mídia, em debates sociais e acadêmicos, ou em conversas entre amigos e familiares, os usos e aplicações da inteligência artificial (IA) são assunto constante. Pessoas mesmerizadas com a “inteligência” e praticidade oferecidas por sistemas de IA, tais como o ChatGPT<sup>2</sup>, especulam os limites e alcances da tecnologia, ao passo em que vislumbram hipóteses de ocorrência do ostracismo humano ao longo do tempo<sup>3</sup>.

A velocidade sem precedentes com que os avanços tecnológicos têm ocorrido dificulta o processo reflexivo pela sociedade. Os indivíduos são tomados por uma avalanche de novidades constantes e sucessivas que os leva à busca de mais informação e incessante atualização. O tempo que poderia ser dedicado à reflexão

---

<sup>1</sup> “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.” DONEDA, Danilo; MENDES, Laura Schertel; SOUZA, Carlos Affonso Pereira de; ANDRADE, Norberto Nuno Gomes de. **Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal**. Pensar, Fortaleza, v. 23, n. 4, p. 1-17, out./dez. 2018, p. 10.

<sup>2</sup> “We’ve trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.” OPENAI. **ChatGPT: optimizing language models for dialogue**. Publicado em: 30.11.2022. Disponível em: <<https://openai.com/blog/chatgpt/>>. Acesso em: 16.02.2023.

<sup>3</sup> “O ChatGPT é bom ou ruim? Depende. Por exemplo, os alunos podem usar essa tecnologia para colar nas provas ou para que ela faça seus deveres de casa. O site norte-americano Atlantic já publicou um artigo prevendo o fim da redação escolar devido à habilidade do ChatGPT em escrever textos - muitos alunos vão deixar de investir tempo de estudo ao usar o robô. Olhando por esse lado, essa tecnologia tende a deixar pessoas mais preguiçosas e “menos inteligentes”. A indústria criativa também pode sofrer danos com a popularização do ChatGPT. É fácil imaginar como essa tecnologia será reaproveitada para automatizar e, eventualmente, substituir a força de trabalho humana na mídia digital. Já do ponto de vista corporativo, é possível ver muita utilidade do ChatGPT. Aplicações comerciais para atendimento e suporte ao cliente podem dar um grande salto com esse atendimento mais próximo ao do ser humano.” MACHADO, Simone. **ChatGPT: o que é e como funciona robô de conversa que vem gerando polêmica**. TILT UOL. INTELIGÊNCIA ARTIFICIAL. Publicado em: 15.12.2022. Disponível em: <<https://www.uol.com.br/tilt/noticias/redacao/2022/12/15/o-que-e-chatgpt-a-plataforma-de-ia-que-vem-gerando-polemica.htm>>. Acesso em: 16.02.2023.

sobre a condição humana atual é, muitas vezes, utilizado na adaptação para o uso de novas ferramentas tecnológicas em um ciclo infindável e exaustivo de conectividade.

Os processos de *machine* e *deep learning* deram às máquinas a capacidade de emular a inteligência humana, e por meio da delegação, que pode se dar de forma consciente ou não, tomar decisões sobre os humanos, inclusive sobre as suas questões existenciais. “Somos livres para escolher os rumos das nossas vidas? Há limites para a liberdade de existir?”<sup>4</sup>

No intuito de buscar soluções pautadas na centralidade humana nos usos e aplicações da tecnologia e nas relações jurídicas desta decorrentes, o presente estudo se propõe a analisar a supervisão humana como mecanismo de tutela da autonomia existencial no contexto de tomada de decisões automatizadas por mecanismos de IA.

Para tanto, no primeiro capítulo, serão analisados: (i) alguns atributos relevantes na caracterização da IA; (ii) a forma de funcionamento do processo automatizado de tomada de decisões; (iii) a importância da centralidade no humano; (iv) a transferência de papéis entre humanos e tecnologia; (v) a autonomia da IA.

No segundo capítulo, serão abordadas: (i) a existência e autonomia humana do ponto de vista filosófico e jurídico; (ii) a relevância da tutela da autonomia existencial no uso da IA; (iii) a delegação de poder no processo decisório; (iv) a questão do controle nas relações sociais e jurídicas e seus reflexos na aplicação da IA.

No terceiro e último capítulo de desenvolvimento do presente estudo, serão analisados os fundamentos para utilização da supervisão humana como instrumento da tutela da autonomia existencial, quais sejam: (i) os princípios que enfatizam a relevância do controle humano e promoção de valores humanos no uso da tecnologia; (ii) as propostas legislativas em andamento no Brasil e União Europeia que visam regular os usos e aplicações da IA; (iii) os dispositivos legais vigentes que

---

<sup>4</sup> VIVEIROS DE CASTRO, Thamís Dalsenter. **A função da cláusula de bons costumes no direito civil e a teoria tríptica da autonomia privada existencial**. Revista Brasileira de Direito Civil – RBDCivil | Belo Horizonte, vol. 14, p. 99-125, out./dez. 2017a, p. 99.

amparam a supervisão humana das decisões automatizadas que digam respeito a questões existenciais e direitos fundamentais.

Cumpre esclarecer que este trabalho não tem por escopo promover um estudo sobre a viabilidade de regulação da IA ou direito comparado. Contudo, por não haver previsão legal específica para a supervisão humana da IA, nem tampouco jurisprudência acerca do tema, optou-se por enriquecer o seu conteúdo por meio de uma análise pontual das propostas legislativas em andamento no Brasil e na União Europeia.

No Brasil, está em tramitação no Senado Federal o Projeto de Lei nº 21, de 2020 (PL nº 21/2020)<sup>5</sup>. O texto aprovado na Câmara dos Deputados e sobre o qual ainda não houve deliberação do Senado, não traz explicitamente a supervisão humana como instrumento de proteção da pessoa humana, mas apresenta fundamentos e princípios que podem justificar a utilização do referido mecanismo de tutela que serão abordados ao longo do terceiro capítulo.

Importa ressaltar que há indícios de uma possível alteração substancial no texto do PL nº 21/2020, já que, em 09.12.2022, foi publicado no Diário do Senado Federal o relatório apresentado por uma comissão de juristas instituída para subsidiar a elaboração de uma minuta de substitutivo aos Projetos de Lei nº 5.051, de 2019; nº 21, de 2020; e nº 872, de 2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil<sup>6</sup>.

Essa minuta de substitutivo do PL nº 21/2020, que foi publicada, mas ainda não tramitou no plenário do Senado, é um documento minucioso e abrangente que, diferentemente do texto aprovado na Câmara dos Deputados, traz, expressamente, a exigência da supervisão humana de sistemas que classifica como de alto risco.

---

<sup>5</sup> BRASIL. **Projeto de lei nº 21, de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e aplicação da inteligência artificial no Brasil, e dá outras providências. Publicação: 30.09.2021. Disponível em: <<https://legis.senado.leg.br/sdleg-getter/documento?dm=9023442&ts=1676314334096&disposition=inline>>. Acesso em: 16.02.2023.

<sup>6</sup> DIÁRIO DO SENADO FEDERAL. **Relatório Final Da Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência Artificial**. Publicado em: 09.12.2022. Disponível em: < <https://legis.senado.leg.br/diarios/ver/111533?sequencia=3>>. Acesso em: 11.02.2023.

Da leitura da referida minuta, depreende-se que o seu texto foi inspirado, dentre outras fontes, na proposta legislativa da União Europeia, que visa estabelecer regras harmonizadas em matéria de inteligência artificial, a EUR-Lex 52021PC0206<sup>7</sup>, encontra-se em fase de discussão no Conselho Europeu e, do mesmo modo, prevê de forma inequívoca a necessidade de supervisão humana de sistemas de IA que classifica como de risco elevado<sup>8</sup>.

Em 03.05.2023, o texto da minuta de substitutivo do PL nº 21/2020 foi apresentada ao plenário do Senado Federal na forma do Projeto de Lei nº 2338/2023<sup>9</sup> que dispõe sobre o uso da inteligência artificial e aguarda despacho inicial daquela casa legislativa. Portanto, as referências que o presente trabalho faz ao texto da minuta de substitutivo do PL nº 21/2020 aplicam-se ao texto do PL nº 2338/2023 que começa a tramitar e tem conteúdo idêntico.

A EUR-Lex 52021PC0206 foi proposta em 22.04.2021, tendo sido antecida por uma série de estudos, realizados desde abril de 2018<sup>10</sup>, que resultaram em um conjunto de documentos e orientações para o uso ético da IA. Alguns destes achados, que são referência na área de uso ético da IA, foram utilizados na fase de pesquisa e serão citados ao longo deste trabalho.

Ressalte-se que parte dos documentos que instruíram a elaboração da EUR-Lex 52021PC0206 somente fora disponibilizada em inglês, eis porque há, no presente texto, referências em ambas as línguas (inglês e português). Outrossim,

---

<sup>7</sup> EUR-Lex 52021PC0206. **Proposta de regulamento do parlamento europeu e do conselho que estabelece regras harmonizadas em matéria de inteligência artificial (regulamento inteligência artificial) e altera determinados atos legislativos da União.** 2021. Disponível em: <[https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-1aa75ed71a1.0004.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-1aa75ed71a1.0004.02/DOC_1&format=PDF)>. Acesso em: 15.02.2023.

<sup>8</sup> A escolha da proposta legislativa da União Europeia como objeto de análise se deu por diferentes razões: (i) pelo fato dela estar em preparação desde 2018 e tramitação desde 2021; (ii) pela identidade de sistema jurídico com Brasil (sistema germano-românico); (iii) pela disponibilização dos principais documentos em português, já que Portugal, nossa pátria mãe, faz parte do conselho da União Europeia; (iv) pelo fato de que a proposta se apresenta como inspiração para a minuta de substitutivo da proposta brasileira apresentada recentemente no Senado Federal.

<sup>9</sup> BRASIL. **Projeto de lei nº 2338, de 2023.** Dispõe sobre o uso da Inteligência Artificial. Publicação: 03.05.2023. Disponível em: <<https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>>. Acesso em: 05.05.2023.

<sup>10</sup> Em abril de 2018, por meio do comunicado Artificial Intelligence for Europe, a Comissão Europeia abordou a necessidade de assegurar o enquadramento ético e legal do uso da IA em consonância com os valores da União Europeia e em alinhamento com os direitos fundamentais. Informações: COMISSÃO EUROPEIA. **Inception impact assessment: proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence.** Disponível em: <[https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=PI\\_COM:Ares\(2020\)3896535&from=EN](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=PI_COM:Ares(2020)3896535&from=EN)>. Acesso em: 20.12.2021.



algumas expressões ou palavras dos trechos citados em português seguem os usos e costumes da ortografia de Portugal (país integrante da União Europeia), que se difere, por vezes, daquela utilizada corriqueiramente no Brasil.

Além das pretensões legislativas mencionadas, há tratados de direitos humanos e dispositivos constitucionais e infraconstitucionais que podem justificar a tutela da autonomia humana por meio da supervisão humana na delegação cada vez mais comum de processos decisórios para a IA.

Decisões que digam respeito à saúde, moradia, trabalho, educação, segurança pública, acesso a serviços essenciais, identidade ou relacionamentos interpessoais devem ser integralmente atribuídas a robôs?<sup>11</sup> No atual contexto tecnológico, social e jurídico, a máquina está apta a ser o juiz do humano? Estes são os questionamentos que se buscará responder ao longo deste trabalho.

---

<sup>11</sup> “Você se submeteria a um procedimento cirúrgico recomendado por um sistema baseado em Inteligência artificial? Se tivesse dúvidas sobre a recomendação, preferiria que a segunda opinião fosse dada por um humano ou por outra máquina? Caso concordasse com a cirurgia, seria melhor que ela fosse realizada por um médico ou por um robô especializado?” SOUZA, Carlos Affonso Pereira de; OLIVEIRA, Jordan Vinicius de. **Sobre os ombros de robôs?** A inteligência artificial: entre fascínios e desilusões. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito: ética, regulação e responsabilidade**. São Paulo: Ed. Revista dos Tribunais, 2020, p. 63-79.

## 2.

### Decisões autônomas e a autonomia da IA.

#### 2.1

#### Aspectos preliminares relevantes na conceituação, caracterização e regulação da IA.

##### 2.1.1

##### Conceituando a “inteligência” da IA.

A tecnologia alcançou um ciclo de desenvolvimento e consequente obsolescência muito dinâmico, tornando difícil a tarefa de se estabelecer um conceito definitivo para o termo “inteligência artificial” (IA).

A exemplo, a tecnologia que é posta em uso e se torna habitual deixa de ser considerada uma IA logo que uma nova tecnologia emerge. É ao que STONE et al. (2016, p. 10) se refere como “*AI effect*” ou “*odd paradox*”<sup>12</sup>.

Deste modo, a IA deixa de ser “inteligente” assim que uma nova tecnologia, mais avançada, que realiza as mesmas tarefas de forma mais eficiente, abrangente e/ou rápida surge. A “recém-nascida” IA herda o atributo “inteligência” outrora pertencente à máquina que se tornou obsoleta.

Assim, a efemeridade da “inteligência” da IA reflete o “estranho paradoxo” no qual a própria existência da máquina, considerada um ente inteligente, propicia que seja criada uma tecnologia mais avançada que a destitua da categoria de IA, em contraponto à inteligência do humano que, em geral, é potencializada em compasso com o tempo de vida e experiências acumuladas.

---

<sup>12</sup> “In fact, the field of AI is a continual endeavor to push forward the frontier of machine intelligence. Ironically, AI suffers the perennial fate of losing claim to its acquisitions, which eventually and inevitably get pulled inside the frontier, a repeating pattern known as the ‘AI effect’ or the ‘odd paradox’—AI brings a new technology into the common fold, people become accustomed to this technology, it stops being considered AI, and newer technology emerges. The same pattern will continue in the future. AI does not “deliver” a life-changing product as a bolt from the blue. Rather, AI technologies continue to get better in a continual, incremental way.” STONE, Peter, et al. **Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence**. In: **One hundred year study on artificial intelligence**: report of the study panel. Stanford University, 2016, p. 10.

Não obstante a consolidação do termo “inteligência” para se referir ao processo cognitivo desenvolvido pela IA, que leva em consideração a sua racionalidade<sup>13</sup>, capacidade de aprendizado e de resolução de problemas<sup>14</sup>, essa cognição artificial ainda não alcançou o nível do intelecto humano<sup>15</sup>.

Para RUSSEL (2019, p. 18-19), antes de entendermos como criar uma inteligência, ajudaria entender o que é inteligência. Segundo ele, tal resposta não é encontrada em testes de QI ou mesmo testes de Turing, mas em uma simples relação entre o que percebemos, o que queremos e o que fazemos. Assim, um ente é inteligente na medida em que o que ele faz é feito para alcançar o que quer, considerando o que percebe ou tem consciência sobre<sup>16</sup>.

Sob outro prisma, a inteligência artificial pode ser definida tanto como a atividade destinada a desenvolver máquinas inteligentes<sup>17</sup>, quanto como a

---

<sup>13</sup> “O termo «IA» contém uma referência explícita ao conceito de inteligência. No entanto, uma vez que a inteligência (tanto nas máquinas como nos seres humanos) é um conceito vago, ainda que longamente estudado por psicólogos, biólogos e neurocientistas, os investigadores de IA utilizam sobretudo o conceito de racionalidade. Este refere-se à capacidade de escolher a melhor ação a adotar para atingir um determinado objetivo, tendo em conta certos critérios que se devem otimizar e os recursos disponíveis. Claro que a racionalidade não é o único ingrediente do conceito de inteligência, mas constitui uma parte importante do mesmo.” GPAN IA. **Uma definição de IA**: principais capacidades e disciplinas científicas. Comissão europeia: 2019b, p. 3. Disponível em: <<https://digital-strategy.ec.europa.eu/pt/node/1950>>. Acesso em: 27.10.2022.

<sup>14</sup> “Notably, the characterization of intelligence as a spectrum grants no special status to the human brain. But to date human intelligence has no match in the biological and artificial worlds for sheer versatility, with the abilities ‘to reason, achieve goals, understand and generate language, perceive and respond to sensory inputs, prove mathematical theorems, play challenging games, synthesize and summarize information, create art and music, and even write histories.’ This makes human intelligence a natural choice for benchmarking the progress of AI. It may even be proposed, as a rule of thumb, that any activity computers once performed should be counted as an instance are able to perform and people of intelligence. But matching any human ability is only a sufficient condition, not a necessary one. There are already many systems that exceed human intelligence, at least in speed, such as scheduling the daily arrivals and departures of thousands of flights in an airport.” STONE et al., 2016, p. 13.

<sup>15</sup> “One important caveat: various authors have imagined a world in which robots or software achieve, or at any rate claim, a human-like consciousness. robots or software achieve, or at any rate claim, a human-like consciousness. Little in the literature gives me confidence that artificial intelligence will approximate human intelligence in the foreseeable future. There are analytic and technical reasons to believe robots will never think like people.” CALO, Ryan. **Robotics and the lessons of cyberlaw**. California Law Review, Berkeley, v. 103, n. 3, p. 513-563, jun. 2015, p. 528.

<sup>16</sup> “Before we can understand how to create intelligence, it helps to understand what it is. The answer is not to be found in IQ tests, or even in Turing tests, but in a simple relationship between what we perceive, what we want, and what we do. Roughly speaking, an entity is intelligent to the extent that what it does is likely to achieve what it wants, given what it has perceived.” RUSSEL, Stuart. **Human compatible**: artificial intelligence and the problem of control. New York: Viking/Penguin Random House LLC, 2019, p. 18-19.

<sup>17</sup> “Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.”. NILSSON, Nils J. **The quest for artificial intelligence**: a history of ideas and achievements. Cambridge, UK: Cambridge University Press, 2010, apud STONE et al., 2016, p. 10.

capacidade de determinado sistema de simular a habilidade cognitiva humana<sup>18</sup>. Como dito por WANG (2008, p. 1), o termo “inteligência artificial” tem diferentes significados para diferentes públicos<sup>19</sup>.

Para o presente estudo, interessa a conceituação e caracterização de sistemas de inteligência artificial<sup>20</sup>, havendo critérios que possibilitam a identificação de um sistema computacional que visa reproduzir o atributo humano denominado “inteligência”<sup>21</sup>.

---

<sup>18</sup> “[...] IA é uma expressão polissêmica. É o ramo da ciência destinado ao estudo de máquinas e sistemas para realizar atividades de modo autônomo. É também qualquer sistema computacional dotado de capacidades que imitam as habilidades intelectivas do humano na execução de tarefas, auxiliado por tecnologias variadas, que envolvem algoritmos. Os algoritmos são uma série de instruções constituídas por fórmulas matemáticas, operações e tratamentos estatísticos que programam a execução de tarefas por uma unidade operacional (qualquer dispositivo tecnológico como, por exemplo, um computador) para, em curto espaço de tempo e com elevado grau de precisão, alcançar um determinado resultado.” SOARES, 2021, p. 45-46.

<sup>19</sup> “It is well known that the phrase ‘Artificial Intelligence (AI)’ means different things to different people.” WANG, Pei. **What do you mean by “AI”?** AGI, vol. 171, 2008, p. 1.

<sup>20</sup> Na definição do art. 3º, 1 da EUR-Lex 52021PC0206: “Sistema de inteligência artificial» (sistema de IA), um programa informático desenvolvido com uma ou várias das técnicas e abordagens enumeradas no anexo I, capaz de, tendo em vista um determinado conjunto de objetivos definidos por seres humanos, criar resultados, tais como conteúdos, previsões, recomendações ou decisões, que influenciam os ambientes com os quais interage.” Na definição do art. 2º do PL nº 21/2020, referente ao texto aprovado na Câmara dos Deputados: “Para os fins desta Lei, considera-se sistema de inteligência artificial o sistema baseado em processo computacional que, a partir de um conjunto de objetivos definidos por humanos, pode, por meio do processamento de dados e de informações, aprender a perceber e a interpretar o ambiente externo, bem como a interagir com ele, fazendo previsões, recomendações, classificações ou decisões, e que utiliza, sem a elas se limitar, técnicas como: I – sistemas de aprendizagem de máquina (*machine learning*), incluída aprendizagem supervisionada, não supervisionada e por reforço; II – sistemas baseados em conhecimento ou em lógica; III – abordagens estatísticas, inferência bayesiana, métodos de pesquisa e de otimização.”. Na definição do art. 4º da minuta apresentada ao Plenário do Senado Federal em 08.12.2022 pela comissão de juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial: “I – sistema de inteligência artificial: sistema computacional, com graus diferentes de autonomia, desenhado para inferir como atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões que possam influenciar o ambiente virtual ou real.”

<sup>21</sup> “Though people have different opinions on how to accurately define ‘Artificial Intelligence’, on a more general level they do agree on what this field is about. Human beings differ from animals and machines mainly in their mental, or cognitive, ability, which is commonly called ‘intelligence’, and AI is the attempt to reproduce this ability in computer systems. This vague consensus sets important constraints on how AI should be defined.” WANG, 2008, p. 1.

## 2.1.2

### Os insumos da IA e a sua regulação.

Segundo MULHOLLAND (2020, p. 330), os dados e os algoritmos são os “insumos da IA”<sup>22</sup>. Logo, a IA é formada pela soma de comandos estabelecidos por meio de uma linguagem numérica, os algoritmos, que são fórmulas matemáticas<sup>23</sup>, e um grande volume de informações que podem ser colhidas de forma estruturada, como um banco de dados<sup>24</sup>, ou de forma não estruturada, a exemplo da captação da imagem de pessoas singulares.

No que se refere aos algoritmos, como ensina FRAZÃO (2019, p. 341), “são verdadeiras caixas pretas”, pois os usuários (incluindo empresas e/ou o Estado, e consumidores finais e/ou cidadãos) não têm acesso ao seu código e comandos neles contidos. Somente os desenvolvedores detêm a informação, que é protegida sob a salvaguarda do sigilo comercial e industrial<sup>25</sup>.

---

<sup>22</sup> “Um sistema de Inteligência Artificial é formado por uma série de técnicas algorítmicas que utilizam e gerenciam dados em grande escala (*big data*) com o objetivo de gerar conclusões – por meio de inferências probabilísticas – baseadas no tratamento daqueles dados. Quanto mais dados forem inseridos, absorvidos ou tratados pela IA, maior é a capacidade de ‘racionalização’ e processamento desses dados e, por consequência, melhores serão os resultados obtidos, beneficiando – em tese – os usuários de tal tecnologia. Os dados e os algoritmos são, portanto, os insumos+ da IA, sem os quais seria inviável desenvolver a tecnologia. Com base na qualidade dos algoritmos e dos dados coletados, será possível proporcionar experiências adequadas para as pessoas que utilizam a IA para as mais diversas funções.” MULLHOLLAND, Caitlin. **Responsabilidade civil e processos decisórios autônomos em sistemas de inteligência artificial (ia)**: autonomia, imputabilidade e responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito**: ética, regulação e responsabilidade. São Paulo: Ed. Revista dos Tribunais, 2020, p. 330-331.

<sup>23</sup> “Algoritmos são fórmulas matemáticas. Basicamente, série de instruções colhidas de símbolos e signos que são solucionados por microprocessadores, gerando novas fórmulas, em ciclo constante de *inputs* e *outputs*, com dados que são recebidos, processados pelo algoritmo, e devolvidos como resultado do processamento. Não se confundem, portanto, com robôs – que correspondem, no máximo, ao corpus no qual se realiza todo o processamento descrito, ou, como anota Carlos Goettebauer, ‘embora nem todo algoritmo seja um programa de computador, todo programa de computador pode ser definido como algoritmo’.” FALEIROS JÚNIOR, José Luiz de Moura. **A evolução da inteligência artificial em breve retrospectiva**. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Fellipe Braga Neto, et al. **Direito digital e inteligência artificial**: diálogos entre Brasil e Europa. Indaiatuba: Editora Foco, 2021, p. 20.

<sup>24</sup> “Um banco de dados é uma coleção de dados relacionados. Os dados são fatos que podem ser gravados e que possuem um significado implícito. Por exemplo, considere nomes, números telefônicos e endereços de pessoas que você conhece. Esses dados podem ter sido escritos em uma agenda de telefones ou armazenados em um computador, por meio de programas como o Microsoft Access ou Excel. Essas informações são uma coleção de dados com um significado implícito, conseqüentemente, um banco de dados.” ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. Rev. tec. Luis Ricardo Figueiredo. São Paulo: Pearson Addison Wesley, 2005, p. 2.

<sup>25</sup> A LGPD, ao tratar do princípio da transparência, estabelece a proteção aos segredos comercial e industrial: “Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: [...] VI - transparência: garantia, aos titulares, de informações claras, precisas

Quanto aos dados, é importante notar que a IA não se restringe a dados pessoais, pois os dados que a alimentam podem abranger também dados de natureza pública, a exemplo: dados históricos, geográficos, meteorológicos e/ou demográficos.

Portanto, a legislação existente em vigor que regula o tratamento de dados pessoais, tanto a brasileira (a Lei Geral de Proteção de Dados Pessoais - LGPD<sup>26</sup>), quanto a europeia (a *General Data Protection Regulation* - GDPR)<sup>27</sup>, apesar da sua relevância para a questão em foco, não abrange os usos e aplicações da IA em toda a sua extensão.

Contudo, cumpre esclarecer que, do ponto de vista regulatório, a fiscalização das decisões autônomas de IA se iniciou com a implementação da GDPR, que em seus artigos 13, 14, 15 e 22 trata do perfilamento de indivíduos, sedimenta o direito à explicação, e os princípios da transparência e da responsabilidade algorítmica (*algorithmic accountability*)<sup>28</sup>.

A LGPD seguiu a mesma linha, a exemplo do art. 20<sup>29</sup>, *caput*, que prevê o direito à explicação por meio da revisão de decisões autônomas “incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade”.

---

e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial.”

<sup>26</sup> Lei nº 13.709, de 14 de agosto de 2018.

<sup>27</sup> Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016.

<sup>28</sup> “There are four Articles of the GDPR that specifically address algorithmic decision-making. Article 22 of the GDPR addresses ‘automated individual decision-making, including profiling.’ Articles 13, 14, and 15 each contain transparency rights around automated decision-making and profiling.” KAMINSKI, Margot E. **The right to explanation, explained.** Berkeley Tech. LJ 34: 189, 2019, p. 4.

<sup>29</sup> Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. § 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial. § 2º Em caso de não oferecimento de informações de que trata o § 1º deste artigo baseado na observância de segredo comercial e industrial, a autoridade nacional poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais.

Os parágrafos 1º e 2º do referido artigo, por sua vez, primam pelo princípio da transparência, também previsto no art. 6º, VI. Já o art. 6º, X<sup>30</sup> e o art. 42, *caput* da LGPD<sup>31</sup> imputam a responsabilidade do “controlador ou operador que, em razão do exercício de atividade de tratamento de dados pessoais, causar a outrem dano” incluindo a *accountability* no panorama nacional.

Deste modo, as legislações citadas, já em vigor, somadas a outros dispositivos constitucionais e infraconstitucionais existentes no cenário nacional<sup>32</sup>, importam para o entendimento do panorama regulatório dos processos decisórios autônomos de IA.

Todavia, as propostas legislativas em andamento na Europa e Brasil que visam regular especificamente o uso e aplicação da IA, implementam mecanismos de escrutínio mais específicos com o fito de remediar a falta de transparência e opacidade dos algoritmos, especialmente no que diz respeito à preservação da autonomia humana como será visto no capítulo pertinente.

---

<sup>30</sup> Art. 6º [...] X - responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

<sup>31</sup> Art. 42. O controlador ou o operador que, em razão do exercício de atividade de tratamento de dados pessoais, causar a outrem dano patrimonial, moral, individual ou coletivo, em violação à legislação de proteção de dados pessoais, é obrigado a repará-lo.

<sup>32</sup> “[...] quando se trata de economia movida a dados, o que está em jogo é muito mais do que a privacidade dos usuários. É o próprio futuro das pessoas que pode estar sendo definido pelos algoritmos, sem que se tenha a possibilidade de conhecer e criticar os dados e correlações que alimentam seus processos decisórios. Sob essa perspectiva, para adequado endereçamento do problema, talvez não seja suficiente uma lei de proteção de dados, embora esta seja certamente uma das providências mais importantes nessa seara. Há que se acionar igualmente o Direito Constitucional, o Direito da Concorrência e o Direito do Consumidor para se avaliar como tais dados estão sendo efetivamente utilizados e se tal utilização está sendo compatível com as garantias e direitos dos cidadãos e dos consumidores.” FRAZÃO, Ana. **Plataformas digitais, big data e riscos para os direitos da personalidade**. In: TEPEDINO, Gustavo; MENEZES, Joyceane Bezerra de (Coord.). **Autonomia privada, liberdade existencial e direitos fundamentais**. Belo Horizonte: Fórum, 2019, p. 344.

### 2.1.3

#### Uso de probabilidade - inexistência de neutralidade – potencial para a ocorrência de discriminação.

A IA funciona lastreada em probabilidade<sup>33</sup> resultante do tratamento dos dados utilizados na sua formulação. Portanto, o processo de tomada de decisões por tais mecanismos depende da qualidade dos dados escolhidos<sup>34</sup>, o que torna o procedimento sujeito às falhas inerentes aos estudos estatísticos.

A depender da escolha dos dados e dos comandos iniciais determinados por meio da formulação do algoritmo, os resultados alcançados podem não refletir a realidade ou apresentar distorções, que, nas suas mais graves versões, geram discriminação<sup>35</sup>.

Assim, a possibilidade de falha da IA no processo de tomada de decisões, afasta a ideia de objetividade e verdade absoluta nos resultados por ela produzidos<sup>36</sup>, colocando em xeque o conceito de neutralidade que povoa o imaginário

---

<sup>33</sup> “The combination of probability theory with an expressive formal language is a fairly new subfield of AI, often called probabilistic programming. Several dozen probabilistic programming languages, or PPLs, have been developed, many of them deriving their expressive power from ordinary programming languages rather than first-order logic. All PPL systems have the capacity to represent and reason with complex, uncertain knowledge. Applications include Microsoft’s TrueSkill system, which rates millions of video game players every day; models for aspects of human cognition that were previously inexplicable by any mechanistic hypothesis, such as the ability to learn new visual categories of objects from single examples;5 and the global seismic monitoring for the Comprehensive Nuclear-Test-Ban Treaty (CTBT), which is responsible for detecting clandestine nuclear explosions.” RUSSEL, 2019, p. 198.

<sup>34</sup> “O viés de seleção se refere à prática de escolher os dados de forma seletiva — conscientemente ou inconscientemente —, de modo que isso leva a uma conclusão que é enganosa ou efêmera.” BRUCE, Peter; BRUCE, Andrew. **Estatística prática para cientistas de dados: 50 conceitos essenciais**. Trad. Luciana Ferraz. Rio de Janeiro: Alta Books, 2019, 2. Recurso eletrônico. Item viés de seleção, n.p.

<sup>35</sup> “Com efeito, não se tem nem mesmo como avaliar o potencial danoso de algoritmos que se baseia em dados e correlações sigilosos e sem qualquer transparência, que podem ser incorretos ou falsos, e que ainda podem ser utilizados para lastrear correlações que não correspondem a causalidades e, o que é mais grave, a reproduzir correlações que podem ser frutos de discriminações e uma série de injustiças da vida social.” FRAZÃO, 2019, p. 343.

<sup>36</sup> “Mesmo nas melhores circunstâncias, a análise estatística raramente revela ‘a verdade’. Em geral construímos um caso circunstancial baseado em dados imperfeitos. Como resultado, há numerosas razões para que indivíduos intelectualmente honestos discordem acerca de resultados estatísticos ou suas implicações. No nível mais básico, podemos discordar sobre a questão que está sendo respondida. Fãs de esportes discutirão por toda a eternidade sobre quem é ‘o melhor jogador de beisebol de todos os tempos’ porque não há definição objetiva de ‘melhor’. Estatísticas descritivas rebuscadas podem fornecer informações sobre essa questão, mas jamais a responderão em definitivo. Como será ressaltado no próximo capítulo, questões mais significativas socialmente viram presas do mesmo desafio básico. O que está acontecendo com a saúde econômica da classe média americana?



popular e leva à crença de que robôs não erram, já que executariam exatamente os comandos aos quais são direcionados.

Para o filósofo Martin Heidegger, a forma como nos relacionamos com a tecnologia nos faz seus prisioneiros e a ideia de neutralidade nos torna cegos para o que representa a sua essência, pois formulamos, aceitamos e lidamos com ela de maneira equivocada<sup>37</sup>.

Essa impressão distorcida se intensifica com a difusão do uso da inteligência artificial no que DONEDA et al. (2018, p.2) descreve como “uma mudança na subjetividade das relações entre as pessoas e a tecnologia”, já que tarefas que antes eram executadas por humanos, inclusive a tomada de decisões, agora o são por máquinas<sup>38</sup>.

É da natureza humana a falha, o cometimento de erros na execução de tarefas, a tomada de decisões maculada por preconceitos pessoais ou estruturais. O humano não é neutro, carrega consigo questões morais e valores que influenciam nas suas decisões. Logo, não é antinatural que a máquina programada por humanos cometa falhas e perpetue a discriminação.

O problema dos vieses da IA reside na forma de elaboração dos comandos (algoritmos) e de escolha dos seus subsídios (dados), que, em última instância,

---

A resposta depende de como se define ‘classe média’ e ‘saúde financeira’. Há limites sobre os dados que podemos reunir e os tipos de experimentos que podemos realizar.” WHEELAN, Charles. **Estatística**: o que é, para que serve, como funciona. Trad. George Schlesinger. Rio de Janeiro: Zahar, 2016, p. 16.

<sup>37</sup> “Technology is not equivalent to the essence of technology. When we are seeking the essence of “tree,” we have to become aware that That which pervades every tree, as tree, is not itself a tree that can be encountered among all the other trees. Likewise, the essence of technology is by no means anything technological. Thus we shall never experience our relationship to the essence of technology so long as we merely conceive and push forward the technological, put up with it, or evade it. Everywhere we remain unfree and chained to technology, whether we passionately affirm or deny it. But we are delivered over to it in the worst possible way when we regard it as something neutral; for this conception of it, to which today we particularly like to do homage, makes us utterly blind to the essence of technology.” HEIDEGGER, Martin. **The question concerning technology and other essays**. Trans. William Lovitt. New York & London: Garland Publishing, Inc.: 1977, p. 4.

<sup>38</sup> “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.” DONEDA et al., p. 2.

mesmo que intermediadas por máquinas, são feitas por humanos. O modo de funcionamento dos mecanismos de IA, que se dá por métodos de *machine* e *deep learning*, também pode contribuir para a perpetuação da discriminação.

*Machine learning* é um processo de tentativa e erro, no qual ocorre a catalogação de resultados (LOGs) visando o aperfeiçoamento do algoritmo para resolução de problemas em um ciclo de treinamento contínuo da IA<sup>39</sup>. Portanto, a máquina aprende com as suas próprias experiências (TEPEDINO et al., 2019, p. 63-64).

*Deep learning* é uma técnica mais avançada, um processo de “aprendizado profundo” no qual ocorre a criação de grandes modelos de redes neurais que se assemelham ao cérebro humano, condizente com o atual momento de desenvolvimento tecnológico<sup>40</sup>. Portanto, a máquina aprende e decide sozinha, partindo das premissas que foram estabelecidas na sua formulação.

Segundo BAROCAS et al. (2017, p. 3), a seleção de dados utilizada para construir modelos computacionais (dados de treinamento) é uma fonte relevante de potencial discriminação, já que amostras populacionais conduzem a modelos que exibem erros sistemáticos facilmente negligenciáveis e, por vezes, impossíveis de serem reconhecidos, o que se agrava com a padronização da validação dos métodos baseados em amostras viciadas<sup>41</sup>.

---

<sup>39</sup> “Diferentemente dos algoritmos, que são fórmulas/programas, o *machine learning* é um processo baseado em tentativa e erro que vai gerando a catalogação de resultados (os *logs*) e tornando algoritmos mais preparados para a solução de problemas, na medida em que a continuidade das tentativas os ‘treina’ para serem melhores na apresentação de resultados.” FALEIROS JUNIOR, 2021, p. 20.

<sup>40</sup> “A complexidade dos sistemas dotados de inteligência artificial incrementa-se exponencialmente a partir dos modelos de *machine learning* (aprendizado de máquina), caracterizados pela aptidão da máquina a adquirir aprendizado a partir das suas próprias experiências. Caso se verifique, ainda, a utilização de modelos baseados em redes neurais à semelhança do funcionamento do cérebro humano, alude-se, no estágio mais atual da evolução tecnológica, ao *deep learning* (aprendizado profundo).” TEPEDINO, Gustavo; SILVA, Rodrigo da Guia. **Desafios da inteligência artificial em matéria de responsabilidade civil**. Revista Brasileira de Direito Civil – RBDCivil | Belo Horizonte, v. 21, p. 61-86, jul./set. 2019, p. 63-64.

<sup>41</sup> “The selection of the data used to build the models—the training data—is an important source of potential bias. Non-representative samples of the population will often lead to models that exhibit systematic errors. Such sampling biases are easy to overlook and sometimes impossible to fully recognize; worse, standard validation methods that depend on hold-out data drawn from the same sample will fail to reveal them. Even representative samples—or datasets that capture the entire population of interest—can fail to ensure that models perform equally well for different parts of the population. When minority groups do not follow the same pattern of behavior as the majority group, machine learning may struggle to model the behavior of the minority as effectively as the majority because there will be proportionally fewer examples of the minority behavior from which to learn. Under these conditions, the dominant group in society may well enjoy relatively higher accuracy

Os algoritmos, por seu turno, podem, proposital ou acidentalmente, de forma consciente ou não<sup>42</sup>, reproduzir tendências preconceituosas ocorridas na fase de mineração de dados, resultando na exacerbação das desigualdades e discriminação de grupos historicamente vulneráveis, mesmo que não tenham sido programados manualmente para tanto (BAROCAS et al., 2016, p. 673-674)<sup>43</sup>.

Como afirma FRAZÃO (2019, p. 342) “os programadores desses algoritmos podem transpor para eles seus próprios vieses e problemas cognitivos”, comprometendo a suposta neutralidade e perpetuando processos discriminatórios.

Foi o que aconteceu em 2014, quando a Amazon, que estava prestes a entrar em uma onda de contratações de empregados devido ao crescimento de vendas, montou um time de engenheiros com o objetivo de desenvolver uma ferramenta de IA que pudesse rastrear com rapidez candidatos na *web* para recrutamento.

Foram criados 500 modelos computacionais com foco em funções e locais específicos. Cada modelo foi ensinado a reconhecer cerca de 50.000 termos que apareceram em currículos de candidatos anteriores. Os algoritmos aprenderam a valorar pouco habilidades comuns entre profissionais de TI, tais como a habilidade de escrever códigos diversificados.

---

rates. Training data may also encode prior prejudicial or biased assessments. A model trained on historical hiring data could easily lead to future hiring decisions that simply replicate the discrimination at work in previous human decision-making upon which such modeling hoped to improve. Tainted training examples might wrongly instruct the machine to see features that actually predict success on the job as indicators of poor performance.” BAROCAS, Solon, et al. **Big data, data science, and civil rights**. *arXiv preprint arXiv:1706.03102*. Computing Community Consortium – Catalyst, 2017, p. 3.

<sup>42</sup> Quando usa o termo “consciente” ou “consciência”, o presente estudo se refere à consciência humana, do programador ou do usuário, já que considera que a IA não é dotada de “consciência”.

<sup>43</sup> “While discrimination certainly endures in part due to decision makers’ prejudices, a great deal of modern-day inequality can be attributed to what sociologists call “institutional” discrimination.8 Unconscious, implicit biases and inertia within society’s institutions, rather than intentional choices, account for a large part of the disparate effects observed. Approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. It can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment. Algorithms10 could exhibit these tendencies even if they have not been manually programmed to do so, whether on purpose or by accident.” BAROCAS, Solon; SELBST, Andrew D. **Big data's disparate impact**. California Law Review, Inc., 104:671, 2016, p. 673-674.

Em vez disso, a IA favoreceu candidatos que se descreveram utilizando verbos frequentemente utilizados em currículos de engenheiros do gênero masculino tais como “executou” (*executed*) ou “capturou” (*captured*)<sup>44</sup>.

Assim, em 2015, a companhia descobriu que o seu sistema de recrutamento não estava ranqueando os candidatos com neutralidade de gênero, pois os modelos haviam sido treinados para vetar candidatos observando os padrões de currículos submetidos em um período de 10 anos, sendo que a maioria deles havia sido enviado por homens, um reflexo da hegemonia masculina na indústria de tecnologia.

O sistema da Amazon, em verdade, aprendeu que candidatos do gênero masculino eram preferíveis, penalizando currículos que incluíam palavras como “feminino” (*women’s*) referenciadas em expressões como “capitã do clube de xadrez feminino” (*women’s chess club captain*), e rebaixando a pontuação de mulheres de duas universidades americanas exclusivamente femininas<sup>45</sup>.

No referido exemplo, observa-se que houve uma classificação discriminatória de pessoas pela IA com base no gênero, já que a máquina estabeleceu um *ranking* diferente para homens e mulheres, colocando as últimas em posição inferior e desfavorável no cenário competitivo.

---

<sup>44</sup> “Amazon’s experiment began at a pivotal moment for the world’s largest online retailer. Machine learning was gaining traction in the technology world, thanks to a surge in low-cost computing power. And Amazon’s Human Resources department was about to embark on a hiring spree: Since June 2015, the company’s global headcount has more than tripled to 575,700 workers, regulatory filings show. So it set up a team in Amazon’s Edinburgh engineering hub that grew to around a dozen people. Their goal was to develop AI that could rapidly crawl the web and spot candidates worth recruiting, the people familiar with the matter said. The group created 500 computer models focused on specific job functions and locations. They taught each to recognize some 50,000 terms that showed up on past candidates’ resumes. The algorithms learned to assign little significance to skills that were common across IT applicants, such as the ability to write various computer codes, the people said. Instead, the technology favored candidates who described themselves using verbs more commonly found on male engineers’ resumes, such as “executed” and “captured,” one person said.” DASTIN, Jeffrey. **Amazon scraps secret AI recruiting tool that showed bias against women**. In: MARTIN, Kirsten. **Ethics of data and analytics concepts and cases**. Boca Raton/London/ New York: CRC Press, Taylor & Francis Group, 2019, p. 298.

<sup>45</sup> “But by 2015, the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way. That is because Amazon’s computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry. In effect, Amazon’s system taught itself that male candidates were preferable. It penalized resumes that included the word ‘women’s,’ as in ‘women’s chess club captain.’ And it downgraded graduates of two all-women’s colleges, according to people familiar with the matter. They did not specify the names of the schools.” DASTIN, 2019, p. 296.

O *scoring* ou ranqueamento é o método utilizado para classificar os indivíduos e definir quais serão os beneficiários das decisões da IA<sup>46</sup>. Nas palavras de CITRON et al. (2014, p. 3), algoritmos preditivos mineram informações pessoais para formular suposições sobre os prováveis comportamentos e riscos individuais.

Assim, as atividades *on* e *offline* das pessoas são transformadas em *scores* (pontos) que os ranqueiam acima ou abaixo dos demais. Tanto entes privados, quanto públicos têm confiado em avaliações algorítmicas para tomar importantes decisões sobre os indivíduos<sup>47</sup>.

Por conseguinte, para além de tomar decisões no lugar dos humanos, substituindo-os nesse papel<sup>48</sup>, as máquinas têm tomado decisões sobre os humanos, sobre a sua dignidade, as suas potencialidades e, conseqüentemente, o seu destino<sup>49</sup>. O juiz máquina determina quem são os humanos merecedores dos benefícios descritos nos códigos do seu algoritmo.

Tais conclusões levam ao questionamento, suscitado na introdução, a ser respondido nos subtópicos a seguir: no atual contexto tecnológico, social e jurídico, a máquina está apta a ser o juiz do humano?

---

<sup>46</sup> “Não seria exagero afirmar que a transferência de processos decisórios para máquinas e algoritmos, tal como vem sendo feita na atualidade, transforma-os em verdadeiros oráculos do nosso tempo, possibilitando que eles possam julgar, classificar e pontuar os cidadãos em verdadeiros *rankings* dos quais pode depender o acesso a empregos, seguros, crédito, dentre outros bens e serviços.” FRAZÃO, 2019, p. 342.

<sup>47</sup> “How are these scores developed? Predictive algorithms mine personal information to make guesses about individuals’ likely actions and risks. A person’s on- and offline activities are turned into scores that rate them above or below others. Private and public entities rely on predictive algorithmic assessments to make important decisions about individuals.” CITRON, Danielle Keats; PASQUALE, Frank. **The scored society**: Due process for automated predictions. Wash. L. Rev., v. 89, p. 1, 2014, p. 3.

<sup>48</sup> “É elemento conceitual do aprendizado por máquinas a sua capacidade de definir e modificar regras de tomadas de decisão de forma autônoma, isto é, sem a necessidade de interferência humana.” MULHOLLAND, 2020, p. 331.

<sup>49</sup> “Por meio de algoritmos aperfeiçoados por técnicas de *machine learning*, é possível fazer inferências, predições, diagnósticos e tomar decisões complexas que, em muitos casos, apresentam alto teor de subjetividade, exigem sofisticadas análises valorativas e envolvem diversos riscos. Tais decisões afetam não apenas a vida dos que criam, usam ou contratam tais sistemas, mas também a vida daqueles que estão sujeitos aos seus efeitos. É o que acontece quando algoritmos decidem que candidato a emprego deve ser contratado, quem deve pagar mais por determinado seguro, qual a probabilidade de reincidência de determinado criminoso ou, no caso dos carros autônomos, quem deve ser atropelado em situações extremas.” FRAZÃO, 2019, p. 341.

## 2.2

### Processo autônomo de tomada de decisões por mecanismos de IA: funcionamento, finalidade e riscos.

A tomada de decisões de forma autônoma por mecanismos de IA vem sendo aplicada em uma constatare crescente. O alcance desse tipo de processo decisório vai desde decisões aparentemente simples, como sugestões de compra nas redes sociais, até decisões com maiores implicações que definem, por exemplo, quem é merecedor de um crédito bancário, uma vaga de emprego, uma vaga na universidade, ou elegível para uma cirurgia ou tratamento médico<sup>50</sup>.

Por meio dos processos de *machine e deep learning*, as máquinas passaram a simular o pensamento humano<sup>51</sup> e a tomar decisões que impactam os humanos e suas relações jurídicas, afetando, consequentemente, a sua liberdade de escolha, o que constitui, nas palavras de MULHOLLAND (2020, p. 332), uma “fonte potencial de danos por violação de direitos personalíssimos”<sup>52</sup>.

---

<sup>50</sup> “The past decade has witnessed an unprecedented acceleration in both the sophistication and uptake of various machine learning systems, including systems that employ ‘deep learning’ algorithms. From music and TV show recommendations, product advertising, and opinion polling to medical diagnostics, university admissions, job placement, and financial services, the range of the potential application of these new AI technologies is truly vast.” ZERILLI, John, et al. **Transparency in algorithmic and human decision-making: is there a double standard?**. Philosophy & Technology 32.4: 661-683, 2019, p. 2.

<sup>51</sup> “Em uma camada da *machine learning* está a espécie denominada de *deep learning*, o qual ‘orienta’ máquinas para executarem determinadas atividades tal qual um ser humano as faria, considerando operações como, por exemplo, a identificação de locais, objetos, plantas, animais ou pessoas por meio de imagens; o reconhecimento de sons (como a fala das pessoas ou o canto de pássaros), ou o estabelecimento de padrões de acordo com decisões relacionadas a dados antecedentes.” SOARES, Flaviana Rampazzo. **Levando os algoritmos a sério**. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Felipe Braga Neto, et al. **Direito digital e inteligência artificial: diálogos entre Brasil e Europa**. Indaiatuba: Editora Foco, 2021, p. 46.

<sup>52</sup> “Por meio do uso dessas técnicas – *machine learning e deep learning* –, uma das aplicações mais interessantes – e, eventualmente, problemática – é a que se concretiza por meio da delegação total de processos decisórios para a IA. Inteligências Artificiais que estabelecem tomadas de decisão podem ser utilizados em ODRs (*online dispute resolution*), que substituem as decisões ‘humanas’ em mediação e conflitos, até os utilizados em sistemas de polícia preditiva que identificam potencialidade criminosa e probabilidade geográfica de atividade criminosa. Existem outros usos menos conhecidos da automação total da IA e que constituem fonte potencial de danos por violação de direitos personalíssimos. São sistemas de IA que podem decidir autonomamente se uma pessoa terá ou não concedido um empréstimo bancário; determinar se, na execução penal, será ou não conferido algum benefício que permita a liberdade condicionada; e recomendar se um paciente é candidato a um tipo de tratamento médico.” MULHOLLAND, 2020, p. 332.

Faz-se necessário um olhar mais atento para o detalhamento do processo decisória da IA, de como é o seu mecanismo de funcionamento e em que situações vem sendo utilizado a fim de se investigar e propor meios de preservação da autonomia humana nesse contexto.

## 2.2.1

### Ciclo de vida da IA – fases de utilização – controle *ex post*.

Importa notar que a IA tem um ciclo de vida<sup>53</sup> dividido em fases: (i) desenvolvimento, (ii) concepção, (iii) implantação e (iv) utilização<sup>54</sup>. Os eventuais vieses que ocorrem no processo decisório autônomo da máquina têm origem nas fases de desenvolvimento e concepção, quando são colhidos os dados iniciais (de construção e teste do modelo) e formulado(s) o(s) algoritmo(s), tornando-se, contudo, evidentes na fase de utilização ou operação.

A EUR-Lex 52021PC0206, que optou por uma abordagem baseada no risco<sup>55</sup>, ao adotar regras específicas para os sistemas de IA de risco elevado (que dizem respeito à saúde, segurança e direitos fundamentais de pessoas singulares), estabeleceu a necessidade de avaliação de conformidade *ex ante*<sup>56</sup> e supervisão *ex*

---

<sup>53</sup> É interessante a escolha do termo técnico “ciclo de vida” para indicar as fases de funcionamento de um ente não orgânico. Estaria a IA sendo equiparada a uma forma de vida? Ela goza do atributo “existência”?

<sup>54</sup> As fases da IA segundo as orientações do GPAN IA (grupo independente de peritos de alto nível sobre a inteligência artificial criado pela Comissão Europeia em junho de 2018): “O ciclo de vida do sistema de IA engloba as suas fases de desenvolvimento (incluindo investigação, conceção, fornecimento de dados e ensaios limitados), implantação (incluindo aplicação) e utilização.” GPAN IA. **Orientações éticas para uma IA de confiança.** Comissão europeia, 2019a, p. 47. Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>>. Acesso em: 26.10.2022.

<sup>55</sup> “Dadas as suas características específicas (por exemplo, a opacidade, a complexidade, a dependência dos dados, o comportamento autónomo), a utilização da inteligência artificial pode afetar negativamente um conjunto de direitos fundamentais consagrados na Carta dos Direitos Fundamentais da UE (a seguir designada por «Carta»). A presente proposta procura assegurar um nível elevado de proteção desses direitos fundamentais e visa fazer face aos vários riscos mediante uma abordagem baseada no risco claramente definida.” EUR-Lex 52021PC0206, 2021, p. 12.

<sup>56</sup> “O título III inclui regras específicas relativas aos sistemas de IA que criam um risco elevado para a saúde e a segurança ou para os direitos fundamentais de pessoas singulares. Em conformidade com uma abordagem baseada no risco, esses sistemas de IA de risco elevado são autorizados no mercado europeu, mas estão sujeitos ao cumprimento de determinados requisitos obrigatórios e a uma avaliação da conformidade *ex ante*.” EUR-Lex 52021PC0206, 2021, p. 14.

*post*<sup>57</sup> dos sistemas, privilegiando, no que se refere aos direitos fundamentais, o controle *ex post*<sup>58</sup>.

Com efeito, o item 1 do artigo 14º da referida proposta legislativa, que trata da supervisão humana, determina que os sistemas de IA de risco elevado “devem ser concebidos e desenvolvidos” com a inclusão de “ferramentas de interface homem-máquina” que permitem que os sistemas sejam “eficazmente supervisionados por pessoas singulares durante o período de utilização”<sup>59</sup>.

Portanto, há no dispositivo citado uma preocupação com a inclusão de ferramentas no desenvolvimento do sistema, o que caracteriza uma conformidade *ex ante*, porém, com o objetivo de propiciar a supervisão *ex post*. Assim, ao pretender positivar a supervisão humana como instrumento de tutela de direitos essenciais e fundamentais, a EUR-Lex 52021PC0206 o faz com enfoque na fase de utilização do sistema, a fase *ex post*.

O item 2 do art. 14º coloca a supervisão humana como mecanismo de prevenção e mitigação de “riscos para a saúde, a segurança ou os direitos fundamentais” que possam surgir quando da utilização de um sistema de IA de risco elevado, “em especial quando esses riscos persistem apesar da aplicação de outros requisitos” estabelecidos no texto da proposta<sup>60</sup>.

---

<sup>57</sup> “Depois de ter efetuado a avaliação da conformidade necessária, o fornecedor deve registrar esses sistemas de IA de risco elevado autónomos numa base de dados da UE que será gerida pela Comissão, a fim de aumentar a transparência e a supervisão públicas e de reforçar a supervisão *ex post* por parte das autoridades competentes.” EUR-Lex 52021PC0206, 2021, p. 16.

<sup>58</sup> “Caso continuem a ocorrer violações dos direitos fundamentais, as pessoas afetadas têm acesso a vias eficazes de recurso graças à garantia da transparência e da rastreabilidade dos sistemas de IA, associadas a fortes controlos *ex post*.” EUR-Lex 52021PC0206, 2021, p. 12.

<sup>59</sup> EUR-Lex 52021PC0206, art. 14º. Supervisão humana. 1: “Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de tal modo, incluindo com ferramentas de interface homem-máquina apropriadas, que possam ser eficazmente supervisionados por pessoas singulares durante o período de utilização do sistema de IA.” É de se notar que a versão em português da proposta legislativa europeia faz uso do termo “homem-máquina”, contudo, a versão original, em inglês, faz uso da expressão “*human-machine*” (humano-máquina). Eis o texto em inglês: “Article 14. Human oversight 1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.” Em busca do uso de uma linguagem mais inclusiva, que ultrapassa a barreira do gênero, não obstante a termo utilizado na versão em português da proposta europeia, o presente trabalho, sempre que possível, fará uso do termo “humano-máquina”, conforme tradução literal da versão em inglês.

<sup>60</sup> EUR-Lex 52021PC0206, art. 14º, item 2: “A supervisão humana deve procurar prevenir ou minimizar os riscos para a saúde, a segurança ou os direitos fundamentais que possam surgir quando um sistema de IA de risco elevado é usado em conformidade com a sua finalidade prevista ou em condições de utilização indevida razoavelmente previsíveis, em especial quando esses riscos persistem apesar da aplicação de outros requisitos estabelecidos neste capítulo.”



Na mesma linha da proposta europeia, a minuta de substitutivo do PL nº 21/2020 adotou uma abordagem baseada no risco<sup>61</sup>, definindo no art. 17 os sistemas de inteligência artificial de alto risco, para os quais, o artigo 20, parágrafo único determina a utilização da supervisão humana como forma de “prevenir ou minimizar os riscos para os direitos e liberdades das pessoas que possam decorrer do seu uso normal ou de seu uso em condições de utilização indevida razoavelmente previsíveis”.

Deste modo, ambas as propostas legislativas tratam do ponto chave do presente estudo, que é a supervisão humana como instrumento de tutela do humano na automatização de decisões, direcionando-o para um controle *ex post* dos sistemas, na sua fase de utilização, fazendo relevante que se entenda como ocorre o mecanismo de tomada de decisão neste momento do ciclo de vida da IA.

Conforme ilustra a figura abaixo, em fase de operação, a máquina racional percebe o ambiente por meio de sensores, recolhe e interpreta os dados, processa as informações obtidas, decide qual a melhor ação a adotar e age no ambiente que lhe provocou o estímulo, podendo ou não o alterar<sup>62</sup>.

---

<sup>61</sup> “Seu objetivo normativo é conciliar uma abordagem baseada em riscos com uma modelagem regulatória baseada em direitos.” DIÁRIO DO SENADO FEDERAL, 2022, p. 10.

<sup>62</sup> “Por conseguinte, um sistema de IA é, antes de mais, racional, segundo afirma um dos manuais de IA mais utilizados. Mas como é que um sistema de IA alcança a racionalidade? Fá-lo, tal como indica a primeira frase da definição operacional de IA acima referida, percecionando o ambiente em que está imerso através de sensores, recolhendo e interpretando dados dessa forma, raciocinando sobre o que foi percecionado ou processando as informações obtidas a partir dos dados e decidindo qual a melhor ação a adotar, e agindo, depois, em conformidade através de atuadores, o que poderá levar a uma alteração do ambiente. Os sistemas de IA podem utilizar regras simbólicas ou aprender um modelo numérico, e também podem adaptar o seu comportamento analisando o modo como o ambiente foi afetado pelas suas ações anteriores. A ilustração de um sistema de IA apresentada na figura 1 pode ser esclarecedora.” GPAN IA, 2019b, p. 3.

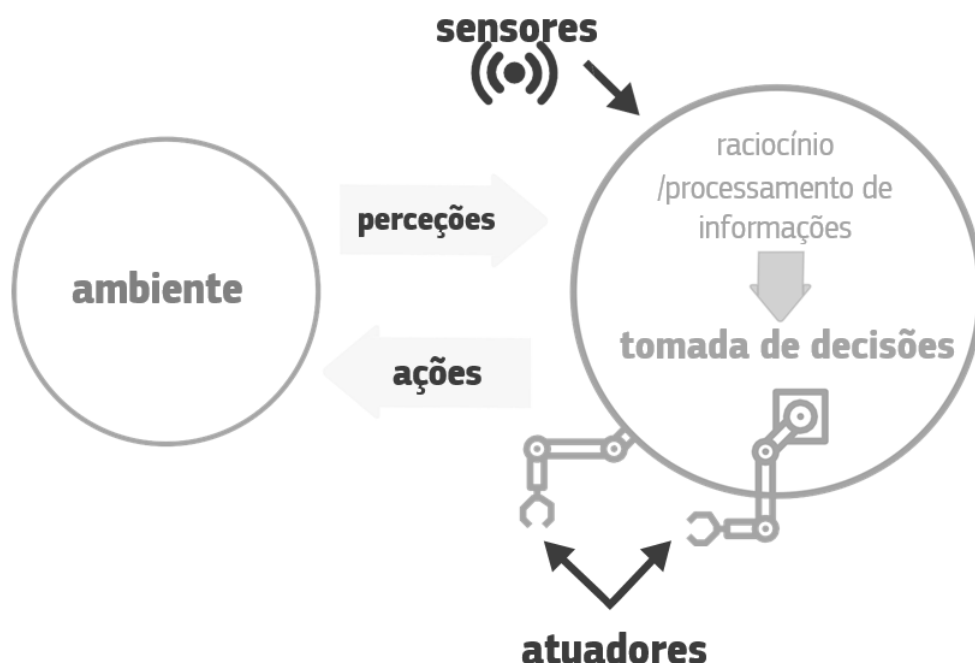


Figura 1: representação esquemática de um sistema de IA<sup>63</sup>.

Os sensores são dispositivos de entrada de dados que podem ser câmeras; microfones; um teclado; um *website*; ou instrumentos que meçam temperatura, pressão, distância, força, e que captam (percebem) os dados do ambiente<sup>64</sup>. Os dados colhidos podem ser estruturados (organizados segundo modelos pré-definidos, como um banco de dados) ou não estruturados (quando não há um sistema de organização prévia, como acontece com uma imagem capturada)<sup>65</sup>.

Os dados captados pelos sensores são transformados em informação inteligível para a máquina (modelos numéricos), o que permite o raciocínio e

<sup>63</sup> GPAN IA, 2019b, p. 2.

<sup>64</sup> Ao explicar a figura apresentada: “Na figura 1 os sensores do sistema são representados por um símbolo de Wi-Fi. Na prática, poderão ser câmaras, microfones, um teclado, um sítio Web ou outros dispositivos de entrada de dados, bem como sensores de quantidades físicas (p. ex., sensores de temperatura, pressão, distância, força/binário ou sensores táteis). De um modo geral, é necessário equipar o sistema de IA com sensores adequados para perceber os dados presentes no ambiente que são pertinentes para o objetivo que lhe foi atribuído pelo seu criador humano.” GPAN IA, 2019b, p. 2.

<sup>65</sup> “No que respeita aos dados recolhidos, muitas vezes é conveniente distinguir entre dados estruturados e dados não estruturados. Os *dados estruturados* são dados organizados segundo modelos pré-definidos (p. ex., numa base de dados relacional), enquanto os *dados não estruturados* não possuem uma organização conhecida (como acontece numa imagem ou no excerto de um texto).” GPAN IA, 2019b, p. 2.

processamento que resultam nas decisões. As decisões, por sua vez, são propostas de ação que visam atender o objetivo a ser atingido conforme a programação do algoritmo e aprendizado constante da IA<sup>66</sup>.

Assim, a IA interpreta e raciocina por meio de fórmulas matemáticas pré-estabelecidas e outras que cria ao longo do seu processo de aprendizado<sup>67</sup>, tomando decisões e produzindo ações que afetam o ambiente do qual colheu as percepções e no qual resultam as suas ações. Nesse ambiente, estão os humanos e suas relações jurídicas e interpessoais.

As percepções são os *inputs*, as decisões ou previsões são os *outputs*, que interferem no ambiente que se vê na figura acima. O ambiente pode ser um outro programa, uma rede social, o ciberespaço, uma realidade virtual ou o mundo físico. Independentemente do tipo de ambiente no qual são implementadas as ações decorrentes das decisões tomadas pela IA, uma série de interações e relações humanas são afetadas pelos seus vereditos.

Segundo FRAZÃO (2019, p. 242), quando se fala em inteligência artificial o “controle sobre os *inputs* e *outputs* dos sistemas torna-se ainda mais precário”, pois o “código está em constante mutação”<sup>68</sup>.

O ambiente é construído, habitado e/ou frequentado por pessoas que o utilizam para trabalhar, estudar, cuidar da saúde, morar, interagir e que não têm controle sobre as decisões da IA, ou, em alguns casos, sequer têm conhecimento de que uma máquina está tomando as decisões.

---

<sup>66</sup> “No centro de um sistema de IA está o seu módulo de raciocínio/processamento da informação, o qual utiliza os dados provenientes dos sensores para propor uma ação, em função do objetivo a atingir. Tal significa que os dados recolhidos pelos sensores têm de ser transformados em informação que o módulo de raciocínio/processamento de informações consiga compreender. Continuando com o exemplo de um sistema de IA de limpeza, a câmara fornece uma fotografia do pavimento ao módulo de raciocínio/processamento de informações e este módulo tem de decidir se deve ou não limpar o pavimento (ou seja, qual é a melhor ação para atingir o objetivo pretendido). Embora para os seres humanos possa parecer fácil passar da fotografia de um pavimento para a decisão sobre se este necessita de ser limpo ou não, para uma máquina será mais difícil, pois uma fotografia é apenas uma sequência de zeros e uns.” GPAN IA, 2019b, p. 2.

<sup>67</sup> “Um sistema racional com aprendizagem é um sistema racional que, depois de executar uma ação, avalia o novo estado do ambiente (através da percepção) para determinar o nível de sucesso da ação e adapta, posteriormente, as suas regras de raciocínio e os seus métodos de decisão.” GPAN IA, 2019b, p. 3.

<sup>68</sup> “Diante da inteligência artificial, o controle sobre os *inputs* e *outputs* dos sistemas torna-se ainda mais precário. De fato, se o próprio código está em constante mutação, não se sabe em que medida será possível compreender a relação entre *inputs* e *outputs*, ainda que haja auditoria e transparência.” FRAZÃO, 2019, p. 342.

Portanto, o impacto dessas decisões na autonomia humana é significativo e precisa ser mitigado por meio do desenvolvimento de instrumentos de controle pelos humanos, além de ações que propiciem a transparência e previsibilidade no funcionamento dos sistemas.

### 2.2.2

#### O funcionamento da IA por meio de regras e padrões.

Segundo MULHOLLAND (2020, p. 331) os processos de *machine learning*: (i) criam “conhecimento e padrões originais” utilizados como base dos modelos de predição da IA; (ii) têm a “capacidade de definir e modificar regras de tomadas de decisão de forma autônoma, isto é, sem a necessidade de interferência humana”<sup>69</sup>.

Deste modo, a IA, autonomamente, cria padrões e define as regras do seu processo decisório, modificando-as quando entende pertinente. Logo, o humano se torna desnecessário do ponto de vista técnico, o que não significa que seja prescindível do ponto de vista ético e jurídico.

Ao discorrer sobre o “devido processo tecnológico” (*technological due process*), CITRON (2007, p. 1302) afirma que regras (*rules*) e padrões (*standards*) têm diferentes propósitos. Enquanto *regras* estabelecem instruções básicas de comportamento, facilitam a previsibilidade, impõe sanções, promovem transparência e adequação, promovem justiça por meio de tratamento igualitário, impedem decisões lastreadas em valores pessoais, e tornam o processo de tomada de decisões mais barato, evitando a necessidade de reconsideração constante das questões; *padrões* permitem aos tomadores de decisão modular um resultado aos fatos, aumentando a probabilidade de uma “decisão ideal”, levando em consideração a mudança

---

69 “A técnica conhecida como *machine learning* (aprendizado por máquinas) se configura como qualquer metodologia e conjunto de técnicas que utilizam dados em grande escala (*input*) para criar conhecimentos e padrões originais e, com base neles, gerar modelos que são usados para a predição a respeito dos dados tratados (*output*). É elemento conceitual do aprendizado por máquinas a sua capacidade de definir e modificar regras de tomadas de decisão de forma autônoma, isto é, sem a necessidade de interferência humana.” MULHOLLAND, 2020, p. 331.

de circunstâncias trazidas pela tecnologia e obrigando os tomadores de decisão a fundamentar suas escolhas, aumentando a transparência e responsabilidade no processo<sup>70</sup>.

Logo, decisões que seriam mais bem tratadas por padrões não deveriam ser automatizadas, que é o caso de políticas que implícita ou explicitamente requeiram o exercício da discricionariedade humana<sup>71</sup>.

Ao sustentar que decisões que necessitam de modulação e envolvem políticas mais sensíveis, e/ou exigem algum grau de discricionariedade no processo de tomada de decisão não devem ser automatizadas, a jurista antecipa preocupações que vieram a ser endereçadas oficialmente quando da elaboração e apresentação da EUR-Lex 52021PC0206 e pelo grupo de juristas que apresentou a minuta de substitutivo do PL nº 21/2020.

Ambas as propostas legislativas, como se verá ao longo deste estudo, dão um tratamento diferenciado a questões sensíveis, relacionadas à existência digna e plúrima dos indivíduos e que apresentam nuances que escapam ao julgamento da máquina, já que, sob o olhar humano, a IA teria melhores condições de validar os padrões que necessitam de modulação para a efetividade das suas decisões.

---

<sup>70</sup> “Rules and standards have different goals. Rules establish basic instructions for behavior. They facilitate predictability. Because rules provide notice of consequences attached to an activity, the public can order its affairs accordingly. Rules promote fairness by requiring consistent treatment of similar cases. They prevent decision makers from basing decisions on parties’ perceived attractive qualities. Rules cut the costs of decision making by eliminating the need to reconsider recurring issues. Standards, on the other hand, permit decision makers to tailor an outcome to the facts, increasing the likelihood of an ‘ideal’ ruling. They can account for changing circumstances brought about by technology. Standards also force decision makers to articulate their choices, increasing the chance that they will ‘make visible and accountable the inevitable weighing process that rules obscure.” CITRON, Danielle Keats. **Technological due process**. *Wash. UL Rev.* 85 (2007): 1249, p. 1302.

<sup>71</sup> “Decisions best addressed with standards should not be automated. Policies that explicitly or implicitly require the exercise of human discretion cannot be automated. For instance, agencies should not automate policies that allow individuals to plead extenuating circumstances that software cannot anticipate.” CITRON, 2007, p. 1304.

### 2.2.3

#### O uso da tecnologia com centralidade no humano.

A existência da IA deve ter por escopo auxiliar a eficiência de processos, otimizar recursos, trazer mais conforto e praticidade aos usuários. Logo, como define HEIDEGGER (1977, p. 4-5), a máquina pode e deve ser utilizada como um “instrumento” a serviço do homem, já que a tecnologia é “um meio para um fim” e “uma atividade humana”<sup>72</sup>.

Segundo BALKIN (2015, p. 49), quando juristas falam de tecnologia é geralmente um atalho para algo mais complexo: (i) como as pessoas interagem com novas invenções; (ii) como as pessoas interagem com outras pessoas utilizando tais invenções ou considerando tais invenções.

Deste modo, o que se chama de efeitos da tecnologia são menos efeitos das coisas e mais características das relações sociais que empregam tais coisas. Essas relações sociais incluem relacionamentos de poder e autoridade, formas de organização e produção humanas, e características de sexualidade e associação humanas e vida em família<sup>73</sup>.

Logo, em se tratando de uma atividade criada e desenvolvida por humanos, para atendê-los, tem-se a centralidade no humano como uma das principais premissas para o desenvolvimento ético da IA.

De acordo com a Comissão Europeia, que em 08.04.2019 emitiu o comunicado denominado *Building Trust in Human-Centric Artificial Intelligence*, a

---

<sup>72</sup> “According to ancient doctrine, the essence of a thing is considered to be what the thing is. We ask the question concerning technology when we ask what it is. Everyone knows the two statements that answer our question. One says: Technology is a means to an end. The other says: Technology is a human activity. The two definitions of technology belong together. For to posit ends and procure and utilize the means to them is a human activity. The manufacture and utilization of equipment, tools, and machines, the manufactured and used things themselves, and the needs and ends that they serve, all belong to what technology is. The whole complex of these contrivances is technology. Technology itself is a contrivance, or, in Latin, an *instrumentum*.” HEIDEGGER, 1977, p. 4-5.

<sup>73</sup> “Indeed, what lawyers call “technology” is usually a shorthand for something far more complex. When we talk about “technology,” we are really talking about (1) how people interact with new inventions and (2) how people interact with other people using those new inventions or presupposing those new inventions. What we call the effects of technology are not so much features of things as they are features of social relations that employ those things. These social relations include relationships of power and authority, forms of human organization and production, and features of human sexuality, association, and family life.” BALKIN, 2015, p. 49.

confiança é um pré-requisito para assegurar uma abordagem da IA com centralidade no humano, considerando que ela não é um fim em si mesma, mas uma ferramenta que deve servir as pessoas com o objetivo final de aumentar o bem-estar humano<sup>74</sup>.

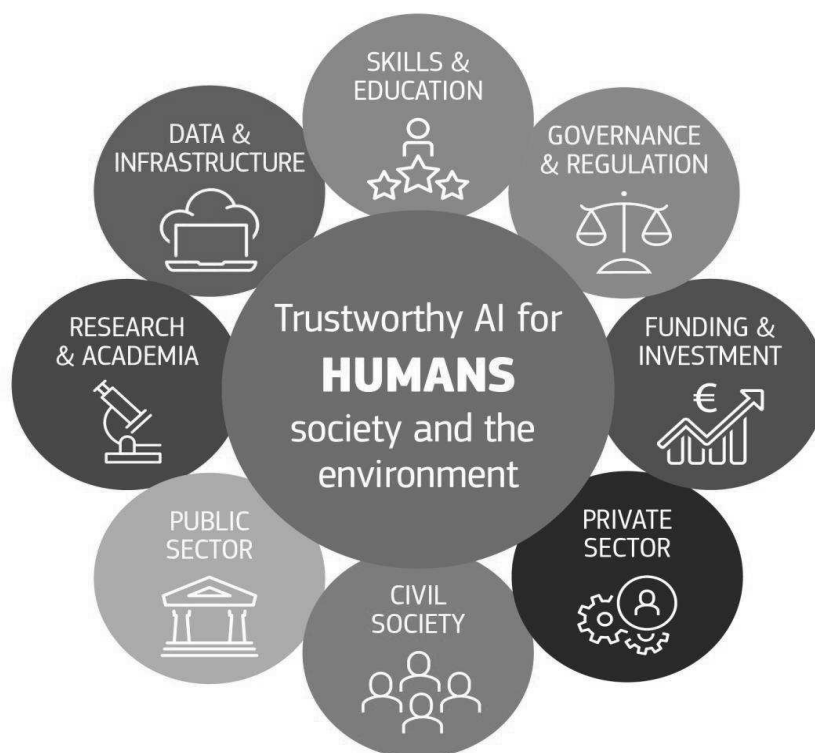
Assim, os valores humanos que fundam a sociedade, tais como a dignidade humana, a liberdade, a democracia, a igualdade, a existência do Estado Democrático de Direito, o respeito pelos direitos humanos, incluindo os direitos das minorias, devem ser plenamente integrados no desenvolvimento das tecnologias de IA por aqueles envolvidos no processo de criação e implementação da tecnologia<sup>75</sup>.

Com efeito, o humano no centro permite o desenvolvimento ético da IA nas áreas de formação profissional e educação, governança e regulação, financiamentos e investimentos, no setor privado, na sociedade civil, no setor público, pesquisa e academia, e dados e infraestrutura, conforme esquematizado na figura a seguir:

---

<sup>74</sup> “The European AI strategy and the coordinated plan make clear that trust is a prerequisite to ensure a human-centric approach to AI: AI is not an end in itself, but a tool that has to serve people with the ultimate aim of increasing human well-being.” EUROPEAN COMMISSION. **Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: Building Trust in Human-Centric Artificial Intelligence**. 2019, p. 1-2. Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>>. Acesso em: 31.10.2022.

<sup>75</sup> “To achieve this, the trustworthiness of AI should be ensured. The values on which our societies are based need to be fully integrated in the way AI develops. The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the societies of all Member States in which pluralism, non-discrimination, tolerance, justice, solidarity and equality prevail. In addition, the EU Charter of Fundamental Rights brings together – in a single text – the personal, civic, political, economic and social rights enjoyed by people within the EU.” EUROPEAN COMMISSION. **Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: Building Trust in Human-Centric Artificial Intelligence**, 2019, p. 2.

Figura 2<sup>76</sup>.

Na exposição de motivos da EUR-Lex 52021PC0206, consta que a IA “deve ser uma ferramenta ao serviço das pessoas e uma força positiva para a sociedade com o objetivo final de aumentar o bem-estar dos seres humanos” e que as regras a ela aplicáveis devem “centrar-se no ser humano”.

Em linha semelhante, o art. 5º do PL nº 21/2020, nos seus incisos I e II, estabelece como princípios para desenvolvimento e aplicação da IA no Brasil “finalidade benéfica” e a “centralidade do ser humano”<sup>77</sup> respectivamente. Também

<sup>76</sup> AI HLEG. **Policy and investment recommendations for trustworthy AI**. 2019, p. 25. Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>>. Acesso em: 26.10.2022. Ressalte-se que o AI HLEG (*independent high-level expert group on artificial intelligence set up by the European Commission*) é a nomenclatura em inglês do GPAN IA (grupo independente de peritos de alto nível sobre a inteligência artificial criado pela Comissão Europeia em junho de 2018). Como nem todos os documentos foram transcritos para o português, faz-se esta referência em inglês.

<sup>77</sup> Art. 5º São princípios para o desenvolvimento e a aplicação da inteligência artificial no Brasil: I – finalidade benéfica: busca de resultados benéficos para a humanidade pelos sistemas de inteligência artificial; II – centralidade do ser humano: respeito à dignidade humana, à privacidade, à proteção



na sua justificativa, o PL nº 21/2020 “faz uma abordagem da IA centrada no ser humano” e tem dentre os seus objetivos principais “melhorar o bem-estar das pessoas”.

Na mesma linha, o texto da minuta do substitutivo do PL nº 21/2020 e do PL nº 2328/2023, em seu art. 2º, I, apresenta a centralidade no humano como o primeiro fundamento para o desenvolvimento, implementação e uso de sistemas de IA no Brasil<sup>78</sup>.

As pretensões legislativas citadas, visando promover a regulação do uso da IA de modo a manter o humano no centro e, conseqüentemente, preservar direitos existenciais e fundamentais, têm razão de ser, já que, com o avanço da aplicação deste tipo de tecnologia em diversos seguimentos de serviços privados e públicos, situações em que a condição humana tem sido relegada a segundo plano têm ocorrido<sup>79</sup>.

---

de dados pessoais e aos direitos fundamentais, quando o sistema tratar de questões relacionadas ao ser humano.

<sup>78</sup> “Art. 2º O desenvolvimento, implementação e uso de sistemas de inteligência artificial no Brasil têm como fundamentos: I – a centralidade da pessoa humana.” Na exposição de motivos da minuta, também aparece a preocupação com a centralidade no humano que aqui registra-se também em homenagem ao jurista Danilo Doneda: “Ao fim e ao cabo, essa proposta de regulação serve a um propósito central, como bem lembrou Danilo Doneda (*in memoriam*), membro da Comissão de Juristas quando da sua instalação, que é da ‘a centralidade do elemento humano [que] deve ser enfatizada em todos os seus aspectos, em todos os pontos necessários.’”

<sup>79</sup> A exemplo: “Em agosto, a Xsolla, filial russa de uma empresa de *software* e serviços interativos com sede em Los Angeles, fez uma reestruturação inovadora de sua equipe, atraindo a atenção de veículos de comunicação do mundo todo. Sem prévio aviso, ela decidiu demitir 150 dos 450 funcionários de seus escritórios em Perm e Moscou, seguindo apenas a recomendação de um algoritmo de eficiência no trabalho que os considerou ‘improdutivos’ e ‘pouco comprometidos’ com os objetivos da empresa. Nem o impacto da pandemia nem as tão citadas ‘razões estruturais’. Desta vez, a causa alegada para justificar as demissões em massa foi o julgamento frio de um programa de Inteligência artificial alimentado com *big data*. A medida foi tão drástica e incomum que o diretor-executivo e fundador da empresa, Alexander Agapitov, apressou-se em declarar à edição russa da *Forbes* que não concordava totalmente com o veredicto da máquina, mas era obrigado a acatá-lo devido aos protocolos internos pactuados com a sua assembleia de acionistas. Ele até se ofereceu para ajudar os trabalhadores demitidos a encontrar novos empregos o mais rápido possível porque, em sua opinião, eles são, na maioria, ‘bons profissionais’.” ECHARRI, Miquel. **150 demissões em um segundo**: os algoritmos que decidem quem deve ser mandado embora. EL PAÍS. Tempos modernos. Publicado em: 10.10.2021. Disponível em: <<https://brasil.elpais.com/tecnologia/2021-10-10/150-demissoes-em-um-segundo-assim-funcionam-os-algoritmos-que-decidem-quem-deve-ser-mandado-embora.html>>. Acesso em: 21.12.2021.

## 2.2.4

### Usurpação de papéis humanos – valor social - efeito substituição.

Atenta para a inversão de valores na relação humano-máquina, e a fim de preservar a centralidade humana no desenvolvimento tecnológico, a literatura, tanto técnica, da área de sistemas, quanto jurídica, tem buscado qualificar o fenômeno e propor soluções.

RUSSEL (2019, p. 92) utiliza o termo “usurpação de papéis humanos” pelas máquinas nas relações que envolvam a prestação de serviços interpessoais, alertando especialmente para os riscos de imitação do humano por robôs com aparência humanoide<sup>80</sup>.

Enfatiza, ainda, alguns dos perigos do aumento de decisões autônomas de IA, quais sejam: i) violação da dignidade humana, ao dar às máquinas autoridade sobre os humanos, colocando-os em *status* inferior, fazendo-os perder o direito de intervir em decisões que afetem a si próprios; ii) o poder de *design* das máquinas em mãos de uma elite intelectual e econômica, aumentando os padrões de discriminação e controle sobre as massas de manobra; iii) a preponderância do fator econômico sobre o humano, por ser mais fácil, mais rápido e mais barato deixar as decisões para as máquinas<sup>81</sup>.

---

<sup>80</sup> “We should think twice before allowing machines to take over roles involving interpersonal services. If being human is our main selling point to other humans, so to speak, then making imitation humans seems like a bad idea. Fortunately for us, we have a distinct advantage over machines when it comes to knowing how other humans feel and how they will react. Nearly every human knows what it’s like to hit one’s thumb with a hammer or to feel unrequited love. Counteracting this natural human advantage is a natural human disadvantage: the tendency to be fooled by appearances—especially human appearances.” RUSSEL, 2019, p. 92.

<sup>81</sup> “In addition to the needless and even absurd elevation of the status of robots, there is a danger that the increased use of machines in decisions affecting people will degrade the status and dignity of humans. This possibility is illustrated perfectly in a scene from the science-fiction movie *Elysium*, when Max (Matt Damon) pleads his case before his ‘parole officer’ to explain why the extension of his sentence is unjustified. Needless to say, Max is unsuccessful. The parole officer even chides him for failing to display a suitably deferential attitude. One can think of such an assault on human dignity in two ways. The first is obvious: by giving machines authority over humans, we relegate ourselves to a second-class status and lose the right to participate in decisions that affect us. (A more extreme form of this is giving machines the authority to kill humans, as discussed earlier in the chapter.) The second is indirect: even if you believe it is not the machines making the decision but those humans who designed and commissioned the machines, the fact that those human designers and commissioners do not consider it worthwhile to weigh the individual circumstances of each human subject in such cases suggests that they attach little value to the lives of others. This is perhaps

Para que a máquina venha a complementar e não usurpar os papéis tipicamente humanos, o cientista de computação propõe a retenção de conhecimento, autoridade e autonomia<sup>82</sup>. Assim, o controle e fiscalização humanas sobre o processo decisório autônomo da IA impediriam a “usurpação” para a qual alerta.

CALO (2015) conceitua o “robô” distinguindo-o de outras tecnologias pela tríade sentido–pensamento–ação<sup>83</sup>, do ponto de vista da utilidade, o que representa o conceito técnico; e pela tríade materialidade (existência física/corpórea) – emergência (autonomia)-valor social, no que se refere às suas “qualidades essenciais”<sup>84</sup>, com enfoque nas questões jurídicas e regulatórias<sup>85</sup>.

Partindo das evoluções legais decorrentes do advento da *internet*, o jurista faz um paralelo que conduz a um possível caminho de regulação do que chama de “robótica”. Apesar de não fazer referência direta à IA, os robôs aos quais se refere, de acordo com a conceituação exposta anteriormente, são mecanismos de inteligência artificial.

No presente item, interessa endereçar o valor social (*social valence*) que CALO (2015) atribui aos robôs. É o que descreve como “tecnologia antropomórfica” e o impacto psicológico que gera nas pessoas, que tendem a interagir com

---

a symptom of the beginning of a great separation between an elite served by humans and a vast underclass served, and controlled, by machines.” RUSSEL, 2019, p. 94-95.

<sup>82</sup> “There are different ways we can design the relationship between humans and machines. If we design it so that humans retain sufficient understanding, authority, and autonomy, the technological parts of the system can greatly magnify human capabilities, allowing each of us to stand on a vast pyramid of capabilities—a demigod, if you like. But consider the worker in an online-shopping fulfillment warehouse. She is more productive than her predecessors because she has a small army of robots bringing her storage bins to pick items from; but she is a part of a larger system controlled by intelligent algorithms that decide where she should stand and which items she should pick and dispatch. She is already partly buried in the pyramid, not standing on top of it. It’s only a matter of time before the sand fills the spaces in the pyramid and her role is eliminated.” RUSSEL, 2019, p. 97.

<sup>83</sup> “Few complex technologies have a single, stable, uncontested definition. Robots are no exception. There is some measure of consensus, however, around the idea that robots are mechanical objects that take the world in, process what they sense, and in turn act upon the world. The utility here of the so-called sense-think-act paradigm lies in distinguishing robots from other technologies.” CALO, 2015, p. 529.

<sup>84</sup> “Embodiment, emergence, and social valence alone, and especially in combination turn out to be relevant to an extraordinarily wide variety of legal contexts: criminal law and procedure, tort, intellectual property, speech, privacy, contract, tax, and maritime law, to name but a few.” CALO, 2015, p. 532.

<sup>85</sup> “To sum up, robots are best thought of as artificial objects or systems that sense, process, and act upon the world to at least some degree. But this is just a technical definition, akin to describing the networks and protocols that comprise the Internet. What turns out to be important for legal and policy discourse is not the precise architecture, but the possibilities and experiences the architecture generates and circumscribes.” CALO, 2015, p. 531.

robôs com características antropomórficas (que comumente são propositalmente projetados dessa maneira para se tornarem mais atraentes para a interação) como se fossem humanos<sup>86</sup>, propiciando uma sensação de acolhimento e reconhecimento de si no outro que, na verdade, é um objeto<sup>87</sup>.

Em contraponto a CALO (2015), BALKIN (2015), que não distingue entre robôs e IA<sup>88</sup>, critica o valor social (*social valence*) proposto pelo colega por entender que as pessoas não necessariamente respondem aos robôs como se fossem humanos. Elas também podem responder como se fosse um animal, sendo que a projeção humana em animais ou objetos inanimados não é uma novidade, nem tampouco está adstrita a robôs, pois são um reflexo do indivíduo no mundo exterior<sup>89</sup>.

---

<sup>86</sup> “There is a robust literature suggesting that people are hardwired to react to anthropomorphic technology such as robots as though they were interacting with a person. The threshold for what stimulates this reaction is low: early psychological studies show how 204 people attribute social roles to shapes that move around on a page. Any robot that exists physically in the world (acts) and appears to navigate the world with a measure of deliberation (senses and thinks) can invoke a social response. But robots are often intentionally designed to be anthropomorphic because it makes them more engaging. The effects, moreover, persist over time, and do not depend on one's familiarity with the technology.” CALO, 2015, p. 545-546.

<sup>87</sup> No mesmo sentido, Stuart Russel afirma que, de forma desonesta, os robôs ultrapassam a consciência humana e apelam diretamente para o lado emocional: “As research tools, the robots may provide insights into how humans interpret robot behavior and communication. As prototypes for future commercial products, they represent a form of dishonesty. They bypass our conscious awareness and appeal directly to our emotional selves, perhaps convincing us that they are endowed with real intelligence. Imagine, for example, how much easier it would be to switch off and recycle a squat, gray box that was malfunctioning—even if it was squawking about not wanting to be switched off—than it would be to do the same for JiaJia or Geminoid DK. Imagine also how confusing and perhaps psychologically disturbing it would be for babies and small children to be cared for by entities that appear to be human, like their parents, but are somehow not; that appear to care about them, like their parents, but in fact do not.” RUSSEL, 2019, p. 92.

<sup>88</sup> “I describe the problems posed by robotics for law a little differently than Calo does, in part because I do not distinguish sharply between robots and artificial intelligence (AI) agents. As innovation proceeds, the distinction between these two kinds of technologies may be far less important to the law than it seems at present; we do not yet know whether the boundaries between these two technologies will increasingly blur (as I expect) or further differentiate.” BALKIN, Jack M. **The path of robotics law**. Calif. L. Rev. Circuit, v. 6, p. 45, 2015, p. 45-46.

<sup>89</sup> “The third feature of robotics that Calo calls essential is the tendency of human beings to respond to robots as if they are interacting with people. He calls this tendency ‘social valence’ and argues that ‘to a greater degree than perhaps any technology in history, robots have a social meaning to people’. Of course, people do not necessarily respond to robots as if they were people. They may respond to them as if they were animals. Consider a robotic dog that one trains to do tricks, or a robotic seal designed for elderly patients to pet. What Calo is describing might be either anthropomorphism (ascribing human features) or zoomorphism (ascribing animal features). The more general point, however, is that human beings project agency on non-living things. Humans may also project emotions, feelings of pleasure and pain, the capacity to form relationships with others, and the capacity to care for others and be cared by them in turn. The projection of human or animal emotions onto inanimate objects is as old as history itself. People hear the wind howl and the ocean roar; they project agency and loyalty onto their ships and cars. The projection of humanity onto what is not human is the reflection of the self on the outside world.” BALKIN, 2015, p. 55-56.

Deste modo, BALKIN (2015) propõe o que denomina “efeito substituição” (*substitution effect*). Para ele, os robôs e mecanismos de IA substituirão seres vivos, especialmente os humanos. Porém, em determinadas maneiras e para determinados propósitos, de forma contextualizada, instável e de acordo com a conveniência e oportunidade<sup>90</sup>.

Segundo o jurista, o problema não seria as pessoas confundirem robôs com seres vivos, mas quererem a substituição de animais e seres humanos por máquinas em situações e circunstâncias específicas. A depender da função que exerçam, as pessoas tratariam o robô ou a IA como um animal ou um humano com propósitos especiais<sup>91</sup>.

Ainda segundo BALKIN (2015), a substituição pode oferecer habilidades e performances melhor que os humanos, disponibilizando um “substituto” que promete ser um “super-humano” citando os seguintes exemplos: (i) robôs de companhia que podem atender melhor às necessidades emocionais humanas e não se cansam; (ii) a IA que faz o papel do árbitro com mais rapidez e eficiência; (iii) ou o robô militar que ceifa o inimigo com mais eficiência e não sofre de *stress* pós-traumático.

Deste modo, o jurista indica situações em que é possível identificar a parcialidade na escolha do “ente substituto” considerado superior ao humano ou um “super-humano”. Logo, a questão, em verdade, seria menos a tecnologia utilizada e mais a interação com as pessoas que implementam essas tecnologias<sup>92</sup>.

---

<sup>90</sup> “I call it the “substitution effect”—the fact that people will substitute robots and AI agents for living things, and especially for humans. But they will do so only in certain ways and only for certain purposes. This substitution is likely to be contextual, unstable, and often opportunistic. The problem of substitution touches many different areas of law, and it promises to confound us for a very long time.” BALKIN, 2015, p. 46.

<sup>91</sup> “The problem is not that people confuse robots for living things, for usually they do not. Rather, the problem is that, through their interactions with robots and AI systems, people are willing to substitute them for animals or human beings in certain contexts and for certain purposes. Call this the substitution effect. People cause an entity to stand in for a human or animal and they treat the entity as such—but only in certain ways. In other words, people treat robots and AI agents as special-purpose animals or as special-purpose human beings.” BALKIN, 2015, p. 57.

<sup>92</sup> “Substitution not only promises the possibility of greater benefits with lower costs, but it may offer skills and performances better than humans can offer. In such cases, the substitute promises to be super human. The robot companion promises to better cater to our emotional needs, the AI trading program engages in arbitrage faster and more efficiently than the floor trader, the robot army mows down the enemy more efficiently, and so on. In fact, it is the very partiality of the substitution that makes it superior. The robot caretaker doesn’t tire of attending to our needs, and the robot soldier doesn’t suffer from posttraumatic stress disorder, because neither of them is fully human. I began this Essay by noting that technology concerns not merely the relationship of persons to things but

Da exposição das preocupações e conclusões dos estudiosos citados, depreende-se que há diferentes formas de abordar e nomear o problema. Seja por meio da indicação dos riscos da usurpação de papéis humanos, da atribuição de valor social aos robôs, ou da substituição dos humanos pela IA em determinadas funções, o alerta para a preservação da autonomia humana frente às máquinas está aceso.

É de se notar que as características antropomórficas englobam, além da aparência física (muito explorada nos robôs com finalidade sexual), outros atributos, tais como a voz e/ou a utilização de linguagem e artifícios que denotam a existência de supostas emoções ou empatia no comportamento da IA.

A exemplo, dispositivos que interagem com pessoas imitando a voz humana, tais como a Siri ou a Alexa, ou *chatbots* que se comunicam e “reagem” ao comportamento humano, que respondem com “humor” ou expressam outros sentimentos e reações tipicamente humanas, também adotam a linha da antropomorfia.

Esse jogo de palavras, de imagens, de sons, de respostas que brincam com o cérebro humano, induzindo o indivíduo a um estado de aceitação daquele ente como um igual e até mesmo de criação de conexões afetivas com os dispositivos<sup>93</sup>,

---

rather the social relationships between people that are mediated by things. When the “things” at issue are substitutes for people, or special-purpose people, the mediation of social relationships between human beings is even further obscured. Our interactions with robots and AI systems are interactions with the people who are deploying these new technologies, even when we do not realize it.” BALKIN, 2015, p. 59.

<sup>93</sup> Um exemplo de criação de conexão afetiva de um humano com uma IA e suas implicações é o caso do homem japonês que se casou com um holograma em 2018 e passou a vivenciar uma crise no casamento em 2022 devido à interrupção do serviço de suporte ao programa pelo fabricante: “O japonês Akihiko Kondo, de 38 anos, se casou com o holograma de uma cantora pop em 2018. A cerimônia que uniu o exótico casal custou 2 milhões de ienes, aproximadamente R\$ 77 mil. Após quatro anos, o professor vive uma crise em seu relacionamento. Segundo o jornal japonês Mainichi, ele não consegue mais se comunicar com a esposa devido a um obstáculo tecnológico. Em 2008, Kondo se apaixonou por Hatsune Miku, que é retratada como uma garota de cabelo azul-turquesa. O sentimento pela personagem começou depois que o bullying sofrido no ambiente de trabalho o fez ficar deprimido. Ele afirma que chegou a assistir 24 horas por dia aos vídeos da amada.” CUNHA, Maria. **Japonês casado com um holograma vive crise no relacionamento por falta de comunicação**: depois de quatro anos, casal passa por um momento difícil após dispositivo usado para conversas deixar de funcionar. R7. Internacional. Publicado em: 28.04.2022. Atualizado em 29.04.2022. Disponível em: <<https://noticias.r7.com/internacional/japones-casado-com-um-holograma-vive-crise-no-relacionamento-por-falta-de-comunicacao-29042022>>. Acesso em: 07.12.2022.

propicia que as máquinas venham a desempenhar atividades essencialmente humanas, tornando os humanos menos necessários para os próprios humanos.

Quando se fala em robôs humanoides, tecnologia antropomórfica e substituição de seres vivos por mecanismos de IA, todas essas definições remetem a uma ideia de presença de alguma forma física, palpável, que torna mais fácil identificar que há uma máquina agindo, interagindo com o mundo exterior, com o ambiente humano.

Porém, o maior perigo reside onde esse fenômeno não é fisicamente visível, onde sequer se tem conhecimento de que um mecanismo de IA está usurpando, interagindo ou substituindo o humano e, para além disso, tomando decisões. Por isso, a presença do humano no controle e fiscalização dos mecanismos decisórios autônomos da IA, especialmente no que diz respeito aos direitos fundamentais e existenciais, se mostra relevante.

## 2.3

### **A capacidade de tomada de decisões automatizadas conduz à da autonomia da IA?**

O presente capítulo tem explorado as especificidades tecnológicas e características da inteligência artificial que compõe o modo de funcionamento dos seus mecanismos de tomada de decisões autônomas, bem como os reflexos deste tipo de processo no ambiente humano.

Pelo que foi dito até então, tem-se que a IA está apta a interpretar, aprender e decidir dentro das possibilidades que lhe são oferecidas pelos dados fornecidos e algoritmos iniciais formulados. Portanto, a inteligência artificial tem “a capacidade de tomadas de decisão absolutamente autônomas em relação à interferência humana” (MULHOLLAND et al., 2020, p. 267-268).

Porém, a habilidade pré-programada para tomar determinados tipos de decisões no modo *solo* não indica, necessariamente, que a IA seja dotada do atributo

autonomia que, juridicamente, vem entrelaçado com direitos a serem exercidos e deveres a serem cumpridos.

Como adentrar na intrincada discussão sobre a eventual existência ou forma de classificação da personalidade jurídica da IA não é objeto do presente estudo, esse texto ater-se-á ao atual estágio tecnológico de desenvolvimento e forma de funcionamento do processo decisório nos mecanismos de IA para dissecar a ideia da autonomia da máquina.

### 2.3.1

#### **Conceito de singularidade tecnológica.**

Aqueles que discorrem sobre singularidade prospectam o futuro utilizando o desenvolvimento tecnológico atual como um degrau para a tentativa de previsão da evolução do comportamento das máquinas no que se refere à sua capacidade de autodeterminação. Trata-se de tentar estabelecer se e quando a IA será plenamente autônoma e quicá superior ao humano.

Segundo FALEIROS JÚNIOR (2021, p. 21), singularidade tecnológica é “o estado de aproximação entre o biológico e o tecnológico que permita ao algoritmo processar dados, formular hipóteses e apresentar soluções, mas também agir de forma arbitrária, livre e autônoma.”

Para WALSH (2017), que apresenta argumentos contra a ideia, a singularidade tecnológica é a hipótese de que, em algum momento no futuro, nós inventaremos máquinas com recursos de autoaperfeiçoamento, o que será um ponto de inflexão que resultará em um crescimento tecnológico descontrolado<sup>94</sup>.

Ao criticar o posicionamento de Ray Kurzweil expressado no livro “*The singularity is near: when humans transcend biology*”<sup>95</sup>, MCDERMOTT (2006)

---

<sup>94</sup> “The technological singularity, often simply called the singularity, is the hypothesis that at some point in the future we will invent machines that can recursively self improve, and that this will be a tipping point resulting in runaway technological growth.” WALSH, Toby. **The Singularity May Never Be Near**. The AI Magazine 38.3 (2017): 58-62. Web, p. 1.

<sup>95</sup> KURZWEIL, Ray. **The Singularity Is Near: When Humans Transcend Biology**. Viking Adult, ISBN 0670033847, September 2005, 672 pages.



pondera que o alarmismo proposto pelo autor, que colocaria o advento da eventual singularidade tecnológica como um evento destruidor do mundo, prejudica o desenvolvimento tecnológico consciente<sup>96</sup>.

Ao tratar da possibilidade de se alcançar a singularidade tecnológica por meio do desenvolvimento de uma inteligência artificial geral em nível humano, GOERTZEL (2007) discorda de MCDERMOTT (2006) por entender que KURZWEIL (2005) fez bem em trazer a público a preocupação com a possibilidade de desenvolvimento do que prefere chamar de inteligência artificial geral (AGI) ainda no século 21, defendendo que a hipótese da singularidade deve ser levada a sério<sup>97</sup>.

PETTA GOMES DA COSTA (2019), por sua vez, acredita que existe o “mito da próxima singularidade” no qual o conceito de singularidade tecnológica foi criado para prever a próxima singularidade, quando máquinas ultrapassariam a inteligência humana e, eventualmente, até se rebelariam contra os humanos.

---

<sup>96</sup> “Although I don’t think Kurzweil has proved his case about AI (and I have had nothing to say about his arguments on genetics, nanotechnology, politics, or ecology), he may be right. But right or wrong, I wish he would stop writing these books! The field of AI is periodically plagued by “hype hurricanes” that seem to surge up from the slightest perturbation. [...] Most of these storms are started or spun up by journalists, venture capitalists, or defense-department bureaucrats. But there are a very small number of people within the field of AI who contribute to the problem, and Kurzweil is the worst offender. He may think that the Singularity is such a world-shattering event that we should be thinking about it even if the chance of its happening is small, and that the possible damage done to AI is a trifle by comparison. Furthermore, if overhyping AI now causes all the funding to be sucked out of the field for a few years, his charts show that such events are negligible glitches in an inexorable exponential march. But his negligible glitch could be somebody else’s wrecked career. This business is tough enough without us shooting ourselves in the foot. Holster your gun, Ray!” MCDERMOTT, Drew. **Kurzweil's argument for the success of AI**. Artificial Intelligence, Volume 170, Issue 18, 2006, Pages 1227-1233, ISSN 0004-3702, p. 1233.

<sup>97</sup> “Though I disagree with Ray Kurzweil on a number of points of emphasis, and a few points of substance, I believe he is doing pretty much exactly the right thing, by getting the general public enthused about the possibilities AGI is very likely going to bring us during the 21st century. I think the time is more than ripe for the mainstream of AI research to shift away from narrowly-constrained problem-solving and back to explicit pursuit of the general-intelligence goals that gave the field its name in the first place. In the last few years I have seen some signs that this may in fact be happening, and I find this quite encouraging. The difficulty of predicting the future is not just a cliché, it’s a basic fact of our existence. And part of the hypothesis of the Singularity is that this difficulty is just going to get worse and worse. One of my main critiques of Kurzweil is that he is simply too pat and confident in his predictions about the intrinsically unpredictable—quite differently from Vinge, who tends to emphasize the sheer unknowability and inscrutability of what’s to come. In spite of all this uncertainty, however, we must still plan our own actions, our own lives and our own research careers. My contention is that the hypothesis of the Singularity is a serious one, worthy of profound practical consideration; and even more emphatically, that the pursuit of human-level AGI deserves to be taken very seriously, at very least as much so as grand pursuits in other areas of science and engineering (gene therapy; building artificial organisms a la Craig Venter’s work on *mycoplasma genitalium*; quantum computing; unifying physics; etc.)” GOERTZEL, Ben. **Human-level Artificial General Intelligence and the Possibility of a Technological Singularity - A Reaction to Ray Kurzweil's The Singularity Is Near, and McDermott's Critique of Kurzweil**. Artificial Intelligence 171.18 (2007): 1161-173. Web, p. 1172.

Acrescenta que a singularidade tecnológica não seria iminente, pois já teria acontecido e novas tecnologias seriam apenas variações e derivações das automações em larga escala ocorridas durante a revolução industrial.

Assim, ele discorda de KURZWEIL (2005) sob o ponto de vista de que tanto a IA, quanto outras tecnologias atuais, seriam o resultado da última singularidade. Além disso, a singularidade não seria uma ameaça aos humanos, mas uma forma de melhorar a sua biologia e inteligência por meio de uma eventual fusão do corpo humano com máquinas<sup>98</sup>.

Essa última formulação apresentada, o “mito da próxima singularidade”, coaduna com a ideia do “*odd paradox*” trazida anteriormente na conceituação da inteligência da IA, na qual a máquina deixa de ser considerada “inteligente” quando uma nova IA, mais eficiente, surge. CALO (2015, p. 528), por seu turno, defende que há razões analíticas e técnicas para acreditar que robôs jamais pensarão como pessoas.

Vê-se que o tema singularidade tecnológica é deveras controverso. Independentemente dos eventuais devaneios filosóficos e/ou utópicos (falar de futuro sempre requer um certo grau de abstração para que se possa projetar algo que ainda não existe ou não se conhece)<sup>99</sup>, a sua abordagem é relevante, especialmente, quando se está tratando da “autonomia” da tecnologia.

---

<sup>98</sup> “The concept of technological singularity was created to predict the next singularity, when machines would surpass human intelligence and perhaps ultimately even rebel against humans. Kurzweil calculated that at some point in the early 21st century, a new singularity would occur through the improvement of AI, and machines would surpass human beings using recent technologies with global impacts, such as the Internet. However, not only can this hypothesis be denied, but we may assume that the Internet, AI and many other new technologies are, in fact, the result of the last singularity. In other words, technological singularity is not imminent because it has already happened. For every recent achievement, the four characteristics that need to be met to constitute a technological singularity and the population explosion that follows it have already been found in the Industrial Revolution. These new technologies appear to be only variations or derivations of the large-scale automation typical of the Industrial Revolution. [...] In fact, instead of posing a superintelligence AI threat against humans, I believe the last singularity (i.e., automation) will change humans at such a level that our biology and intelligence will greatly improve—and possibly even include a merging of the human body and machines.” PETTA GOMES DA COSTA, De Leon. **Re-viewing the Concept of Technological Singularities: How Can It Explain Human Evolution?** *Nanoethics* 13.2 (2019): 119-30. Web, p. 128.

<sup>99</sup> Ao tratar dos limites e possibilidade de existência de uma IA superinteligente, Stuart Russel considera ser impossível prever os passos do desenvolvimento tecnológico: “While stretching your imagination, don’t stretch it too far. A common mistake is to attribute godlike powers of omniscience to superintelligent AI systems—complete and perfect knowledge not just of the present but also of the future. This is quite implausible because it requires an unphysical ability to determine the exact current state of the world as well as an unrealizable ability to simulate, much faster than real time,

Portanto, a preservação da autonomia humana na utilização da tecnologia perpassa, necessariamente, pela ideia da capacidade de autodeterminação da máquina no seu processo decisório. A exemplo, a hipótese de singularidade proposta por Ray Kurzweil somente se concretizaria se não houvesse a precaução em impor condutas éticas e legais que impedissem a sobreposição da vontade das máquinas sobre a vontade humana.

Entretanto, como visto ao longo deste capítulo, tem havido uma preocupação da sociedade e uma movimentação no campo do direito no sentido de colocar o humano no centro, como objetivo final da utilização da tecnologia. Logo, os limites da evolução tecnológica e suas implicações, assim como em qualquer outro ramo da ciência, estão sendo delineados a fim de impedir que a existência humana seja ameaçada.

### 2.3.2

#### **Níveis de automação da IA: exemplo dos carros autônomos.**

Considerando que a singularidade ainda não foi alcançada, resta entender em que estágio tecnológico se encontra a capacidade de automação da IA. Como os modelos dos programas são muito diferenciados a depender da extensão e objetivos de aplicabilidade da IA, tem-se utilizado o exemplo dos carros autônomos como parâmetro para compreensão dos seus estágios de desenvolvimento.

Segundo RUSSEL (2019), a Sociedade de Engenheiros Automotivos (Society of Automotive Engineers – SAE<sup>100</sup>) define seis níveis de automação<sup>101</sup>, sendo

---

the operation of a world that includes the machine itself (not to mention billions of brains, which would still be the second-most-complex objects in the universe). This is not to say that it is impossible to predict some aspects of the future with a reasonable degree of certainty [...]”. RUSSEL, 2019, p. 72.

<sup>100</sup> A SAE é uma organização internacional com filial no Brasil. Informações disponíveis em: <<https://saebrasil.org.br/quem-somos/>>. Acesso em: 08.12.2022.

<sup>101</sup> Nos termos da norma da SAE J3016\_202104, revisada em 30.04.2021, os seis níveis de sistemas de automação de direção para veículos rodoviários são: (i) nível 0 – nenhuma automação; (ii) nível 1 – assistência ao motorista; (iii) nível 2 – automação de direção parcial; (iv) nível 3 – automação de direção condicional; (v) nível 4 – automação de direção alta; (vi) nível 5 – automação de direção completa. Fonte: SAE International. **Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles J3016\_202104**. Disponível em: <[https://www.sae.org/standards/content/j3016\\_202104/](https://www.sae.org/standards/content/j3016_202104/)>. Acesso em: 08.12.2022.

o nível 0 nenhuma automação, e o nível 5 automação total, na qual o sistema é capaz de performar em tempo integral no modo direção automática considerando todos os aspectos da dinâmica de conduzir o veículo sob todas as condições diferenciadas da pista e ambientais, tal como um motorista humano<sup>102</sup>.

Os projetos atuais em andamento buscam atingir o nível 4 de autonomia, que seria o veículo plenamente capaz de dirigir autonomamente e parar com segurança, sujeito a limites geográficos e condições climáticas. Considerando que o veículo com autonomia de nível 4 não consegue lidar com mudanças nas condições climáticas e de tráfego, além de outras circunstâncias incomuns que podem surgir, é necessária a presença de um humano dentro do veículo, pronto para assumir o controle se preciso. Já o nível 5, autonomia irrestrita, não requereria um condutor humano, sendo muito mais difícil de ser alcançada.

Não obstante a imprescindibilidade do humano dentro do veículo, o nível 4 de autonomia vai muito além das tarefas simples e reflexas de seguir linhas brancas e evitar obstáculos. O veículo precisa avaliar a intenção e prováveis trajetos futuros de todos os objetos relevantes no caminho, incluindo objetos que podem não estar visíveis, baseado em observações presentes e passadas.

Assim, usando a ferramenta de busca antecipada, o veículo precisa encontrar o trajeto que otimize alguma combinação de segurança e progresso. Alguns projetos estão tentando abordagens mais diretas baseadas em aprendizado por reforço (principalmente em simulação) e aprendizado supervisionado de gravações de centenas de condutores humanos, mas parece improvável que essas abordagens alcancem o nível necessário de segurança<sup>103</sup>.

---

<sup>102</sup> “The Society of Automotive Engineers (SAE) defines six levels of automation, where Level 0 is none at all and Level 5 is full automation: “The full-time performance by an automatic driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver.” RUSSEL, 2019, p. 219.

<sup>103</sup> “Current projects are aiming at SAE Level 4 autonomy,<sup>6</sup> which means that the vehicle must at all times be capable of driving autonomously or stopping safely, subject to geographical limits and weather conditions. Because weather and traffic conditions can change, and because unusual circumstances can arise that a Level 4 vehicle cannot handle, a human has to be in the vehicle and ready to take over if needed. (Level 5—unrestricted autonomy—does not require a human driver but is even more difficult to achieve.) Level 4 autonomy goes far beyond the simple, reflex tasks of following white lines and avoiding obstacles. The vehicle has to assess the intent and probable future trajectories of all relevant objects, including objects that may not be visible, based on both current and past observations. Then, using lookahead search, the vehicle has to find a trajectory that optimizes some combination of safety and progress. Some projects are trying more direct approaches based on reinforcement learning (mainly in simulation, of course) and supervised learning from

Segundo GOMES (2021, p. 591), “atualmente é possível adquirir um automóvel como nível 3 de automação, o que significa dizer que, em determinadas situações, o próprio fabricante informa que não há necessidade de monitoramento da condução por parte do motorista.”

Essa classificação importa, pois, já que máquinas estão tomando decisões de forma autônoma sobre pessoas, é necessário que se entenda qual o grau de autonomia exercido nesse processo decisório. Partindo do exemplo do carro autônomo, percebe-se que hoje a IA não tem condições de tomar decisões que podem impactar na vida e segurança dos indivíduos sem a supervisão humana.

Assim, questiona-se se a capacidade de interpretar informações e definir ações que correspondam aos dados processados como um exercício efetivo de comportamento autônomo pela IA. Por essa razão, como se verá a seguir, há juristas que preferem classificar a “autonomia da IA” de modo diverso.

### 2.3.3

#### **Autonomia ou comportamento emergente?**

Consoante exposto anteriormente, CALO (2015, p. 532) coloca a “emergência” como uma das “qualidades essenciais” para a conceituação do robô. A “emergência” ou “comportamento emergente” (*emergent behavior*) é o termo utilizado pelo jurista para se referir à autonomia do robô<sup>104</sup>.

Segundo ele, o termo “autonomia” sugeriria que os robôs estão, de alguma forma, tomando decisões para agir de uma maneira específica, com uma intenção; enquanto o termo “emergência” se adequaria à junção da complexidade e utilidade, ou seja, regras simples aplicadas para realização de tarefas de aparente alta

---

recordings of hundreds of human drivers, but these approaches seem unlikely to reach the required level of safety.” RUSSEL, 2019, p. 52-53.

<sup>104</sup> “Emergent behavior is a clearly stated goal of robotics and artificial intelligence, going directly to the ‘think’ component of our earlier definition of the technology. Kenneth Anderson and Matthew Waxman describe why this capacity would be useful in a military setting. Robotic systems might be faster than people at reacting to battlefield developments, especially ones initiated by other machines. As Ronald Arkin’s work explores, a machine that is versatile enough to ‘learn’ from mistakes could stop itself (and people) from committing those mistakes in the future.” CALO, 2015, p. 538.

sofisticação. Utiliza como exemplo a forma como as formigas seguem regras simples para realizar tarefas complexas e aparentemente inteligentes<sup>105</sup>.

Deste modo, CALO (2015) enfatiza a diferença de autonomia, que estaria ligada à vontade de praticar determinado ato, do comportamento emergente, que seria um comportamento automático, por meio da realização de tarefas determinadas por regras simples, mas que apresentam um resultado complexo, e, às vezes, admirável.

Para DONEDA et al. (2018, p. 7) por “*comportamento emergente* se entende a capacidade do robô de aprender e de se adaptar às circunstâncias”, e ao comentar CALO (2015), pondera se, diante da imprevisibilidade da IA, é correto se afirmar que ela efetivamente tomou uma decisão autônoma<sup>106</sup>. Nas palavras de MULLHOLLAND et al. (2021, p. 71) emergência “representa a aptidão do robô em obedecer a comandos”.

Segundo BALKIN (2015), os problemas jurídicos causados pela IA se devem ao fato de que nem sempre é possível prever o que ela fará quando interage com o ambiente. Por isso, concorda com o termo “emergência” ou “comportamento emergente” utilizado por CALO (2015) no lugar de “autonomia”, já que esta última levanta questões difíceis de serem respondidas sobre o *status* da inteligência artificial.

---

<sup>105</sup> “I use the term ‘emergence’ instead of ‘autonomy’ by design. Autonomy suggests that robots are somehow making a decision to act in a particular way. Little is gained, and much is arguably lost, by pretending contemporary robots exhibit anything like intent. Instead, I would draw on Stephen Johnson’s work in emergence spanning across a number of disciplines. Johnson sees the essence of emergence as the coupling of complexity and usefulness, the movement of low-level rules to tasks of apparently high sophistication. A common example is the way ants follow simple rules to accomplish complex, seemingly intelligent tasks. Artificial intelligence pioneer Alan Turing understood the utility of emergence and, according to Johnson, contemporary designers of intelligent systems rely on the principles of emergence with greater and greater frequency.” CALO, 2015, p. 539-540.

<sup>106</sup> “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.” DONEDA et al., 2018, p. 7.

Ainda de acordo com BALKIN (2015), as características de um sistema de auto-aprendizagem (*machine learning*) aumentam a possibilidade de resolução de problemas e realização de tarefas de maneiras que não se possa prever ou esperar, o que seria mais uma justificativa para se utilizar o termo “emergência” em vez de “autonomia”<sup>107</sup>.

Essa distinção é relevante no campo da responsabilidade civil<sup>108</sup>, pois, a partir do momento em que se considera que uma máquina não é dotada de autonomia plena, ela não pode ser responsabilizada por seus atos, que são imprevisíveis. Logo, os humanos responsáveis pela sua criação, implementação, comercialização e operação seriam os entes autônomos responsáveis pelo comportamento emergente da máquina.

Assim, *mutatis mutandis*, se a máquina não é autônoma o suficiente para ser responsabilizada, também não se pode entender que seja autônoma para decidir questões relevantes e que afetam direitos existenciais de forma livre, sem supervisão humana (no direito há que se observar tanto os ônus quanto os bônus). Consequentemente, se a IA não tem a obrigação de responder, não pode ter o poder (direito) de decidir.

Por conseguinte, entender a capacidade tecnológica da IA de tomar decisões de forma autônoma como autonomia parece ser um equívoco. Automação não significa autonomia. Eis porque, mostra-se apropriada a utilização do termo “comportamento emergente”, deixando o instituto da autonomia para o exercício da liberdade humana por meio da expressão da sua vontade.

---

<sup>107</sup> “I would look at the issue differently. I would associate the interconnected nature of the Internet, digital generativity, and the multiple layers of innovation with the problem of emergence. That is because from the standpoint of law—as opposed to the standpoint of engineering—the problem posed by emergence is the problem of assigning responsibility for the unpredictable behavior of robots and AI systems. Calo emphasizes that this unpredictability comes from the complexity of algorithms, including those that learn from experience.<sup>32</sup> That is certainly an important cause of the problem for law, but it is not the only one. Multiple layers of innovation, the generativity of digital systems, and the easy flow of data are also reasons why we may be unable to predict what self-learning or interactive systems will do once we set them loose on the world. This is yet another reason why “emergence” is a better term than “autonomy.” A self learning system is not made more autonomous because of the fact that it is structured in layers, generative, modular, buggy, and the work of many hands. Rather, these features are related to the reasons why the law cares about emergence: they enhance the possibility that self-learning systems will solve problems and perform tasks in ways that we will not foresee or expect.” BALKIN, 2015, p. 54-55.

<sup>108</sup>Faz-se aqui a ressalva que o presente texto não tem por objeto o estudo da responsabilidade civil de atos praticados por mecanismos de IA, mas a referência se mostrou oportuna.

### 3.

## “Existirmos: a que será que se destina?”

### 3.1

#### A existência e a autonomia humana.

Pergunta de difícil resposta<sup>109</sup>, a busca do significado do termo “existência” vem sendo discutida pela filosofia ocidental desde a antiguidade<sup>110</sup>, não tendo o presente estudo, direcionado à seara jurídica, a pretensão de respondê-la.

Contudo, para fins de construção lógica do pensamento, a abordagem do conceito de existência, ainda que de forma superficial, parece ser um degrau necessário ao aprofundamento do conhecimento em questões ligadas à condição humana e, conseqüentemente, aos direitos existenciais.

Os filósofos existencialistas Jean Paul Sartre e Martin Heidegger adotaram conceitos de existência atrelados à autonomia humana<sup>111</sup>.

Segundo Sartre, a existência precede a essência, pois "o homem primeiramente existe, se descobre, surge no mundo; e somente depois se define"<sup>112</sup>. Assim, por não haver um Deus a determinar o seu destino, o homem é livre, sendo

---

<sup>109</sup> Charles H. Kahn, ao discorrer sobre a terminologia para existência no período clássico da filosofia grega: “A terminologia para existência é muito mais complicada e só posso fazer um esboço tosco do problema.” KAHN, Charles H. **Sobre o verbo grego ser e o conceito do ser**. Coord. Mauro Iglésias. Série filosofia antiga 1. Rio de Janeiro: Núcleo de estudos de filosofia antiga, departamento de filosofia da Puc-Rio, 1997, p. 81.

<sup>110</sup> “A história do termo ‘existência’ parece consistir amplamente de questões ainda não respondidas. [...] Na extensa discussão do conceito (ou dos conceitos) de Ser na filosofia grega de Parmênides a Aristóteles, o tema da existência não figura como um tópico distinto da reflexão filosófica. Meu objetivo aqui é defender e ilustrar essa afirmação, e ao mesmo tempo sugerir algumas razões por que o conceito de existência não é isolado como um tópico em si. Finalmente, vou levantar, de modo tentativo, a questão de se a negligência desse tópico representou ou não necessariamente uma desvantagem filosófica.” KAHN, 1997, p. 90-91.

<sup>111</sup> Martin Heidegger é um filósofo alemão existencialista que, apesar das controvérsias que o cercam devido à sua filiação ao partido nazista durante a Segunda Guerra Mundial, publicou um livro em 1954 que trata da tecnologia e sua relação com o humano cujo título em inglês é “*The question concerning technology and other essays*” (HEIDEGGER, 1977), já citado anteriormente, e que traz conceitos que têm afinidade com os temas explorados ao longo deste trabalho. Como se vê na nota de rodapé nº 112, Jean Paul Sartre, filósofo francês, também existencialista, cita Martin Heidegger ao definir existência. Por tais razões, a visão de ambos sobre existência foi utilizada como parâmetro no presente tópico.

<sup>112</sup> SARTRE, Jean Paul. **O existencialismo é um humanismo**. Os pensadores, XLV. São Paulo: Abril S.A. Cultural e Industrial/Paris: Les Éditions Nagel/Lisboa: Editorial Presença Ltca., 1973, p. 12.



responsável por suas escolhas, e, somente após estabelecer o traçado do seu caminho, descobre a sua essência<sup>113</sup>.

Para Heidegger, “a essência do *Dasein* está em sua existência”<sup>114</sup>. Todas as coisas, os entes, são, mas não existem, apenas o homem, que Heidegger denomina como *Dasein*, existe, justificando o conceito de *Existenz*<sup>115</sup>, que diz respeito às relações estabelecidas entre o *Dasein*, o ente privilegiado, e os demais entes<sup>116</sup>. Portanto, o único a gozar do plano da existência seria o ser humano.

Por conseguinte, os referidos filósofos defendiam que a liberdade de escolha e de pensamento, bem como a capacidade cognitiva do humano, definem a relação do homem com os outros seres, com o mundo e com a sociedade do seu tempo e de tempos vindouros<sup>117</sup>.

---

<sup>113</sup> “O existencialismo ateu, que eu represento, é mais coerente. Declara ele que, se Deus não existe, há pelo menos um ser no qual a existência precede a essência, um ser que existe antes de poder ser definido por qualquer conceito, e que este ser é o homem ou, como diz Heidegger, a realidade humana. Que significará aqui o dizer-se que a existência precede a essência? Significa que o homem primeiramente existe, se descobre, surge no mundo; e que só depois se define. O homem, tal como o concebe o existencialista, se não é definível, é porque primeiramente não é nada. Só depois será alguma coisa e tal como a si próprio se fizer. Assim, não há natureza humana, visto que não há Deus para a conceber. O homem é, não apenas como ele se concebe, mas como ele quer que seja, como ele se concebe depois da existência, como ele se deseja após este impulso para a existência; o homem não é mais que o que ele faz. Tal é o primeiro princípio do existencialismo.” SARTRE, 1973, p. 12.

<sup>114</sup> “Diferentemente da tradição escolástica - e assim como fizeram os existencialistas franceses e alemães - que diferenciava os conceitos de existência (*existentia*) e essência (*essentia*), Heidegger ‘une’ estes dois termos e afirma que a essência do *Dasein* está em sua existência (*‘Das ‘Wesen’ des Dasein liegt in seiner Existenz’*).” MARÇAL, J. C. “Existência” e “verdade” em Martin Heidegger. ISSN 1984-3879, SABERES, Natal – RN, v. 1, n.6, fev. 2011, p. 145.

<sup>115</sup> “Martin Heidegger, na sua obra *Ser e Tempo* (*Sein und Zeit*) de 1927, elabora um novo conceito para o termo existência. Heidegger faz a distinção entre existência (*Existenz*) e ser simplesmente dado (*Vorhandenheit*). Todas as coisas, os entes, são, mas não existem; apenas o homem, que Heidegger denomina *Dasein*, existe. *Vorhandenheit* fornece o sentido espacial e temporal que designa o modo de ser da coisa ‘enquanto o que se dá simplesmente antes e diante de qualquer especificação’. As coisas ‘são’, ou seja, a pedra ‘é’, a casa ‘é’, Deus ‘é’, mas apenas o homem, o *Dasein*, possui o privilégio ontológico de existir. *Existenz*, aqui, é entendida como o modo próprio que surge das relações recíprocas estabelecidas entre *Dasein* e ser, entre o *Dasein* e todas as entificações – e a entificação privilegiada é o homem, daí a afirmação de que só o homem existe, pois apenas este ente está posto na clareira do ser.” MARÇAL, 2011, p. 142-143.

<sup>116</sup> “These analyses (constituting a *Dasein* analytic) must be understood as an attempt at a ‘non-categorical’ ontology. Heidegger uses the word ‘Existenz’ to refer to the unique kind of being that *Dasein* ‘is’. Rather than categories, the basic ontological concepts to be developed here are accordingly called ‘existentiala’, among the most basic of which is being-in-the-world. As an ontological concept, this notion would function in place of, but in a manner not unlike, the concepts of substance and accident in the region of objects or things.” STAPLETON, Timothy. **Dasein as being-in-the-world**. Martin Heidegger key concepts. Edited by Bret W. Davis. First published in 2010 by Acumen Publishing Limited, p. 47.

<sup>117</sup> “Quando dizemos que o homem se escolhe a si, queremos dizer que cada um de nós se escolhe a si próprio; mas com isso queremos também dizer que, ao escolher-se a si próprio, ele escolhe todos os homens. Com efeito, não há dos nossos atos um sequer que, ao criar o homem que desejamos ser, não crie ao mesmo tempo uma imagem do homem como julgamos que deve ser. [...] Se a existência, por outro lado, precede a essência e se quisermos existir, ao mesmo tempo que construímos a nossa

Sob tal prisma, existência e autonomia estão diretamente ligadas. A liberdade de escolha, inerente ao ser pensante, justificaria a sua existência. Eis a relevância da tutela da autonomia existencial em um cenário no qual o poder decisório (que engloba a liberdade de escolha e pensamento), e a capacidade cognitiva (inteligência) vem sendo delegadas a entes não humanos.

### 3.1.1

#### Existência digna, solidária e igual.

Ao conceituar a existência sob a perspectiva civil-constitucional, BODIN DE MORAES (2010, p. 240) atrela o seu sentido ao da coexistência, ou seja, a existência de forma solidária. Assim, “o indivíduo existe enquanto em relação com outros (sentido de alteridade) e com o mundo a ele externo”<sup>118</sup>.

Segundo TEIXEIRA (2018, p. 80-81), “concretizar a dignidade é atribuir a cada pessoa a ampla liberdade para que ela construa a própria vida” e “adone-se de sua existência”, sendo essa liberdade traduzida em uma “autonomia condicionada à responsabilidade”<sup>119</sup>.

---

imagem, esta imagem é válida para todos e para toda a nossa época. Assim, a nossa responsabilidade é muito maior do que poderíamos supor, porque ela envolve toda a humanidade. [...] Assim sou responsável por mim e por todos, e crio uma certa imagem do homem por mim escolhida; escolhendo-me, escolho o homem.” SARTRE, 1973, p. 12-13.

<sup>118</sup> “O ser humano existe apenas enquanto integrante de uma espécie que precisa de outro(s) para existir (*rectius*, coexistir). A concepção outrora dominante teve, por longo tempo, o homem como um ser hermeticamente fechado ao mundo exterior, isolado, solitário em seu mundo interior, como se fosse uma ilha: era o chamado *homo clausus*. Esta concepção foi abandonada em prol da compreensão a ela oposta, isto é, aquela segundo a qual o indivíduo existe enquanto em relação com outros (o sentido de alteridade) e com o mundo a ele externo.” BODIN DE MORAES, Maria Celina. **Na medida da pessoa humana**: estudos de direito civil-constitucional. Rio de Janeiro: Renovar, 2010, p. 240.

<sup>119</sup> “[...] concretizar a dignidade é atribuir a cada pessoa a ampla liberdade para que ela construa a própria vida, realize suas necessidades, faça suas escolhas e “adone-se” de sua existência, dirigindo-a da forma como entender que lhe traga maior realização, pois as concepções de cada um devem ser consideradas uma vez que todos os valores são possíveis no Estado Democrático de Direito, que, como visto, tem o pluralismo como um dos pilares fundamentais. [...] Além dessa abertura e da ampla possibilidade para concretizar a dignidade, também deve ser considerada a liberdade que a pessoa tem, nos limites de suas particularidades. Não se trata, portanto, da liberdade no sentido liberal, mas de uma autonomia condicionada à responsabilidade e, por isso, dependente das condições materiais, vulnerabilidade individual, informação que cada pessoa tem sobre a situação existencial em jogo que demanda sua decisão. Assim, o diálogo deve ser estabelecido entre dignidade, autonomia e responsabilidade. É nessa trilogia que será possível uma efetiva possibilidade de cada pessoa construir, de forma livre, a própria personalidade, desenvolvê-la em todas as suas potencialidades, pois na base de toda e qualquer relação humana deve estar sempre presente o respeito à

Com efeito, a existência digna extrapola o conceito individual de existência para alcançar a existência fraterna, pautada na solidariedade e igualdade. Trata-se da liberdade de escolha e suas consequências sociais enfatizadas por Sartre e Heidegger. O homem existe em si e em relação ao próximo e ao mundo.

Lastreada na “construção kantiana”<sup>120</sup>, que norteou o conceito de dignidade “como valor intrínseco às pessoas humanas”<sup>121</sup>, BODIN DE MORAES (2010) desdobra o “substrato material da dignidade” em quatro postulados<sup>122</sup> que dizem respeito à igualdade, integridade física e moral – psicofísica, liberdade e solidariedade<sup>123</sup>.

Tais valores permitem o pleno exercício da autonomia e autodeterminação (o humano decidindo os rumos da sua vida), e a preservação da coexistência no exercício desta autonomia (liberdade dos indivíduos ou instituições encontrando limites na liberdade alheia, e fomentando, assim, a reciprocidade).

Tem-se assim, a necessidade de tutela da autonomia existencial de forma individual, na qual as especificidades relativas aos indivíduos devem ser respeitadas de modo a preservar a dignidade (art. 1º, III, CF) e a igualdade (art. 5º, *caput*); e de forma coletiva, na qual o bem comum, pautado na solidariedade (art. 3º, I da CF), deve prevalecer.

No contexto de tomada de decisões por mecanismos de IA, tanto a tutela da dignidade do indivíduo afetado pelos vereditos da máquina, quanto dos reflexos de

---

dignidade.” TEIXEIRA, Ana Carolina Brochado. **Autonomia Existencial**. Revista Brasileira de Direito Civil – RBDCivil | Belo Horizonte, vol. 16, p. 75-104, abr./jun. 2018, p. 80-81.

<sup>120</sup> “O imperativo categórico está contido na sentença: ‘Age de tal modo que a máxima de tua vontade possa sempre valer simultaneamente como um princípio para uma legislação geral’. Esta formulação foi desdobrada por Kant em três máximas morais. São elas: i) ‘Age como se a máxima de tua ação devesse ser erigida por tua vontade em lei universal da natureza’, que corresponde à universalidade da conduta ética, válida em todo tempo e lugar; ii) ‘Age de tal maneira que sempre trates a humanidade, tanto na tua pessoa como na pessoa de outrem, como um fim e nunca como um meio’, que representa o cerne do imperativo, pois afirma a dignidade dos seres humanos como pessoas; iii) ‘Age como se a máxima de tua ação devesse servir de lei universal para todos os seres racionais’, que exprime a separação entre o reino natural das causas e o reino humano dos fins, atribuindo à vontade humana uma vontade legisladora geral.” BODIN DE MORAES, 2010, p. 80-81.

<sup>121</sup> BODIN DE MORAES, 2010, p. 85.

<sup>122</sup> “O substrato material da dignidade deste modo entendida pode ser desdobrado em quatro postulados: i) o sujeito moral (ético) reconhece a existência dos outros como sujeitos iguais a ele; ii) merecedores do mesmo respeito à integridade psicofísica de que é titular, iii) é dotado de vontade livre, de autodeterminação, iv) é parte do grupo social, em relação ao qual tem a garantia de não vir a ser marginalizado.” BODIN DE MORAES, 2010, p. 85.

<sup>123</sup> BODIN DE MORAES, 2010, p. 85.

tais decisões na coletividade e na forma com que os indivíduos se relacionam importam.

## 3.2

### **A relevância da tutela da autonomia existencial no contexto de tomada de decisões autônomas.**

#### 3.2.1

##### **Projeto de vida e pluralidade.**

Para VIVEIROS DE CASTRO (2017a), a autonomia<sup>124</sup> é um desdobramento do princípio da dignidade, devendo o projeto de desenvolvimento pessoal do sujeito partir da sua relação com outros sujeitos e com a comunidade em geral<sup>125</sup>, transformando o papel atual da autonomia em guia das relações sociais “de tal modo que o reconhecimento recíproco da condição de sujeitos torne possível que a sociedade goze democraticamente de esferas autônomas de desenvolvimento pessoal.”<sup>126</sup>

Nessa senda, os valores jurídicos da liberdade e da solidariedade aparecem se contrapondo e se equilibrando, o que faz com que a autonomia assuma papel central na análise do contexto de tomada de decisões autônomas por IA, especialmente quando tais decisões afetam aspectos existenciais e direitos fundamentais dos indivíduos e da coletividade.

---

<sup>124</sup> VIVEIROS DE CASTRO ao conceituar a autonomia existencial: “Em breves linhas, é possível afirmar que a autonomia existencial é espécie do gênero autonomia privada e se configura como instrumento da liberdade individual para realização das potencialidades da pessoa humana e de seus interesses não patrimoniais, incidindo nas situações jurídicas subjetivas situadas na esfera extrapatrimonial, cujo referencial objetivo é o próprio titular no espaço de livre desenvolvimento da personalidade.” VIVEIROS DE CASTRO, 2017a, p. 101.

<sup>125</sup> “[...] a autonomia que se desdobra do princípio da dignidade da pessoa humana demanda que o projeto de livre desenvolvimento pessoal seja concebido a partir da relação entre o sujeito e os demais interesses que decorrem da sua relação com os outros sujeitos e com a comunidade em geral.” VIVEIROS DE CASTRO, Thamís Dalsenter. **Bons costumes no direito civil brasileiro**. São Paulo: Almedina, 2017b. E-book Kindle, capítulo 1, item 1.1, n.p.

<sup>126</sup> VIVEIROS DE CASTRO, 2017b, capítulo 1, item 1.1, n.p.

Tem-se, deste modo, a proteção pelo ordenamento jurídico não só da autonomia individual, mas também da autonomia interpessoal e coletiva como classificado por VIVEIROS DE CASTRO (2017b) ao formular a teoria tríplice da autonomia existencial<sup>127</sup>.

Defendendo a tutela da pessoa como princípio geral da ordem pública constitucional, PERLINGIERI (2008) eleva a exigência do respeito à personalidade e seu livre desenvolvimento a requisito para a definição da interpretação dos atos, conceituação do ilícito, configuração das relações jurídicas (inclusive patrimoniais), incidindo assim sobre toda a organização da vida em comunidade<sup>128</sup>.

Por conseguinte, no processo de despatrimonialização do direito civil, no qual o “ser” passa a ser valorado em detrimento do “ter”, a autonomia existencial ganha papel de destaque, restando valorizados o pluralismo e a função promocional do direito<sup>129</sup>.

Por essa perspectiva, questiona-se se é possível haver sociedade plúrima e efetivação da função promocional do direito quando algoritmos cartesianos passam a tomar as decisões não somente no lugar do humano, mas sobre o humano.

A IA não é infalível, assim como não o é o humano. Porém, a capacidade de cognição e empatia humanas tornam o processo de tomada de decisões diferenciado, pois tais características são ferramentas que propiciam avaliações mais específicas, considerando as peculiaridades dos indivíduos e das circunstâncias<sup>130</sup>.

---

<sup>127</sup> Partindo da sua definição de autonomia privada, VIVEIROS DE CASTRO (2017b) formula a teoria tríplice da autonomia existencial, classificando os atos de autonomia em: i) *de eficácia pessoal*, que apresentam consequências relevantes unicamente na esfera jurídica do seu titular; ii) *de eficácia interpessoal*, que geram repercussão em esferas jurídicas distintas do titular da situação, alcançando pessoas que não praticaram o ato de autonomia (lesão ou risco real de lesão); iii) *de eficácia social*, que apresentam efeitos jurídicos diretos e imediatos que geram ou podem gerar lesão a direitos de um número indeterminado de pessoas (risco real de ofensa de direitos de pessoas não necessariamente identificadas, pode se dar em abstrato). VIVEIROS DE CASTRO, 2017b, capítulo 1, item 1.3.1, n.p.

<sup>128</sup> PERLINGIERI, Pietro. **O direito civil na legalidade constitucional**. Trad. Maria Cristina De Cicco. Rio de Janeiro: Renovar, 2008, p. 168-169.

<sup>129</sup> “[...] a releitura do Direito Civil, com a passagem de seus princípios fundadores do Código à Constituição, revela processo de profunda transformação social, em que a autonomia privada passa a ser remodelada por valores não patrimoniais, de cunho existencial inseridos na noção de ordem pública. O indivíduo, elemento subjetivo basilar e neutro do Direito Civil codificado, deu lugar, no cenário das relações de Direito Privado, à pessoa humana, para cuja promoção se volta a ordem jurídica como um todo.” TEPEDINO, Gustavo; TEIXEIRA, Ana Carolina Brochado; ALMEIDA, Vitor. **O direito civil entre o sujeito e a pessoa: estudos em homenagem ao professor Stefano Rodotà**. Belo Horizonte: Fórum, 2016, p. 17.

<sup>130</sup> “Diferentemente dos robôs, os humanos são capazes de avaliar criticamente uma conduta ou ideia que, mesmo que guiada por razões capazes de explicá-las, nem sempre vão ser boas o suficiente para

Segundo Hannah Arendt<sup>131</sup>, “a pluralidade humana é a paradoxal pluralidade de seres singulares”<sup>132</sup>. Logo, o pluralismo, inerente à condição humana, caminha de mãos dadas com a ideia de individualidade, já que cada ser é único e tem características próprias que desenham o seu projeto individual e social.

Em decorrência do universo de peculiaridades característico da condição humana, assim como é impossível ao legislador prever todas as situações jurídicas presentes e futuras a serem regidas por uma determinada norma (que tem cunho genérico e abstrato), é impossível ao programador do algoritmo (que está adstrito ao banco de dados escolhido para alimentação de cada programa específico) prever todos os cenários que, eventualmente, possam se apresentar no processo decisório da IA<sup>133</sup>.

---

justificá-las. [...] Apesar de o conceito de Inteligência Artificial (IA) depender de uma forte associação com a inteligência humana, os processos cognitivos de artefatos não-humanos e humanos são bastante diferentes, diante da forma como eles apreendem conhecimento. Ao mesmo tempo que esta diferença transforma radicalmente a eficiência destas máquinas, ela limita a sua capacidade de interpretar determinados contextos e “pensar e agir como um humano.” MULHOLLAND, Caitlin; FRAJHOF, Isabella Z. **Entre as leis da robótica e a ética**: regulação para o adequado desenvolvimento da inteligência artificial. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Felipe Braga Neto, et al. **Direito digital e inteligência artificial**: diálogos entre Brasil e Europa. Indaiatuba: Editora Foco, 2021, p. 66.

<sup>131</sup> Hannah Arendt é citada por BODIN DE MORAES (2010) ao tratar do conceito filosófico-político de dignidade: “Se não fossem iguais, os homens não seriam capazes de se compreenderem entre si e aos seus ancestrais, nem de prever as necessidades das gerações futuras. Se não fossem diferentes, os homens dispensariam o discurso ou a ação para se fazerem entender, pois com simples sinais e sons poderiam comunicar suas necessidades imediatas e idênticas. A pluralidade humana, afirma Hannah Arendt, tem este duplo aspecto: o da igualdade e o da diferença.” BONDIN DE MORAES, 2010, p. 76.

<sup>132</sup> “A pluralidade humana, condição básica da ação e do discurso, tem o duplo aspecto de igualdade e diferença. Se não fossem iguais, os homens seriam incapazes de compreender-se entre si e seus ancestrais, ou de fazer planos para o futuro e prever as necessidades das gerações vindouras. Se não fossem diferentes, se cada ser humano não diferisse de todos os que existiram, existem ou virão a existir, os homens não precisariam do discurso ou da ação para se fazerem entender. Com simples sinais e sons, poderiam comunicar suas necessidades imediatas e idênticas. Ser diferente não equivale a ser outro – ou seja, não equivale a possuir essa curiosa qualidade de «alteridade», comum a tudo o que existe e que, para a filosofia medieval, é uma das quatro características básicas e universais que transcendem todas as qualidades particulares. A alteridade é, sem dúvida, aspecto importante da pluralidade; é a razão pela qual todas as nossas definições são distinções e o motivo pelo qual não podemos dizer o que uma coisa é sem distingui-la de outra. Em sua forma mais abstrata, a alteridade está presente somente na mera multiplicação de objetos inorgânicos, ao passo que toda vida orgânica já exibe variações e diferenças, inclusive entre indivíduos da mesma espécie. Só o homem, porém, é capaz de exprimir essa diferença e distinguir-se; só ele é capaz de comunicar a si próprio e não apenas comunicar alguma coisa – como sede, fome, afeto, hostilidade ou medo. No homem, a alteridade, que ele tem em comum com tudo o que existe, e a distinção, que ele partilha com tudo o que vive, tornam-se singularidade, e a pluralidade humana é a paradoxal pluralidade de seres singulares.” ARENDT, Hannah. **A condição humana**. Trad. Roberto Raposo. Rio de Janeiro: Forense Universitária, 2007, p. 188-189.

<sup>133</sup> “As diretrizes algorítmicas não se comparam às disposições legislativas ou regulamentares, pois estas são preceitos genéricos e abstratos de alcance geral, desprovidas de discriminações indevidas,

Há que se acrescentar a esse raciocínio, o já mencionado fator de autoaprendizagem das máquinas, que pode gerar certa imprevisibilidade<sup>134</sup>. A possibilidade de falha não indica que a IA seja ineficiente ou não deva ser utilizada em processos decisórios. A eventual ocorrência de erros ou imprecisões decorrentes do uso da IA parece ser consequência natural dos processos de tomada de decisão e que, por meio da experiência e dos precedentes que vão sendo criados, pode propiciar um constante aprimoramento do sistema.

Todavia, tal perspectiva conduz a reflexões quanto aos riscos de se atribuir à IA o poder autônomo de decisão sobre questões existenciais, que digam respeito aos direitos fundamentais previstos na Declaração Universal dos Direitos Humanos de 1948, tais como o direito à vida, liberdade, igualdade, segurança, moradia, educação, trabalho, saúde, reproduzidos na Constituição Federal de 1988 em seus artigos 5º, 6º e 7º.

### 3.2.2

#### **Direitos fundamentais e questões existenciais.**

Referindo-se ao sistema público norte-americano, CITRON (2007) afirma que no passado, sistemas computacionais ajudavam humanos a aplicar regras para casos individuais, mas agora, os sistemas automatizados se tornaram os principais tomadores de decisão.

Com efeito, frequentemente, tais sistemas eliminariam a atuação humana nos processos de tomada de decisão que julgam, por exemplo: a exclusão de indivíduos de programas públicos de saúde ou alimentação; quem pode ou não fazer uso de transporte aéreo por questões de segurança; quem são os pais que estariam utilizando indevidamente auxílios públicos destinados aos filhos menores e

---

decididas no âmbito estatal, que muito pouco levam em conta, em prol da igualdade e da imparcialidade, as situações particulares dos seus destinatários.” SOARES, 2021, p. 49.

<sup>134</sup> “Assim, além de uma atuação de algoritmo com compasso predeterminado e previsível, os sistemas de inteligência artificial são desenvolvidos com uma *capacidade de autoaprendizado* e, por isso, estão aptos a executar operações autonomamente em face de prévias escolhas automáticas, ou, em linguagem coloquial, de *tomar decisões* ‘resultantes de uma combinação de *inputs* de programação não originária’, que podem não ser necessariamente as mesmas para situações simulares, porque esse caminho será trilhado pelo sistema de acordo com o resultado de uma experiência anterior.” SOARES, 2021, p. 46.

determinando que sejam adotados procedimentos de cobrança contra eles; quais eleitores estão impedidos de votar sem emitir notificação prévia; quem são as pequenas empresas consideradas inelegíveis para contratos com o governo federal<sup>135</sup>.

Conclui que a automação gera problemas imprevisíveis de cerceio de importantes direitos fundamentais. Alguns sistemas cerceiam em segredo, enquanto outros não dispõem de registros auditáveis, tornando impossível a revisão das questões jurídicas e fáticas que tenham embasado as suas decisões<sup>136</sup>.

SOUZA et. al. (2020, p. 74), “apresenta um agrupamento de soluções tecnológicas fundadas em Inteligência Artificial (doravante IA) e seus respectivos contrapontos” que incluem a possibilidade de utilização de tomada de decisões por IA em diferentes searas relevantes para a autonomia existencial tais como saúde, segurança, trabalho, justiça, transporte e relações sexuais<sup>137</sup>.

Ao tratar da utilização da IA nas relações sexuais, pondera que o uso pode “potencializar a sexualidade humana” e ajudar no “desenvolvimento das potencialidades sexuais de sujeitos que apresentem dificuldades, de ordem psicológica ou física, para desenvolver relações sexuais humanas”.

Como contraponto, ressalta a preocupação da utilização de robôs sexuais em “comportamentos sexuais considerados como socialmente reprováveis, a exemplo da pedofilia e do estupro”, o que representaria um ato “moralmente rejeitável e que pode implicar repercussões externas de desprezo à dignidade e à figura humana em geral.”<sup>138</sup>

STONE et al. (2016) descreve os domínios nos quais acredita que haverá transformações significativas devido à utilização das tecnologias de IA até 2030, quais sejam: transporte, serviços domésticos, saúde, educação, populações de baixa

---

<sup>135</sup> “In the past, computer systems helped humans apply rules do individual cases. Now, automated systems have become the *primary* decision makers. These systems often take human decision making out of the process of terminating individuals’ Medicaid, food stamp, and other welfare benefits. Another of these systems targets people for exclusion from air travel. Computer programs identify parents believed to owe child support and instruct state agencies do file collection proceedings against those individuals. Voters are purged from the rolls without notice, and small businesses are deemed ineligible for federal contracts.” CITRON, 2007, p. 1252.

<sup>136</sup> “Automation generates unforeseen problems for the adjudication of important individual rights. Some systems adjudicate in secret, while others lack recordkeeping audit trails, making review of the law and facts supporting a system’s decisions impossible.” CITRON, 2007, p. 1253.

<sup>137</sup> SOUZA et al., 2020, p. 63-79.

<sup>138</sup> SOUZA, et. al., 2020, p. 75.



renda, segurança pública e policiamento, emprego e espaço de trabalho e entretenimento<sup>139</sup>.

No campo da saúde, no que se refere mais especificamente aos cuidados com a saúde dos idosos, propõe que haverá interesse crescente e mercado para tecnologias já disponíveis e em desenvolvimento com o fim de auxiliar na saúde física, emocional, social e mental dos idosos.

Cita como exemplos: (i) dispositivos que monitorarão os movimentos e atividades do indivíduo e serão acoplados a plataformas de mídia social, que, por seu turno, poderão fazer recomendações sobre a saúde mental e física do idoso; (ii) dispositivos de monitoramento de saúde domésticos capazes de detectar mudanças de humor ou comportamento e mandar alertas aos cuidadores<sup>140</sup>.

Imagine-se, assim, um mundo em que o Instagram<sup>141</sup>, lastreado em dados que colhe do monitoramento de atividades e movimento de uma idosa, está autorizado, oficialmente, e respaldado legalmente a fazer recomendações de saúde mental e física para ela; ou que câmeras ou outros dispositivos de vigilância e detecção de movimentos, calor etc., instalados dentro da casa de um idoso, enviem alertas de mudança de humor e comportamento para familiares.

Na primeira hipótese, observa-se a substituição de profissionais da área da saúde (médicos, psicoterapeutas, fisioterapeutas) por uma IA controlada por uma empresa de mídia social. Tem-se dados extremamente sensíveis e pessoais do indivíduo sendo tratados e disponibilizados para uma *big tech* que, por meio de um

---

<sup>139</sup> “Though different instances of AI research and practice share common technologies, such as machine learning, they also vary considerably in different sectors of the economy and society. We call these sectors “domains,” and in this section describe the different states of AI research and implementation, as well as impacts and distinct challenges, in eight of them: transportation; home/service robotics; healthcare; education; low-resource communities; public safety and security; employment and workplace; and entertainment. Based on these analyses, we also predict trends in a typical North American city over the next fifteen years. Contrary to AI’s typical depiction in popular culture, we seek to offer a balanced overview of the ways in which AI is already beginning to transform everyday life, and how those transformations are likely to grow by the year 2030.” STONE et. al., 2016, p. 18.

<sup>140</sup> “Currently, someone who is seventy was born in 1946 and may have first experienced some form of personalized IT in middle age or later, while a fifty-year-old today is far more technology-friendly and savvy. As a result, there will be a growing interest and market for already available and maturing technologies to support physical, emotional, social, and mental health. Here are a few likely examples by category: [...] Health and wellness • Mobile applications that monitor movement and activities, coupled with social platforms, will be able to make recommendations to maintain mental and physical health. • In-home health monitoring and health information access will be able to detect changes in mood or behavior and alert caregivers.” STONE et. al., 2016, p. 30.

<sup>141</sup> Trata-se da escolha de uma mídia social de forma meramente ilustrativa.

algoritmo, decidirá se a idosa deve sair para caminhar, ler um livro ou ter mais horas de sono.

Na segunda hipótese, vê-se um ser humano que fica sujeito ao julgamento de uma máquina para definir qual o comportamento adequado e esperado dele, que, em uma inversão de papéis, apresenta-se quase que como um humano robô, ao qual é cerceado o direito de privacidade, liberdade e exercício da sua individualidade, não podendo se aborrecer, se entristecer, ficar eufórico ou cansado sem ter que dar satisfação a terceiros.

Em tal contexto, não existe privacidade, liberdade, solidariedade, um olhar individualizado para as necessidades de cada um ou um cuidado por profissionais especializados. Nesse mundo mais prático, mais barato e sob total controle, não há espaço para a autonomia existencial.

A utilização da IA para cuidados com a saúde é uma realidade e pode trazer uma série de benefícios para as pessoas, inclusive os idosos. Porém, os limites éticos e jurídicos precisam ser estabelecidos e respeitados. É importante que o humano seja colocado no centro, que haja transparência, um consentimento consciente que propicie o exercício da autodeterminação informativa<sup>142</sup> e o controle sobre a sua própria existência.

### 3.2.3

#### **A delegação da autonomia no uso da IA.**

A premissa estabelecida no capítulo anterior define que a IA não é plenamente autônoma, em verdade, apresenta um comportamento emergente. Logo, ela

---

<sup>142</sup> “No âmbito da União Europeia, graças a diversas diretivas e suas transposições aos ordenamentos nacionais, nasceu um modelo diverso. O seu ponto de partida é representado pelo reconhecimento do direito à autodeterminação informativa, o que não significa simplesmente atribuir a cada um o poder de impedir determinados usos das informações a si relacionadas, segundo a ótica originária do direito a ser deixado só. Significa acima de tudo o poder de controlar, a cada momento, o uso que outros façam das minhas informações. Os elementos-chave desse modelo, portanto, são o consentimento do interessado e o seu direito de acesso a todas as coletâneas de informações. Trata-se de um poder difuso, que não requer mediações burocráticas, visto que o acesso é exercido diretamente pelo interessado frente a todos os sujeitos, públicos e privados, que recolhem dados pessoais.” RODOTÀ, Stefano. **A vida na sociedade da vigilância: a privacidade hoje**. Organização, seleção e apresentação: Maria Celina Bodin de Moraes. Rio de Janeiro: Renovar, 2008, p. 148-149.

“existe” como objeto, não como sujeito<sup>143</sup>. Contudo, é um objeto que, sob um poder delegado pelo homem<sup>144</sup>, pode interferir na autonomia existencial individual e coletiva.

Em vez de haver uma pessoa, física ou jurídica, agindo de forma direta de um lado e alterando a esfera de autonomia do outro indivíduo, há uma máquina, desprovida do atributo dignidade, ocupando a posição de ação e interferência na existência humana. Portanto, não há reciprocidade direta.

Ao delegar o poder decisório à IA, os indivíduos ou as instituições exercem uma escolha autônoma. Esta escolha, praticada por sujeitos de direito, atribui a um ente que não é sujeito de direito, e, conseqüentemente, não é plenamente autônomo, um poder decisório sobre pessoas titulares de autonomia, inclusive sobre as suas questões existenciais.

Como dito no capítulo anterior, a forma de utilização da tecnologia importa, já que o seu modo de emprego é fruto de escolhas e das relações humanas. Segundo HEIDEGGER (1977), tudo depende da forma como ela é manipulada como meio e a necessidade de dominá-la se torna mais urgente quando a tecnologia ameaça fugir do controle humano<sup>145</sup>.

Devido ao tipo de tecnologia utilizada no funcionamento da IA (*machine e deep learning*), pode haver uma imprevisibilidade nos resultados. Os vieses nos processos decisórios não se dão por ato voluntário da máquina, já que o ente mecânico não é dotado de autonomia da vontade.

---

<sup>143</sup> “*Personalidade jurídica*, atributo essencial ao ser humano, é a *aptidão para possuir direitos e deveres, que a ordem jurídica reconhece a todas as pessoas*. Em nosso Direito, esse reconhecimento é feito pelo art. 1º do Código Civil: ‘Toda pessoa é capaz de direitos e deveres na ordem civil.’ Todo fato regulado por norma jurídica constitui sempre um vínculo entre pessoas. Sujeito ou *titular* é o portador de direitos ou deveres em uma relação jurídica.” NADER, Paulo. **Introdução ao estudo do direito**. Rio de Janeiro: Forense, 2006, 26ª edição, p. 288.

<sup>144</sup> Sobre a forma de delegação do poder decisório à IA: “Os sistemas de IA atuais são orientados para objetivos, ou seja, recebem de um ser humano a especificação de um objetivo e utilizam algumas técnicas para o atingir. Não definem os seus próprios objetivos. No entanto, alguns sistemas de IA (tais como os baseados em certas técnicas de aprendizagem automática) podem ter mais liberdade para decidir o caminho a seguir para atingirem o objetivo estabelecido.” GPAN IA, 2019b, p. 6.

<sup>145</sup> “But this much remains correct: modern technology too is a means to an end. That is why the instrumental conception of technology conditions every attempt to bring man into the right relation to technology. Everything depends on our manipulating technology in the proper manner as a means. We will, as we say, “get” technology “spiritually in hand.” We will master it. The will to mastery becomes all the more urgent the more technology threatens to slip from human control.” HEIDEGGER, 1977, p. 5.

Logo, as discriminações ocorridas são fruto de um ato voluntário de quem fez a programação (de forma consciente ou não); ou decorrem de distorções que surgem ao longo do processo de autoaprendizagem e não passaram por qualquer tipo de escrutínio humano.

Assim, ambos os polos da relação, tanto quem programa e/ou faz uso da IA, quanto quem está sujeito às decisões por ela produzidas, podem ter a sua esfera de autonomia afetada pelo uso indevido da IA.

Para BALKIN (2015, p. 49), falar de tecnologia, significa falar de interação entre humanos e dos humanos com as máquinas, referindo-se às relações humanas de poder e autoridade inseridas na forma de organização, instituições e modo de produção que interferem em aspectos existenciais como sexualidade, associação e vida em família.

Portanto, quando se fala da autonomia da IA, não está se falando da relação humano-máquina, pois o artefato não tem o nível de consciência necessário para estabelecer esse tipo de vínculo (uma relação jurídica ou um negócio jurídico). Trata-se, na maioria das vezes, da relação instituição-homem, a exemplo: poder público-cidadão, agente financeiro-cliente, hospital-paciente, universidade-aluno, empregador-empregado, e tantas outras instituições que representam o modelo de organização da sociedade hodierna.

### **3.3**

#### **A questão do controle.**

##### **3.3.1**

#### **Vigilância, classificação e punição.**

Tratar de processo decisório significa tratar de poder: poder de decisão, poder de estabelecer regras, poder de imprimir padrões, poder de fiscalizar, poder de classificar, poder de categorizar, poder de conceder, poder de retirar, poder de

premiar, poder de punir. O controle pertence a quem detém o poder que sujeita os demais.

A delegação do poder decisório às máquinas decorre do exercício inicial de autonomia de quem delega, mas, não necessariamente, do consentimento consciente de quem é o objeto das decisões. Mesmo que haja consentimento, a depender do contexto, “existem condicionamentos tais que excluem uma real possibilidade de escolha”<sup>146</sup>.

Segundo RODOTÀ (2008, p. 149) ao discorrer sobre a necessidade de proteção dos dados sensíveis<sup>147</sup>, há um “desnível de poder entre cidadãos e os que recolhem informações”, e a depender da dimensão dessa desigualdade, o indivíduo pode estar sujeito a “pressões e condicionamentos que tornam vãs as possibilidades efetivas de controle por parte dos cidadãos”, transformando o seu consentimento em “um mero requisito formal”<sup>148</sup>.

Ao propor um caminho para uma “cidadania eletrônica”<sup>149</sup> em “uma sociedade do controle, da vigilância e da classificação” em expansão, RODOTÀ (2008, p. 146) alerta para a existência do que chama de “Pequenos Irmãos” (*Small Brothers* - um paralelo à ideia do *Big Brother* de George Orwell), representados por diversos artefatos de vigilância cotidiana tais como câmeras, programas de coletas de dados, de perfilamento de indivíduos, e a interconexão entre diversos bancos de dados.

Nos dispositivos que podem ser classificados como *Small Brothers*, há em comum o uso da tecnologia da inteligência artificial e a delegação de poder para que tais mecanismos tomem decisões pelos humanos (a exemplo, “recomendar” que produto seria mais “adequado” para um consumidor) ou sobre os humanos (a

---

<sup>146</sup> RODOTÀ, 2008, p. 76.

<sup>147</sup> “Para evitar que a coleta das informações se transforme em violação da dignidade das pessoas, são previstas algumas formas de ‘indisponibilidade’ ou ‘inalienabilidade’ de certas categorias de dados. É o caso, em especial, dos dados sensíveis (relacionados às opiniões, à raça, à saúde, à vida sexual), cujo tratamento não é suficiente nem mesmo a autorização por escrito do interessado, sendo exigida a autorização do Garante para a Proteção dos Dados Pessoais.” RODOTÀ, 2008, p. 156.

<sup>148</sup> RODOTÀ, 2008, p. 149.

<sup>149</sup> “Rumo a uma Cidadania Eletrônica” é o título de um dos textos do livro “A vida na sociedade da vigilância: a privacidade hoje” de Stefano Rodotà, no qual o autor discorre sobre liberdade existencial, direito à autodeterminação informativa e necessidade de proteção de dados sensíveis no *ciberespaço*. RODOTÀ, 2008, p. 141-163.

exemplo, uma câmera que pode, por meio do reconhecimento facial, indicar quem cometeu um crime).

Trata-se do controle fragmentado pelos mecanismos de vigilância, que resulta, segundo RODOTÀ (2008, p. 156), na pessoa cuja unidade está igualmente “fragmentada”. São as “pessoas eletrônicas” presentes em dezenas de bancos de dados, “indivíduos multiplicados” que se tornam “abstrações no ciberespaço”.

Deste modo, Stefano Rodotà apresenta a ideia do corpo fragmentado<sup>150</sup> relacionado ao uso da tecnologia moderna que conversa com a ideia do “corpo produtivo” e do “corpo submetido” retratado por Michel Foucault na obra “Vigiar e Punir: Nascimento da Prisão”<sup>151</sup>. Tem-se o “corpo dócil” do “homem-máquina” passível de submissão, análise e manipulação<sup>152</sup>.

Para FOUCAULT (2013b), o “corpo político” teria sido historicamente utilizado pelas instituições de poder, inicialmente na forma dos suplícios, pelos

---

<sup>150</sup> “No campo do Direito Civil, as transformações do corpo e os problemas jurídicos decorrentes da fragmentação e da perda da sua unidade foram analisadas pioneiramente por Stefano Rodotà. Bem do jurista italiano a advertência de que o corpo que se fragmenta também experimenta os efeitos da sua crescente perda de materialidade, processo que teve seu ponto alto com a contraposição do corpo físico ao eletrônico. Tem-se, assim, uma multiplicidade de dimensões do corpo a serem tuteladas, todas elas complementares e essenciais à personalidade, e a consequente ampliação de questionamentos sobre os limites e possibilidades de proteção da integridade psicofísica nesse contexto. Como proteger a autonomia corporal na *Information Age*, quando corpo se transforma em sistema informativo, gerando incontáveis dilemas jurídicos, não só pela crescente exposição da personalidade no ambiente virtual e das questões que envolvem também da tutela da privacidade, mas também por conta das violações geradas pela coleta e circulação de dados sensíveis?” VIVEIROS DE CASTO, Thamis Dalsenter. **Questões atuais sobre o direito ao próprio corpo na legalidade constitucional**. In: MATOS, Ana Carla Harmatiuk; TEIXEIRA, Ana Carolina Brochado; TEPEDINO, Gustavo. (Org.). *Direito Civil, Constituição e unidade do sistema*. 1 ed. Belo Horizonte: Fórum, 2019, v. 1, p. 33-46, p. 35.

<sup>151</sup> “Mas o corpo está também diretamente mergulhado num campo político; as relações de poder operam sobre ele um efeito imediato; investem-no, marcam-no, controlam-no, supliciam-no, sujeitam-no a trabalhos, obrigam-no a cerimônias, exigem-lhe sinais. Este investimento político do corpo está ligado, segundo relações complexas e recíprocas, à sua utilização econômica; em boa parte, é como força de produção, que o corpo é investido de relações de poder e de domínio; mas, em contrapartida, a sua constituição como força de trabalho só é possível se estiver integrado num sistema de sujeição (em que a necessidade é também um instrumento político cuidadosamente organizado, calculado e utilizado); o corpo só se torna força útil se for simultaneamente corpo produtivo e corpo submetido. Esta sujeição não é obtida apenas pelos instrumentos da violência ou da ideologia; pode muito bem ser direta, física, usar a força contra força, incidir sobre elementos materiais e, porém, não ser violenta; pode ser calculada, organizada, tecnicamente refletida; pode ser subtil, não recorrer a armas nem ao terror e, porém, ser de ordem física.” FOUCAULT, Michel. **Vigiar e punir: nascimento da prisão**. Introdução: António Fernando Cascais. Lisboa: Edições 70 / Edições Almedina, S.A., 2013b, p. 66-67.

<sup>152</sup> “O *Homem-Máquina* de La Mettrie é, em simultâneo, uma redução materialista da alma e uma teoria geral do adestramento, no centro das quais reina a ideia de «docilidade», que junto o corpo analisável ao corpo manipulável. É dócil o corpo que pode ser submetido, que pode ser utilizado, que pode ser transformado e aperfeiçoado.” FOUCAULT, 2013b, p. 286.

regimes monárquicos, e posteriormente como objeto de controle do tempo e produtividade dos indivíduos pelo capitalismo, sempre como ponto de apoio “às relações de poder e de saber que investem os corpos humanos e os sujeitam, fazendo deles objetos de saber”<sup>153</sup>.

Deste modo, a sujeição do corpo físico e psíquico (a alma<sup>154</sup>) tem sido utilizada como instrumento de controle para manutenção da ordem política, social e econômica instituída, resultando, muitas vezes, no vilipêndio da integridade psicofísica dos indivíduos e, conseqüentemente, da sua dignidade existencial. Na atualidade, esse controle vem sendo exercido por meio do uso da inteligência artificial.

Baseado na figura arquitetônica do *Panopticon* de Jeremy Bentham<sup>155</sup>, Michel Foucault retrata o “Panoptismo”<sup>156</sup> que seria a forma com a qual as instituições modernas exerciam/exercem o poder sobre os indivíduos por meio do uso de mecanismos de vigilância, controle e punição.

O *Panopticon* de Bentham é uma estrutura física, uma ideia arquitetônica que permite a sensação de vigília constante do indivíduo institucionalizado que, em

---

<sup>153</sup> FOUCAULT, 2013b, p. 67.

<sup>154</sup> “Não se deve afirmar que a alma é uma ilusão ou um efeito ideológico. Mas antes que existe, que tem uma realidade, que é produzida permanentemente, em redor, na superfície, no interior do corpo pelo funcionamento de um poder que se exerce sobre os que são punidos – de uma forma mais geral, sobre aqueles que são vigiados, adestrados e corrigidos, sobre os loucos, as crianças, os aprendizes, os colonizados, sobre aqueles que são fixados a um aparelho de produção e que são controlados durante toda a sua existência.” FOUCAULT, 2013b, p. 73.

<sup>155</sup> “Entramos assim na idade do que eu chamaria de ortopedia social. Trata-se de uma forma de poder, de um tipo de sociedade que classifico de sociedade disciplinar por oposição às sociedades propriamente penais que conhecíamos anteriormente. É a idade de controle social. Entre os teóricos que há pouco citei, alguém de certa forma previu e apresentou como que um esquema desta sociedade de vigilância, da grande ortopedia social. Trata-se de Bentham. Peço desculpas aos historiadores da filosofia por esta afirmação, mas acredito que Bentham seja mais importante para nossa sociedade do que Kant, Hegel etc. Ele deveria ser homenageado em cada uma de nossas sociedades. Foi ele que programou, definiu e descreveu da maneira mais precisa as formas de poder em que vivemos e que apresentou um maravilhoso e célebre pequeno modelo desta sociedade da ortopedia generalizada: o famoso *Panopticon*. Uma forma de arquitetura que permite um tipo de poder espírito sobre o espírito; uma espécie de instituição que deve valer para escolas, hospitais, prisões, casas de correção, hospícios, fábricas, etc.” FOUCAULT, Michel. **A verdade e as formas jurídicas**. Trad. Eduardo Jardim e Roberto Machado. Rio de Janeiro: Nau, 2013a, p. 87-88.

<sup>156</sup> “O panoptismo é um dos traços característicos da nossa sociedade. É uma forma de poder que se exerce sobre os indivíduos em forma de vigilância individual e contínua, em forma de controle de punição e recompensa e em forma de correção, isto é, de formação e transformação dos indivíduos em função de certas normas. Este triplice aspecto do panoptismo – vigilância, controle e correção – parece ser uma dimensão fundamental e característica das relações de poder que existem em nossa sociedade.” FOUCAULT, 2013a, p. 103.

contrapartida, não consegue enxergar quem ocupa ou o que se passa dentro da torre central, da qual há visão panorâmica para toda a estrutura.

Do centro tudo se vê, os objetos de escrutínio estão constantemente expostos, portanto há transparência; do anel periférico, não se vê o detentor do poder de vigiar e punir, de fiscalizar e cobrar, de educar e avaliar, de observar e medicar. Logo, há opacidade. A desigualdade é a base do mecanismo de controle.

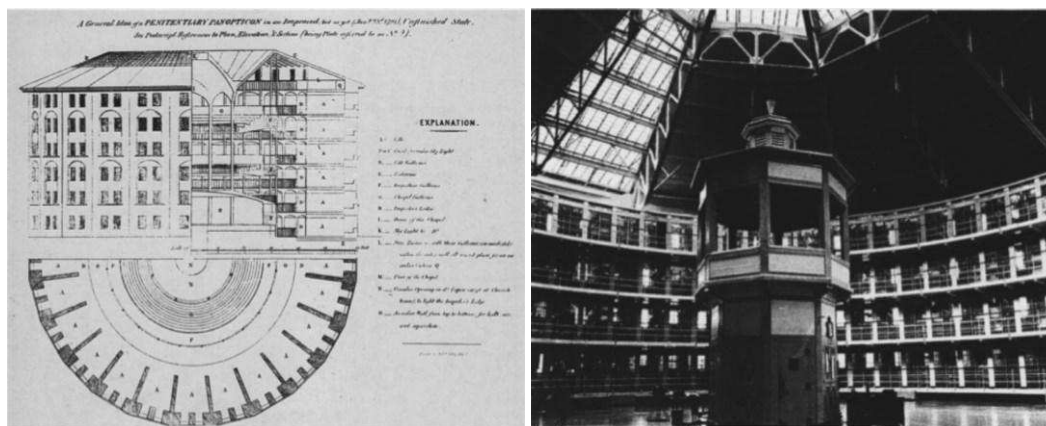


Figura 3: Planta do panóptico de Jeremy Bentham<sup>157</sup>.

Figura 4: Interior da penitenciária de Stateville, EUA, século XX<sup>158</sup>.

A prisão, no sentido literal, é o ponto de partida da ideia, contudo, FOUCAULT (2013a, p. 88) defende que o modelo de controle panóptico é uma “forma de arquitetura que permite um tipo de poder do espírito sobre o espírito; uma espécie de instituição que deve valer para escolas, hospitais, prisões, casas de correção, hospícios, fábricas, etc.”.

Não obstante o entendimento de ZUBOFF (2015, p. 82)<sup>159</sup>, para quem o *design* do panóptico de Bentham, por se tratar de uma estrutura física, com um único ponto de observação, seria deveras prosaico considerando essa nova arquitetura da sociedade de informação, preferindo a utilização do termo *Big Other*<sup>160</sup>; pode-se

<sup>157</sup> FOUCAULT, 2013b, p. 254.

<sup>158</sup> FOUCAULT, 2013b, p. 262.

<sup>159</sup> “Power can no longer be summarized by that totalitarian symbol of centralized command and control. Even the panopticon of Bentham’s design, which I used as a central metaphor in my earlier work (Zuboff, 1988, Ch. 9,10), is prosaic compared to this new architecture. The panopticon was a physical design that privileged a single point of observation.” ZUBOFF, Shoshana. **Big other**: surveillance capitalism and the prospects of an information civilization. *Journal of information technology*, 30.1 (2015): 75-89, p. 82.

<sup>160</sup> Shoshana Zuboff utiliza o termo *Big Other* (o Grande Outro) que seria a forma de exercício do controle na sociedade da informação de forma descentralizada em substituição ao *Big Brother*. “Big



afirmar que o panoptismo descrito por FOUCAULT (2013b) se assemelha aos métodos de emprego da IA.

A abstração feita por Foucault, retirada de um modelo de controle físico, aplica-se ao modelo de organização das instituições sociais e econômicas hodiernas que fazem uso da IA como instrumento de vigilância e decisão, que despersonaliza, descentraliza e permite uma automação do poder que, não incomum, sequer é identificado pelos usuários dos sistemas <sup>161</sup>.

No caso da IA, não se trata de uma torre física aparente e central, mas de um poder muitas vezes opaco que centraliza informações ao mesmo tempo em que atua de forma delegada e descentralizada, na qual, cada um dos componentes (sejam eles câmeras, programas, robôs ou aplicativos) atuam como “juizes da normalidade” que “estão presentes em toda parte”<sup>162</sup>.

HARCOURT (2015, p. 91-92), faz uma crítica à ideia da aplicação do panoptismo de Michel Foucault na atual sociedade de exposição. Pondera que uma sociedade na qual as pessoas nem sempre têm ciência de que estão sendo observadas e escolhem se expor como forma de recriar a própria identidade (ninguém é

---

Other is the 21st-century incarnation of the electronic text that aspires to encompass and reveal the comprehensive immanent facts of market, social, physical, and biological behaviors. The institutional processes that constitute the architecture of Big Other can be imagined as the material instantiation of Hayek’s ‘extended order’ come to life in the explicated transparency of computer-mediation. These processes reconfigure the structure of power, conformity, and resistance inherited from mass society and symbolized for over half a century as Big Brother. Power can no longer be summarized by that totalitarian symbol of centralized command and control. Even the panopticon of Bentham’s design, which I used as a central metaphor in my earlier work (Zuboff, 1988, Ch. 9,10), is prosaic compared to this new architecture. The panopticon was a physical design that privileged a single point of observation.” ZUBOFF, 2015, p. 82.

<sup>161</sup> “O Panóptico é uma máquina de dissociar o par ver-ser visto: no anel periférico, os indivíduos são totalmente vistos, sem nunca verem; na torre central, vê-se tudo, sem nunca se ser visto. É um dispositivo importante, uma vez que automatiza e desindividualiza o poder. Este tem o seu princípio menos numa pessoa do que numa certa distribuição concertada dos corpos, das superfícies, das luzes e dos olhares; numa aparelhagem cujos mecanismos internos produzem a relação na qual estão presos indivíduos. As cerimónias, os rituais, as marcas pelas quais o poder excessivo se manifesta no soberano são inúteis. Há uma maquinaria que assegura assimetria o desequilíbrio, a diferença. Pouco importa, portanto, quem exerce o poder. Qualquer indivíduo, escolhido ao acaso, pode fazer funcionar a máquina: na falta do diretor, a sua família, o seu círculo, os seus amigos, as suas visitas e até os seus criados.” FOUCAULT, 2013b, p. 405-406.

<sup>162</sup> “Os juizes da normalidade estão presentes em toda a parte. Estamos na sociedade do professor-juiz, do médico -juiz, do educador-juiz, do «trabalhador social»-juiz; todos eles fazem reinar a universalidade do normativo; e cada indivíduo, no ponto onde se encontra, submete-lhe o corpo, os gestos, os comportamentos, as condutas, as aptidões e os desempenhos. A rede prisional, nas suas formas compactas ou disseminadas, com os seus sistemas de inserção, de distribuição, de vigilância, de observação, foi o grande suporte, na sociedade moderna, do poder normalizador.” FOUCAULT, 2013b, p. 598.

encarcerado ou forçado a essa situação) não se enquadra na estrutura central de poder do panóptico.<sup>163</sup>

Entretanto, como dito anteriormente, o “Panóptico é uma máquina de dissociar o par ver-ser visto”<sup>164</sup> pois, do ponto de vista da periferia, há uma opacidade, não se enxerga o que se passa dentro da torre e o indivíduo não tem certeza se que está, de fato, a ser observado, se há alguém ali dentro e, em havendo, quem seria(m) essa(s) pessoa(s)<sup>165</sup>.

No caso da IA, apesar do uso constante e extensivo da tecnologia na vida dos indivíduos, há sim situações nas quais não se tem ciência da vigília constante. A título exemplificativo, é comum indivíduos que fazem uso de redes sociais não terem noção de que estão sendo submetidos ao perfilamento<sup>166</sup> e mineração de dados.

---

<sup>163</sup> “In the end, the metaphor of panoptic power is somewhat inexact today – for these and other reasons. As Siva Vaidhyanatham argues in *The Googlization of Everything*, many of us are not even aware of the myriad ways in which we are being tracked. ‘Unlike Bentham’s prisoners,’ Vaidhyanatham emphasizes, ‘we don’t know all the ways in which we are being watched or profiled.’ Our situation resembles perhaps more a ‘cryptopticon,’ in his words – reflecting the Kafkaesque dimension of our times. Recall that Foucault, knowing visibility was one of the most important features of panoptic power. For the panopticon to function, the inmate had to be aware he was being watched – he had to always ‘[know] himself to be observed.’ That too, it seems, is no longer the case. In addition, many of us ‘don’t seem to care’ about being watched – which is far different from the condition of those in the panopticon. For the panopticon to function, the watching gaze must be internalized, not ignored. But overall, the most striking inversion is that today we expose ourselves so freely and passionately and anxiously – at least, so many of us in advanced capitalist liberal democracies. It is an expository form of power we face at the present time, one in which we use digital mediums to tell stories about ourselves and create and reshape our identities. One in which we participate fully – willingly or not – through all of our ordinary, mundane, daily actions: carrying a cell phone, sending emails and texts, taking the subway. No, we are not being forced into this situation, we are not enclosed, it is not just disciplinary.” HARCOURT, Bernard E., **Exposed: desire and disobedience in the digital age**. Cambridge, Massachusetts, London, England: Harvard University Press, 2015, p. 91-92.

<sup>164</sup> FOUCAULT, 2013b, p. 405.

<sup>165</sup> “Foi por isso que Bentham formulou o princípio segundo o qual o poder deve ser visível e inverificável. Visível: o recluso terá incessantemente diante de si a alta silhueta da torre central, de onde é espiado. Inverificável: o recluso nunca deve saber se está realmente a ser observado; mas deve ter a certeza de que pode estar sempre a ser vigiado.” FOUCAULT, 2013b, p. 405.

<sup>166</sup> Michel Foucault aborda a questão do perfilamento social de indivíduos e “esquemas de exclusão”: “A divisão constante do normal e do anormal, a que todos os indivíduos estão sujeitos, reconduz até nós, e aplicando-as a objetos diferentes, a marcação binária e o exílio do leproso; a existência do todo um conjunto de técnicas e de instituições que assumem a tarefa de avaliar, controlar e corrigir os anormais, faz funcionarem os dispositivos disciplinares originados pelo medo da peste. Todos os mecanismos de poder que, ainda hoje se dispõem em redor do anormal, tanto para o marcar como para o modificar, compõem estas duas formas das quais derivam remotamente.” FOUCAULT, 2013b, p. 401-402.

Contudo, há circunstâncias em que a vigilância é evidente e ostensiva: uso de câmeras em espaços públicos e particulares, identificação biométrica, programas utilizados por empregadores para fiscalizar o rendimento de empregados, dentre outras. Em tais casos, a metáfora no panóptico se aplica perfeitamente.

Outrossim, mesmo nos casos em que não há ciência da vigília constante por meio da tecnologia, a simbologia do panóptico pode ser aplicada, pois a despersonalização e descentralização que ocorre na delegação de poder decisório para a IA possibilitam uma relação assimétrica, de controle, na qual quem é submetido às decisões não consegue verificar ou questionar os mecanismos de tomada de decisão.

Como se insurgir? Contra quem se insurgir? O controle está posto e a caixa preta do algoritmo se assemelha à ideia da torre visível, mas inverificável. “Há uma maquinaria que assegura a assimetria, o desequilíbrio, a diferença. Pouco importa, portanto, quem exerce o poder. Qualquer indivíduo escolhido ao acaso pode fazer funcionar a máquina [...]”.<sup>167</sup>

No que se refere à questão da exposição, ao contrário do que HARCOURT (2015) defende, ela não é necessariamente uma escolha. Como será explorado no tópico a seguir, ZOUBOFF (2015) pontua que no “capitalismo de vigilância” a exposição se tornou uma imposição, uma necessidade para que o indivíduo funcione social e profissionalmente em uma sociedade digital totalmente conectada. “A visibilidade é uma armadilha.”<sup>168</sup>

### 3.3.2

#### **Mediação e persuasão.**

Segundo CALO (2015, p. 519), as experiências no ciberespaço são mediadas, o que significa dizer que ocorrem por meio da tecnologia. Assim, a mediação tecnológica permite que duas pessoas coexistam em um ciberespaço comum para

---

<sup>167</sup> FOUCAULT, 2013b, p. 406.

<sup>168</sup> FOUCAULT, 2013b, p. 403.

consumir e renovar o seu conteúdo, mas, por outro lado, traz a perspectiva de interferência, persuasão e controle<sup>169</sup>.

Segundo o jurista, ao mesmo tempo em que indica novos e animadores caminhos para o fomento das comunidades, a mediação também cria peculiares formas de vigilância e controle. Consequentemente, a maior parte dos debates entre reguladores, indústria, ativistas e acadêmicos discute quem poderia exercer o controle sobre esses espaços e objetos compartilhados, sobre a rede e outras camadas de tecnologia que os permeiam, ou sobre as informações criadas.

Logo, a compreensão de como ocorre o controle no mundo das novas tecnologias permitiria uma abordagem adequada de tópicos fundamentais no direito cibernético, incluindo propriedade intelectual e virtual, a neutralidade da rede e a privacidade<sup>170</sup>.

Ao discorrer sobre a nova lógica de acumulação que denomina “capitalismo de vigilância”<sup>171</sup>, ZOUBOFF (2015, p. 85) defende que a arquitetura desse regime é lastreada na arquitetura global de mediação por computadores que transforma o texto eletrônico da organização delimitada em um organismo inteligente de abrangência mundial que chama de *Big Other* (O Grande Outro).

Assim, novas possibilidades de subjugação são produzidas enquanto essa lógica insitucional inovadora prospera em mecanismos inesperados e ilegíveis de extração e controle que exilam as pessoas de seu próprio comportamento<sup>172</sup>.

---

<sup>169</sup> “Experiences in cyberspace are mediated, meaning they take place *through* technology. The fact of common mediation is precisely what enables two people to coexist in a common cyberspace, to consume but also “remix” content. But, as Joel Reidenberg, Lawrence Lessig, and others have catalogued, mediation also introduces the prospect of interference, persuasion, or control.” CALO, 2015, p. 519.

<sup>170</sup> “I have so far only described one side of the proverbial coin. Common mediation opens exciting new paths for communities but also invites exquisite new forms of surveillance and control. Many of the central debates among policymakers, industry, activists, and academics involve understanding precisely who can exert control over these shared spaces and objects, over the network and other ‘layers’ of technology that underlie them, or over the information each generates. Control is a way to understand key topics in cyberlaw, including intellectual and virtual property, net neutrality, and privacy.” CALO, 2015, p. 522.

<sup>171</sup> Para Shoshana Zuboff, o “capitalismo de vigilância” estabelece uma nova forma de poder na qual o contrato e o imperativo legal são suplantados por prêmios e punições de um novo tipo de mão invisível. “Surveillance capitalism establishes a new form of power in which contract and the rule of law are supplanted by the rewards and punishments of a new kind of invisible hand.” ZUBOFF, 2015, p. 82.

<sup>172</sup> “These developments became the basis for a fully institutionalized new logic of accumulation that I have called surveillance capitalism. In this new regime, a global architecture of computer mediation turns the electronic text of the bounded organization into an intelligent world-spanning organism that I call Big Other. New possibilities of subjugation are produced as this innovative

Destarte, além do julgamento, faz parte do poder decisório da IA o processo de intermediação das relações entre humanos e entre humanos e as instituições públicas e privadas. A IA decide quem vai ter acesso a determinada informação (exemplo: sugestão de compra), quem vai ter contato com determinada pessoa (exemplo: aplicativos de namoro), quem ou que tipo de conteúdo pode ser exposto aos demais membros de uma comunidade (exemplo: mídias sociais).

Os mecanismos de auxílio ao humano rapidamente podem se tornar mecanismos de controle dependendo da forma como a mediação pela tecnologia é feita, pois, como dito por CALO (2015), há mecanismos de interferência e persuasão que afetam a esfera da autonomia existencial.

Segundo ZOUBOFF (2015), nada no passado preparou as pessoas para essas novas práticas, o que leva à existência de poucas barreiras de proteção do humano. Os indivíduos rapidamente passaram a depender de novas ferramentas de informação e comunicação como recursos necessários na luta cada vez mais estressante, competitiva e estratificada por uma vida efetiva. As novas ferramentas, redes, aplicativos, plataformas e mídias tornaram-se, assim, requisitos para a participação social<sup>173</sup>.

Esses instrumentos de mediação passam a ter um poder inédito e diferenciado. Segundo o filósofo francês Eric Sadin, as máquinas são chamadas a dizer a verdade aos humanos, sendo dotadas de poder de fala. Isso ocorre, por exemplo, por meio de *chatbots* e assistentes digitais pessoais (Alexa ou Siri).

A indústria envolvida nessa nova formulação econômico-social busca estar presente e influenciar os humanos apontando o que supostamente é o mais indicado para cada um, havendo uma disputa entre as *big techs* para exercerem o papel de mediadores e serem detentores do poder de dizer a “verdade”<sup>174</sup>.

---

institutional logic thrives on unexpected and illegible mechanisms of extraction and control that exile persons from their own behavior.” ZUBOFF, 2015, p. 85.

<sup>173</sup>“Nothing in past experience prepared people for these new practices, and so there were few defensive barriers for protection. Individuals quickly came to depend upon the new information and communication tools as necessary resources in the increasingly stressful, competitive, and stratified struggle for effective life. The new tools, networks, apps, platforms, and media thus became requirements for social participation.” ZUBOFF, 2015, p. 85.

<sup>174</sup> Trecho de entrevista concedida por Eric Sadin em 2019: P. “You are talking about an almighty power in the making that is capable of stating ‘the truth’ for us.” R. “What characterizes AI is that it is a power of expertise that is constantly being improved. Its self-learning systems are able to analyze ever more varied situations and reveal to us facts of which we were unaware in some cases. They do so at speeds that exceed our cognitive abilities. We are experiencing a change in the status

### 3.4

#### Reflexos das decisões autônomas da IA na autonomia existencial – análise casuística.

Ao longo do capítulo anterior foram apresentados três exemplos de situações fáticas que demonstram o menosprezo ou aviltamento da autonomia humana nas esferas pessoal, interpessoal e social<sup>175</sup> devido à utilização displicente da IA.

No primeiro, um caso de vilipêndio da autonomia existencial na esfera social, foi descrita a aplicação desastrosa de um mecanismo de IA pela Amazon, entre 2014 e 2015, para seleção de candidatos às vagas de emprego que resultou na eliminação de candidatas do gênero feminino.

No segundo, um caso de desrespeito à autonomia existencial nas esferas pessoal e interpessoal, foi narrado o episódio ocorrido em 2021 de demissão coletiva instantânea de centenas de empregados da empresa Xsolla devido a uma decisão seguida de ação realizada exclusivamente por um mecanismo de IA.

No terceiro, um caso de desdém da autonomia existencial na esfera pessoal, foi exposta a história do japonês Mainichi Kondo, que se apaixonara e se casara com um holograma de aparência feminina em 2018, mas que enfrentava uma crise de relacionamento em 2022 devido à cessação de atualização do software pela empresa responsável pela criação e utilização da IA.

---

of digital technologies: They are no longer intended to only give us access to information but to reveal to us the reality of the phenomena beyond appearances. Basically, these computational systems have a singular and disturbing vocation: to state the truth. Technology is given new prerogatives: to illuminate the course of our existence with its lights. That is a major fact.” P. “How does this techne logos, which you define as a technology capable of uttering verbs, manifest itself in our daily lives?” R. “At the moment, when these machines are called upon to tell us the truth, they find themselves endowed with speech, just like these connected speakers with whom we can dialogue. This provision is also at work in chatbots, conversational agents, or personal digital assistants designed to guide us in our daily lives. We will increasingly be surrounded by spectra in charge of administering our lives. This is what I call ‘power-kairos’ — the will of the digital industry to be continuously at our side in order to seek and, as soon as the opportunity presents itself, to influence our actions by stating what is supposed to suit us. The upcoming economic battle between Google, Facebook, Amazon, Baidu [the Chinese ‘Google’], and others will result in fierce competition for this spectral presence, with each actor struggling to impose its hold at the expense of all others.” SADIN, Eric. **Artificial Intelligence: A Radical Anti-Humanism**. Artificial intelligence and ethics, two subjects that go hand-in-hand in current literature, are often confined to the spectrum of the subject of privacy. DZone. Thomas Jardinnet. Publicado em: 12.02.2019. Disponível em: <<https://dzone.com/articles/artificial-intelligence-a-radical-anti-humanism>>. Acesso em: 09.01.2023.

<sup>175</sup> Classificação que segue a anteriormente citada teoria tríplice da autonomia existencial de Thamis Dalsenter Viveiros de Castro.

Os episódios narrados têm em comum o desrespeito à liberdade existencial de indivíduos, de uma coletividade ou da sociedade como um todo. Em verdade, em uma aplicação dos princípios da dignidade e solidariedade, os três exemplos citados têm impactos sociais.

As pautas de gênero, trabalho e emprego dizem respeito a todos que pretendem viver em uma sociedade livre, igual, solidária e democrática; assim como questões referentes aos sentimentos, sexualidade e relacionamentos interpessoais, já que o ser humano existe na coexistência.

A fim de enriquecer o conteúdo do presente estudo, será apresentado a seguir um último exemplo, a ser abordado de uma maneira mais aprofundada e direcionada, que engloba dois temas que vem sendo objeto de relevantes debates sociais e acadêmicos: identidade de gênero e identificação biométrica e de pessoas singulares pela IA.

### 3.4.1

#### **Exemplo de discriminação de gênero por biometria facial – contexto fático.**

Em 2018, após a implementação do sistema de biometria facial em 100% das catracas dos ônibus do Distrito Federal com o intuito de inibir fraudes<sup>176</sup>, Maria Eduarda, uma universitária de 23 anos, teve o seu “Bilhete Único” bloqueado pelo fato de ser uma mulher trans que não foi reconhecida pelo sistema, restando assim impedida de utilizar o benefício por 20 dias<sup>177</sup>.

---

<sup>176</sup> “Acima dos validadores, onde os passageiros passam o cartão, há câmeras que captam imagens de quem passa pela catraca. Por meio de um software, elas são comparadas com as fotos cadastradas no sistema. Quando o programa automaticamente detecta divergências — ou seja, alerta que as imagens não coincidem —, faz-se uma análise visual para confirmar se é caso de fraude. Se confirmada a irregularidade, o benefício é suspenso e se abre um processo administrativo em que o usuário tem direito ao contraditório e à ampla defesa. Se ainda assim os esclarecimentos forem insatisfatórios, o cartão é bloqueado. ‘Após esse bloqueio, se for estudante, o benefício só poderá ser pedido novamente no semestre letivo seguinte e, se for pessoa com deficiência, em 12 meses’, detalha o diretor-geral do DFTrans.” SEMOB. **Biometria facial é obrigatória em 100% dos ônibus do DF**. Disponível em: <<https://semob.df.gov.br/biometria-facial-e-obrigatoria-em-100-dos-onibus-do-df/>>. Publicado em: 03.05.2018. Atualizado em: 07.11.2018. Acesso em: 10.08.2022.

<sup>177</sup> ALVES, Sarah. **Além do racismo, reconhecimento facial erra mais em pessoas trans**. Tilt UOL. Inteligência artificial. Publicado em: 14.02.2021. Atualizado em: 22.06.2021. Disponível em:

A autonomia, a privacidade, a liberdade de expressão, o direito de ir e vir e o direito à educação da estudante foram maculados devido a uma decisão instantânea tomada por uma máquina segundos após a captura da imagem do seu rosto por uma câmera instalada dentro do ônibus.

Não havia nenhum humano para fiscalizar ou intermediar a situação, não havia nenhum botão ou outro meio de comunicação para solicitar ajuda ou esclarecer o erro da máquina. Coube à estudante enfrentar semanas de burocracia para reafirmar sua identidade e gênero, e ver restabelecido o seu benefício legal.

A identidade construída e escolhida por Maria Eduarda, fruto do exercício da sua liberdade existencial, é merecedora de tutela, pois se trata do seu direito ao livre desenvolvimento da personalidade<sup>178</sup>.

É razoável e justo que um *small brother* instalado em um coletivo decida se Maria Eduarda é de fato Maria Eduarda, e, conseqüentemente, se ela pode ou não fazer uso de um benefício estudantil de transporte público?

Mecanismos de reconhecimento facial e reconhecimento automatizado de gênero vêm sendo utilizados e afetam a liberdade existencial de indivíduos como a mulher transgênero do exemplo citado. É essencial que se entenda a forma de funcionamento de tais tecnologias a fim de se avaliar a razoabilidade da sua aplicação.

---

<<https://www.uol.com.br/tilt/noticias/redacao/2021/02/14/nao-e-so-racismo-reconhecimento-facial-tambem-erra-mais-em-pessoas-trans.htm>>. Acesso em: 10.08.2022.

<sup>178</sup> “No âmbito da comunicação eletrônica, ela pode se exprimir sobretudo como uma necessidade de anonimato ou, melhor dizendo, como exigência de assumir a identidade preferida, apresentando-se com um nome, um sexo, uma idade que podem ser diferentes daqueles efetivamente correspondentes aos dados do indivíduo. Requer-se assim a tutela de uma identidade nova, de uma intimidade construída, como condição necessária para o desenvolver a própria personalidade, para alcançar plenamente a liberdade existencial.” RODOTÀ, 2008, p. 116.



### 3.4.2

#### Funcionamento de sistemas de detecção e reconhecimento facial e reconhecimento automatizado de gênero (AGR).

O reconhecimento facial<sup>179</sup> é feito por meio de sistemas de IA que realizam a leitura automática da face humana. Características da identidade dos indivíduos vêm sendo operacionalizadas e computadorizadas e servem de insumo de treinamento para que máquinas sejam capazes de responder a essas características realizando basicamente duas tarefas: detectar e reconhecer faces humanas<sup>180</sup>.

Ligado ao processo de reconhecimento facial está o reconhecimento automatizado de gênero (*Automatic Gender Recognition* – AGR) que utiliza não somente imagens e vídeos (que podem considerar tanto características faciais, quanto corporais), mas também dados de áudio para identificar o gênero dos indivíduos<sup>181</sup>.

Os métodos de AGR se amparam na tecnologia de reconhecimento facial, também utilizada para o reconhecimento corporal, para analisar as características

---

<sup>179</sup> “Reconhecimento facial consiste no uso de tecnologias por meio das quais se permite a identificação automática de pessoas, como também de objetos. Inserida no campo da denominada ‘computação visual’, consiste na análise comparativa entre imagens produzidas por pessoas e objetos. Assim, quando houver fortes semelhanças entre a imagem prévia e a imagem capturada para comparação, ativa-se o comando de que houve o reconhecimento.” TOMASEVICIUS FILHO, Eduardo. **Reconhecimento facial e lesões aos direitos da personalidade**. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Felipe Braga Neto, et al. **Direito digital e inteligência artificial: diálogos entre Brasil e Europa**. Indaiatuba: Editora Foco, 2021, p.130.

<sup>180</sup> “Image detection and classification, in particular, represents a pertinent domain where we see a tight coupling of human identity and computation. Perhaps the most salient example is *automated facial analysis technology* (FA), an umbrella term for computer vision methods that use machine learning (ML) techniques to automate problems related to reading the human face. FA is often discussed in the context of two specific tasks: facial recognition. Both use computational methods to measure the human face, whether simply to detect that a face is present (i.e., facial detection) or to detect a specific individual’s face (i.e., facial recognition).” SCHEUERMAN, Morgan Klaus; PAUL, Jacob M.; BRUBAKER, Jed R. **How computers see gender: an evaluation of gender classification in commercial facial analysis and image labeling services**. Proceedings of the ACM on Human-Computer Interaction 3. CSCW (2019): 1-33, p. 2.

<sup>181</sup> “*Automatic Gender Recognition* (AGR) (also known as *gender classification*) refers to algorithmic methods, including automatic facial recognition and body recognition technologies, that extract features from images, videos, or audio of one or more individuals in order to identify their gender. AGR often leverages computer vision algorithms and/or voice recognition modules. A common method is to extract features (e.g., facial hair) from an individual’s visual and/or audio data (e.g., a video showing their face) and compare them with ground-truth samples (e.g. videos of faces for which the gender is known) in a existing database. If the input features are found to be similar to those in the database, a match is declared.” HAMIDI, Foad; SCHEUERMAN, Morgan Klaus; BRANHAM, Stacy M. **Gender recognition or gender reductionism? The social implications of embedded gender recognition systems**. In Proceedings of the 2018 chi conference on human factors in computing systems, pp. 1-13, 2018, p. 2.

das pessoas e assim definir o seu gênero. Deste modo, traços como a voz, o comprimento dos cabelos, pelos faciais, contorno e alinhamento facial, volume dos lábios, posição dos olhos, presença de maquiagem, vestuário, seios, quadril e ombros são analisados para categorização de gênero<sup>182</sup>.

Não obstante a complexidade da identidade humana, que carrega aspectos pessoais, sociais, profissionais e culturais, a IA, por meio de conceitos pré-estabelecidos de gênero, nos quais não se enquadram as pessoas transgênero e não-binárias, tem simplificado tais atributos para fins de padronização e atendimento de interesses do mercado.

Há uma massificação, perfilamento e classificação padronizada dos indivíduos, “uma discriminação entre as pessoas baseada em critérios *standard*, em perfis automatizados”<sup>183</sup>, o “juízo qualitativo”<sup>184</sup>, a imposição de estampas ou modelos aos seus corpos e almas representam o que DONEDA et al. (2018, p. 12) denomina como a “erosão da autodeterminação humana”<sup>185</sup>.

Em um sistema de ponderação de interesses, as facilidades práticas e econômicas que podem, eventualmente, ser trazidas por sistemas de reconhecimento facial e AGR não servem para justificar a mitigação de direitos fundamentais e a estigmatização de indivíduos que assumem a sua identidade de gênero de forma desvinculada do seu sexo biológico, em pleno exercício do direito do livre desenvolvimento da sua personalidade.

Ao se considerar que o gênero pré-estabelecido de maneira binária é uma construção cultural<sup>186</sup>, o que tem sido objeto de intensas discussões na sociedade

---

<sup>182</sup> HAMIDI et al., 2018.

<sup>183</sup> RODOTÀ, 2008, p. 121.

<sup>184</sup> “O juízo ‘qualitativo’ sobre as relações assim estabelecidas não pode fundar um juízo de realidade, negando que se esteja diante de uma formação social. E a valoração negativa daquelas relações não pode levar à conclusão de que é preferível desencorajá-las, para evitar distorções no processo formativo da personalidade. Como são baseadas em um dado objetivo – a existência de uma entidade social na qual um sujeito decide se inserir – as garantias constitucionais devem ser ampliadas também a essas novas realidades, reconhecendo a legitimidade das escolhas dos indivíduos que privilegiam a presença nas redes e fazem desta um momento significativo para a definição global de sua identidade.” RODOTÀ, 2008, p. 117.

<sup>185</sup> “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’, mesmo que não intencional, devem fomentar, e nunca minar a dignidade humana e a autodeterminação (YANG, 2018).” DONEDA et al., 2018, p. 12.

<sup>186</sup> A feminista norte-americana Judith Butler pondera que se o gênero (construção social) é visto de forma independente do sexo biológico, ele se torna um artifício flutuante, logo, homens podem estar

atual, a programação e treinamento da IA para que mantenha esses padrões é um mecanismo de manutenção do controle de uma ordem social estabelecida e consequente perpetuação de padrões discriminatórios.

### 3.4.3

#### **Análise de dois estudos que tratam dos impactos dos sistemas de AGR nos indivíduos transgênero e não-binários.**

No primeiro estudo, publicado em 2018, um grupo de pesquisadores da Universidade de Maryland em Baltimore County (HAMIDI, et. al, 2018), nos EUA, discutiu as implicações sociais de sistemas de reconhecimento automatizado de gênero (AGR), buscando repostas sobre a percepção e comportamento das partes interessadas no que diz respeito à referida tecnologia e de como ela poderia afetar grupos de vulneráveis.

Foram entrevistados 13 indivíduos transgênero, incluindo 3 *designers* de tecnologia. Constatou-se que indivíduos transgênero têm comportamentos extremamente negativos em relação à AGR e questionam se esse tipo de tecnologia pode reconhecer de forma precisa aspectos tão subjetivos da sua identidade. Assim, apontam preocupações com a privacidade e perigos em potencial que podem resultar da classificação incorreta ou equivocada do gênero pela tecnologia<sup>187</sup>.

---

em corpos femininos e mulheres em corpos masculinos: “If gender is cultural meanings that the sexed body assumes, then gender cannot be said to follow from a sex in any one way. Taken to its logical limit, the sex/gender distinction suggests a radical discontinuity between sexed bodies and culturally constructed genders, assuming for the moment the stability of binary sex, it does not follow that the construction of ‘men’ will accrue exclusively to the bodies of males or that ‘women’ will interpret only female bodies. Further, even if the sexes appear to be unproblematically binary in their morphology and constitution (which will become a question), there is no reason to assume that genders ought also to remain as two. The presumption of a binary gender system implicitly retains the belief in a mimetic relation of gender to sex whereby gender mirror sex or is otherwise restricted by it. When the constructed status of gender is theorized as radically independent of sex, gender itself becomes a free-floating artifice, with the consequence that *man* and *masculine* might just as easily signify a female body as a male one, and *woman* and *feminine* a male body as easily as a female one.” BUTLER, Judith. **Gender trouble: feminism and the subversion of identity**. New York/London: 2007, p. 9.

<sup>187</sup> “Applications of AGR are increasingly being explored in domains such as security, marketing, and social robotics. However, little is known about stakeholders’ perceptions and attitudes towards AGR and how this technology might disproportionately affect vulnerable communities. To begin to address these gaps, we interviewed 13 transgender individuals, including three transgender technology designers about their perceptions and attitudes towards AGR. We found that transgender individuals have overwhelmingly negative attitudes towards AGR and fundamentally question whether it can accurately recognize such a subjective aspect of their identity. They raised concerns about

Foram levantadas pela maioria dos entrevistados as seguintes preocupações: i) o gênero, uma questão de identidade pessoal, não necessariamente está atrelada à aparência física, que é o principal dado de análise do mecanismo de AGR<sup>188</sup>; ii) que os sistemas classifiquem e estigmatizem pessoas trans como mal intencionadas ou com intuito de induzir o outro a erro sobre o seu gênero para fins de relacionamentos pessoais<sup>189</sup>; iii) ter o gênero erroneamente classificado por uma máquina pode ser pior do que a discriminação por um humano, pois seria mais uma fonte de invalidação que intensificaria os efeitos negativos da discriminação<sup>190</sup>; iv) impossibilidade de corrigir a máquina, ao contrário do humano<sup>191</sup>; v) ter o gênero equivocadamente classificado por uma máquina implica em ser discriminado não por uma pessoa, mas por um grupo de pessoas<sup>192</sup>; vi) que a tecnologia da AGR reforce os padrões de gênero que indivíduos trans poderiam internalizar e carregar consigo pela percepção de que sistemas são mais objetivos, levando a uma interpretação de gênero massificada<sup>193</sup>.

As informações coletadas demonstram a insegurança das pessoas que não se enquadram no padrão de gênero binário pré-estabelecido quanto ao uso de mecanismos de AGR. Mesmo indivíduos inseridos na área de desenvolvimento de

---

privacy and potential harms that can result from being incorrectly gendered, or recommendations on how to accommodate gender diversity when designing new digital systems.” HAMIDI et al., 2018, p. 1.

<sup>188</sup> “Some participants (P2, P8, T2, P10, T3) disagree with the assumption AGR systems make about the nature of gender as something that can be classified using external features; they stated that gender is an internal identity not necessarily tied to physical features.” HAMIDI et al., 2018, p. 5.

<sup>189</sup> “Finally, P7 noted that trans people are often mistakenly accused of ‘catfishing’, or luring others into relationships using false and constructed online personas. He was therefore worried that AGR systems might flag transgender individuals, lumping them in with ill-intentioned people trying to commit fraud or deception.” HAMIDI et al., 2018, p. 5.

<sup>190</sup> “Most of the participants (P1, P7, P8, T1, P9, P10, T3) considered misgendered by AGR *worse* than being misgendered by another human being. AGR simply introduced another potential source of invalidation. [...] Participants also foresaw that increased misgendering would lead to an increase in its negative effects.” HAMIDI et al., 2018, p. 5.

<sup>191</sup> “One set of concerns was rooted in the belief that AGR systems might not allow users to perceive and therefore correct gender classification errors. P9 and T3 expressed being more tolerant of human mistakes, because ‘people you can correct’ (T3).” HAMIDI et al., 2018, p. 5.

<sup>192</sup> “P8 was aware that computers carry the biases of their human developers. But, as a result, he interpreted being misgendered by a computer as a more severe act committed by many people as opposed to just one.” HAMIDI et al., 2018, p. 5.

<sup>193</sup> “Finally, both P1 and P7 thought being misgendered by AGR technology would reinforce gendered standards that transgender individuals would then internalize and hold themselves to. We saw that the perception of computers as somehow being more ‘objective’ or as a synthesis of general human standards led participants to a sort of insult-to-injury mindset because they interpreted the gender label assigned by the computer as more definitive and exacting.” HAMIDI et al., 2018, p. 5.

tecnologia, temem sofrer discriminação ou ver a discriminação intensificada pelo uso da referida tecnologia.

No segundo estudo, pesquisadores do departamento de ciência da informação da Universidade do Colorado em Boulder (SCHEUERMAN et al., 2019), também nos EUA, procurou avaliar como os computadores veem o gênero e constataram que sistemas de análise facial tem pior funcionamento com indivíduos transgênero e são incapazes de classificar gêneros não-binários.

Os pesquisadores coletaram 2,450 imagens de faces do *Instagram*, sendo que cada uma delas teve a identidade de gênero categorizada pelo próprio indivíduo por meio de uma *hashtag*. As fotos foram então divididas em 7 grupos de 350 imagens (#mulher, #homem, #mulher trans, #homem trans, #sem gênero, #sem gênero *queer*, #não-binário) e analisadas por quatro dos maiores provedores de serviço de análise facial (IBM, Amazon, Microsoft e Clarifai). O Google não foi incluído, pois não oferece serviço de reconhecimento de gênero.

No geral, os sistemas foram mais precisos com fotos de mulheres cisgênero, que tiveram o gênero detectado corretamente 98,3% das vezes. Todavia, os homens trans foram erroneamente identificados como mulheres em até 38% das vezes. E aqueles que se identificaram sem gênero, sem gênero *queer*, ou não-binários (o que indica que eles não se identificam nem como masculino, nem como feminino) foram caracterizados de forma equivocada 100% das vezes.

Segundo Jed R. Brubaker, um dos pesquisadores responsáveis pelo estudo, esses sistemas só conhecem a linguagem do masculino e feminino, por isso, havendo muitas identidades de gênero, não é possível para eles acertar a classificação<sup>194</sup>. Portanto, a pluralidade não é levada em consideração.

---

<sup>194</sup> “Researchers collected 2,450 images of faces from Instagram, each of which had been labeled by its owner with a hashtag indicating their gender identity. The pictures were then divided into seven groups of 350 images (#women, #man, #transwoman, #transman, #agender, #genderqueer, #non-binary) and analyzed by four of the largest providers of facial analysis services (IBM, Amazon, Microsoft and Clarifai). As our vision and our cultural understanding of what gender is has evolved, the algorithms driving our technological future have not. That’s deeply problematic.” - Jed Brubaker. Notably, Google was not included because it does not offer gender recognition services. On average, the systems were most accurate with photos of cisgender women (those born female and identifying as female), getting their gender right 98.3% of the time. They categorized cisgender men accurately 97.6% of the time. But trans men were wrongly identified as women up to 38% of the time. And those who identified as agender, genderqueer or nonbinary—indicating that they identify as neither male or female—were mischaracterized 100% of the time. ‘These systems don’t know any other

### 3.4.4

#### Abordagem das propostas legislativas sobre a identificação biométrica e categorização de pessoas singulares.

As preocupações e resultados expostos nos referidos estudos não são exclusividade da sociedade norte-americana e vêm sendo objeto de debate em diferentes países por aqueles que pretendem regular o uso da IA seja por meio de *soft law* ou *hard law*.

No caso da União Europeia, dentre os sistemas de IA de risco elevado listados no art. 6º, 2 e anexo III da EUR-Lex 52021PC0206, o primeiro domínio elencado trata da identificação biométrica e categorização de pessoas singulares<sup>195</sup>.

O art. 14º, que prevê a obrigatoriedade de supervisão humana sobre as decisões autônomas de IA de risco elevado, especialmente no que se refere à identificação biométrica e categorização de pessoas singulares, traz no seu inciso 5 a obrigatoriedade de verificação e confirmação da decisão do sistema por duas pessoas singulares, sendo vedada qualquer ação ou decisão lastreada somente na identificação do sistema<sup>196</sup>.

Enquanto nos demais sistemas de risco elevado bastaria apenas uma etapa de verificação humana, no caso de identificação biométrica ou de pessoas singulares seria necessária uma dupla verificação.

A minuta do substitutivo do PL nº 21/2020 e o PL nº 2328/2023, no art. 17, inciso X, também coloca a identificação biométrica no rol dos sistemas de alto risco, considerando a necessidade de cuidado especial com este tipo de aplicação, além de, no mesmo artigo, em seu inciso IV, prever como de alto risco sistemas que

---

language but male or female, so for many gender identities it is not possible for them to be correct,' said Brubaker." COLORADO.EDU. CU Boulder Today. **Facial recognition software has a gender problem.** Lisa Marshall. Disponível em: < <https://www.colorado.edu/today/2019/10/08/facial-recognition-software-has-gender-problem>>. Publicado em: 08.10.2019. Acesso em: 10.08.2022.

<sup>195</sup> Anexo III. Sistemas de Inteligência Artificial de risco elevado a que se refere o artigo 6º, nº 2, são os sistemas de IA incluídos num dos domínios a seguir enumerados: 1. Identificação biométrica e categorização de pessoas singulares: a) Sistemas de IA concebidos para serem utilizados para a identificação biométrica à distância «em tempo real» e «em diferido» de pessoas singulares.

<sup>196</sup> 5. Em relação aos sistemas de IA de risco elevado a que se refere o anexo III, ponto 1, alínea a), as medidas referidas no nº 3 devem, além disso, permitir assegurar que nenhuma ação ou decisão seja tomada pelo utilizador com base na identificação resultante do sistema, salvo se a mesma tiver sido verificada e confirmada por, pelo menos, duas pessoas singulares.

impliquem em revogação de serviços públicos “que sejam considerados essenciais”<sup>197</sup>, que também se aplicaria ao caso da estudante trans do DF.

Destarte, as implicações do reconhecimento facial e, consequentemente, decisões automatizadas de reconhecimento de gênero têm sido tratadas com ênfase por aqueles que pretendem regular o tema como uma forma de preservação da autonomia humana face a “autonomia” das máquinas.

Se tais proposições legais vierem a ser aprovadas, uma estudante transgênero em situação semelhante ao exemplo citado não ficaria semanas sem acesso ao transporte público devido à uma decisão instantânea de um sistema de reconhecimento facial instalado dentro de um ônibus.

Essa proposição legislativa é apenas um exemplo de tutela da situação, pois, além das tentativas de regulação da IA tanto na Europa quanto no Brasil, há princípios e dispositivos legais já vigentes que podem ser utilizados como justificativa para fundamentar a aplicação da supervisão humana como instrumento de tutela da autonomia existencial.

Com efeito, até o momento, o presente estudo investigou as características e especificidades da IA que podem ter reflexos relevantes na autonomia existencial, demonstrando a relevância deste instituto jurídico a fim de justificar a sua proteção por meio da utilização da supervisão humana, que será explorada no capítulo seguinte.

---

<sup>197</sup> Art. 17. São considerados sistemas de inteligência artificial de alto risco aqueles utilizados para as seguintes finalidades: [...] IV – avaliação de critérios de acesso, elegibilidade, concessão, revisão, redução ou revogação de serviços privados e públicos que sejam considerados essenciais, incluindo sistemas utilizados para avaliar a elegibilidade de pessoas naturais quanto a prestações de serviços públicos de assistência e de seguridade; [...] X – sistemas biométricos de identificação.

## 4.

### **A supervisão humana como instrumento de tutela da autonomia existencial no contexto de utilização da IA.**

Este capítulo, desenvolvimento final do presente trabalho, tem por escopo enfatizar a relevância da utilização da supervisão humana como forma de tutela da autonomia existencial no processo de tomada de decisões pela IA, bem como apresentar embasamento jurídico a justificar a sua aplicação.

Após a definição do tipo de supervisão humana tratado, serão analisados os princípios relevantes para a matéria, a forma de inserção da supervisão humana como ferramenta nas propostas legislativas do Brasil e União Europeia e, por fim, os dispositivos legais vigentes que apoiam a premissa.

#### 4.1

#### **Finalidade e forma de implementação da supervisão humana.**

A presença de um humano no circuito decisório da IA visa assegurar a promoção de valores humanos e evitar as ameaças à dignidade resultantes de decisões puramente automatizadas<sup>198</sup>. O olhar humano é uma maneira de impedir que na relação humano-máquina ocorra a desumanização do indivíduo<sup>199</sup>.

Segundo o guia de orientações éticas do GPAN IA (2019a) os “sistemas de IA devem apoiar a autonomia e a tomada de decisões dos seres humanos”

---

<sup>198</sup> “Values related to fairness, transparency, accuracy and accountability are not the only ones implicated when algorithms are addressed as objects of political scrutiny; there is, in addition, the indignity that results from automated decision making, or the lack of a human in the loop. I argue that the way in which the legal communities have constructed personhood in relation to computer automation provides insight into fascinating, various cultural concepts of people and algorithms. The US and EU regions negotiate the interplay between computer automation and personhood in shared ways, but Europeans have drawn important lines regarding what it means to be a human, based on how computational technologies must adhere to certain restrictions on fully automated processing – injecting a human in the loop.” JONES, Meg Leta. **The right to a human in the loop: Political constructions of computer automation and personhood.** *Social Studies of Science* 47.2 (2017): 216-239, p. 217.

<sup>199</sup> “European political culture holds that treating an individual in a wholly automated way, or to provide only automated treatment, is to dehumanize the individual, because a machine can only treat a human in a computational manner. Thus, to treat a human in a wholly computational manner reduces the individual’s dignity and restoration of dignity can be provided by a human in the loop.” JONES, p. 231-232.



permitindo a supervisão humana<sup>200</sup>, que pode ser realizada por meio de 3 abordagens, sendo as duas primeiras: (i) intervenção humana (*human-in-the-loop* - HITL), “refere-se à capacidade de intervenção humana em todos os ciclos de decisão do sistema, a qual, em muitos casos, não é possível nem desejável”<sup>201</sup>; (ii) fiscalização humana (*human-on-the-loop* – HOTL), “refere-se à capacidade de intervenção humana durante o ciclo de concepção do sistema e de acompanhamento do funcionamento do sistema”.

A supervisão tratada no presente estudo e abordada nas propostas legislativas analisadas diz respeito à fase de utilização da IA sendo, portanto, um mecanismo, como dito anteriormente, de controle *ex post*. Logo, as abordagens HITL (intervenção) e HOTL (fiscalização) não se mostram adequadas para o conceito de supervisão humana que aqui se aplica.

A última forma de abordagem, a do controle humano (*human-in-command* — HIC), diz respeito à supervisão de todos os espectros das atividades do sistema considerando seus impactos econômicos, sociais, jurídicos e éticos, e dá ao supervisor o poder de decisão de quando e de como utilizar o sistema, de não o utilizar ou de anular uma decisão por ele tomada<sup>202</sup>.

De acordo com as descrições, a abordagem HIC, que propõe que o humano esteja no comando, no controle, se mostra mais aderente ao conteúdo das propostas legislativas analisadas e aos princípios abordados nos tópicos a seguir.

---

<sup>200</sup> GPAN IA, 2019a, p. 18-19.

<sup>201</sup> GPAN IA, 2019a, p. 19.

<sup>202</sup> “A abordagem HIC refere-se à capacidade de supervisionar toda a atividade do sistema de IA (incluindo o seu impacto económico, societal, jurídico e ético mais geral) e de decidir quando e como utilizar o sistema em qualquer situação específica. Tal pode incluir a decisão de não utilizar um sistema de IA numa determinada situação, de estabelecer níveis de apreciação humana durante a utilização do sistema, ou de assegurar a capacidade de anular uma decisão tomada por um sistema. Além disso, deve ser garantido que as autoridades públicas responsáveis pela aplicação da lei têm a possibilidade de exercer a supervisão em conformidade com o seu mandato. Podem ser necessários mecanismos de supervisão em graus variáveis para apoiar outras medidas de segurança e controlo, dependendo do domínio de aplicação e do potencial risco do sistema de IA. Não havendo alteração das demais condições, quanto menor for a supervisão que um ser humano pode exercer sobre um sistema de IA, maior será a necessidade de sujeitar o mesmo a amplos testes e a uma governação rigorosa.” GPAN IA, 2019a, p. 19.

## 4.2

### A relevância do controle humano no uso da tecnologia.

Consoante exposto no capítulo anterior, as decisões produzidas por mecanismos de IA produzem efeitos na esfera da autonomia humana e afetam as relações jurídicas estabelecidas entre sujeitos de direito para os quais a tecnologia funciona como uma ferramenta de intermediação.

Idealmente, por meio de utilização da tecnologia ocorreria a concretização das vontades de ambos os polos da relação. Entretanto, devido à dinâmica de funcionamento dos processos de *machine* e *deep learning*<sup>203</sup>, que pode vir a produzir resultados imprevisíveis, tanto a autonomia do desenvolvedor/fornecedor, quanto a autonomia do usuário do sistema, podem ser afetadas negativamente.

A necessidade de centralidade no humano, que é um dos princípios da utilização da IA tanto na proposta legislativa brasileira, quanto na europeia<sup>204</sup>, requer a criação, implementação e utilização de mecanismos que propiciem o controle humano no uso da tecnologia, afinal, o “permanente objetivo do Direito, em suas manifestações diversas, é o *ser humano*.”<sup>205</sup>

Antes das tentativas de regulação legislativas mencionadas, inúmeros guias éticos e de boas práticas foram criados, produzindo um compilado de princípios associados ao rápido desenvolvimento da inteligência artificial<sup>206</sup> que desempenham um importante papel norteador das práticas relacionadas à IA.

Um estudo publicado em 2020 pelo Berkman Klein Center, centro de estudos da Universidade de Harvard, identificou oito temas chave recorrentes e

---

<sup>203</sup> “Algumas técnicas de aprendizagem automática, embora muito bem-sucedidas do ponto de vista da exatidão, são muito opacas quanto à forma como tomam decisões. O conceito de *IA de caixa negra* refere-se aos cenários em que não é possível identificar a razão que está na origem de determinadas decisões.” GPAN IA, 2019b, p. 6.

<sup>204</sup> Vide tópico 2.2.3.

<sup>205</sup> NADER, p. 287.

<sup>206</sup> “Alongside the rapid development of artificial intelligence (AI) technology, we have witnessed a proliferation of “principles” documents aimed at providing normative guidance regarding AI-based systems. Our desire for a way to compare these documents – and the individual principles they contain – side by side, to assess them and identify trends, and to uncover the hidden momentum in a fractured, global conversation around the future of AI, resulted in this white paper and the associated data visualization.” FJELD, Jessica, et al. **Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI**. Berkman Klein Center Research Publication 2020-1 (2020), p. 3.

comuns nos diversos guias de boas práticas e éticos no uso da IA analisados<sup>207</sup>. Dentre eles, está o controle humano da tecnologia (*human control of technology*)<sup>208</sup> que se desdobra em três princípios: (i) revisão humana das decisões automatizadas; (ii) capacidade de recusar a decisão automatizada; (iii) controle humano da tecnologia no sentido *lato sensu*.

O princípio da revisão humana das decisões automatizadas defende a ideia de que quando sistemas de IA são implementados, as pessoas que estão sujeitas às suas decisões devem estar aptas a requerer e obter a revisão humana de tais decisões. Trata-se de um mecanismo de controle *ex post*, na fase de implementação, propiciando a oportunidade de remediar um resultado questionável<sup>209</sup>.

O fundamento de tal princípio é a necessidade de proteção da capacidade de autodeterminação dos humanos ao interagirem com sistemas de IA<sup>210</sup>. Portanto, o que ele resguarda é a autonomia existencial, a capacidade consciente de fazer escolhas, de livre desenvolvimento da personalidade e de realização de um projeto de vida individual.

Nas orientações éticas para uma IA de confiança do GPAN IA (2019a), este princípio, assim como aquele que será descrito a seguir, aparece vinculado ao “princípio do respeito da autonomia humana”<sup>211</sup> que é “um dos quatro princípios éticos enraizados nos direitos fundamentais, que devem ser respeitados para

---

<sup>207</sup> “The first substantial aspect of our findings are the eight key themes themselves: Privacy. [...] Accountability. [...] Safety and Security. [...] Transparency and Explainability. [...] Fairness and Non-discrimination. [...] Human Control of Technology. [...] Professional Responsibility. [...] Promotion of Human Values.” FJELD, 2020, p. 4-5.

<sup>208</sup> Segundo o referido estudo, nas orientações éticas para uma IA de confiança produzidas pelo GPAN IA, o tema *human review of automated decisions* é abordado com ampla cobertura. Vide gráfico em FJELD, 2020, p. 8-9.

<sup>209</sup> “The principle of ‘human review of automated decision’ stands for the idea that where AI systems are implemented, people who are subject to their decisions should be able to request and receive human review of those decisions. In contrast to other principles under this theme, the “human review of automated decision” principle is always *ex post* in its implementation, providing the opportunity to remedy an objectionable result.” FJELD et al., 2020, p. 53.

<sup>210</sup> “Although the documents in our dataset are situated in a variety of contexts, there is remarkable commonality between them in the articulation of this principle. The underlying rationale, when explicit, is that ‘Humans interacting with AI systems must be able to keep full and effective self-determination over themselves.’ FJELD et al., 2020, p. 53.

<sup>211</sup> Sobre o princípio do respeito da autonomia humana: “Os direitos fundamentais em que a UE se alicerça visam garantir o respeito da liberdade e da autonomia dos seres humanos. Os seres humanos que interajam com sistemas de IA devem poder manter uma autodeterminação plena e efetiva sobre si próprios e participar no processo democrático. Os sistemas de IA não devem subordinar, coagir, enganar, manipular, condicionar ou arregimentar injustificadamente os seres humanos. Em vez disso, os sistemas de IA devem ser concebidos para aumentar, complementar e capacitar as competências cognitivas, sociais e culturais dos seres humanos.” GPAN IA, 2019a, p. 14-15.

assegurar que os sistemas de IA são desenvolvidos, implantados e utilizados de forma confiável”<sup>212</sup>.

Na minuta do substitutivo do PL nº 21/2020 e no PL nº 2338/2023, o princípio aparece: (i) como um direito descrito em uma seção específica da minuta<sup>213</sup>; (ii) dentre os fundamentos para o desenvolvimento, implementação e uso da IA no Brasil descritos no art. 2º, em especial no inciso VIII<sup>214</sup>; (ii) dentre os princípios para o desenvolvimento, implantação e uso da IA no Brasil estabelecidos no art. 3º, em especial nos incisos II, VI, VIII, IX e X<sup>215</sup>; (iv) dentre as medidas de governança presentes no art. 21, em especial no inciso IV<sup>216</sup>.

Saliente-se que a revisão humana se difere da supervisão humana pelo momento em que é aplicada no processo decisório. Ambas configuram formas de controle *ex post* e visam garantir o controle humano no uso da tecnologia dando a palavra final à uma pessoa ou um grupo de pessoas.

Contudo, enquanto a revisão visa propiciar um remédio para que humano tenha a oportunidade de contestar um veredito que já foi dado pela máquina, a supervisão impede que o veredito seja dado pela IA sem a sua anuência, garantindo o direito primordial de decisão ao humano que analisará as percepções sugeridas pela tecnologia.

O princípio da capacidade de recusar a decisão automatizada permite aos indivíduos, como o próprio título sugere, a oportunidade e escolha de não se sujeitar

---

<sup>212</sup> GPAN IA, 2019a, p. 14.

<sup>213</sup> Capítulo II. Dos direitos. Seção III. Do direito de contestar decisão e solicitar intervenção humana. Artigos 9º ao 11.

<sup>214</sup> Art. 2º O desenvolvimento, implementação e uso de sistemas de inteligência artificial no Brasil têm como fundamentos: [...] VIII – a privacidade, a proteção de dados e a autodeterminação informativa.

<sup>215</sup> Art. 3º O desenvolvimento, implementação e uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios: I – crescimento inclusivo, desenvolvimento sustentável e bem-estar; [...] II – autodeterminação e liberdade de decisão e de escolha; [...] VI – transparência, explicabilidade, inteligibilidade e auditabilidade; [...] VIII – devido processo legal, contestabilidade e contraditório; IX – rastreabilidade das decisões durante o ciclo de vida de sistemas de inteligência artificial como meio de prestação de contas e atribuição de responsabilidades a uma pessoa natural ou jurídica; X – prestação de contas, responsabilização e reparação integral de danos.

<sup>216</sup> Art. 21. Adicionalmente às medidas de governança estabelecidas neste capítulo, órgãos e entidades do poder público da União, Estados, Distrito Federal e Municípios, ao contratar, desenvolver ou utilizar sistemas de inteligência artificial considerados de alto risco, adotarão as seguintes medidas: [...] IV – garantia facilitada e efetiva ao cidadão, perante o poder público, de direito à explicação e revisão humanas de decisão por sistemas de inteligência artificial que gerem efeitos jurídicos relevantes ou que impactem significativamente os interesses do afetado, a ser realizada pelo agente público competente.

a decisões de sistemas de IA aonde eles são implementados<sup>217</sup>. Este é mais um princípio que visa assegurar o direito de escolha, de livre existência e respeito à autonomia humana.

Considerado como um corolário do direito à notificação, segundo o qual o humano deve ser informado de que interage com uma IA, há uma certa controvérsia quanto à sua forma de aplicação e extensão. A IA está presente em muitos aspectos da vida cotidiana e decisões são tomadas por tais mecanismos a todo instante, há que se pesar a viabilidade de propiciar este tipo de escolha aos indivíduos de forma indiscriminada<sup>218</sup>.

A exemplo, talvez fique inviável que empresas que forneçam produtos e serviços, e trabalhem com tecnologia de IA para realizar vendas e entregas, deem a opção ao consumidor de escolher não ser atendido pela máquina e sim por uma pessoa. Isso poderia encarecer sobremaneira a infraestrutura e logística e inviabilizar o negócio.

Contudo, rechaçar a ideia de plano, com base apenas na análise econômica da questão, é negar a centralidade no humano, que é princípio basilar da utilização ética da IA. Há que se ponderar quais são os direitos relevantes, ligados diretamente à humanidade do indivíduo e dizem respeito à preservação da autonomia existencial individual e coletiva, os direitos fundamentais que são merecedores desse tipo de tutela<sup>219</sup>.

---

<sup>217</sup> “The ‘ability to opt out of automated decision’ principle is defined, as its title suggests, as affording individuals the opportunity and choice not to be subject to AI systems where they are implemented.” FJELD et al., 2020, p. 54.

<sup>218</sup> “Of course, individuals interact with AI systems in numerous ways: their information may be used as training data; they may be indirectly impacted by systemic deployments of AI, and they may be personally subject to automated decisions. Perhaps because these principles are articulated with relative brevity, or perhaps because of the significant challenges in implementation, only three documents contained this principle: AI in the UK, the European High Level Expert Group guidelines, and the Smart Dubai AI principles. All documents articulated this principle as a natural corollary of the right to notification when interacting with an AI system. The latter two documents disagree about the extent of the principle’s implementation, with Smart Dubai saying that entities should “consider” providing the ability to opt out ‘where appropriate’ and the European document standing for a ‘meaningful opportunity for human choice’.” FJELD et al., 2020, p. 54.

<sup>219</sup> “[...] 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

Ao descrever o princípio do respeito à autonomia humana, acima citado, o GPAN IA (2019a) defende que, na distribuição de funções entre humanos e máquinas, deve ser deixada “uma oportunidade significativa para a escolha humana”.

Partindo-se de uma interpretação teleológica do documento e do que é dito logo em seguida<sup>220</sup>, pode-se entender que a preocupação do grupo de peritos é mais voltada para a preservação da autonomia existencial do que para uma aplicação genérica e irrestrita do princípio, dando espaço para a ponderação de interesses e se referindo à supervisão humana como forma de controle das decisões.

Na minuta de substitutivo do PL nº 21/2020 e no PL nº 2338/2023, o princípio encontra respaldo no art. 3º, em especial nos incisos II, VI e VII<sup>221</sup>; no art. 5º, em especial, nos incisos II e III<sup>222</sup>; e art. 8º, *caput*<sup>223</sup>. Note-se que no art. 5º, III a minuta se refere a decisões ou previsões “que impactem de maneira significativa os interesses do afetado” sugerindo uma aplicação do princípio pautada na razoabilidade.

O controle humano da tecnologia no sentido *lato sensu*, que está sob o guarda-chuva do tema homônimo segundo o estudo citado, é o princípio que requer que sistemas de IA sejam designados e implementados de forma que as pessoas tenham capacidade de intervir nas suas ações<sup>224</sup>.

---

princípios éticos da inteligência artificial e das decisões automatizadas amparadas no uso de algoritmos, como será abordado ao final deste artigo.” DONEDA et al., 2018, p. 6.

<sup>220</sup> “A distribuição de funções entre os seres humanos e os sistemas de IA devem seguir princípios de concepção centrados no ser humano e deixar uma oportunidade significativa para a escolha humana. Isto implica que se garanta a supervisão e o controle por parte de seres humanos sobre os processos de trabalho dos sistemas de IA.” GPAN IA, 2019a, p. 15.

<sup>221</sup> Art. 3º O desenvolvimento, implementação e uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios: [...] II – autodeterminação e liberdade de decisão e de escolha; [...] VI – transparência, explicabilidade, inteligibilidade e auditabilidade; [...] VIII – devido processo legal, contestabilidade e contraditório.

<sup>222</sup> Art. 5º Pessoas afetadas por sistemas de inteligência artificial têm os seguintes direitos, a serem exercidos na forma e nas condições descritas neste Capítulo: [...] II – direito a explicação sobre a decisão, recomendação ou previsão tomada por sistemas de inteligência artificial; III – direito de contestar decisões ou previsões de sistemas de inteligência artificial que produzam efeitos jurídicos ou que impactem de maneira significativa os interesses do afetado.

<sup>223</sup> Art. 8º A pessoa afetada por sistema de inteligência artificial poderá solicitar explicação sobre a decisão, previsão ou recomendação, com informações a respeito dos critérios e dos procedimentos utilizados, assim como sobre os principais fatores que afetam tal previsão ou decisão específica [...].

<sup>224</sup> “The ‘human control of technology (other/general)’ principle requires that AI systems are designed and implemented with the capacity for people to intervene in their actions. This was the most commonly referenced principle under the theme of Human Control of Technology, and most of the documents that included it framed it broadly, as in our definition.” FJELD et al., 2020, p. 54.

Ele diz respeito à capacidade de escolha do ser humano para: (i) definir em que circunstâncias utilizar a IA; (ii) de quando delegar à IA o poder decisório; (iii) promover valores humanos com foco na qualidade de vida humana; (iv) determinar quando o sistema deve transferir o controle ao humano; (v) promover o monitoramento contínuo da IA com capacidade de intervenção humana; (vi) manutenção da tecnologia com o *status* legal de ferramenta, deixando para os sujeitos de direito a retenção do controle sobre e responsabilidade pelas máquinas<sup>225</sup>.

Este princípio geral acaba por englobar uma série de outros princípios relevantes na criação e implementação de sistemas de IA tais como: promoção de valores humanos, acesso à tecnologia, utilização benéfica da IA para a sociedade, responsabilidade profissional, diversidade e não discriminação, equidade, transparência, explicabilidade, prevenção de danos, privacidade e responsabilização (*accountability*).

Todos estes princípios visam preservar a centralidade do humano e promover a tutela da pessoa humana por meio de diferentes formas de controle sobre o desenvolvimento, implementação e utilização da tecnologia. Uma das formas de se garantir esse domínio do humano sobre a máquina diz respeito ao tópico central ora analisado: a supervisão humana das decisões autônomas de IA.

### 4.3

#### **A importância da promoção de valores humanos no uso da IA.**

Outro tema abordado no estudo publicado em 2020 pelo Berkman Klein Center, e que se mostra relevante para o presente trabalho, é o da promoção de valores humanos. Ele se relaciona com o controle humano da tecnologia, porém, traz uma abordagem mais universal, transportando para o contexto de criação e utilização da IA valores globais que servem de fundamento para o desenvolvimento da sociedade atual.

---

<sup>225</sup> Resumo da descrição do princípio. FJELD et al., 2020, p. 54-55.

Assim, o tópico está diretamente relacionado à defesa dos direitos humanos, direitos humanos internacionais, direitos fundamentais, dignidade da pessoa humana e autonomia humana<sup>226</sup>.

Trata-se da ideia de fomento e preservação de valores para as presentes e futuras gerações por meio da observância da dignidade humana, integridade, liberdade, privacidade, diversidade cultural e de gênero, direitos fundamentais e proteção do meio ambiente e biodiversidade<sup>227</sup>.

Três são os princípios que, segundo o referido estudo, decorrem da preocupação com a manutenção e promoção de valores universais na criação, implementação e uso da IA que andam de mãos dadas com a solidariedade social e autonomia existencial individual, interpessoal e coletiva, quais sejam: (i) valores humanos e florescimento humano; (ii) acesso à tecnologia; (iii) máximo proveito para a sociedade.

O princípio dos valores humanos e florescimento humano define o desenvolvimento e uso da IA com observância das normas sociais, crenças culturais intrínsecas e o melhor interesse da humanidade. Traduz a preocupação de utilização da IA com vistas ao progresso da civilização humana<sup>228</sup>.

A aplicação do princípio tem sido sugerida não somente na fase de desenvolvimento dos sistemas, mas também na fase de implementação e durante todo o

---

<sup>226</sup> “While principles under this theme were coded distinctly from explicit references to human rights and international instruments of human rights law, there is a strong and clear connection. References to human values and human rights were often adjacent to one another, and where the documents provided more specific articulations of human values, they are largely congruous with existing guarantees found in international human rights law. Moreover, principles that refer to human values often include explicit references to fundamental human rights or international human rights, or mention concepts from human rights frameworks and jurisprudence such as human dignity or autonomy. The OECD and G20 AI principles also add ‘internationally recognized labor rights’ to this list.” FJELD et al., 2020, p. 60.

<sup>227</sup> “There is also an overlap between articulations of the Promotion of Human Values and social, economic, or environmental concepts that are outside the boundaries of political and civil rights, including among documents coded under the principle of AI ‘leveraged to benefit society.’ Principle 3, ‘Make AI Serve People and Planet,’ from the UNI Global Union’s AI principles, is emblematic, calling for: ‘throughout their entire operational process, AI systems [to] remain compatible and increase the principles of human dignity, integrity, freedom, privacy and cultural and gender diversity, as well as ... fundamental human rights. In addition, AI systems must protect and even improve our planet’s ecosystems and biodiversity.’ FJELD et al., 2020, p. 60-61.

<sup>228</sup> “The principle of ‘human values and human flourishing’ is defined as the development and use of AI with reference to prevailing social norms, core cultural beliefs, and humanity’s best interests. As the Chinese AI Governance Principles put it, this principle means that AI should ‘serve the progress of human civilization.’ FJELD et al., 2020, p. 61.



ciclo de vida da IA<sup>229</sup>. Este princípio aparece nas orientações éticas do GPAN IA (2019a) na forma do princípio da prevenção de danos<sup>230</sup>, do princípio da equidade<sup>231</sup> e como um dos requisitos para uma IA de confiança, o bem-estar societal e ambiental<sup>232</sup>.

No texto atual do PL nº 21/2020, a preocupação com os valores humanos e florescimento humano aparece: (i) dentre os objetivos a aplicação da IA no Brasil descritos no art. 3º, em especial nos incisos I, IV e VI<sup>233</sup>; (ii) dentre os fundamentos para o desenvolvimento e aplicação da IA no Brasil descritos no art. 4º em especial nos incisos I, II, III, IV e V<sup>234</sup>; (iii) dentre os princípios para o desenvolvimento e

---

<sup>229</sup> “Many of the documents that refer to the theme of ‘human values and human flourishing’ are concerned with how the societal impacts of AI can be managed through AI system design. [...] However, most, if not all, of these documents also acknowledge that human values will also need to be promoted in the implementation of AI systems and ‘throughout the AI system lifecycle.’ FJELD et al., 2020, p. 61.

<sup>230</sup> “Os sistemas de IA não devem causar danos ou agravá-los nem afetar negativamente os seres humanos de qualquer outra forma. Isto implica a proteção da dignidade, bem como da integridade mental e física, do ser humano. Os sistemas de IA e os ambientes em que operam devem ser seguros e protegidos. Devem ser tecnicamente sólidos e deve garantir-se que não estão abertos a utilizações malévolas. As pessoas vulneráveis devem receber maior atenção e ser incluídas no desenvolvimento e na implantação dos sistemas de IA. Há também que prestar especial atenção às situações em que os sistemas de IA podem causar ou agravar impactos negativos devido a assimetrias de poder ou de informação, nomeadamente entre empregadores e trabalhadores, empresas e consumidores ou governos e cidadãos. A prevenção dos danos implica também ter em consideração o ambiente natural e todos os seres vivos.” GPAN IA, 2019a, p. 15.

<sup>231</sup> “O desenvolvimento, a implantação e a utilização dos sistemas de IA devem ser equitativos. Embora reconheçamos que há muitas interpretações diferentes de equidade, consideramos que esta tem uma dimensão substantiva e processual. A dimensão substantiva implica um compromisso com: a garantia de uma distribuição equitativa e justa dos benefícios e dos custos, bem como de inexistência de enviesamentos injustos, discriminação e estigmatização contra pessoas e grupos. Se for possível evitar os enviesamentos, os sistemas de IA podem até aumentar a equidade societal. A igualdade de oportunidades em termos de acesso à educação, aos bens e serviços e à tecnologia deve ser igualmente promovida.” GPAN IA, 2019a, p. 15.

<sup>232</sup> “Em conformidade com os *princípios da equidade e da prevenção de danos*, a sociedade em geral, outros seres sensíveis e o ambiente também devem ser considerados partes interessadas ao longo do ciclo de vida da IA. A sustentabilidade e a responsabilidade ecológica dos sistemas de IA devem ser incentivadas e deve ser promovida a investigação em soluções de IA direcionadas para áreas de interesse global, como, por exemplo, os Objetivos de Desenvolvimento Sustentável. Idealmente, a IA deve ser utilizada em benefício de todos os seres humanos, incluindo as gerações futuras.” GPAN IA, 2019a, p. 23.

<sup>233</sup> Art. 3º A aplicação de inteligência artificial no Brasil tem por objetivo o desenvolvimento científico e tecnológico, bem como: I – a promoção do desenvolvimento econômico sustentável e inclusivo e do bem-estar da sociedade; [...] IV – a melhoria na prestação de serviços públicos e na implementação de políticas públicas; [...] VI – a proteção e a preservação do meio ambiente.

<sup>234</sup> Art. 4º O desenvolvimento e a aplicação da inteligência artificial no Brasil têm como fundamentos:

I – o desenvolvimento científico e tecnológico e a inovação; II – a livre iniciativa e a livre concorrência; III – o respeito à ética, aos direitos humanos e aos valores democráticos; IV – a livre manifestação de pensamento e a livre expressão da atividade intelectual, artística, científica e de comunicação; V – a não discriminação, a pluralidade, o respeito às diversidades regionais, a inclusão e o respeito aos direitos e garantias fundamentais do cidadão;

aplicação da IA no Brasil estabelecidos no art. 5º, especialmente naqueles estabelecidos nos incisos I, II, III, VI e VII<sup>235</sup>.

Na minuta de substitutivo do PL nº 21/2020, o princípio se apresenta: (i) dentre os fundamentos para o desenvolvimento, implementação e uso da IA no Brasil descritos no art. 2º, em especial, nos incisos II, IV, VII e X<sup>236</sup>; (ii) dentre os princípios para o desenvolvimento, implantação e uso da IA no Brasil estabelecidos no art. 3º, em especial, nos incisos I, IV, V, XI e XII<sup>237</sup>.

Por trazer uma conotação cultural e social forte, o princípio conversa com a ideia de controle humano da tecnologia e, consequentemente, com a proposta de manter as decisões relevantes, que digam respeito aos valores humanos e a preocupação com a prosperidade humana, sob o controle de pessoas. Diferentemente da máquina, o ser humano traz consigo o instinto de preservação e perpetuação da espécie que não deve ser ignorado.

O princípio do acesso à tecnologia faz emergir uma preocupação genuína em não aumentar a desigualdade entre indivíduos de modo a garantir igual acesso das pessoas ao amplo espectro de tecnologia oferecido, bem como aos seus benefícios. É uma tentativa de universalizar o uso da tecnologia, evitando-se a exclusão

---

<sup>235</sup> Art. 5º São princípios para o desenvolvimento e a aplicação da inteligência artificial no Brasil: I – finalidade benéfica: busca de resultados benéficos para a humanidade pelos sistemas de inteligência artificial; II – centralidade do ser humano: respeito à dignidade humana, à privacidade, à proteção de dados pessoais e aos direitos fundamentais, quando o sistema tratar de questões relacionadas ao ser humano; III – não discriminação: mitigação da possibilidade de uso dos sistemas para fins discriminatórios, ilícitos ou abusivos; [...] VI – segurança e prevenção: utilização de medidas técnicas, organizacionais e administrativas, considerando o uso de meios razoáveis e disponíveis na ocasião, compatíveis com as melhores práticas, os padrões internacionais e a viabilidade econômica, direcionadas a permitir o gerenciamento e a mitigação de riscos oriundos da operação de sistemas de inteligência artificial durante todo o seu ciclo de vida e o seu contínuo funcionamento; VII – inovação responsável: garantia de adoção do disposto nesta Lei, pelos agentes que atuam na cadeia de desenvolvimento e operação de sistemas de inteligência artificial que estejam em uso, documentando seu processo interno de gestão e responsabilizando-se, nos limites de sua respectiva participação, do contexto e das tecnologias disponíveis, pelos resultados do funcionamento desses sistemas.

<sup>236</sup> Art. 2º O desenvolvimento, implementação e uso de sistemas de inteligência artificial no Brasil têm como fundamentos: [...] II – o respeito aos direitos humanos e aos valores democráticos; [...] IV – a proteção ao meio ambiente e o desenvolvimento sustentável; [...] VII – a livre iniciativa, a livre concorrência e a defesa do consumidor; [...] X – o acesso à informação e à educação, bem como a conscientização sobre os sistemas de inteligência artificial e suas aplicações.

<sup>237</sup> Art. 3º O desenvolvimento, implementação e uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios: I – crescimento inclusivo, desenvolvimento sustentável e bem-estar; [...] IV – não discriminação; V – justiça, equidade e inclusão; [...] XI – prevenção, precaução e mitigação de riscos sistêmicos derivados de usos intencionais ou não intencionais e efeitos não previstos de sistemas de inteligência artificial; e XII – não maleficência e proporcionalidade entre os métodos empregados e as finalidades determinadas e legítimas dos sistemas de inteligência artificial.

dos indivíduos ou de grupos de indivíduos quanto ao acesso aos proveitos advindos do uso da IA<sup>238</sup>.

Trata-se da preocupação com políticas de educação, treinamento e programação da IA que permitam que todos tenham igual oportunidade de acesso à tecnologia, independentemente das diferenças culturais, econômicas ou no nível de escolaridade e vulnerabilidade que vivenciam.

Nas orientações éticas do GPAN IA (2019a), este princípio aparece repetidamente: (i) dentre os direitos fundamentais para uma IA de confiança, na liberdade do indivíduo<sup>239</sup>; (ii) no já citado princípio da equidade que estabelece que a “igualdade de oportunidades em termos de acesso à educação, aos bens e serviços e à tecnologia deve ser igualmente promovida”<sup>240</sup>; (iii) dentro do requisito “diversidade, não discriminação e equidade” no item “acessibilidade e conceção universal”<sup>241</sup>.

No texto atual do PL nº 21/2020, o princípio aparece: (i) dentre os objetivos da aplicação da IA no Brasil descritos no art. 3º, em especial nos incisos I e IV, já citados; (ii) dentre os fundamentos para o desenvolvimento e aplicação da IA no Brasil descritos no art. 4º, em especial, no inciso V, já citado; (iii) dentre os princípios para o desenvolvimento e aplicação da IA no Brasil estabelecidos no art. 5º, especialmente naquele estabelecido no inciso III, já citado.

Na minuta de substitutivo do PL nº 21/2020, o princípio se apresenta: (i) dentre os fundamentos para o desenvolvimento, implementação e uso da IA no

---

<sup>238</sup> “The ‘access to technology’ principle represents statements that the broad availability of AI technology, and the benefits thereof, is a vital element of ethical and rights-respecting AI. Given the significant transformational potential of AI, documents that include this principle worry that AI might contribute to the growth of inequality.” FJELD et al., 2020, p. 61.

<sup>239</sup> “Os seres humanos devem ser livres de tomar decisões sobre as suas próprias vidas. Tal implica liberdade de intrusão soberana, mas também exige a intervenção de organizações governamentais e não governamentais para garantir que os indivíduos ou as pessoas em risco de exclusão têm igual acesso aos benefícios e oportunidades da IA.” GPAN IA, 2019a, p. 13.

<sup>240</sup> GPAN IA, 2019a, p. 15.

<sup>241</sup> “Nos domínios das relações entre as empresas e os consumidores, em especial, os sistemas devem centrar-se no utilizador e ser concebidos por forma a permitir que todas as pessoas utilizem os produtos ou serviços de IA, independentemente da sua idade, do seu género, das suas capacidades ou das suas características. A possibilidade de acesso a esta tecnologia por parte das pessoas com deficiência, presentes em todos os grupos da sociedade, reveste-se de especial importância. Os sistemas de IA não devem seguir uma abordagem única para todos os casos e devem tomar em consideração os princípios de conceção universal, que visam abranger a maior variedade possível de utilizadores, respeitando as normas de acessibilidade pertinentes<sup>45</sup>. Possibilitar-se-á, assim, um acesso equitativo e uma participação ativa de todas as pessoas em atividades humanas já existentes ou emergentes que utilizam computadores, nomeadamente em tecnologias de apoio.” GPAN IA, 2019a, p. 23.

Brasil descritos no art. 2º, em especial, no inciso X já citado; (ii) dentre os princípios para o desenvolvimento, implantação e uso da IA no Brasil estabelecidos no art. 3º, em especial, nos incisos I, II e IV já citados; (iii) no art. 7º, § 3º<sup>242</sup>.

Note-se que o texto do § 3º do art. 7º é especialmente relevante por preconizar que grupos de vulneráveis (crianças, idosos e deficientes) tenham acesso protegido, na medida do necessário, e consciente aos sistemas de IA. Tal dispositivo é imbuído do espírito do princípio de acesso que prega o respeito à diversidade e individualidades dos seres humanos de modo que todos tenham iguais oportunidade de acesso à tecnologia respeitadas as suas diferenças.

A proposta de utilizar a supervisão humana como forma de preservação da autonomia existencial é também uma forma de garantir a aplicação prática deste princípio, impedindo que vieses discriminatórios que eventualmente possam ser praticados pela IA afetem grupos de vulneráveis ou restrinjam o acesso dos indivíduos aos sistemas perpetuando distorções socioeconômicas e culturais existentes na sociedade.

O princípio do máximo proveito para a sociedade preconiza que os sistemas de IA devem servir aos interesses coletivos. Dignidade, liberdade, democracia, justiça, equidade, igualdade, solidariedade, proteção ao meio ambiente, desenvolvimento sustentável são valores e objetivos sociais que fazem parte da promoção de valores humanos que precisam estar em primeiro lugar no contexto das relações jurídicas, inclusive no uso da IA<sup>243</sup>.

---

<sup>242</sup> Art. 7º Pessoas afetadas por sistemas de inteligência artificial têm o direito de receber, previamente à contratação ou utilização do sistema de inteligência artificial, informações claras e adequadas quanto aos seguintes aspectos: [...] § 3º Os sistemas de inteligência artificial que se destinem a grupos vulneráveis, tais como crianças, adolescentes, idosos e pessoas com deficiência, serão desenvolvidos de tal modo que essas pessoas consigam entender o seu funcionamento e seus direitos em face dos agentes de inteligência artificial.

<sup>243</sup> “The principle that AI be leveraged to benefit society’ stands for the notion that AI systems should be employed in service of public-spirited goals. The documents vary in the specificity with which they articulate goals. Where they are specific in the goals they list, they may include social, political, and economic factors. Examples of beneficial ends in the European High Level Expert Group guidelines include: ‘Respect for human dignity... Freedom of the individual... Respect for democracy, justice and the rule of law... Equality, non-discrimination and solidarity - including the rights of persons at risk of exclusion... Citizens’ rights... including the right to vote, the right to good administration or access to public documents, and the right to petition the administration.’ The High Level Expert Group and the German AI strategy were the two documents to explicitly include the environment and sustainable development as factors in their determination of AI that is ‘leveraged to benefit society.’ FJELD et al., 2020, p. 62.

A ideia principal é que fatores econômicos não se sobreponham aos valores humanos e, conseqüentemente, aos interesses coletivos, como já exposto na apreciação dos princípios anteriores, que coadunam com a ideia de preservação da autonomia existencial, seja no âmbito individual, interpessoal ou coletivo.

A supervisão humana é um mecanismo que visa garantir que o controle humano da tecnologia e a promoção de valores humanos nos usos e aplicações da IA prevaleçam e, conseqüentemente, a sociedade tenha o máximo proveito na utilização da tecnologia.

#### **4.4**

#### **A supervisão humana nos textos das propostas legislativas em andamento.**

Como visto, a supervisão humana é um mecanismo de tutela da autonomia existencial que encontra respaldo nos princípios supramencionados, que foram incorporados nas propostas legislativas em andamento. Diante da possibilidade de normatização da ferramenta, é importante uma análise mais minuciosa tanto da abordagem feita nos textos das propostas legais, quanto de como seria a sua aplicação prática.

##### **4.4.1**

##### **Abordagem.**

Quando o grupo de peritos de alto nível sobre a inteligência artificial (GPNA IA) apresentou o conjunto de orientações éticas para um IA de confiança<sup>244</sup>

---

<sup>244</sup> GPAN IA, 2019a.

em 08.04.2019, elencou quatro princípios éticos<sup>245</sup> e sete requisitos essenciais<sup>246</sup>. O primeiro princípio listado é o respeito da autonomia humana<sup>247</sup> e o primeiro requisito elencado é o da ação e supervisão humanas<sup>248</sup>.

Portanto, no referido guia ético, houve uma preocupação em privilegiar a discricionariedade humana, assegurando a palavra final ao indivíduo no que diz respeito à proteção da sua autonomia. A orientação dos *experts* foi considerada na formulação da proposta legislativa europeia que, no seu art. 14º, prevê a obrigatoriedade de supervisão humana sobre as decisões autônomas de IA de risco elevado<sup>249</sup>.

Os sistemas de IA de risco elevado listados no anexo III da EUR-Lex 52021PC0206<sup>250</sup> em rol não taxativo<sup>251</sup>, dizem respeito a direitos fundamentais e questões existenciais tais como o direito à identidade, à privacidade, à educação, ao

---

<sup>245</sup> Eis os princípios éticos listados no documento: (i) respeito da autonomia humana; (ii) prevenção de danos; (iii) equidade; (iv) explicabilidade.

<sup>246</sup> Eis os requisitos essenciais listados no documento: (i) ação e supervisão humanas; (ii) solidez técnica e segurança; (iii) privacidade e governação de dados; (iv) transparência; (v) diversidade, não discriminação e equidade; (vi) bem-estar societal e ambiental; (vii) responsabilização.

<sup>247</sup> GPAN IA, 2019a, p. 14-15.

<sup>248</sup> Sobre o requisito da ação e supervisão humanas: “Os sistemas de IA devem apoiar a autonomia e a tomada de decisões dos seres humanos, tal como prescrito pelo princípio de *respeito da autonomia humana*. Isto exige que os sistemas de IA funcionem como facilitadores de uma sociedade democrática, próspera e equitativa, apoiando a ação do utilizador e a promoção dos direitos fundamentais, e que permitam a supervisão humana.” GPAN IA, 2019a, p. 18-19.

<sup>249</sup> Artigo 14.º. Supervisão humana. 1. Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de tal modo, incluindo com ferramentas de interface homem-máquina apropriadas, que possam ser eficazmente supervisionados por pessoas singulares durante o período de utilização do sistema de IA. 2. A supervisão humana deve procurar prevenir ou minimizar os riscos para a saúde, a segurança ou os direitos fundamentais que possam surgir quando um sistema de IA de risco elevado é usado em conformidade com a sua finalidade prevista ou em condições de utilização indevida razoavelmente previsíveis, em especial quando esses riscos persistem apesar da aplicação de outros requisitos estabelecidos neste capítulo.

<sup>250</sup> Os sistemas de IA de risco elevado estão descritos no art. 6º, 2 da EUR-Lex 52021PC0206, que, por sua vez, remete ao anexo III da proposta que define como de risco elevado os sistemas de IA incluídos nos seguintes domínios: i) identificação biométrica e categorização de pessoas singulares; ii) gestão e funcionamento de infraestruturas críticas; iii) educação e formação profissional; iv) emprego, gestão de trabalhadores e acesso ao emprego por conta própria; v) acesso a serviços privados e a serviços e prestações públicas essenciais, bem como o usufruto dos mesmos; vi) manutenção da ordem pública; vii) gestão da migração, do asilo e do controle das fronteiras; viii) administração da justiça e processos democráticos.

<sup>251</sup> Nos termos do art. 7º, 1 da EUR-Lex 52021PC0206, o rol de sistemas de risco elevado pode ser atualizado. *In verbis*: “A Comissão fica habilitada a adotar atos delegados nos termos do artigo 73.º para atualizar a lista do anexo III, aditando sistemas de IA de risco elevado que preencham ambas as condições que se seguem: a) Os sistemas de IA destinam-se a ser utilizados em qualquer um dos domínios enumerados no anexo III, pontos 1 a 8; b) Os sistemas de IA representam um risco de danos para a saúde e a segurança ou um risco de impacto adverso nos direitos fundamentais que, em termos de gravidade e probabilidade de ocorrência, é equivalente ou superior ao risco de danos ou impacto adverso representado pelos sistemas de IA de risco elevado já referidos no anexo III.”

emprego e trabalho, ao acesso a serviços essenciais, à segurança, ao acesso à justiça e manutenção da democracia.

O texto atual do PL nº 21/2020 não faz menção à supervisão humana especificamente. Contudo, o artigo 5º traz alguns princípios que poderiam vir a sustentar a utilização da supervisão humana em determinados processos decisórios previstos nos incisos: II - centralidade do ser humano, III – não discriminação, IV - busca pela neutralidade, VI – segurança e prevenção, e VII – inovação responsável.

É de se notar que o inciso VI, que trata do princípio da segurança e prevenção, estabelece a utilização de medidas “compatíveis com as melhores práticas, os padrões internacionais” de modo a mitigar riscos na utilização da IA<sup>252</sup>. Assim, os guias éticos e de boas práticas, bem como os projetos legislativos estrangeiros (a exemplo da EUR-Lex 52021PC0206) que vierem a se tornar regulação, poderiam servir de subsídio para a utilização da supervisão humana na aplicação do referido princípio.

O texto da minuta do substitutivo do PL nº 21/2020 e do PL nº 2338/2023 adota a mesma linha da proposta europeia no que se refere aos sistemas de alto risco, elencando-os no art. 17<sup>253</sup> da seguinte forma: I- funcionamento de estruturas críticas; II- educação e formação profissional; III- emprego, gestão de trabalhadores e acesso ao emprego; IV- serviços privados e públicos que sejam considerados essenciais; V- capacidade de endividamento e classificação de crédito de pessoas naturais; VI- serviços de respostas a emergências; VII- administração da justiça; VIII- veículos autônomos; IX- área da saúde; X- sistemas biométricos de identificação; XI- investigação criminal e segurança pública; XII- estudo analítico de crimes

---

<sup>252</sup> VI – segurança e prevenção: utilização de medidas técnicas, organizacionais e administrativas, considerando o uso de meios razoáveis e disponíveis na ocasião, compatíveis com as melhores práticas, os padrões internacionais e a viabilidade econômica, direcionadas a permitir o gerenciamento e a mitigação de riscos oriundos da operação de sistemas de inteligência artificial durante todo o seu ciclo de vida e o seu contínuo funcionamento.

<sup>253</sup> Os sistemas de IA de alto risco estão descritos no art. 17 da minuta de substitutivo do PL Nº 21/2020, sendo considerados como tais os sistemas que atuem nas seguintes esferas: I- funcionamento de estruturas críticas; II- educação e formação profissional; III- emprego, gestão de trabalhadores e acesso ao emprego; IV- serviços privados e públicos que sejam considerados essenciais; V- capacidade de endividamento e classificação de crédito de pessoas naturais; VI- serviços de respostas a emergências; VII- administração da justiça; VIII- veículos autônomos; IX- área da saúde; X- sistemas biométricos de identificação; XI- investigação criminal e segurança pública; XII- estudo analítico de crimes relativos a pessoas naturais; XIII- investigação por autoridades administrativas; XIV- gestão da migração e controle de fronteiras.

relativos a pessoas naturais; XIII- investigação por autoridades administrativas; XIV- gestão da migração e controle de fronteiras.

Note-se que, assim como na EUR-Lex 52021PC0206, os direitos listados na minuta de substitutivo do PL nº 21/2020 como de alto risco tratam de direitos fundamentais, tais como acesso à água, luz, educação, emprego, saúde, igualdade, acesso à justiça, segurança, dignidade da pessoa humana, privacidade, todos afeitos à autonomia existencial.

O art. 18 da minuta de substitutivo do PL nº 21/2020 e do PL nº 2338/2023<sup>254</sup> estabelece que o rol do art. 17 não é taxativo, enquanto o parágrafo único do art. 20 determina a necessidade de supervisão humana para sistemas de IA de alto risco.

Deste modo, observa-se uma preocupação tanto dos legisladores europeus, quanto dos legisladores brasileiros em não engessar o rol de direitos fundamentais e existenciais que mereceriam uma tutela diferenciada. Parece um passo acertado, já que a codificação (no sentido legal do termo) da existência parece ir de encontro à função promocional do direito.

#### **4.4.2**

#### **Formas de implementação e operação.**

Os itens nº 3 e 4 do art. 14 da EUR-Lex 52021PC0206 estabelecem a forma de implementação e operação da supervisão humana da proposta, determinando como, quando e por quem o mecanismo de controle deve ser integrado no sistema<sup>255</sup> e quais as capacidades, atribuições e prerrogativas que os humanos responsáveis pela tarefa devem ter.

---

<sup>254</sup> Art. 18. Caberá à autoridade competente atualizar a lista dos sistemas de inteligência artificial de risco excessivo ou de alto risco, identificando novas hipóteses [...].

<sup>255</sup> 3. A supervisão humana deve ser assegurada por meio de um ou de todos os seguintes tipos de medidas: a) Medidas identificadas e integradas, quando tecnicamente viável, pelo fornecedor no sistema de IA de risco elevado antes de este ser colocado no mercado ou colocado em serviço; b) Medidas identificadas pelo fornecedor antes de o sistema de IA de risco elevado ser colocado no mercado ou colocado em serviço e que sejam adequadas para implantação por parte do utilizador.



Assim, para que a supervisão seja efetiva e resguarde a autonomia humana, a proposta europeia determina que o(s) supervisor(es) tenha(m): (i) compreensão das capacidades e limitações do sistema; (ii) consciência da possibilidade de tendência a confiar automática e/ou excessivamente na máquina; (iii) conhecimento das ferramentas e métodos de interpretação necessários para interpretar o sistema; (iv) capacidade de decidir quando não usar o sistema, ou de ignorar, anular ou reverter o resultado; (v) capacidade de intervir ou parar o sistema por meio de um “botão de paragem”<sup>256</sup>.

Como dito anteriormente<sup>257</sup>, o item nº 5 do art. 14 ainda acrescenta uma exigência de dupla verificação e confirmação (por duas pessoas singulares) quando os sistemas tratarem de identificação biométrica e categorização de pessoas singulares<sup>258</sup>.

A minuta de substitutivo do PL nº 21/2020 trata, na seção II, das medidas de governança para sistemas de IA de alto risco, incluindo a regra do art. 20 que determina as medidas que os agentes que forneçam ou operem tais sistemas adotem. Dentre elas, está a supervisão humana, prevista no parágrafo único.

Em linha semelhante ao disposto na EUR-Lex 52021PC0206, o parágrafo único do art. 20 determina que o(s) supervisor(es) humano(s) dos sistemas de alto risco devem ter: (i) compreensão das capacidades e limitações do sistema para controlá-lo; (ii) ciência da possível tendência de confiar automática ou excessivamente

---

<sup>256</sup> 4. As medidas a que se refere o n.º 3 devem permitir que as pessoas responsáveis pela supervisão humana façam o seguinte, em função das circunstâncias: a) Compreendam completamente as capacidades e limitações do sistema de IA de risco elevado e sejam capazes de controlar devidamente o seu funcionamento, de modo que os sinais de anomalias, disfuncionalidades e desempenho inesperado possam ser detetados e resolvidos o mais rapidamente possível; b) Estejam conscientes da possível tendência para confiar automaticamente ou confiar excessivamente no resultado produzido pelo sistema de IA de risco elevado («enviesamento da automatização»), em especial relativamente aos sistemas de IA de risco elevado usados para fornecer informações ou recomendações com vista à tomada de decisões por pessoas singulares; c) Sejam capazes de interpretar corretamente o resultado do sistema de IA de risco elevado, tendo em conta, nomeadamente, as características do sistema e as ferramentas e os métodos de interpretação disponíveis; d) Sejam capazes de decidir, em qualquer situação específica, não usar o sistema de IA de risco elevado ou ignorar, anular ou reverter o resultado do sistema de IA de risco elevado; e) Serem capazes de intervir no funcionamento do sistema de IA de risco elevado ou interromper o sistema por meio de um botão de «paragem» ou procedimento similar.

<sup>257</sup> Vide item 3.3.4.

<sup>258</sup> 5. Em relação aos sistemas de IA de risco elevado a que se refere o anexo III, ponto 1, alínea a), as medidas referidas no n.º 3 devem, além disso, permitir assegurar que nenhuma ação ou decisão seja tomada pelo utilizador com base na identificação resultante do sistema, salvo se a mesma tiver sido verificada e confirmada por, pelo menos, duas pessoas singulares.

no sistema; (iii) ter conhecimento das ferramentas e métodos de interpretação disponíveis para interpretar corretamente o sistema; (iv) capacidade de optar pela não utilização do sistema ou ignorar, anular ou reverter o resultado; (v) intervir no funcionamento do sistema podendo interromper o seu funcionamento<sup>259</sup>.

Ambas as propostas apresentam um grau considerável de detalhamento na forma de implementação e operação da supervisão humana em sistemas de IA de alto ou elevado risco. A prerrogativa do supervisor humano de poder optar por não utilizar o sistema, ignorá-lo ou parar o seu funcionamento é uma clara expressão de respeito à autonomia humana e uma forma de primar pela centralidade do humano no uso da tecnologia, cimentando a classificação da tecnologia como mero instrumento de realização das capacidades humanas.

Contudo, há que se ter cautela quanto à efetivação das medidas para assegurar uma eficaz supervisão humana dos sistemas caso as propostas sejam aprovadas da forma como estão postas. Da leitura dos dispositivos supracitados, depreende-se que será necessária mão de obra altamente qualificada e treinamento específico para a realização da tarefa, considerando as especificidades do sistema a ser supervisionado e a gama de usuários afetados.

A EUR-Lex 52021PC0206 contém dispositivos de atribuição de deveres e etapas no uso da IA que, dentre outras funções atinentes aos sistemas de risco elevado, devem criar o ambiente propício para a supervisão humana de tais sistemas caso o texto seja aprovado. São elas, dentre outras: obrigação de criação de sistemas de gestão de qualidade (art. 17º); obrigação de elaboração de documentação técnica (art. 18º); avaliação de conformidade (art. 19º); manutenção de registros

---

<sup>259</sup> Art. 20. Além das medidas indicadas no art. 19, os agentes de inteligência artificial que forneçam ou operem sistemas de alto risco adotarão as seguintes medidas de governança e processos internos: [...] Parágrafo único. A supervisão humana de sistemas de inteligência artificial de alto risco buscará prevenir ou minimizar os riscos para direitos e liberdades das pessoas que possam decorrer de seu uso normal ou de seu uso em condições de utilização indevida razoavelmente previsíveis, viabilizando que as pessoas responsáveis pela supervisão humana possam: I – compreender as capacidades e limitações do sistema de inteligência artificial e controlar devidamente o seu funcionamento, de modo que sinais de anomalias, disfuncionalidades e desempenho inesperado possam ser identificados e resolvidos o mais rapidamente possível; II – ter ciência da possível tendência para confiar automaticamente ou confiar excessivamente no resultado produzido pelo sistema de inteligência artificial; III – interpretar corretamente o resultado do sistema de inteligência artificial tendo em conta as características do sistema e as ferramentas e os métodos de interpretação disponíveis; IV – decidir, em qualquer situação específica, por não usar o sistema de inteligência artificial de alto risco ou ignorar, anular ou reverter seu resultado; e V – intervir no funcionamento do sistema de inteligência artificial de alto risco ou interromper seu funcionamento.

automáticos para auditoria (art. 20º); medidas corretivas (art. 21º); cooperação das autoridades competentes (art. 23º); avaliação de conformidade (art. 43º); certificados (art. 44º); ambientes de testagem (art. 53º); criação de um comitê europeu para IA (art. 56º); códigos de conduta (art. 69º); sanções (art. 71º).

O texto da minuta do substitutivo do PL nº 21/2020 e do PL nº 2328/2023 prevê medidas de governança, controle e regulação *a posteriori* que podem ajudar a tornar a supervisão desenhada no texto realidade caso venha a ser aprovado, tais como: (i) avaliação de impacto algorítmico (art. 22 e seguintes); (ii) estabelecimento de responsabilidade civil (art. 27 e seguintes), (iii) criação de códigos de boas práticas e governança (art. 30 e seguintes), comunicação de incidentes graves (art. 31); (iv) designação de autoridade competente pelo poder executivo para fiscalizar a implementação da lei (art. 32 e seguintes); (v) sanções administrativas (art. 36); (vi) criação de *sandboxes* regulatórios (art. 38 e seguintes).

Logo, partindo-se de uma interpretação sistemática das propostas legais analisadas, a ferramenta de supervisão humana encontra respaldo em uma série de dispositivos que buscam a viabilidade da implementação e aprimoramento da supervisão humana da IA ao longo das experiências que vierem a ser construídas, o que não significa que os mecanismos propostos pelos legisladores garantam que sejam sanadas todas as questões no uso da IA que envolvam afetem a autonomia existencial.

Por se tratar de uma ideia que será objeto de debates e, se aprovada, passará pelos percalços naturais da aplicação prática de uma nova regulação, a implementação da supervisão humana, nos termos sugeridos, deverá enfrentar desafios e precisará de aprimoramento ao longo do processo de aplicação.

Elementos como a definição dos indivíduos que farão parte da engrenagem de supervisão humana da IA, seu treinamento, a prestação de contas do trabalho realizado, os possíveis custos humanos e financeiros da implementação da ferramenta, apontam a complexidade do que está por vir.

## 4.5

### **Dispositivos legais vigentes que justificam a necessidade de implementação da supervisão humana nas decisões autônomas de IA.**

Nos tópicos anteriores, discorreu-se sobre princípios específicos do campo de estudo da IA e propostas legislativas que ainda estão em discussão no Brasil e União Europeia. Porém, além deles, há normas positivadas que servem de fundamento para justificar a utilização da supervisão humana das decisões autônomas de IA como instrumento de tutela da autonomia existencial<sup>260</sup>.

A exemplo, a utilização da IA com vistas a preservar a autonomia existencial dos indivíduos e da coletividade encontra amparo no campo dos direitos humanos. Prova disso são os princípios analisados no início deste capítulo, que têm por escopo principal a centralidade no humano por meio da manutenção do controle humano da tecnologia e promoção de valores humanos.

Com o avanço no uso deste tipo de tecnologia, as preocupações com a proteção dos direitos humanos no desenvolvimento e aplicação da IA têm sido crescentes<sup>261</sup>.

Apesar do Brasil não a ter ratificado, a Declaração Universal de Direitos Humanos (DUDH) tem preponderante influência na positivação do princípio da dignidade da pessoa humana na constituição brasileira<sup>262</sup> e seus reflexos. Além da

---

<sup>260</sup> O presente tópico discorrerá sobre alguns dos dispositivos legais vigentes que podem servir de fundamento para a aplicação da supervisão humana como instrumento de tutela da autonomia existencial no uso da IA em processos decisórios. É de se ressaltar, contudo, que a lista apresenta não é exaustiva, ressaltando-se que, a depender do caso concreto, pode-se aplicar princípios contidos em legislações especiais tais como o Estatuto da Criança e do Adolescente (ECA), o Estatuto do Idoso, ou a Consolidação das Leis Trabalhistas (CLT) que não serão especificamente tratadas.

<sup>261</sup> “In recent years, the human rights community has become more engaged with digital rights, and with the impacts of AI technology in particular. Even outside of human rights specialists, there has been an increasing appreciation for the relevance of international human rights law and standards to the governance of artificial intelligence.<sup>327</sup> To an area of technology governance that is slippery and fast-moving, human rights law offers an appealingly well-established core set of concepts, against which emerging technologies can be judged. To the broad guarantees of human rights law, principles documents offer a tailored vision of the specific – and in some cases potentially novel – concerns that AI raises.” FJELD et al., 2020, p. 64.

<sup>262</sup> “O respeito à dignidade da pessoa humana, fundamento do imperativo categórico kantiano, de ordem moral, tornou-se um comando jurídico no Brasil como o advento da Constituição Federal de 1988, do mesmo modo que já havia ocorrido em outras partes. Em particular, após o término da Segunda Grande Guerra, em reação às atrocidades cometidas pelo nazi-fascismo, a Declaração Universal dos Direitos Humanos, proclamada pelas Nações Unidas em 1948, enunciava em seu art. 1º:

dignidade, o art. 1º da DUDH também preconiza os princípios da liberdade, igualdade e solidariedade<sup>263</sup>, enquanto o art. 2º trata da não discriminação<sup>264</sup>.

O texto da DUDH apresenta outros dispositivos que dizem respeito aos sistemas de alto ou elevado risco listados nas propostas legislativas tratadas, tais como o art. 23º, que trata do direito ao trabalho<sup>265</sup>; o art. 26º, que trata do direito à educação<sup>266</sup>; e o art. 29º, que trata do livre desenvolvimento da personalidade dentro da solidariedade social<sup>267</sup>.

A Convenção Americana de Direitos Humanos, Pacto de São José da Costa Rica, do qual o Brasil é signatário, estabelece no art. 5º, o direito à integridade física, psíquica e moral<sup>268</sup>; e no art. 24, o direito à igualdade<sup>269</sup>.

A Constituição Federal, por sua vez, apresenta uma série de dispositivos que podem servir de subsídio para a implementação da supervisão humana nas decisões autônomas de IA que interfiram na autonomia existencial. Além dos princípios da dignidade da pessoa humana (art. 1º, III) e da solidariedade (art. 3º, I), os direitos fundamentais individuais e sociais, previstos nos artigos 5º e 6º, respectivamente, também se aplicam<sup>270</sup>.

---

‘Todas as pessoas nascem livres e iguais em dignidade e direitos’.” BODIN DE MORAES, 2010, p. 82.

<sup>263</sup> Artigo 1º Todos os seres humanos nascem livres e iguais em dignidade e em direitos. Dotados de razão e de consciência, devem agir uns para com os outros em espírito de fraternidade.

<sup>264</sup> Artigo 2º Todos os seres humanos podem invocar os direitos e as liberdades proclamados na presente Declaração, sem distinção alguma, nomeadamente de raça, de cor, de sexo, de língua, de religião, de opinião política ou outra, de origem nacional ou social, de fortuna, de nascimento ou de qualquer outra situação. Além disso, não será feita nenhuma distinção fundada no estatuto político, jurídico ou internacional do país ou do território da naturalidade da pessoa, seja esse país ou território independente, sob tutela, autônomo ou sujeito a alguma limitação de soberania.

<sup>265</sup> Artigo 23º 1. Toda a pessoa tem direito ao trabalho, à livre escolha do trabalho, a condições equitativas e satisfatórias de trabalho e à protecção contra o desemprego.

<sup>266</sup> Artigo 26º 1. Toda a pessoa tem direito à educação. [...] 2. A educação deve visar à plena expansão da personalidade humana e ao reforço dos direitos do Homem e das liberdades fundamentais e deve favorecer a compreensão, a tolerância e a amizade entre todas as nações e todos os grupos raciais ou religiosos, bem como o desenvolvimento das actividades das Nações Unidas para a manutenção da paz.

<sup>267</sup> Artigo 29º 1. O indivíduo tem deveres para com a comunidade, fora da qual não é possível o livre e pleno desenvolvimento da sua personalidade.

<sup>268</sup> Artigo 5º Direito à integridade pessoal 1. Toda pessoa tem direito a que se respeite sua integridade física, psíquica e moral.

<sup>269</sup> Artigo 24. Igualdade perante a lei Todas as pessoas são iguais perante a lei. Por conseguinte, têm direito, sem discriminação alguma, à igual protecção da lei.

<sup>270</sup> “Considerando que cada vez mais estas aplicações têm um forte impacto em áreas sensíveis da sociedade, tais como o uso de dados para o desenvolvimento de ajuda humanitária, auxílio no diagnóstico médico correto ou proporcionar racionalidade a decisões judiciais (ALMEIDA; DONEDA, p. 143), é certo o potencial impacto que estas decisões automatizadas poderão causar aos direitos individuais e coletivos (art. 5º da CFRB) dos titulares de dados, mas também aos seus direitos sociais

É de se notar que a “eficácia irradiante” dos direitos fundamentais, como leciona SARMENTO (2006), promove a “humanização da ordem jurídica” com observância da dignidade humana e igualdade substantiva<sup>271</sup>. Portanto, o ordenamento jurídico brasileiro contém mecanismos de proteção de tais direitos em situações diversas, inclusive no uso da tecnologia. Eis alguns exemplos:

(i) há um dispositivo, recentemente inserido na Constituição Federal por meio da Emenda Constitucional nº 115/2022, que alçou a proteção de dados pessoais à condição de direito fundamental, o art. 5º, LXXIX da CF<sup>272</sup>, que passou a tutelar os direitos existenciais vilipendiados pelos vieses nos processos decisórios que sejam decorrentes do mau tratamento e/ou escolha equivocada de dados na formulação e uso da IA devem. Sob esta ótica, a supervisão humana pode se encaixar como uma das formas de controle lastreadas neste direito fundamental com o fito de se evitar discriminações<sup>273</sup>.

(ii) o princípio do devido processo legal, direito fundamental presente na regra do art. 5º, LIV da CF, determina que “ninguém será privado da liberdade ou de seus bens sem o devido processo legal”. A liberdade, valor essencial da república, é a base do princípio; o art. 5º, LV da CF, por sua vez assegura aos litigantes “em processo judicial ou administrativo, e aos acusados em geral” o contraditório e ampla defesa “com os meios e recursos a ela inerentes”. Portanto, como

---

(artigo 6º da CRFB). Em casos envolvendo decisões automatizadas que irão conceder ou negar determinado bem jurídico, é essencial avaliar a natureza de tal bem.” MULLHOLLAND, Caitlin; FRAJHOF, Isabella Z. **Inteligência artificial e a lei geral de proteção de dados pessoais**: breves anotações sobre o direito à explicabilidade perante a tomada de decisões por meio de *machine learning*. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito**: ética, regulação e responsabilidade. São Paulo: Ed. Revista dos Tribunais, 2020, 273.

<sup>271</sup> “Uma das mais importantes consequências da dimensão objetiva dos direitos fundamentais é o reconhecimento da sua eficácia irradiante. Esta significa que os valores que dão lastro aos direitos fundamentais penetram por todo o ordenamento jurídico, condicionando a interpretação das normas legais e atuando como impulsos e diretrizes para o legislador, a administração e o Judiciário. A eficácia irradiante, neste sentido, enseja a “humanização” da ordem jurídica, ao exigir que todas as suas normas sejam, no momento de aplicação, reexaminadas pelo operador do direito com novas lentes, que terão as cores da dignidade humana, da igualdade substantiva e da justiça social impressas no tecido constitucional.” SARMENTO, Daniel. **Direitos fundamentais e relações privadas**. Rio de Janeiro: Lumen Juris, 2006, p. 124.

<sup>272</sup> Art. 5º Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade, nos termos seguintes: [...] LXXIX - é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais.

<sup>273</sup> Ao tratar do perfilamento de indivíduos antes da inserção do inciso LXXIX no art. 5º da CF pela EC nº 115/2022: “Neste cenário fica evidente a necessidade de que existam formas de controle destas práticas, a fim de evitar e até mitigar riscos de potenciais discriminações, e ilicitude ou abuso no tratamento de dados pessoais. O cuidado deve ser ainda maior por estarmos diante de um Direito Fundamental.” MULHOLLAND et al., 2020, p. 272.

instrumento de tutela da liberdade e autonomia humanas, a Constituição Federal assegurou o direito à ciência dos processos decisórios, à transparência nos ritos processuais, à resposta e à revisão de decisões<sup>274</sup>.

Logo, a privação de liberdade no ordenamento jurídico brasileiro, seja ela física ou relativa ao exercício da autonomia privada, é exceção. Tal regra vale tanto para as relações jurídicas tradicionais, como para aquelas decorrentes do uso da tecnologia, pois o ordenamento jurídico vige não somente no meio físico, mas também no meio digital.

O que se vê nos exemplos citados ao longo deste trabalho é a privação de liberdade plena disfarçada em decisões aparentemente corriqueiras, normalizadas pela intermediação de robôs, que, segundo CALO (2015) são desprovidos de uma alma para culpar ou de um corpo para chutar<sup>275</sup>.

A legislação infraconstitucional também apresenta uma série de caminhos que propiciam a tutela de direitos existenciais feridos por decisões autônomas de IA. Por exemplo, decisões que atinjam o livre exercício da personalidade e digam respeito à autonomia existencial, tais como as que envolvam questões ligadas à identidade, gênero, projeto de vida, encontram limites no art. 11 do Código Civil<sup>276</sup>.

O princípio da boa-fé objetiva, previsto no art. 422 do Código Civil<sup>277</sup> e presente na minuta de substitutivo do PL nº 21/2020<sup>278</sup>, também justifica a supervisão por um humano, pois a boa-fé implica em um exercício de autonomia que, em determinadas circunstâncias, não pode ser delegada a uma máquina.

---

<sup>274</sup> Em um paralelo com o devido processo legal, CITRON (2007) propõe a aplicação de um devido processo tecnológico nos procedimentos de decisão automatizada de entes governamentais norteamericanos como forma de proteção de importantes direitos individuais. Assim, pondera a necessidade de implementação de passos procedimentais e mecanismos de controle tais como a notificação, direito de ser ouvido (contraditório), publicidade e transparência nas informações e processos internos de delegação de poder, proteção da expertise e discricionariedade humanas.

<sup>275</sup> “The prospect of punishing corporations has already confronted courts and lawmakers with the problem that robots have ‘no soul to damn [and] no body to kick’ when they violate the law.” CALO, 2015, p. 554.

<sup>276</sup> Art. 11. Com exceção dos casos previstos em lei, os direitos da personalidade são intransmissíveis e irrenunciáveis, não podendo o seu exercício sofrer limitação voluntária.

<sup>277</sup> Art. 422. Os contratantes são obrigados a guardar, assim na conclusão do contrato, como em sua execução, os princípios de probidade e boa-fé.

<sup>278</sup> Ao estabelecer os princípios da IA no Brasil, o art. 3º preconiza a observância da boa-fé: “Art. 3º O desenvolvimento, implementação e uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios [...]” A boa-fé do agente também deverá ser considerada na aplicação de sanções nos termos do art. 30, § 3º da minuta.

No que se refere às relações consumeristas, além do direito de defesa do consumidor ser considerado um direito fundamental (art. 5º, XXXII, CF), o Código de Defesa do Consumidor (CDC) estabelece a política nacional das relações de consumo, reconhecendo a vulnerabilidade do consumidor no mercado de consumo no art. 4º, I.

Por haver um desequilíbrio de poder na relação, na qual o fornecedor detém o poder econômico<sup>279</sup>, a transferência do poder decisório para a IA de forma massificada, sem um mecanismo de supervisão humana, coloca a autonomia do consumidor em risco.

Tal preocupação se apresenta no texto da minuta do substitutivo do PL nº 21/2020 e do PL nº 2328/2023, que indica como fundamento para a aplicação da IA no Brasil a defesa do consumidor (art. 2º, VII) que também aparece em outros dispositivos ao longo do texto da minuta, que inclusive fazem referência ao CDC (a exemplo dos artigos 24, I e 29).

Consoante exposto anteriormente<sup>280</sup>, o art. 20 da Lei Geral de Proteção de Dados Pessoais – LGPD, inspirado no art. 22 da *General Data Protection Regulation* – GDPR, estabeleceu o direito à explicação para que o titular de dados pessoais possa “solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais”.

Portanto, o referido dispositivo trata do direito de revisão pós veredito da máquina, o que difere da proposta da supervisão humana, que se dá antes da decisão. Contudo, este artigo é relevante ao tema, pois positiva a defesa da autonomia humana frente a “tirania dos julgamentos automatizados”<sup>281</sup>.

---

<sup>279</sup> “A opção do legislador brasileiro, como já referimos, foi pelo estabelecimento de uma presunção de vulnerabilidade do consumidor, de modo que todos os consumidores sejam considerados vulneráveis, uma vez que a princípio não possuem o poder de direção da relação de consumo, estando expostos às práticas comerciais dos fornecedores no mercado.” MIRAGEM, Bruno. **Curso de direito do consumidor**. 3. ed ver., atual. e ampl. São Paulo: Editora Revista dos Tribunais, 2012, p. 100.

<sup>280</sup> Vide item 2.1.2.

<sup>281</sup> “Ao analisar este dispositivo, Ana Frazão o descreve como uma espécie de devido processo legal para salvaguardar as pessoas daquilo que denominou muito corretamente de ‘tirania dos julgamentos automatizados’[...].” MULHOLLAND et al., 2020, p. 274-275.



Há que se observar ainda na LGPD o art. 6º, IV – direito de livre acesso; art. 6º, VI – transparência dos dados; art. 6º, IX – não discriminação<sup>282</sup>; e o art. 42 – não responsabilização.

A Lei nº 12.965/2014, o Marco Civil da Internet, estabelece como fundamentos do uso da internet no Brasil dentre outros: (i) os direitos humanos e o desenvolvimento da personalidade (art. 2º, II); (ii) a pluralidade e a diversidade (art. 2º, III); (iii) e a finalidade social da rede (art. 2º, VI).

Como a internet é um dos principais meios de utilização dos sistemas de IA, mecanismos de proteção da autonomia existencial (individual, interpessoal e coletiva) presentes da citada norma importam para a consideração da relevância da presença de um humano no circuito decisório da tecnologia.

---

<sup>282</sup> “[...] o princípio da não discriminação (artigo 6º, IX) deve ser refletido em todas as circunstâncias em que o uso de dados, sejam sensíveis ou não, gere algum tipo de desvalor ou indução a resultados que seriam inequitativos. Esse princípio deve servir como base de sustentação da tutela dos dados sensíveis, especialmente quando estamos diante do exercício democrático e do acesso a direitos sociais, tais como o direito ao trabalho, à saúde e à moradia.” MULHOLLAND et al., 2020, p. 271.

## 5

### **Conclusão.**

Os avanços tecnológicos que a sociedade hodierna tem vivenciado causam fascínio. A possibilidade de praticar atos ou realizar desejos de maneira quase que instantânea, com o apertar de um botão, encanta ao passo em que gera angústias e incertezas sobre o futuro da humanidade.

Pedir a comida favorita por meio de um clique, ciente do tempo estimado de chegada e com acompanhamento das etapas de preparo e entrega; ter a rota com menos tráfego disponível em segundos para chegada mais rápida ao aeroporto; acessar um aplicativo que pré-selecione pessoas que combinem entre si para um relacionamento amoroso, todas essas experiências incríveis, que não se poderia imaginar como viáveis há poucos anos, e se tornaram corriqueiras, somente são possíveis por intermédio da tecnologia.

As tarefas citadas são realizadas pela tecnologia da IA que, a partir de dados e algoritmos, tem a capacidade de aprender com os seus erros e por meio deles se aperfeiçoar. Ela auxilia, facilita, otimiza, sugere, ajuda a decidir ou acaba por decidir pelo indivíduo quando, por exemplo, na compra seguinte via aplicativo, pergunta: “vai repetir seu pedido anterior?”. Por que não? Ela conhece os gostos do usuário.

Fazendo uso de técnicas de perfilamento e mineração de dados em larga escala, a IA aprende sobre as preferências das pessoas, as categoriza e define os produtos ou serviços a serem consumidos por elas. Também é capaz de estabelecer as capacidades ou aptidões individuais a serem exploradas.

Inserido nesse ciclo, o indivíduo continua e gradativamente renuncia à sua autonomia e deixa, por comodidade ou de forma inconsciente, das decisões mais simples, como fazer uma compra, às mais complexas, como se relacionar com outro indivíduo, a critério da máquina. Assim, o ser humano passa a delegar o seu poder decisório e em contrapartida ter menos controle sobre o seu destino.

Sob a perspectiva das relações jurídicas, a IA atua como um instrumento de mediação entre os dois polos, fazendo a ponte de conexão entre, por exemplo, o fornecedor e o consumidor, o patrão e o empregado, ou poder o público e o cidadão. Considerando que a IA é desprovida do atributo jurídico “autonomia”, quando há uma decisão automatizada, uma das partes da relação delegou a ela seu poder decisório, enquanto a outra é se torna o sujeito afetado pela decisão.

Nesse contexto, a supervisão humana se apresenta como uma ferramenta que permite a avaliação dos resultados sugeridos pela IA antes que eles se tornem decisões. Assim, eventuais erros na coleta de dados ou formulação do algoritmo, ou a ocorrência de vieses discriminatórios, podem ser constatados em tempo, evitando-se que se cause prejuízo à autonomia existencial dos indivíduos.

No caso utilizado como exemplo, da estudante transgênero do Distrito Federal<sup>283</sup>, se o reconhecimento facial tivesse emitido uma notificação para a autoridade competente verificar uma informação inconsistente com os dados do sistema, ou até mesmo para que a usuária prestasse informações adicionais, em vez de bloquear o seu bilhete único, possivelmente, não teria ocorrido o cerceio de vários dos seus direitos fundamentais.

A colocação do humano no comando permite que se detecte e corrija inconsistências e erros antes do veredito. Desta forma, a IA pode ser treinada e aprimorada em consonância com os valores humanos e, conseqüentemente, desempenhar as suas funções sem perpetuar discriminações, servindo como uma ferramenta para o bem da coletividade.

Logo, o controle humano da tecnologia e o estabelecimento de limites para o poder decisório da máquina são essenciais para que haja um “devido processo tecnológico”<sup>284</sup>. Por meio da prestação de contas, transparência, explicabilidade, responsabilização e auditabilidade, que tornam visíveis os processos decisórios, promove-se a liberdade existencial e a pluralidade.

Tanto nos tratados de direitos humanos citados, quanto na Constituição Federal, há princípios que justificam a implementação da supervisão humana nos processos decisórios da IA, pois não é razoável, dentro da moral constitucional

---

<sup>283</sup> Vide item 3.4.1.

<sup>284</sup> Vide nota de rodapé nº 270.

vigente, pensar-se em atribuir aos sistemas decisões que devem ser objeto da discricionariedade humana.

Por conseguinte, a dignidade humana, a solidariedade, a igualdade, a liberdade, o direito à saúde, à vida, à educação, à moradia, ao trabalho, à não discriminação, à proteção de dados pessoais, dentre outros princípios e direitos fundamentais, indicam a necessidade do olhar humano, dotado de sensibilidade e empatia.

A legislação infraconstitucional citada também justifica o emprego da supervisão humana, em especial o art. 20 da LGPD que traz o direito de revisão, que coloca o controle humano *ex post* no circuito dos processos decisórios da IA no contexto de tratamento de dados pessoais.

Apesar de propor um controle humano após o veredito da máquina (diferentemente da supervisão, que se dá antes da decisão, e na qual cabe ao humano a palavra final), o art. 20 da LGPD dá ao usuário um direito que representa um poder, uma oportunidade de escapar da vigília ostensiva e incansável da tecnologia, e ao mesmo tempo enxergar a forma e as razões pelas quais as decisões são tomadas.

A sociedade está se transformando e o direito tenta exercer a sua tarefa civilizatória<sup>285</sup>. Há um movimento de adaptação que se percebe por meio da formulação de princípios, guias de boas práticas, alteração de dispositivos legais e apresentação de propostas legislativas específicas para tratar de um tema extremamente relevante para o modo de vida atual.

Não por acaso, a obrigatoriedade da supervisão humana das decisões autônomas de IA produzidas por sistemas de alto risco, que afetam direitos existenciais e fundamentais, está posta nas propostas legislativas do Brasil e Europa como um mecanismo de controle e preservação da autoridade humana.

---

<sup>285</sup> Ao tratar do princípio da dignidade humana, Maria Celina Bodin de Moraes discorre sobre a “tarefa civilizatória” do direito: “Nos Estados democráticos, porém, é na esfera política que são reconhecidos os valores comuns e estabelecidos os princípios fundamentais. O direito constitucional representa o conjunto de valores sobre os quais se constrói, na atualidade, o pacto da convivência coletiva, função outrora exercida pelos códigos civis. A primazia da ordem jurídica consubstancia-se na primazia das decisões políticas em face da natureza das coisas, devendo o legislador não só se esforçar por refletir a realidade, mas também, conscientemente, buscar a sua transformação. O direito é justamente isto, uma força de transformação da realidade. É sua a tarefa ‘civilizatória’, reconhecida através de uma intrínseca função promocional, a par da tradicional função repressiva, mantenedora do *status quo*.” BODIN DE MORAES, Maria Celina. **Princípios do direito civil contemporâneo**. Coord. Maria Celina Bodin de Moraes. Rio de Janeiro: Renovar, 2006, p. 5.

Doravante, é importante acompanhar a tramitação das referidas propostas, bem como as discussões que envolvem o tema e, caso a obrigatoriedade da supervisão humana venha a ser positivada de forma expressa, observar a adaptação da sociedade em geral, do poder público, das *big techs* e do poder judiciário ao comando legal, e como se dará a sua aplicação e efetividade.

É o momento de se refletir sobre “que sociedade vamos construir sobre os ombros de robôs”<sup>286</sup> considerando que, pelo menos por enquanto, a máquina não tem a capacidade de perceber a dimensão da alma humana ou de retribuir um olhar empático e dotado de sentimentos em “intacta retina”<sup>287</sup>.

---

<sup>286</sup> “[...] aquilo que nós chamamos de efeitos da tecnologia não são exatamente características das mesmas, mas sim questionamentos sobre como as pessoas usam as tecnologias e que relações sociais são construídas a partir dessas ferramentas. Fica então a pergunta: que sociedade vamos construir sobre os ombros de robôs?” SOUZA et al., 2020, p. 77.

<sup>287</sup> Vide nota de rodapé nº 1.

## 6

**Referências bibliográficas.**

AI HLEG. **Policy and investment recommendations for trustworthy AI**. 2019, p. 25. Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>>. Acesso em: 26.10.2022.

ALVES, Sarah. **Além do racismo, reconhecimento facial erra mais em pessoas trans**. Tilt UOL. Inteligência artificial. Publicado em: 14.02.2021. Atualizado em: 22.06.2021. Disponível em: <<https://www.uol.com.br/tilt/noticias/redacao/2021/02/14/nao-e-so-racismo-reconhecimento-facial-tambem-erra-mais-em-pessoas-trans.htm>>. Acesso em: 10.08.2022.

ARENDT, Hannah. **A condição humana**. Trad. Roberto Raposo. Rio de Janeiro: Forense Universitária, 2007, p. 188-189.

BALKIN, Jack M. **The path of robotics law**. Calif. L. Rev. Circuit, v. 6, p. 45, 2015.

BAROCAS, Solon, et al. **Big data, data science, and civil rights**. *arXiv preprint arXiv:1706.03102*. Computing Community Consortium – Catalyst, 2017, p. 3.

BAROCAS, Solon; SELBST, Andrew D. **Big data's disparate impact**. California Law Review, Inc., 104:671, 2016, p. 673-674.

BODIN DE MORAES, Maria Celina. **Na medida da pessoa humana**: estudos de direito civil-constitucional. Rio de Janeiro: Renovar, 2010.

BODIN DE MORAES, Maria Celina. **Princípios do direito civil contemporâneo**. Coord. Maria Celina Bodin de Moraes. Rio de Janeiro: Renovar, 2006, p. 5.

BRASIL. **Projeto de lei nº 21, de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e aplicação da inteligência artificial no Brasil, e dá outras providências. Publicação: 30.09.2021. Disponível em: <<https://legis.senado.leg.br/sdleg-getter/documento?dm=9023442&ts=1676314334096&disposition=inline>>. Acesso em: 16.02.2023.

BRASIL. **Projeto de lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Publicação: 03.05.2023. Disponível em: <<https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>>. Acesso em: 05.05.2023.

BRUCE, Peter; BRUCE, Andrew. **Estatística prática para cientistas de dados: 50 conceitos essenciais**. Trad. Luciana Ferraz. Rio de Janeiro: Alta Books, 2019, 2. Recurso eletrônico. Item viés de seleção, n.p.

BUTLER, Judith. **Gender trouble: feminism and the subversion of identity**. New York/London: 2007, p. 9.

CALO, Ryan. **Robotics and the lessons of cyberlaw**. California Law Review, Berkeley, v. 103, n. 3, p. 513-563, jun. 2015.

CITRON, Danielle Keats. **Technological due process**. *Wash. UL Rev.* 85 (2007): 1249, p. 1302.

CITRON, Danielle Keats; PASQUALE, Frank. **The scored society: Due process for automated predictions**. *Wash. L. Rev.*, v. 89, p. 1, 2014, p. 3.

COLORADO.EDU. CU Boulder Today. **Facial recognition software has a gender problem**. Lisa Marshall. Disponível em: < <https://www.colorado.edu/today/2019/10/08/facial-recognition-software-has-gender-problem>>. Publicado em: 08.10.2019. Acesso em: 10.08.2022.

COMISSÃO EUROPEIA. **Inception impact assessment: proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence**. Disponível em: <[https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=PI\\_COM:Ares\(2020\)3896535&from=EN](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=PI_COM:Ares(2020)3896535&from=EN)>. Acesso em: 20.12.2021.

CORREIO BRAZILIENSE. Publicado em: 08.11.2014. **Morto aos 28 anos, o piauiense Torquato Neto completaria 70 anos no domingo**. Disponível em: <[https://www.correiobraziliense.com.br/app/noticia/diversao-e-arte/2014/11/08/interna\\_diversao\\_arte,456582/morto-aos-28-anos-o-piauiense-torquato-neto-completaria-70-anos-no-domingo.shtml](https://www.correiobraziliense.com.br/app/noticia/diversao-e-arte/2014/11/08/interna_diversao_arte,456582/morto-aos-28-anos-o-piauiense-torquato-neto-completaria-70-anos-no-domingo.shtml)>. Acesso em: 23.11.2021.

CUNHA, Maria. **Japonês casado com um holograma vive crise no relacionamento por falta de comunicação**: depois de quatro anos, casal passa por um momento difícil após dispositivo usado para conversas deixar de funcionar. R7. Internacional. Publicado em: 28.04.2022. Atualizado em 29.04.2022. Disponível em: <<https://noticias.r7.com/internacional/japones-casado-com-um-holograma-vive-crise-no-relacionamento-por-falta-de-comunicacao-29042022>>. Acesso em: 07.12.2022.

DASTIN, Jeffrey. **Amazon scraps secret AI recruiting tool that showed bias against women**. In: MARTIN, Kirsten. **Ethics of data and analytics concepts and cases**. Boca Raton/ London/ New York: CRC Press, Taylor & Francis Group, 2019.

DE OLIVEIRA, Ana. **Tropicalia: Identifisignificados movimento**. Disponível em: <<http://tropicalia.com.br/identifisignificados/movimento>>. Acesso em: 23.11.2021.

DIÁRIO DO SENADO FEDERAL. **Relatório Final Da Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência Artificial**. Publicado em: 09.12.2022. Disponível em: < <https://legis.senado.leg.br/diarios/ver/111533?sequencia=3>>. Acesso em: 11.02.2023.

DONEDA, Danilo; MENDES, Laura Schertel; SOUZA, Carlos Affonso Pereira de; ANDRADE, Norberto Nuno Gomes de. **Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal**. Pensar, Fortaleza, v. 23, n. 4, p. 1-17, out./dez. 2018.

ECHARRI, Miquel. **150 demissões em um segundo**: os algoritmos que decidem quem deve ser mandado embora. EL PAÍS. Tempos modernos. Publicado em: 10.10.2021. Disponível em: <<https://brasil.elpais.com/tecnologia/2021-10-10/150-demissoes-em-um-segundo-assim-funcionam-os-algoritmos-que-decidem-quem-deve-ser-mandado-embora.html>>. Acesso em: 21.12.2021.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. Rev. tec. Luis Ricardo Figueiredo. São Paulo: Pearson Addison Wesley, 2005, p. 2.

EUR-Lex 52021PC0206. **Proposta de regulamento do parlamento europeu e do conselho que estabelece regras harmonizadas em matéria de inteligência artificial (regulamento inteligência artificial) e altera determinados atos legislativos da União**. 2021. Disponível em: <[https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-1aa75ed71a1.0004.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-1aa75ed71a1.0004.02/DOC_1&format=PDF)>. Acesso em: 15.02.2023.

EUROPEAN COMMISSION. **Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: Building Trust in Human-Centric Artificial Intelligence**. 2019. Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>>. Acesso em: 31.10.2022.

FALEIROS JÚNIOR, José Luiz de Moura. **A evolução da inteligência artificial em breve retrospectiva**. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Fellipe Braga Neto, et al. **Direito digital e inteligência artificial**: diálogos entre Brasil e Europa. Indaiatuba: Editora Foco, 2021, p. 20.

FJELD, Jessica, et al. **Principled artificial intelligence**: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication 2020-1 (2020).

FOUCAULT, Michel. **A verdade e as formas jurídicas**. Trad. Eduardo Jardim e Roberto Machado. Rio de Janeiro: Nau, 2013a.

FOUCAULT, Michel. **Vigiar e punir**: nascimento da prisão. Introdução: António Fernando Cascais. Lisboa: Edições 70 / Edições Almedina, S.A., 2013b.

FRAZÃO, Ana. **Plataformas digitais, big data e riscos para os direitos da personalidade**. In: TEPEDINO, Gustavo; MENEZES, Joyceane Bezerra de (Coord.). **Autonomia privada, liberdade existencial e direitos fundamentais**. Belo Horizonte: Fórum, 2019.

GOERTZEL, Ben. **Human-level Artificial General Intelligence and the Possibility of a Technological Singularity - A Reaction to Ray Kurzweil's The**



**Singularity Is Near, and McDermott's Critique of Kurzweil.** Artificial Intelligence 171.18 (2007): 1161-173. Web, p. 1172.

GOMES, Rodrigo Dias de Pinho. **Carros autônomos e os desafios impostos pelo ordenamento jurídico:** uma breve análise sobre a responsabilidade civil envolvendo veículos inteligentes. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito:** ética, regulação e responsabilidade. São Paulo: Ed. Revista dos Tribunais, 2020, p. 591.

GPAN IA. **Orientações éticas para uma IA de confiança.** Comissão europeia, 2019a. Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>>. Acesso em: 26.10.2022.

GPAN IA. **Uma definição de IA:** principais capacidades e disciplinas científicas. Comissão europeia: 2019b. Disponível em: <<https://digital-strategy.ec.europa.eu/pt/node/1950>>. Acesso em: 27.10.2022.

HAMIDI, Foad; SCHEUERMAN, Morgan Klaus; BRANHAM, Stacy M. **Gender recognition or gender reductionism?** The social implications of embedded gender recognition systems. In Proceedings of the 2018 chi conference on human factors in computing systems, pp. 1-13, 2018, p. 2.

HARCOURT, Bernard E., **Exposed:** desire and disobedience in the digital age. Cambridge, Massachusetts, London, England: Harvard University Press, 2015, p. 91-92.

HEIDEGGER, Martin. **The question concerning technology and other essays.** Trans. William Lovitt. New York & London: Garland Publishing, Inc.: 1977.

JONES, Meg Leta. **The right to a human in the loop:** Political constructions of computer automation and personhood. Social Studies of Science 47.2 (2017): 216-239.

KAHN, Charles H. **Sobre o verbo grego ser e o conceito do ser.** Coord. Mauro Iglésias. Série filosofia antiga 1. Rio de Janeiro: Núcleo de estudos de filosofia antiga, departamento de filosofia da Puc-Rio, 1997.

KAMINSKI, Margot E. **The right to explanation, explained.** Berkeley Tech. LJ 34: 189, 2019, p. 4.

KURZWEIL, Ray. **The Singularity Is Near:** When Humans Transcend Biology. Viking Adult, ISBN 0670033847, September 2005, 672 pages.

MACHADO, Simone. **ChatGPT:** o que é e como funciona robô de conversa que vem gerando polêmica. Tilt UOL. Inteligência artificial. Publicado em: 15.12.2022. Disponível em: < <https://www.uol.com.br/tilt/noticias/redacao/2022/12/15/o-que-e-chatgpt-a-plataforma-de-ia-que-vem-gerando-polemica.htm>>. Acesso em: 16.02.2023.

MARÇAL, J. C. **“Existência” e “verdade” em Martin Heidegger**. ISSN 1984-3879, SABERES, Natal – RN, v. 1, n.6, fev. 2011.

MCDERMOTT, Drew. **Kurzweil's argument for the success of AI**. Artificial Intelligence, Volume 170, Issue 18, 2006, Pages 1227-1233, ISSN 0004-3702, p. 1233.

MIRAGEM, Bruno. **Curso de direito do consumidor**. 3. ed ver., atual. e ampl. São Paulo: Editora Revista dos Tribunais, 2012, p. 100.

MULLHOLLAND, Caitlin. **Responsabilidade civil e processos decisórios autônomos em sistemas de inteligência artificial (ia)**: autonomia, imputabilidade e responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito**: ética, regulação e responsabilidade. São Paulo: Ed. Revista dos Tribunais, 2020.

MULHOLLAND, Caitlin; FRAJHOF, Isabella Z. **Entre as leis da robótica e a ética**: regulação para o adequado desenvolvimento da inteligência artificial. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Felipe Braga Neto, et al. **Direito digital e inteligência artificial**: diálogos entre Brasil e Europa. Indaiatuba: Editora Foco, 2021, p. 66.

MULLHOLLAND, Caitlin; FRAJHOF, Isabella Z. **Inteligência artificial e a lei geral de proteção de dados pessoais**: breves anotações sobre o direito à explicabilidade perante a tomada de decisões por meio de *machine learning*. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito**: ética, regulação e responsabilidade. São Paulo: Ed. Revista dos Tribunais, 2020.

NADER, Paulo. **Introdução ao estudo do direito**. Rio de Janeiro: Forense, 2006, 26ª edição, p. 288.

NILSSON, Nils J. **The quest for artificial intelligence**: a history of ideas and achievements. Cambridge, UK: Cambridge University Press, 2010, apud STONE, Peter, et al. **Artificial intelligence and life in 2030**: the one hundred year study on artificial intelligence. In: **One hundred year study on artificial intelligence**: report of the study panel. Stanford University, 2016, p. 10.

OPENAI. **ChatGPT**: optimizing language models for dialogue. Publicado em: 30.11.2022. Disponível em: <<https://openai.com/blog/chatgpt/>>. Acesso em: 16.02.2023.

PETTA GOMES DA COSTA, De Leon. **Reviewing the Concept of Technological Singularities**: How Can It Explain Human Evolution? Nanoethics 13.2 (2019): 119-30. Web, p. 128.

PERLINGIERI, Pietro. **O direito civil na legalidade constitucional**. Trad. Maria Cristina De Cicco. Rio de Janeiro: Renovar, 2008, p. 168-169.

RODOTÀ, Stefano. **A vida na sociedade da vigilância**: a privacidade hoje. Organização, seleção e apresentação: Maria Celina Bodin de Moraes. Rio de Janeiro: Renovar, 2008.

RUSSEL, Stuart. **Human compatible**: artificial intelligence and the problem of control. New York: Viking/Penguin Random House LLC, 2019.

SADIN, Eric. **Artificial Intelligence**: A Radical Anti-Humanism. Artificial intelligence and ethics, two subjects that go hand-in-hand in current literature, are often confined to the spectrum of the subject of privacy. DZone. Thomas Jardinet. Publicado em: 12.02.2019. Disponível em: < <https://dzone.com/articles/artificial-intelligence-a-radical-anti-humanism>>. Acesso em: 09.01.2023.

SAE International. **Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles J3016\_202104**. Disponível em: <[https://www.sae.org/standards/content/j3016\\_202104/](https://www.sae.org/standards/content/j3016_202104/)>. Acesso em: 08.12.2022.

SARMENTO, Daniel. **Direitos fundamentais e relações privadas**. Rio de Janeiro: Lumen Jurins, 2006, p. 124.

SARTRE, Jean Paul. **O existencialismo é um humanismo**. Os pensadores, XLV. São Paulo: Abril S.A. Cultural e Industrial/Paris: Les Éditions Nagel/Lisboa: Editorial Presença Ltcla., 1973.

SCHEUERMAN, Morgan Klaus; PAUL, Jacob M.; BRUBAKER, Jed R. **How computers see gender**: an evaluation of gender classification in commercial facial analysis and image labeling services. Proceedings of the ACM on Human-Computer Interaction 3. CSCW (2019): 1-33, p. 2.

SEMOB. **Biometria facial é obrigatória em 100% dos ônibus do DF**. Disponível em: <<https://semob.df.gov.br/biometria-facial-e-obrigatoria-em-100-dos-onibus-do-df/>>. Publicado em: 03.05.2018. Atualizado em: 07.11.2018. Acesso em: 10.08.2022.

SOARES, Flaviana Rampazzo. **Levando os algoritmos a sério**. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Fellipe Braga Neto, et al. **Direito digital e inteligência artificial**: diálogos entre Brasil e Europa. Indaiatuba: Editora Foco, 2021, p. 46.

SOUZA, Carlos Affonso Pereira de; OLIVEIRA, Jordan Vinícius de. **Sobre os ombros de robôs?** A inteligência artificial: entre fascínios e desilusões. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito**: ética, regulação e responsabilidade. São Paulo: Ed. Revista dos Tribunais, 2020.

STAPLETON, Timothy. **Dasein as being-in-the-world**. Martin Heidegger key concepts. Edited by Bret W. Davis. First published in 2010 by Acumen Publishing Limited, p. 47.

STONE, Peter, et al. **Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence**. In: **One hundred year study on artificial intelligence: report of the study panel**. Stanford University, 2016.

TEIXEIRA, Ana Carolina Brochado. **Autonomia Existencial**. Revista Brasileira de Direito Civil – RBDCivil | Belo Horizonte, vol. 16, p. 75-104, abr./jun. 2018, p. 80-81.

TEPEDINO, Gustavo; SILVA, Rodrigo da Guia. **Desafios da inteligência artificial em matéria de responsabilidade civil**. Revista Brasileira de Direito Civil – RBDCivil | Belo Horizonte, v. 21, p. 61-86, jul./set. 2019, p. 63-64.

TEPEDINO, Gustavo; TEIXEIRA, Ana Carlolina Brochado; ALMEIDA, Vitor. **O direito civil entre o sujeito e a pessoa: estudos em homenagem ao professor Stefano Rodotà**. Belo Horizonte: Fórum, 2016, p. 17.

TOMASEVICIUS FILHO, Eduardo. **Reconhecimento facial e lesões aos direitos da personalidade**. In: CORDEIRO, A. Barreto Menezes, et al. Coord. Fellipe Braga Neto, et al. **Direito digital e inteligência artificial: diálogos entre Brasil e Europa**. Indaiatuba: Editora Foco, 2021, p.130.

VELOSO, Caetano. **Caetano Veloso conta a história da Música “Cajuína”**. Globoplay. Altas Horas. Serginho Groisman. Publicado em: 01.12.2014. Disponível em: <<https://globoplay.globo.com/v/3119899/>>. Acesso em: 22.11.2021.

VELOSO, Caetano. **Cajuína**. LETRAS. Disponível em: <<https://www.lettras.mus.br/caetano-veloso/44704/>>. Acesso em: 22.11.2021.

VIVEIROS DE CASTRO, Thamís Dalsenter. **A função da cláusula de bons costumes no direito civil e a teoria tríplice da autonomia privada existencial**. Revista Brasileira de Direito Civil – RBDCivil | Belo Horizonte, vol. 14, p. 99-125, out./dez. 2017a.

VIVEIROS DE CASTRO, Thamís Dalsenter. **Bons costumes no direito civil brasileiro**. São Paulo: Almedina, 2017b. E-book Kindle, n.p.

VIVEIROS DE CASTO, Thamís Dalsenter. **Questões atuais sobre o direito ao próprio corpo na legalidade constitucional**. In: MATOS, Ana Carla Harmatiuk; TEIXEIRA, Ana Carolina Brochado; TEPEDINO, Gustavo. (Org.). **Direito Civil, Constituição e unidade do sistema**. 1 ed. Belo Horizonte: Fórum, 2019, v. 1, p. 33-46, p. 35.

WALSH, Toby. **The Singularity May Never Be Near**. The AI Magazine 38.3 (2017): 58-62. Web, p. 1.

WANG, Pei. **What do you mean by “AI”?** AGI, vol. 171, 2008, p. 1.

WHEELAN, Charles. **Estatística: o que é, para que serve, como funciona**. Trad. George Schlesinger. Rio de Janeiro: Zahar, 2016, p. 16.

ZERILLI, John, et al. **Transparency in algorithmic and human decision-making:** is there a double standard?. *Philosophy & Technology* 32.4: 661-683, 2019, p. 2.

ZUBOFF, Shoshana. **Big other:** surveillance capitalism and the prospects of an information civilization. *Journal of information technology* 30.1 (2015): 75-89.